Tecnun Universidad de Navarra

Proyecto Fin de Máster

INGENIERIA INDUSTRIAL

Fostering sustainable, vibrant, and walkable communities through the amenity lens. Use Case: Kendall Square

Pablo Barrenechea Bachiller

San Sebastián, septiembre de 2022

ACKNOWLEDGEMENTS

This Master's Thesis could not have been accomplished without the help of so many people.

To my family, for their constant support and counsel.

To the CS group and its members, for the guidance & making me feel at home, for which I will ever be grateful.

To Kent, for the opportunity and the valuable learning.

To Luis, Markus & Ronan, whose advice and challenge have been proven essential for the completion of this project.

To Leire, for her time and esteemed feedback.

To my university friends, who have become family.

To my Boston friends, with whom I have shared memorable moments.

To my longtime friends, whose encouragement I shall always have

CONTENTS

LIST OF FIGURES

LIST OF TABLES

RESUMEN

En las próximas décadas se espera que se mantenga el ritmo del éxodo hacia las ciudades. Con una población en aumento en las áreas metropolitanas, la planificación urbana juega un papel fundamental para garantizar que las zonas urbanas tengan una escala humana y sean vibrantes y sostenibles. La expansión urbana ha impactado negativamente en la diversidad, densidad y proximidad de las comunidades, trinomio de atributos urbanos indispensable para fomentar la innovación. El grupo City Science (CS) del MIT Media Lab (ML) tiene como propósito evaluar estrategias que ayuden a desarrollar las comunidades habitables del futuro.

Los distritos caminables se postulan como la intervención clave para lograr la simetría vida-trabajo. El objetivo de esta intervención es fomentar políticas que permitan que las personas puedan vivir en la misma comunidad en la que trabajan. Con este fin, es imprescindible que las personas que viven y trabajan en un barrio tengan acceso a las *amenities* cotidianas a una distancia caminable. Los distritos caminables tienen el potencial de mejorar las áreas urbanas y los distritos de negocios al acercar las *amenities* a los ciudadanos. Las *amenities* caminables son el elemento urbano clave que permite a las personas no viajar innecesariamente para disfrutar de las actividades que conforman su vida cotidiana, al mismo tiempo que se reduce el impacto ambiental y se fomenta la innovación.

En el presente Proyecto Fin de Máster, se hace uso de técnicas de *big data* y *machine learning* para analizar, cuantificar y evaluar las *amenities* necesarias en un distrito en función del perfil socioeconómico de los residentes; con el fin de lograr la simetría vida-trabajo y así desarrollar una comunidad equilibrada, innovadora, vibrante y caminable.

ABSTRACT

The unstoppable exodus to cities is expected to keep its pace in the decades to come. With an increasing population in cities, urban planning plays a key role in ensuring that the urban hubs have a human scale, are vibrant, and are sustainable. Urban sprawl has had negative impacts on diversity, density, and proximity, an indispensable trinomial set of urban attributes to foster innovation. The MIT Media Lab (ML) City Science (CS) Group's commitment is to explore strategies that would help achieve livable communities of the future.

Walkable communities are the key intervention to achieve live-work symmetry. The goal of this intervention is to foster policies so people can live in the same community where they work. To accomplish this, it is crucial for people living and working in the neighborhood to have access to their everyday amenities within a walkable distance. Walkable amenities have the potential to enhance urban areas and business districts by bringing everyday amenities into proximity to citizens. Walkable amenities are a key urban element allowing people not to unnecessarily commute to enjoy their daily life activities, reducing environmental impacts, and increasing innovation.

This thesis will use big data and machine learning techniques to analyze, quantify, and evaluate the required amenities to be deployed in a district depending on the socio-economic profile of the residents; in order to achieve live-work symmetry and to develop a well-balanced walkable, vibrant, and innovative community.

1. INTRODUCTION

1.1. City Science

The City Science (CS) group is part of the MIT Media Lab (ML); a leading interdisciplinary research organization, which aims to bring together people from very different backgrounds to research broad areas of interest. MIT ML is formed by more than 20 research groups working on a great variety of topics; from the bio-mechatronics lab in which state-of-the-art prostheses are developed, to the opera of the future where the intersection between music and emotions is explored. This diverse culture full of innovation is the ideal environment for breakthroughs to occur.

In this context, the CS group researches on the future of cities. Under Kent Larson's leadership and following the motto "enabling dynamic, evolving places that respond to the complexities of life", the CS group proposes to look beyond smart cities. It tries to answer the future evolution of cities, and which strategies should be adopted to face the coming challenges in order to build vibrant and innovative urban ecosystems. CS research is based on three main ideas:

- Live-work symmetry: people should live within a walkable distance from their jobs and daily amenities
- Changing of mobility behavior
- Underuse of spaces

In that sense, the line of research performed at the Group is divided into three central branches [1]:

- Urban Planning
- Mobility
- Changing places (Transformable architecture)

On the Urban Planning side, the CS Group has developed a digital platform called CityScope, which follows a data-driven approach for urban planning, in which the impact of every decision can be measured. The main objective of this tool is to reach a consensus among different stakeholders. From mayors and urban planners to communities and residents, the idea is to bring people together around a table where every representative's voice can be heard and a beneficial urban planning for all parties can be obtained.

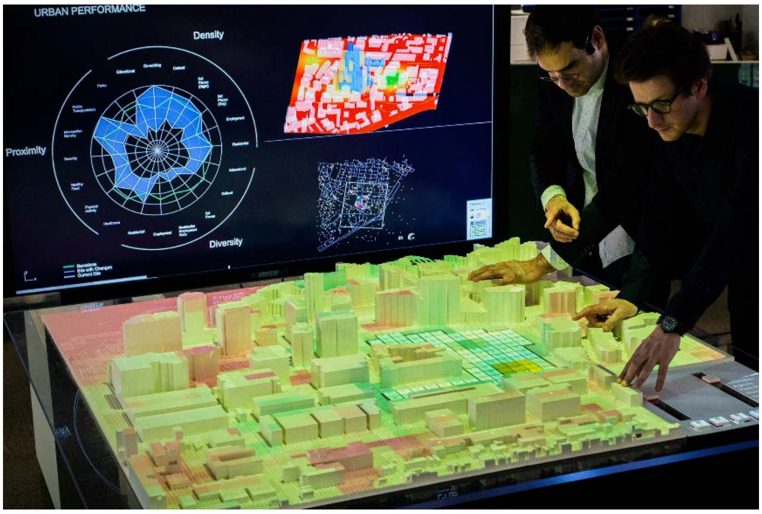

Figure 1. CityScope Volpe: a data-driven interactive simulation tool for urban design

The Volpe Table is located at CS's lab at the ML (Figure 1), and is just an example of the potential of this tool. This Lego table was built in order to analyze the impact of urban planning decisions at Volpe, an area in Kendall Square waiting to be developed. The stakeholders gathered around this table can modify land uses and building characteristics, as well as mobility features, and evaluate the consequences of each of the interventions throughout different layers (e.g. pollution, traffic, proximity to parks, etc.). Measuring the impact of each intervention is one of the greatest contributions, and it is done so by the Urban Performance metric, which is divided into three main indicators offering a holistic view: Diversity, Density, and Proximity; which at the same time are composed by several indicators each.

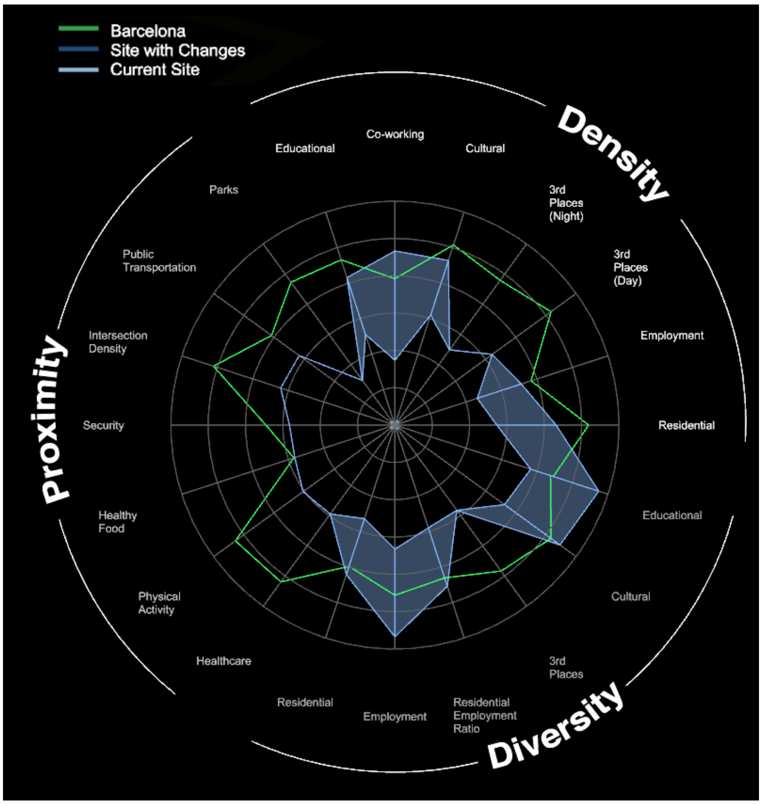

Figure 2. Urban performance radar plot

The image above (Figure 2) shows the radar plot used to measure the Urban Performance of Kendall Square. The inner blue line represents the baseline while the outer one represents the situation with changes. Additionally, the green line represents the benchmark, which is Barcelona, as it is known for its well-functioning urban design.

Assessing the suitability of each decision jointly with an urban performance radar plot and a multi-layered heat map permits stakeholders to evaluate each intervention fully. For instance, a stakeholder might decide to increase the density of people in the area, which has been proven to increase social interaction and thus, foster innovation [2] [3]. However, it may also increase traffic, and commute times can get larger. Having the whole view helps making better informed decisions.

In short, the research question the CityScope is trying to understand is the impact of how Diversity and Density of people, combined with proximity between places, affect the overall Urban Performance. An urban environment that counts with the appropriate amount of density will foster social interaction [2] [3], which is key for the innovation and vibrancy of a region. Diversity will facilitate access to diverse environments, either social, economic or spatial, facilitating creativity and consequently, innovation [4]. Finally, proximity eases the accessibility to any type of amenities, which increases the livability of the district. These indicators are a trinomial, thus requiring to work in equilibrium in order to obtain a good Urban Performance.

Additionally, it is believed that changes in the current mobility systems will be accelerated in the coming years [5] . At the moment, cities are flooded with cars which not only provoke noise and traffic but also, have a great environmental impact. On one hand, based on a proper urban design that fosters live-work symmetry, a reduced need for mobility is thought to be achieved. On the other hand, shifting towards lightweight electric mobility is thought to have a highly positive impact [6]. Under this umbrella, the group has developed many hardware solutions, such as the CityCar, a foldable urban vehicle, or the PEV, a persuasive electric autonomous vehicle (Figure 3). The PEV is a tricycle able to carry a person that incorporates two LIDAR sensors that recognize the surroundings and autonomously adapt its trajectory based on the obstacles found on the way. The making of such a vehicle is englobed under the directive of offering solutions for a transition towards lightweight vehicles.

Figure 3. The PEV autonomously riding in Cambridge, MA

Usability of space is also a great concern for the group. Designing multifunctional transformable spaces is the goal of the Changing Places branch. The reasoning behind this idea is that spaces that maintain the same functionality could be obtained with smaller spaces that incorporate architectural robotics. This has some positive aspects. On one side, more people could be allocated in the same area, with a direct reduction of the environmental impact, as more dense buildings have proven so [7]. Also, smaller spaces could equally make living more affordable. In consequence, these systems can attract more people from different backgrounds to one same place increasing density, diversity and innovation. Under this umbrella, was born Ori Systems, founded by Hasier Larrea, where the incorporation of architectural robotics permits the transformation of spaces.

The following image (Figure 4) shows some of ORI Systems' products in display. On the left side there is the "Cloud Bed", a retractable bed that can slide in or out the ceiling, maximizing the usage of space depending on the time of the day. On the right side, there is the "Pocket Closet", which enables to compartmentalize the space.

Figure 4. ORI Systems' products in a studio

At the CS Group there is a will to have a positive impact on society, and it is done so by trying to build cities better. It is acknowledged that cities have a big room for improvement and there is still a long way to go. In that sense, the CS group offers consultancy services to different government institutions in their pursuit of improving urban ecosystems. However, the impact the CS group can have by doing just so is limited. This is why all the work carried out at the group is open access enabling anyone from anywhere to make use of it.

1.2. Cities

The city-state comes into being for the sake of living, but it exists for the sake of living well
Aristotle

1.2.1. Evolution of cities

History

As urban economist Edward Glaeser pictures it [8], cities are humanity's greatest invention since they go through long periods of stability; they are resilient and innovative. But, what are its origins? The preurban societies were basically nomads. Often referred to as hunter-gatherers they moved from place to place looking for food. The settlement in cities started centuries ago, historians locate this date around 3,500 B.C, with the birth of the first cities in Mesopotamia [9].

Figure 5. Map of the Fertile Crescent Mesopotamia and Egypt and location of first towns

The location is not random, since the first cities were formed in the proximity of natural resources, such as rivers, which made the land around much more fertile. The master of crops and soil along with the domestication of animals, allowed the society to have food surplus and start forming civilizations. At the same time, this excess of food allowed new jobs to prosper, new modes of transport to develop, and commerce to surge, enriching the social interaction between individuals [9] [10].

As time went by, and attacks between different groups started occurring, cities evolved to be confined inside a wall protected from outsiders. The size of the cities was, therefore, limited to the boundary walls [10]. Later on, when attacks started diminishing, cities started sprawling. This urban design is indeed representative of European cities, which contain a historic old part and a later development [11].

The greatest breakthrough came with the Industrial Revolution. The new technologies enabled new sources of energy that eventually led to new modes of transportation and infrastructure to arise. In the US, in the early 20th century, many people migrated from cities to suburban areas in search of fulfillment of the American Dream, often referred to as the ownership of a detached house and a personal car for transport [12].

<u>The origin of cities</u>

There are three main theories related to economic activity creation, thus, city formation [13]:

- Location fundamentals
- Economies of scale
- Agglomeration economies.

The former defends that depending on the location, some regions are more susceptible to human activity. This was the case even in the Iron Age, where regions close to the Mediterranean coast, which counted with an advantage when it comes to trading, had a greater possibility of human presence [14]. The economies of scale theory defends greater benefits can be achieved with a larger production/concentration. Lastly, the agglomeration economies theory is based on the idea that the population wants to settle in closeness to other individuals with the consequent benefits: greater social interaction and share of knowledge [13].

Where are cities heading?

Today, more than half of the world's population lives in urban areas. However, it has not been until one century and a half ago that the exodus to cities increased significantly. At the end of the 19th century, just about 20% of the population lived in cities; number which has been steadily increasing and is projected to locate around 70% by mid-21st century [15]. The numbers are even more impressive for the US, with an 83% of the population currently living in urban areas and a projected 89% for 2050 [16]. In addition, the worldwide population is expected to increase from the actual almost 8 billion, up to 10 billion by the end of the century. This growth on the population is expected to occur mainly in urban areas, which raises a flag on whether the current urbanism approach is the appropriate one to tackle this challenge.

To develop future urbanism, many argue that the current urban development approach needs to be rethought since citizens have been displaced from the streets in sake of transport modes. Car-oriented cities are a past thing; the new way-of-doing should be more person-centric giving people access to the services that cover their daily needs just as ancient cities did [17].

In this context, Carlos Moreno proposes the 15-minute city [18] [19]. This concept, with people at the core of each decision, defends residents should have their everyday needs covered by the services offered within a 15-minute walkable distance. In short, citizens should be able to work and live, understood in all its forms, in the vicinity of their homes. In the chapter 4.1 a broader definition of this method as well as its implications is given,

1.2.2. What is a city?

The city fosters art and is art
Lewis Mumford

There might be a consensus on what a city is on the physical side. As Roger W. Caves points out in the Encyclopedia of the City, cities are "A permanent and densely settled place with boundaries that are administratively defined, a city is the accomplishment of a population whose members work primarily on non-agricultural tasks" [20]. However, cities are not only a physical layout and a given density. Cities are much more. The culture and the urban way of life are also characteristic of cities. In this sense, Lewis Mumford threw an open question about a century ago: What is a city? "The city [...] is a geographic plexus, an economic organization, an institutional process, a theater of social action, and an esthetic symbol of collective unity." The conclusion he got to is the very systemic definition of a city, in which not only the physical aspect is considered, but also the social interaction happening between the different actors [21].

1.2.3. Importance of cities

Cities have come to represent a key cornerstone in today's society. It has already been highlighted that world's population is highly concentrated in cities. Just with half of the population, cities are responsible for more than the 80% of the global Gross Domestic Product (GDP). This metric is even more concentrated, since just 600 urban centers account for 60% of the global GDP with just a fifth of the population [22]. Data for the US shows a greater congregation with metropolitan areas representing 91% of the country's GDP [16].

Additionally, cities are home for innovation. They are responsible for the 90% of innovation as measured by patent applications and R&D investments, fostered by the social interaction between educated and creative people [23]. Nonetheless, innovation takes place largely in big cities. The ten most innovative cities in the United States, account for 48% of its patents, 33% of its GDP but just for the 23% of the total national population [24].

However, cities are also responsible for some negative effects. As reported by United Nations, the world's cities occupy just the 3% of Earth's total land, but still account for approximately the 70% of the world's energy consumption and 75% of the CO2 emissions [25].

These metrics do only confirm the already known importance of cities in today's society. Therefore, when tackling such big issues as climate change, it makes sense to shine the spotlight on cities as the center of the problem, but also the solution.

1.3. Amenities

Amenities are defined by the CS group as the elements, such as parks, restaurants, retail stores and so on, found in the proximity of a residence that help enrich the urban experience.

Since the beginning, amenities' birth and economic activity prosperity has been highly related to trade, which consequently brought the expansion of cities. With the formation of the first cities, new jobs started to prosper. Even if at first people started offering their products and services at their own homes, marketplaces were finally developed. These marketplaces were able to offer a diverse variety of products for citizens. In addition, it also helped foster social interactions and exchange of ideas between people [13]

At the beginning of the past century, the centric amenities had a high relevance. Local business accessible for people on foot were the trend type of commerce. The situation at the moment is quite different. The possibilities new modes of mobility allow have abruptly transformed the commerce scenario.

The literature on amenities [26] [27] [28] [29], defines the current commerce situation as a dual landscape. On one hand, there are big commercial surfaces in the periphery with good car accessibility; while, on the other hand, there are independent businesses in the city centers accessible on foot. Depending on the world region we look at, this dual scenario is tilted towards any of the extremes. In Europe, city centers still hold an important commercial role, despite a liberalizing wave of commerce happening in the decades of the 70s and 80s. This led to the start of the shift of the sector from local businesses to big chains. It is not the case for the US, where suburban and periphery areas have gained a key part, with big commercial surfaces becoming predominant.

A well-functioning-mix of amenities has been proven to have a large positive impact. It is an attraction for people to come live in the area [30], it can influence the mobility patterns of the residents shifting towards more environmentally friendly solutions [31], and the attractiveness of it can determine the economic growth of the region [30] [32]. All in all, they play an important role in determining human behavior not to go unnoticed.

2. CHALLENGE

In the present chapter, the urbanization errant model of US cities, a first iteration of an amenity-dense neighborhood and the Kendall Square situation is assessed.

2.1. Urbanization model of the US

The urban planning carried out in US cities has been highly characterized by the high urban sprawl, understood as the horizontal spreading of a city into its surroundings. In this way of doing, low-density areas with good car access are predominant, with the inevitable consequences this brings. It has been proven that urban sprawl increases traffic and energy use and has a direct impact on worsening air and water quality [33]. Los Angeles is the clearest example of this urbanization model.

Figure 6 shows how the transport per capita energy consumption drastically diminishes with urban density. This does not mean that by increasing urban density, a decline in transport energy consumption will be directly obtained, but if the correct set of policies is implemented this might as well be the case [34].

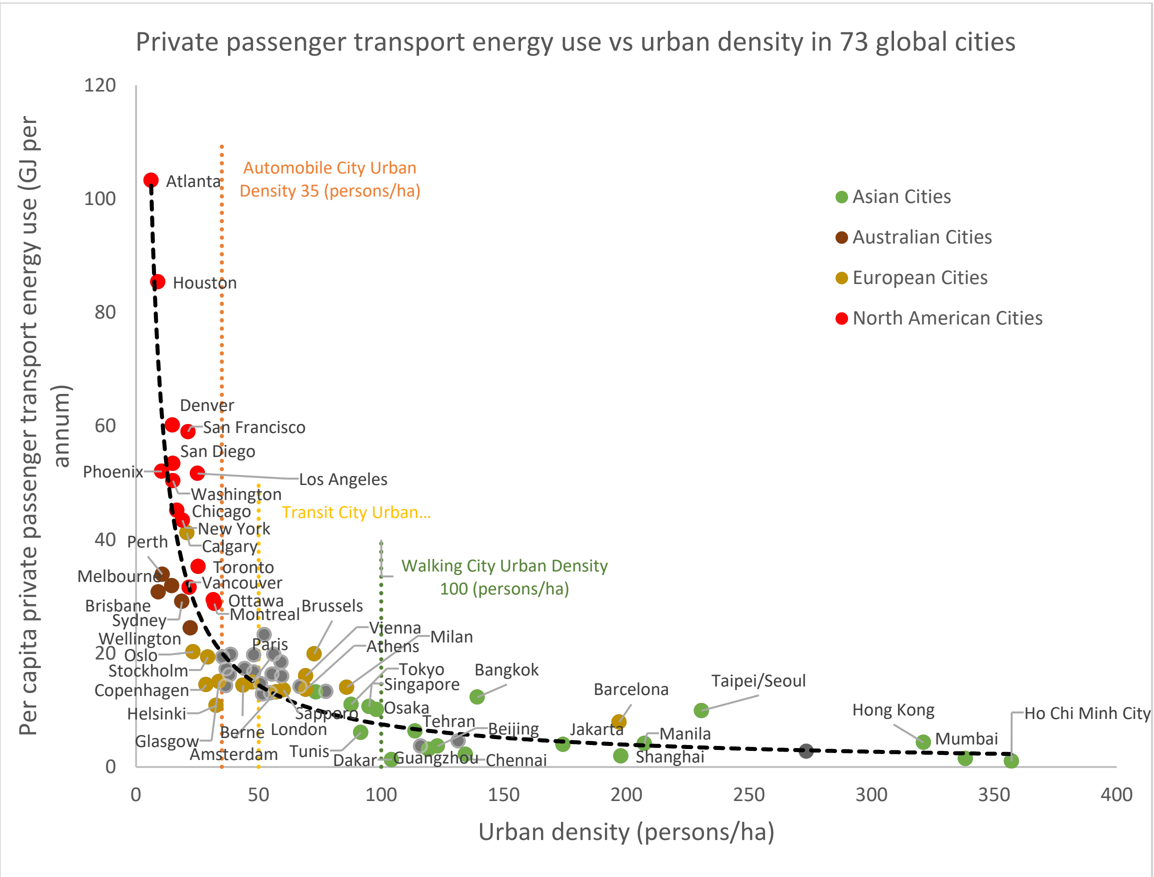

Figure 6. Per capita private passenger transport energy use vs urban density in 73 global cities [34]

Another impactful conclusion that can be driven from the figure is the position of US cities. They are all located on the upper left side of the graph, meaning a low density and high per capita private passenger transport energy use. This is a clear representation of how US urbanization has happened.

Lack of access to amenities by non-car transport modes is a fact in the US. In an effort to parametrize this issue, the American Survey Center conducted a survey inquiring Americans to report if they lived within a walking distance of several amenities [35]. The results were impactful, as long distances and the dependence on the car did indeed have a prevalence. Only 23% of the surveyed people said they were within a walking distance of a grocery store. Similar numbers apply for other amenities: 15% for a fitness center, 18% for a library, 34% for a green space, 14% for the favorite restaurant/bar/coffee shop, and 6% for a place of entertainment. Commuting to work presents some equally striking numbers, with people commuting for about 25 minutes on average per way trip [36].

Having such inaccessibility to everyday needs, leaves no other choice but to commute long distances and to rely on less sustainable mobility options [31]. This is something that has to change. The City Science Group has been advocating for proximity amenities for a long time and the 15-minute city concept has a strong supportive position on this [19].

2.2. Amenities

Amenities foster urban growth [37]. In addition, having the correct amenity-mix can definitely help increase the quality of life given the broad variety of options available within easy reach [18]. It is sensible then, that accessibility to amenities is a requirement when searching for real estate properties. However, it has already been stated that this has not been a priority for urban planners of US cities, in which many amenities have been located in suburban areas leaving residential and commercial areas greatly differentiated [26].

In this context, it is believed that the current trend of amenity location should be challenged. Based on Carlos Moreno's 15-minute city, it will be tried to answer which amenities should be located in the vicinity of people's residences. Within that 15-minute radius, not all amenities should be located at the same proximity, with aspects such as frequency becoming of great importance to help make this decision.

In *A multiscale approach in regional and urban planning strategies*, Claudia Yamu et al. already approach this topic by dividing a series of amenities by frequency of usage. This information can shed a light on which are the amenities that need to be closer localized to a person's residence. The amenities that are used daily should be located in the vicinity of the individual's residence; whereas when the frequency of usage is smaller, these amenities could be further away [36].

Temporal scale	Services and shopping amenities	Leisure amenities	Transport
	Central public administrative amenities	Ski resorts	
Rarely and monthly	Operas, theatres and museums	Forests	Railway
	Hospitals and health care centres	Alluvial forests	
	Local public administrative amenities	Recreational areas	
Weekly	Banks	Swimming pools	Railway, bus, cycling
	Supermarkets	Tennis courts	
	Corner shops	Small playgrounds	
Daily	Bakeries	Neighbourhood parks	Bus, cycling, walking
	Cash machines	Small sports fields	

Figure 7. Typical frequency of usage of amenities and used transport [36]

In addition, information from the American Time Use Survey (ATUS) has been gathered in order to detect which are the activities in which Americans, those of which that actually complete the activity, spend more time commuting. Figure 8 shows that the average working American spends 48 minutes just to commute to work; the consumers of goods and services, as well as the leisure and sports users, need to travel for approximately 40 minutes. But In addition, students spend more than half an hour commuting every day. This survey is actually very representative of how American urbanism has been developed, where long distances have not had importance in the list of priorities [38].

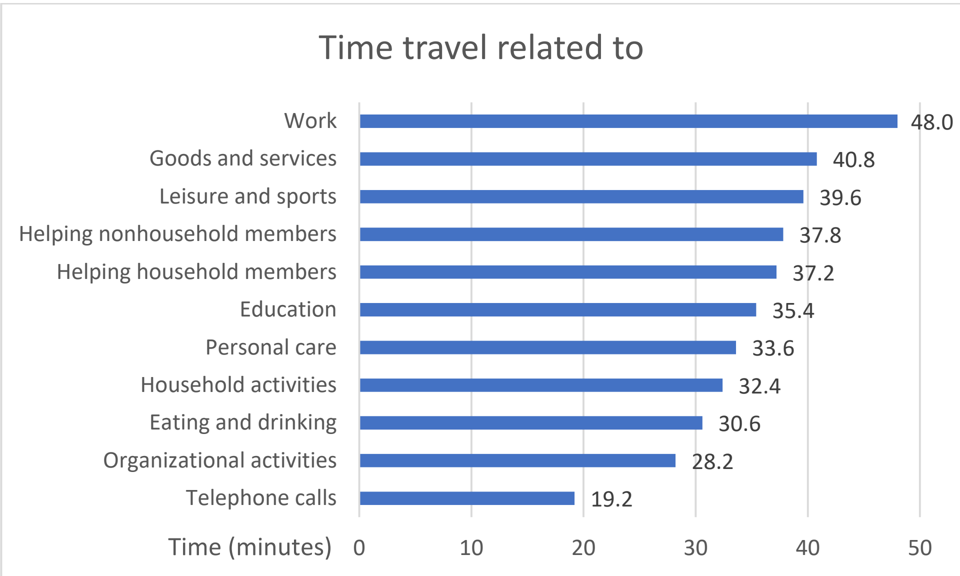

Figure 8. Time travel related to different activities from the people engaging in each activity, weekday average (2019) [38]

Combining the places people visit more frequently with the time it takes to get to those, gives insight on which are the actions that would have the greatest impact on reducing commuting time and hence improving people's quality of life [39].

In this sense, it looks clear that by bringing closer workplaces and frequently used amenities (such as grocery stores, restaurants and sport facilities) to residences commuting time would be highly reduced. This information helps determine the activities that would have the highest impact, although the scope is not just restricted to these.

At the academia, there is enough information and research to do a first iteration of how an ideal 15-minute walkable radius cities should look like [18] [19] [36]. The amenities gathered in Table 1 are the ones identified by the CS group to offer a systemic view of the amenity-mix, while Figure 9 represents the distance at which these amenities should be located from a residence. It is acknowledged that some less important amenities are being omitted.

Distance (min)	Amenity	
5	Restaurants	Shopping centers
5	Leisure and wellness	Parks
10	Night live	Educational
10	Banks	
15	Hotel	Offices
15	Culture	Health care
15	Transportation hubs	Nature
15	Public safety	
15+	Manufacturing	

Table 1. Furthest distance of different amenities from residences in a 15-minute walkable city

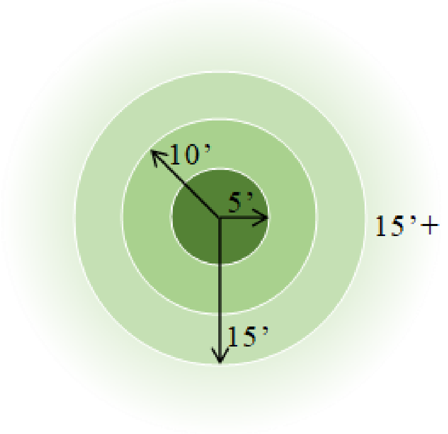

Figure 9, Inner radiuses of a 15-minute walkable city, with the residence at the center

However, amenity facilities might not be optimally located. A deeper research needs to be carried out in order to offer people the closeness to the amenities that they need. This thesis will try to provide an answer to the quantity of amenities to be located in the proximity of people living and working in the district

2.3. Kendall Square

Kendall Square is a region of the City of Cambridge (Figure 10), well known for its innovation potential. It has been referenced by Boston Consulting Group (BCG) as the "most innovative square mile on the planet" given the concentration of brilliant minds and the diverse batch of well-known companies [40]. It is also a place for a great number of startups revolutionizing the tech, biotech, and life sciences landscape. In addition, it comprises some of the buildings of MIT, recently awarded as the best university in the world for the 11th year in a row by QS ranking [41]. All in all, Kendall Square brings together highly-educated people and breakthrough companies.

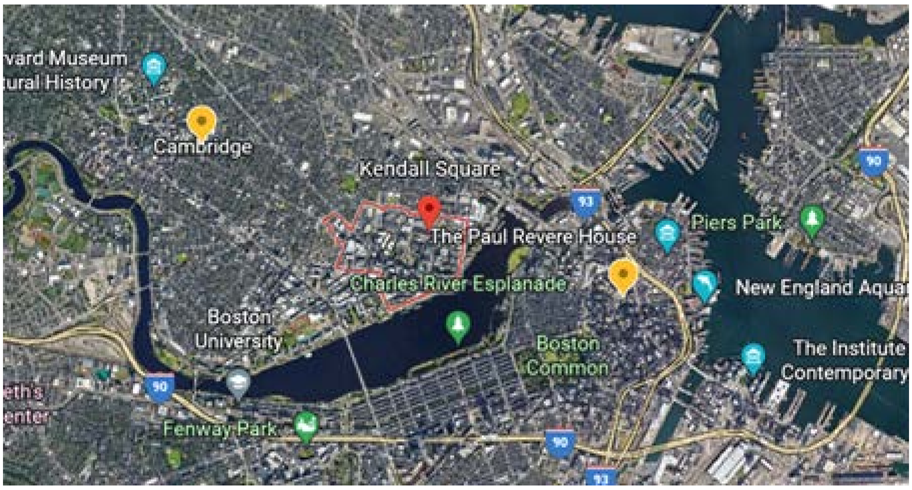

Figure 10. Satellite map of Kendall Square and its surroundings

Kendall Square, though, faces an important challenge. The housing offer in the area is very low, there is just 1837 housing units, without taking dorms into account, for nearly 50,000 workers. This has two clear consequences. People that work in the area need to commute every day with the consequent time loss and the area is not thriving during the evening since there are few residents and also there is a lack of amenities.

Regarding the commute trips, the Figure 11 is indeed eye-opening. It shows the average kilometers travelled by car per person depending on the commute status. Commute_in refers to the people living outside Kendall Square and working in the area, commute_out to the people living in KSQ and working outside, and live_work represents the people living and working in Kendall Square. Commute trips represent trips from home to work and vice-versa while other home-based and work-based, make reference to the ones that have an origin or destination at home/work that are not commute trips, mostly amenities.

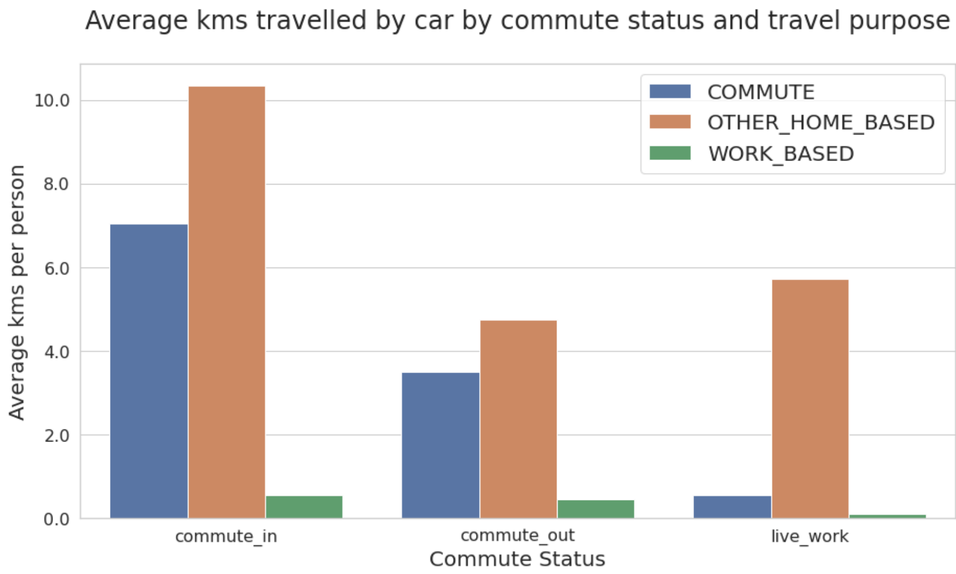

Figure 11. Average kilometers travelled by car by commute status and trip type for Kendall Square

Live-work symmetry has a huge impact on commute trips, with people's trips falling under an average of 1 km. However, other home-based are alarming. It makes sense for these to be shorter than the ones commuting in, since we might be comparing trips of people living in suburban areas (commute_in) with people living in Kendall Square (urban). Still, the average kilometers travelled is too high, confirming the already known insufficient amenity offer in the area which forces people to make long trips to visit amenities.

Observing the impact live-work symmetry has on commute trips, increasing walkable amenities leads to believe that could have a similar impact. Here relies the potential of the work presented in this thesis.

3. GOALS

As explained above, due to current US urbanization, citizens are forced to travel long distances. Upcoming urbanization trends invite to reconsider current urban planning models. Among a broad variety of solutions, proper amenity location could help to ease this problem. It is therefore, the main purpose of this thesis to offer an answer to the following question:

- Where and how many amenities should be located in a district to foster a vibrant, innovative, and walkable community?

Additionally, it is also intended to analyze Kendall Square's current situation and utilize this area as a use case to test our theories, policies and models.

4. STATE OF THE ART

The state of the art is composed of two sections. In the first one, the concept of a 15-minute city is introduced, its principles and implications are studied and some applications are analyzed. In the second part, a literature review on amenities is carried out, in which the main findings of the latest research are exposed.

4.1. 15-minute city

When time is saved, more
time is created

Urbanism development is focused on developing an area granting residents access to the services that cover their daily needs. It has had, though, different approaches. The US approach, has been car-oriented, in which residents have a car-dependence to develop their daily life [42].

However, as it has already been exposed in section 2.1, this way-of doing is very energy intensive and is not environmentally sustainable. Cities should be built for people, either pedestrians or cyclists, rather than cars. Cars have changed the dynamics of urban planning, negatively impacting the biodiversity, quality of life, and distancing even more the citizen from its daily life [18]. In short, urban development needs to be rethought in order to give the people the importance they deserve within a city and learn from the mistakes that have been made.

For instance, Le Corbusier, through its "Unité d'habitation" (Figure 12) tried to propose a solution for the accessibility issue [43]. He developed a concept where in the same building, people could live and access some of their daily needs. However, in the places this was implanted, it was shown how segregation increased [44] and how the reliance on automobiles became even more present. Jane Jacobs, one of the most famous urbanists of the 20th century and proposer of the diversity, density proximity equilibrium triangle, described it as the most dramatic idea, since it favored the isolation of people [45].

Figure 12. Unité d'Habitation, Marseille, Le Corbusier

With the 15-minute city and 20-minute city concept, Carlos Moreno and Kent Larson respectively try to rescue the city development that prevailed centuries ago, in which people had access to everyday needs at a walkable distance [46]. This concept is referred as "chrono-urbanism", since it defends that the improvement of the quality of life is inversely proportional to the time invested in transportation. It is based on the following principles [18]:

- Humans should be at the core of each urban intervention
- Spaces should be multifunctional
- Neighborhoods should be places to live, work and thrive, thus reducing the need for commute

In addition to the principles, this concept defends six essential urban social functions to sustain a decent urban life. Those include [18]:

- Living
- Working
- Commerce
- Healthcare
- Education
- Entertainment

This revolutionary idea aims to question the foundations on which the urbanization concept has been built in the last decades. The human scale has dramatically decreased in the order of prioritization, with car lanes and parking lots occupying a great share of the public space Citizens have interiorized the displacement to the sidewalks, noise and pollutions that cars provoke [18]. This has happened especially in the US, where there are 3.4 parking spaces for every car [47]. The following figure shows Downtown Detroit in 1962, where parking lots and roads took most of the surface. The situation has slightly improved since then, although there is still a high car space prevalence.

Figure 13. Downtown Detroit (1962) showing space devoted to parking lots and streets

At the CS group, we think that the time has come to reclaim that space and displace the car out of the city. To do so, cities should be designed or redesigned so that in the distance of a 15-minute walk or bike ride people can live the essence of what constitutes the urban experience: access to work, housing, food, health, education, culture and leisure. This action, will unequivocally have an impact on people's well-being and the climate's, through a decreasing need for commuting. Also, in the search of pleasant, agile, healthy, flexible cities with economic dynamism and innovation where urban life is vibrant and creative. Four are the drivers on which this idea is sustained [18].

- Ecology: green and sustainable city
- Proximity: live within a reduced distance from other activities
- Solidarity: create links between people
- Participation: involve citizens in the transformation of their neighborhood

Figure 14 is a representation of what is trying to be achieved with the 15-minute city: accessibility to daily needs and engagement within the community, all under the sustainability lens.

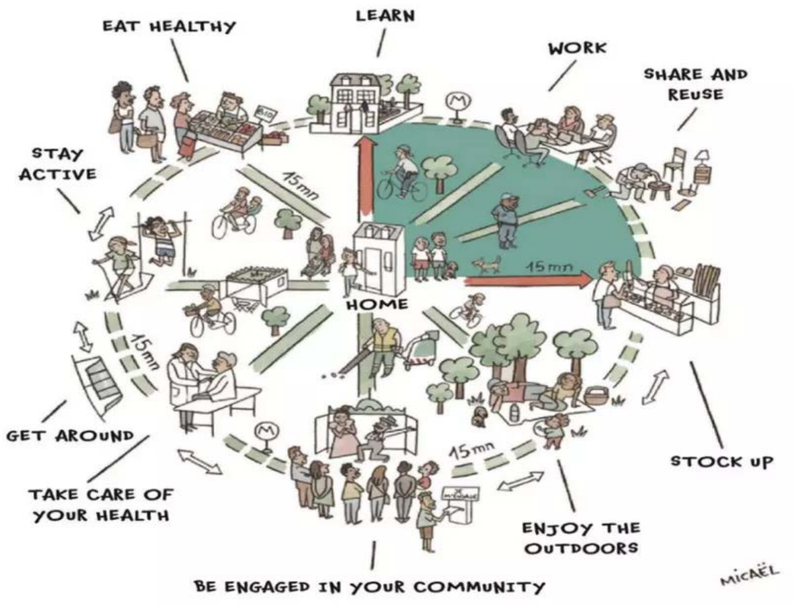

Figure 14. Scheme of the 15-minute city

Diversity is an additional characteristic of this proposal. It has been proven how the interaction of people from different backgrounds help foster innovation [48]. It is therefore sensible that this proposal considers diversity of people within the 15-minute city; additionally, it also intends to enhance the social interactions between these people by spreading a feeling of belonging to the community and promoting engagement.

It is not an easy endeavor to fit the needs from different society profiles in a 15-minute city, since the social needs of each group are different. However, this is what it is trying to be accomplished with the 15-minute walkable city: define the actual combination (or set of combinations) of spaces that would maximize the urban performance of the city, measured by the diversity, density and proximity indicators.

Moreno acknowledges that this is not an easy process and requires not only a change in mentality but also, an implementation of several policies [18] [19]. New services should be developed in districts; bike lanes ought to be increased, to enhance micro mobility, and existing buildings should be transformable, so as to accommodate several uses. In an attempt to increase the usability of space, as a golden rule, Moreno proposes that every building already built should have different usages. It makes little sense that some spaces are empty most of the time. As an example, by order of Anne Hidalgo, schools in Paris ought to open their playgrounds as a public space after school hours. Another impactful example happened at Amazon's headquarters. During the quarantine, Amazon's workers had to stay home and instead of locking down the place, they reconfigured the building so it could serve as a homeless shelter [49].

The accessibility to a wide variety of amenities is not the sole advantage of this theory. As a consequence of planning spaces within the reach of a walk or bike ride, commuting get shorter as well as more sustainable. Not only the distance travelled is decreased, but also alternative mobility choices are presented, which lead to the decrease of the carbon footprint of the commute. Social interactions are also promoted through third places, colloquially known as meeting points [18].

Quality of life could be further improved by focusing on four converging vectors [18]:

- Promotion of social inclusion
- Reinvention of Urban infrastructure
- Building on the digital revolution
- Account for Major environmental issues

The decentralization of services is an additional implication of the 15-minute city. Improving accessibility to services makes it no longer required to displace oneself to acquire basic needs.

Since the popularization of the 15-minute city proposed by Moreno, new concepts have arisen.

Oftentimes, the concept of Smart Cities has been associated with 15-minute cities. The Smart City concept was trying to make technology-enabled-self-consumption and sustainable cities as its first goal. However, this goal has been altered over the time. Places that have adopted this approach, have experienced sustained price spikes, consequently becoming unaffordable for many urban residents. The smart city concept did not meet with its social and economic commitment although it has laid the foundation for future theories to come [18].

Melbourne's government, following the pathway set by Carlos Moreno, are establishing 20-minute neighborhoods [50]. These urban hubs, embrace a set of features that respond to the resident's daily needs. They should, at least, contain local health facilities and services, schools, and shopping centers. In addition, these neighborhoods are planned to be interconnected so as to expand the offer between neighborhoods. The objective pursued through this concept is to obtain more accessible, high-quality places where services that support local living are provided so as to facilitate local thriving and vibrant economies. The points into which this theory is sustained are shown in the Figure 15.

Figure 15. Features of a livable 20-minute neighborhood, Melbourne's government [50]

Project H1 is an urban development in Seoul, South Korea, comprised of more than 500k square meters in which the chrono-urbanist concept is reduced to a 10-minute city. This project aims to transform and industrial site into a green, mixed-use neighborhood where residents can find the amenities that respond to their daily needs [51].

Taken the 15-minute city concept to a next level, Sweden wants now to focus on a hyper-local scale streets. Under "1-min city" name, they plan to redesign every street in the country after having launched the first experiments in Stockholm. However, the scope of this project is not focused on meeting every single need of the residents within a street, since this would be a waste of resources, but, instead, involving the citizen in the decision-making. Some of the first streets to have done so, have included parklets allowing to regain some space to the road and substituting it by public space where people can gather and socialize [52].

Many European cities are taking action related to the presence of the car in the city center. The center of Ljubljana has been car-free for the last 15 years, with the only vehicles permitted being buses, special-purposed electric taxis and bikes. This has had a great impact in the city overall traffic, decreasing by 12% [53] [54]. In addition, in contrast with the initial belief, businesses and tourism have increased. As stated, Paris, Copenhagen, Brussels, Munich, Oslo and even San Sebastian are examples of cities that are implementing public policies to push cars out of the city centers [54] [55] [56].

4.2. Amenities

In the following chapter the impact of amenities in different aspects is exposed

4.2.1. Amenities and walkable districts

The value of a real estate does not only lie on the property itself, but also in its surroundings. There are some elements, what at the CS group we call amenities, that add value to both the area and the property.

- Shopping destinations: people will use them more or less frequently but still want them to be closely located. Places like supermarkets, pharmacies, shopping and convenience stores are highly valued [57].
- Entertainment venues. Even if it is just going out to a restaurant, or a cultural visit, having a closely located diverse offer of entertainment venues definitely increases the value of the property [58].
- Transportation elements: depending on the area an individual lives transportation elements could become greatly important. Elements, such as highway, parking space and/or public transport access are to be considered [59].
- Schools and playgrounds: these are elements that do not only ease the process of raising a child but also define the profile of the people living in the community [60].
- Religious venues, police stations, and sport and health facilities: these complement the elements sought for in the vicinity of a property [61].

Low precipitation frequency, lower presence of manufacturing industries, a high share of coastal shore, low property taxes, better schools, shorter commutes and the like are perceived as attractive attributes for a city to have [62]. In addition it has been proven that the public and private investments in monumental public spaces, parks, street beautification and classical architecture has also a very positive impact in a city attractiveness [62]. In this sense, these findings suggest, all else equal, population and employment growth were 10 percentage points higher in a metro area twice as picturesque. So, it can be inferred that beautiful cities attract highly educated individuals followed by an increase in jobs and fast appreciation of housing prices. It is concluded that the amenities that attract visitors are also responsible of attracting long-term residents [62].

Third places

Amenities are more than just an easy-access facility. They are a meeting point for people from diverse background to exchange ideas, experiences and innovation to surge. Ray Oldenburg, in his book "Celebrating The Third Place", introduced the concept of third places which refers to those public places on neutral ground where people can meet and interact with diverse people. These third places "host the regular, voluntary, informal, and happily anticipated gatherings of individuals beyond the realms of home (1st places) and work (2nd places)" [63].

These third places provide opportunities to socialize and connect with people and are the cornerstone to build communities. In 2021, 44% of Americans, 38% in urban areas, reported not having a Third Place to which attend, which raises the question on whether a different urban design approach could help provoke more interaction between people. Moreover, having local neighborhood spots to attend to, stimulates the Cheers Effect, referred to the benefit of being around people you know. This can also have a community building effect, since it has been proven that local amenities boost sentiments of solidarity, trust and belonging within a neighborhood [64].

Public spaces such as parks, libraries and community centers are important places of urban design; however, cafés, bars, and restaurants may serve as well as effective community hubs, and should not be, therefore, underestimated [63].

4.2.2. Relation between amenities

Complex economic activities tend to concentrate in larger cities, mainly due to being required a deeper multisector knowledge and the coordination of a series of actors. Not even the spread of internet has been able to diminish the importance of this concentration of knowledge [24].

The principle of relatedness states that the probability of an economic activity to enter or exit a region is related to the economic activities already present in that location. In other words, the existence of amenities attracts other amenities, either competitor or complementary [65].

Trying to get an answer to the number of amenities needed in a city, in the Amenity Mix of Urban Neighborhoods, César A. Hidalgo et al., do a recompilation of the existing main amenities and the number of points of each in 47 US cities based on Google Maps API. Google Maps counts with the Google Places API product which helps obtain the number of points in a region by doing API calls. Each point has a category associated to it (e.g. when looking for the category Restaurants in a region, Google Maps will filter all those points that meet that category). The authors obtained all those points by category for a list of US cities. The authors acknowledge that there might be some bias in this data, however they consider it to be an attractive source from which insights can be obtained. From this data, an average per 1000 inhabitants is obtained (Table 2). This approach is mainly focused on large urban areas, discarding, therefore, rural areas in which density, diversity and proximity aspects have different conceptions. This research sets the baseline for a further analysis carried out in the present document [65].

Amenity	Points	Points per 1000	Amenity	Points	Points per 1000
Doctor	153772	2.52	University	6597	0.108
Restaurant	112430	1.84	Gym	5934	0.097
Bus station	110642	1.81	Storage	5849	0.096
Construction contractor	86044	1.41	Night Club	5675	0.093
Religious Centers	58468	0.96	Parking	5527	0.091
School	46516	0.76	Veterinary Care	5373	0.088
Beauty salon	41851	0.69	Art gallery	5358	0.088
Car repair	40215	0.66	Florist	5102	0.084
Real Estate Agency	39484	0.65	Hardware store	4595	0.075
Lawyer	37611	0.62	Department store	3515	0.058
Finance	32221	0.53	Library	3466	0.057
ATM	30753	0.5	Book store	3417	0.056
Clothing store	29806	0.49	Car wash	3202	0.052
Home goods store	29537	0.48	Car rental	2968	0.049

Insurance agency	27866	0.46	Spa	2843	0.047
Dentist	26071	0.43	Funeral home	2761	0.045
Park	25723	0.42	Post Office	2723	0.045
Bar	21506	0.35	Gas station	2552	0.042
Accounting	17280	0.28	Cemetery	2386	0.039
Grocery or supermarket	15206	0.25	Pet store	2270	0.037
Pharmacy	15204	0.25	Locksmith	2182	0.036
Laundry	14391	0.24	Museum	2161	0.035
Convenience Store	13818	0.23	Fire station	2050	0.034
Moving Company	12744	0.21	Police	1613	0.026
Furniture store	12379	0.2	Airport	1535	0.025
Electronics store	11876	0.19	Bicycle store	1409	0.023
Car dealer	11603	0.19	Train	1262	0.021
Hotel and lodging	11452	0.19	Stadium	1245	0.02
Local Government Office	10081	0.17	Movie Theater	1232	0.02
Cafe	9485	0.16	Amusement park	1017	0.017
Bakery	9255	0.15	Courthouse	717	0.012
Shoe store	8612	0.14	Embassy	688	0.011
Liquor store	7948	0.13	Aquarium	492	0.008
Hospital	7942	0.13	Bowling alley	366	0.006
Physiotherapist	7929	0.13	Casino	172	0.003
Travel Agency	7394	0.12	City hall	140	0.002
Jewelry store	6751	0.11	Zoo	114	0.002

Table 2. Total number of each type of amenity in the Google Places data set and the points per 1000 residents in 47 US cities

4.2.3. Amenities and mobility

It has been proven a strong correlation between local amenity accessibility and mobility patterns. Local amenities promote walking and biking attitudes with grocery stores and preschools being critical in triggering change [66]. Næss et al. [67] proved that self-sufficient neighborhoods, in terms of the supply of unspecialized amenities, likely exhibit more intra-neighborhood-oriented travel patterns and lower car use compared with areas with poor local accessibility. Additionally, 150 accessible amenities is defined as the threshold at which people walking and biking outnumber car drivers. Above this reference, as number of amenities increases, although a positive mobility effect, the marginal contribution is smaller [66].

Nonetheless, assessing which amenity has the greatest positive impact is, in fact, challenging. Evidence suggests that, when it comes to urban amenities, a well-functioning package of amenities may have more impact that the individual sum of its parts. Amenities help reinforce the effect of the individual [66].

Income segregation has more to do with the mobility patterns of individuals rather than with the census area they live in. Since most neighborhoods are home to diverse inhabitants regarding income, the census area is not a good predictor of its income segregation profile. However, mobility patterns might be completely different among people from different levels of income. Most encounters in the city happen far from people's neighborhoods and it is in these other places that segregation is experienced [68].

The atlas of inequality assesses the socioeconomic profile of people visiting some amenities, proving that inequality is highly associated with the visiting patterns of individuals (Figure 16). In this sense, they have studied whether an amenity is equal based on the diversity of different income-profile people visits. And they have done so for several US cities. The map for the Boston Area looks as follows, where blue dots represent equal places, with people from different backgrounds visiting the place, while red dots represent just the opposite.

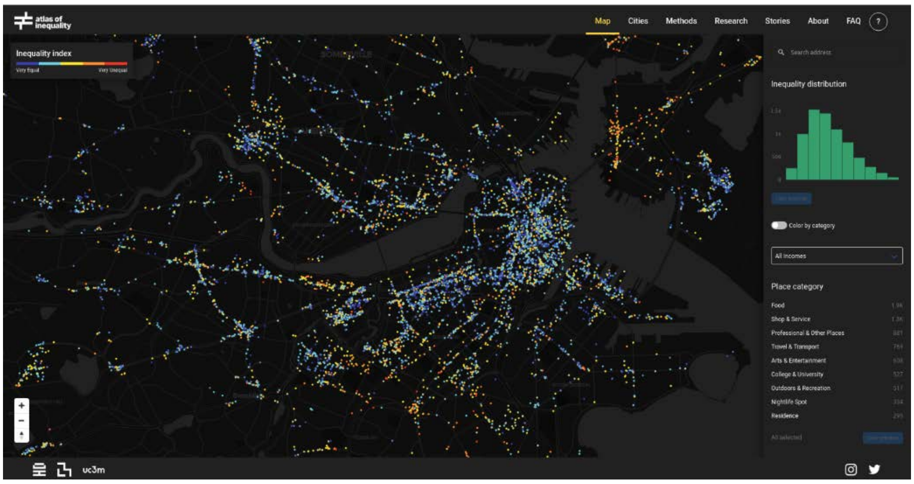

Figure 16. Atlas of inequality of the city of Boston and its Surroundings [69]

All in all, the intention of a public policy should be to fill the map of blue dots, as it has been proven that the positive aspects of the social interaction of people from different economic backgrounds [70].

High residential density, a high degree of spatial diversity and a high level of job accessibility are connected to short commutes by residents in the area. On the other hand, a high job density is related to long commute by employees. In Europe prevails the sustainable urban development model, where cities are tailor-made for pedestrians and cyclists. In the same sense, distance travelled for commute can be used as a proxy for sustainability of the commute [71].

Density, diversity and proximity, first introduced by Jane Jacobs [45], are important spatial factors that affect the travel behavior of individuals. At the same time, denser cities, with more inhabitants per square kilometer, have been proven to consume less fuel per capita for transportation needs [34]. In other words, there seems to be a link between higher density residential places and shorter commute times. On the other hand, a higher density of jobs over housing, a concentration of work in an area, seems to be linked with higher commuting times. A conclusion can be drawn that the smallest commuting time occurs when the housing-jobs balance is close to 1, when there is indeed a live-work symmetry [71].

4.2.4. Amenities and economy

The Atlas of Opportunity is a tool developed by the MIT Media Lab Human Dynamics group alongside business and government stakeholders in Adelaide, Australia [32]. It intends to allow communities to extract insights from open government and commercial data, so as to conduct data-driven strategies based on the understanding of how human behavior shapes economic prosperity. The data is represented through an interactive map divided into SA2s (Statistical Areas level 2), and includes sociodemographic, economic, spending, and mobility data.

Additionally, they have developed some useful indicators, such as the neighborhood attractiveness measure. This indicator relies on the diversity of amenities inside the bounds of a neighborhood to predict both the diversity and volume of the human flow into this same neighborhood. This prediction can be performed given the existing high correlation between number of unique amenity categories and the total volume of flow of people into the neighborhood. This, consecutively, can be used to predict economic productivity and growth.

There is a dynamic relationship between the neighborhood attractiveness and its economic growth, which can be categorized as a virtuous circle. The higher the attractiveness of a neighborhood, the larger the inflow of visitors which in turn encourages entrepreneurs to offer a yet more diverse amenity-mix.

The authors of the Atlas have concluded that the attractiveness of local amenities and services, provokes a more diverse movement patterns, which consequently strongly favors economic growth in the near-future [32].

To a similar conclusion reached other MIT researchers that tried to predict economic outcomes based on diversity of amenities in city neighborhoods. It can be accurately predicted mobility patterns based on this amenity diversity, which, in turn, can help predict economic outcomes. In conclusion, an insightful outcome suggesting that the "diversity of goods and services within a city neighborhood is the largest single factor driving both human mobility and economic growth" is reached [30].

Moreover, an interesting insight was also observed by Glaeser et al. [72], detailing that cities with a larger goods and services offer exhibit a higher growth potential, with a rising demand for living.

4.2.5. Amenities and Sustainability

The more our world functions like the natural world, the more likely we are to endure on this home that is ours, but not ours alone
Janine Benyus

"Meeting the needs of the present without compromising the ability of future generations to meet their own needs", is the very own definition of sustainability. If cities want to be more sustainable, they need to favor local walkable amenities which can lead to environmental savings from two sources:

On one hand, it has already been stated that walkable amenities reduce commute time [18], which is oftentimes used as a proxy to assess the sustainability of the trip [71]. Additionally, shorter commute times do favor the adoption of more environmentally friendly mobility choices [31].

On the other hand, additional savings can be obtained from shifting from large commercial surfaces (e.g. Malls) to local stores. The savings, once again, would come from two sources. Local stores are more energetically efficient on a per square foot basis [73], and local stores can serve the same demand with smaller surfaces, since they do not need large common space (e.g. hallways in Malls)

Walkable amenities are just one leg of the sustainability aspect. Well-designed urban spaces have also the power to influence behavioral changes. Urban spaces can not only have an impact on social interaction but also in helping people become healthier by creating more sustainable communities. It is proven that environmental design has a strong influence on human behavior and attitude, so reshaping and advancing urban spaces could have a positive impact on influencing people on carrying a healthier everyday routine [74].

Well-designed environments can strongly influence what people think and do, given the continued dynamic interaction between a person, a particular behavior and its environment. In that sense, the architecture and design of urban places can definitely have an impact on human behavior. Unlike monetary incentives, that just produce a behavioral change while the incentive is provided, persuasive systems actually encourage a shift in attitude and behavior [75].

But, in order to be sustainable, cities must be resilient. The latter acquires high importance in order not to fall again in past vices but instead having the abilities to withstand natural disasters, economic recessions or political malfeasance [76].

4.2.6. Amenities and urban evolution

The Alonso-Muth model states that, if inhabitants only differ by their wealth, at equilibrium, central locations are occupied by poor households and the suburbs by middle class and rich households [77].

Two main factors are considered in the analysis of the dynamic of the city structure. When transportations costs are very high and/ or the attractiveness for historical amenities is high, rich households will tend to locate in the city center, as it occurs in many European cities; whereas when transportations costs are low and the necessary amenities are generated in the periphery, rich households will decide to locate in the periphery, as seen in American and South-American cities. In addition, once rich households have chosen its preferred location, the required amenities are generated in the area, resulting in a spatial "lock-in" effect, by which the rich go on occupying the same areas occupied by their predecessors. This type of evolution does not foster the sought diversity in which innovation is fostered. On the contrary, it favors income segregation [77].

Urban evolution implies growth of metro area sizes, mainly due to technological change favored by local human capital accumulation. Sites with competitive advantages, such as better geography or market potential, are more prone to grow faster than those sites that lack of these advantages [78].

4.3. Expansion of knowledge

This thesis is intended to expand the knowledge in this area. It has been stated that amenities have the power to influence on people's mobility behaviors [66] [67], define a region's economic growth

[30] [32] [72] and also make it more vibrant and livable [18]. However, little is known on what amenity-mix maximizes these features.

Having the work of Cesar Hidalgo et al. as a starting point [65], this thesis intends not only to extract the number of amenities needed in an area but also get a sense of how these needs change over different age and income profiles. With this information at hand, many are the questions that can be addressed. Newly developed areas could have this tool as a reference for the number of amenities needed in the area, current regions could be studied to analyze what their strengths and weaknesses are, and policies to foster these amenities could be pursued. Additionally, a Point of Interest (POI) based analysis could be carried out regarding age-income profiles of visitors, similar to the one performed in the Atlas of Inequality [69].

5. METHODOLOGY

In the following chapter, the methodology followed during this research is explained, as well as the data sources used to accomplish the studies.

5.1. Data sources

For the purpose of this project, three are the database sources that have been used:

- US census
- Safegraph
- Replica

5.1.1. US Census

US Census is an open-source data provider managed by the US government. They provide all types of census data with a block group capillarity. Information regarding population characteristics has been gathered from this source, in addition to shapefiles, a data format for Geographic Information System analysis (GIS), of selected areas.

5.1.2. Safegraph

Safegraph is a Point of Interest (POI) data provider. Safegraph gathers the information by tracking personal devices when using some partner apps. They count with three distinct products:

- Places: basic information, such as name, category and location (Table 3)
- Patterns: aggregated visitation patterns of people to each POI by block group (Table 4)
- Geometry: shape data of each POI (.shp file)

placekey	location_name	top_category	naics_code	latitude	longitude
222-222@62j-sgb-3qz	Apartments at 33 Forbes St	Lessors of Real Estate	531110	42.321	-71.107
222-22m@62j-shy-zs5	Appworks Technologies	Wired and Wireless Telco Carriers	517312	42.35	-71.057
225-222@62j-sgg-ckf	House of Blues Boston	Restaurants and Other Eating Places	722511	42.347	-71.095

Table 3. Safegraph's Places database sample with selected columns

placekey	raw_visit_counts	visits_by_day	visits_by_each_hour	visitor_home_cbgs	visitor_daytime_cbgs	normalized_visits_by_state_scaling
222-222@62j-sgh-9j9	401	[71,75,70,74,69,23,19]	[0,0 ... 4,1,1 ... 0,0]	{"250173662021":6, ... }	{"250250103002":9, ... }	7188
223-223@62j-shy-p9z	7	[1,1,0,2,3,0,0]	[0,0 ... 0,0,1 ... 0,0]	{"250250904002":4}	{"250251703001":4}	125
227-222@62j-sj7-2tv	84	[8,9,9,12,13,20,13]	[0,0 ... 2,1,1 ... 0,3]	{"250251606011":5, ... }	{"250173501042":7... }	1506

Table 4. Safegraph's Patterns database sample with selected columns

Additionally, they do also offer Spend data, which offers valuable insight into the money spent in each POI, although this information was not accessible by the data plan.

Safegraph's databases have some advantages and downsides. On the positive side, the collected data is real, although anonymized, and it offers a good POI breakdown, which helps make informed decisions regarding specific POIs. However, they are somewhat limited. Safegraph does not track a person throughout the whole day, making the information incomplete. As a consequence, the total visits a person makes to each amenity nor the total visits an amenity receives cannot be calculated. However, this undercounting is fair to be assumed to be equal. In this sense, the number of amenities needed to meet the demand can be calculated. Formulas (1)(2)(3)(4) help understand this rationale.

$$Visits_{person} = C_1 \cdot visits_{registered_{pp}} \tag{1}$$

$Visits_{person}$: *total visits of a person in a month*
C_1: *constant to be applied to the registered visits per person to obtain the total visits per person*
$visits_{registered_{pp}}$: *Safegraph registered vists per person*

$$Visits_{POI} = C_2 \cdot visits_{registered_{POI}} \tag{2}$$

$Visits_{POI}$: *total visits to a POI in a month*
C_2: *constant to be applied to the registered visits to a POI to obtain the total visits to a POI*
$visits_{registered_{POI}}$: *Safegraph registered vists to a POI*

$$Assumption: C_1 = C_2 \tag{3}$$

$$N_{POIs} = \frac{Visits_{POI}}{Visits_{person}} \cdot population = \frac{visits_{registered_{pp}}}{visits_{registered_{POI}}} \cdot population \tag{4}$$

N_{POIs}: *Number of POIs needed to meet the visiting demand of a population*

Additionally, Safegraph's tracking is not even across the whole country. Some regions are undercounted while others have a high share of their population tracked. All in all, they track an average of nearly 5% of the US population. Even though the data is normalized, as it will be explained later on, the level of accuracy that can be reached differs with varying tracking percentages. Another limitation to be aware of is the misassignment of a visits to a block group. It has been seen how this linkage was not entirely precise on some occasions, which could lead to partially false conclusions.

5.1.3. Replica

Replica is a data platform for the built environment. Their data products include mobility, demographics, economic activity, and land use. For this project, just the former two have been used. Replica's databases are synthetic. In other words, the data they commercialize is simulated although it is based on real data. However, this is considered highly beneficial. The algorithms they use, make the data high-fidelity and they preserve the privacy of the citizens, which is a top priority for the group.

Replica's mobility database (Table 5) includes an entry for each trip performed inside a predefined region. The information included for each trip is the following: person id, mode, purpose, origin, destination, distance, duration, and complete routing information. Regarding the demographics database (Table 6), each row represents a person. For each person, the following attributes are included: age, race, ethnicity, income, employment status, and household size.

activity_ id	person_ id	distance_m	duration_sec	mode	travel_ purpose	origin_ bgrp	destination_ bgrp
1489...	1642...	9012	480	ON_DEMAND_AUTO	WORK	250173393002	250173526001
5250...	1353...	56810	2940	CARPOOL	WORK	250173117002	250173526001
6206...	9262...	19151	1260	PRIVATE_ AUTO	WORK	250173564008	250173526001

Table 5. Replica's Trip database sample with selected columns

person_id	age	employment	race	bgrp_home	bgrp_work	hh_id	hh_income	hh_ size
1642...	36	employed	white	250251403006	250173528002	6443...	93225	1
1353...	40	employed	white	250173399004	250173399004	7415...	96360	1
9262...	39	employed	white	250173393001	250173684001	1700...	133623	5

Table 6. Replica's Demographic database sample with selected columns

Replica's data sets are very useful to analyze the trip patterns and demographic profiles although they have some limitations. On one hand, they just offer trip data for one typical weekday and another weekend day. Since individuals' trips are highly variable (one day is not representative of what a person does in a whole week/month/year), it is difficult to extrapolate the data for a representative time period. At the same time, another limitation is the travel purpose Replica offers. The travel purposes are categorized as follows: Home, Shop, Social, Eat, Work, Errands, Commercial, Recreation, and Other. Replica data has very broad travel purpose definitions and does not offer a POI breakdown which would be extremely helpful. An additional factor to take into account is the usage of different databases to build their model. Since they might show slightly different numbers, it is important to be aware of this not to drive false conclusions.

Features	Safegraph	Replica
POI breakdown	🟢	🔴
Demographic profiling	🟡	🟢
Trip data	🟡	🟢
Individual tracking	🟡	🟢
Anonymized data	🟡	🟢

Table 7. Pros and cons of Safegraph and Replica databases

5.2. Python libraries

To perform the analysis, many previously developed python libraries have been used that have eased the path.

- Pandas: used to analyze data in a structured way
- Geopandas: used to analyze geospatial data
- Pandana: library to perform network analysis

- Numpy: useful to work with arrays
- Matplotlib: for graph visualization purposes
- Seaborn: for graph visualization purposes
- Folium: used to create maps
- Scipy: it has optimization algorithms
- Shapely: to work with geometries
- Skicit: for clustering analysis purposes

5.3. Evolution of analysis

The analysis performed in this project has always had the purpose of answering which amenities, how many of them, and where these should be located. The scope to get to that result has varied along the project, as outcomes of the analysis were made available and more data was obtained.

First of all, it was deemed necessary to parametrize the area. With Kendall Square and Volpe as base cases, how these areas are performing at the moment was obtained making use of the three abovementioned data sources. Aspects such as the number of workers, residents, housing stock, mobility patterns, and so forth were calculated in order to have a current photograph of the area. This analysis brings insight on the strengths and weaknesses of the area and helps further evaluate policies to be carried out in place. Additionally, Volpe Redevelopment Plan has been studied.

Having this view, it was later proceeded to evaluate the amenity offer of the area. Lacking of a good reference to which compare the outcome obtained, having benchmarks was made very much needed. Taking vibrant and livable business districts as references, it was gathered a panel of experts to decide on which were the most vivid examples of these districts that could be used to compare what is trying to be achieved in Volpe.

The analysis, then, shifted towards another approach. Leaving Volpe-like areas behind, it was proceeded to think on the people that are going to fill the streets of Volpe. To do so, it was first analyzed the residents of the state of Massachusetts. Focusing just on the urban population and on those block groups that had a minimum tracking by Safegraph, it was intended to predict the visiting patterns to different amenities based on the block group age and income profile.

At this point, Safegraph's databases had already been squeezed enough and no additional valuable information could arguably be obtained. The path was paved for Replica. Even if the Replica's databases' limitations were acknowledged, the day-specific-exhaustive data recollection was considered to have insightful potential outcomes.

To continue with, a demographic clustering analysis was performed aiming to obtain differentiated visiting patterns to amenities depending on the person-specific characteristics.

All of this led to the final tool, where the best of both worlds was tried to bring together by associating Replica and Safegraph's databases. Seizing Replica's good trip data breakdown, it was proceeded to assign a Safegraph POI to each trip. To do so, it was required a hand-to-hand collaboration with Replica data scientists; after which they improved the quality of their data by delivering latitude and longitude coordinates for each trip in addition to a NAICS code of the destination. With this newly developed data, synthetic Safegraph-like databases were built on top of Replica's databases, with the additional advantage that each trip had a binding person profile rather than a whole block group.

Lastly, as part of a lab-wide project, quantification of live-work symmetry and walkable amenities and its implication on mobility was performed. This way, estimates regarding the time saved and number of kilometers travelled reduced were given, which could later be traduced into a reduction of energy used and the consequent reduction of CO_2 emissions.

6. RESULTS

In the following section, chapter by chapter, the results obtained with each approach are detailed.

6.1. Current situation of Kendall Square and Volpe

Assessing the current situation of Kendall Square and how Volpe's Redevelopment Plan is framed is key to be aware of the weaknesses and strengths of the region. As a consequence, solutions that can lead to Kendall Square becoming a walkable and vibrant district can be firmly proposed.

6.1.1. Kendall Square

Kendall Square is one of the regions where the City Science group wants to influence and propose improvements so as to achieve the abovementioned 15-minute city, applied to a zone. In this sense, the first step is to parametrize the area so as to know the actual situation. Using the geopandas library from Python, a boundary has been set that englobes not only Kendall Square but also MIT block groups (Figure 17).

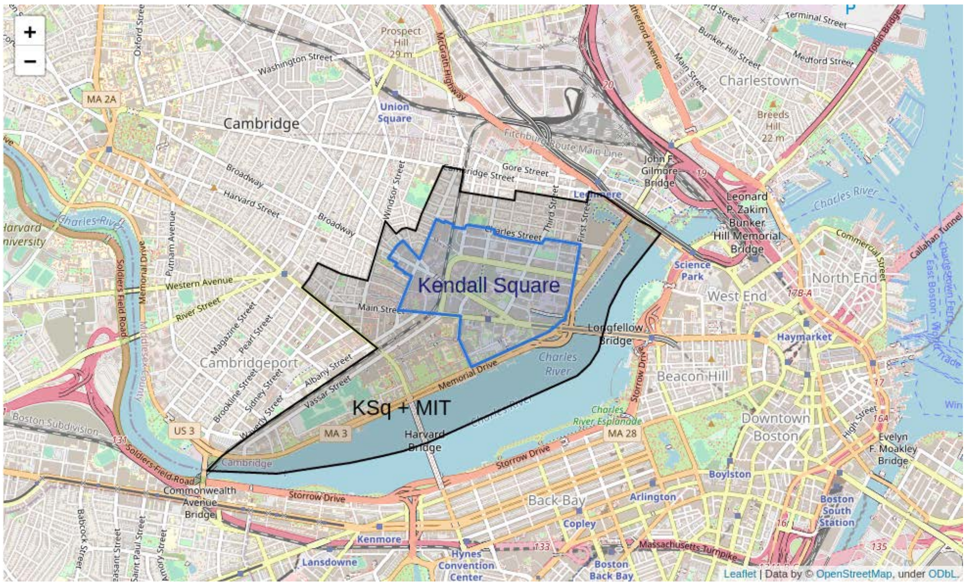

Figure 17. Kendall Square (blue) and Kendall Square + MIT (gray) regions

<u>Process</u>

Using Replica's databases, with its limitations in mind, firstly, the number of workers that work in this boundary has been calculated. At this time, it is encountered a problem, since the Kendall Square political boundary does not coincide with the boundaries of the block groups. In that sense, although there is a specific block group that belongs halfway to Kendall Square, it has been included. The number of workers in the Kendall Square area totals to 49800 with data from the second quarter from 2021. If compared with 2019's data the numbers differ by just a 0.3%, which can lead to believe that the COVID-19 global pandemic has not had a huge effect in Kendall's working population. Additionally, it has been calculated the KSQ + MIT population. The numbers do not differ much, since the block group mentioned before contained most of MIT workers. In that sense, the working population of this area accounts to 51525, 200 workers less than 2019's numbers.

Following that, the housing stock has been studied. With an objective of a live-work symmetry close to 1 for every area, it is studied if there is enough house offer in a 15-minute radius from the working area. Bearing in mind that people are encouraged to walk when the trips are smaller than 15 minutes [6], it is wanted that all the people that have to commute to KS + MIT every day, live within a walking distance. In that sense, an additional boundary is calculated 900 meters away from the KS + MIT area. It is acknowledged that a 15-minute walking city might accept a larger walking distance, although it should also be acknowledged that there is some distance to be traveled inside the predefined working area.

Later on, it is proceeded to overlap the boundary layer with the Block Groups layer in order to obtain which Block Groups correspond in that area. Some block groups might fully belong within the boundary, while other might just do so partially. In those cases, a threshold of the area of the block group belonging inside the boundary has been set on 0.5. Additionally, some of the block groups have been discarded, since these have fallen inside the buffer boundary due to the closeness to the sporting area of MIT, but not really with the working nuclei. All in all, the new map looks like this (Figure 18):

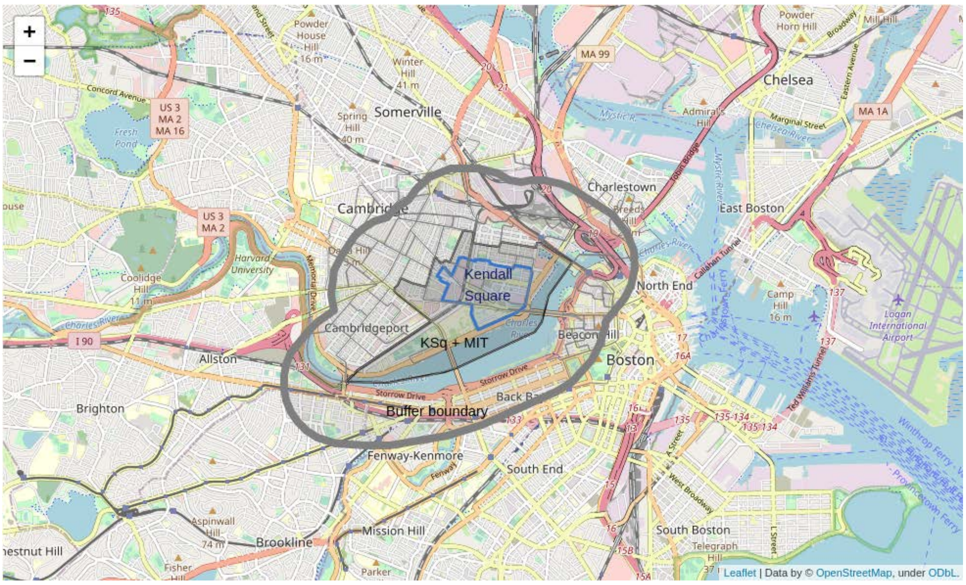

Figure 18. Kendall Square + MIT buffer boundary and its belonging block groups

<u>Results</u>

Having selected the Block Groups that should offer house options to the people living in KS + MIT, it is then calculated the housing and bedroom units' distribution. To do so, open access census data has been leveraged.

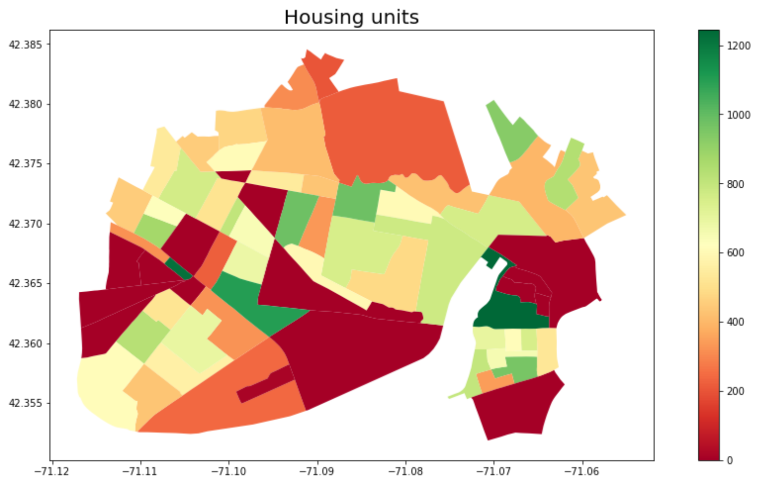

Figure 19. Housing units inside KSQ + MIT buffer boundary

In Figure 19 a heat map with the housing units per block group inside the predefined buffer boundary is presented. There are 32,477 housing units in the selected area. Some Block groups have little to no housing offer. These block groups can either be university buildings with no residence (lower side), hospitals (Massachusetts General Hospital on the lower right side), or commercial venues.

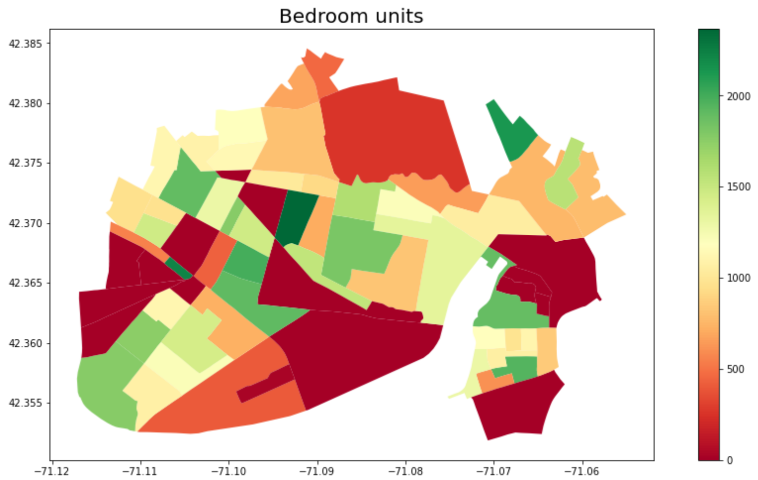

Figure 20. Bedrooms units inside KSQ + MIT buffer boundary

Same as before, in Figure 20 the bedroom units inside the buffer boundary are plotted. The available bedroom units in the selected area account to more than 64K, averaging nearly 2 rooms per house. It is believed that the actual housing unit offer is insufficient to allocate all Kendall Square and MIT workers, plus their families, in addition to complementary services and students.

It has also been calculated the Ownership Rate of each Block Group (Figure 21) since this metric gives insight of the accessibility of newcomers, which is the case for many workers in KS.

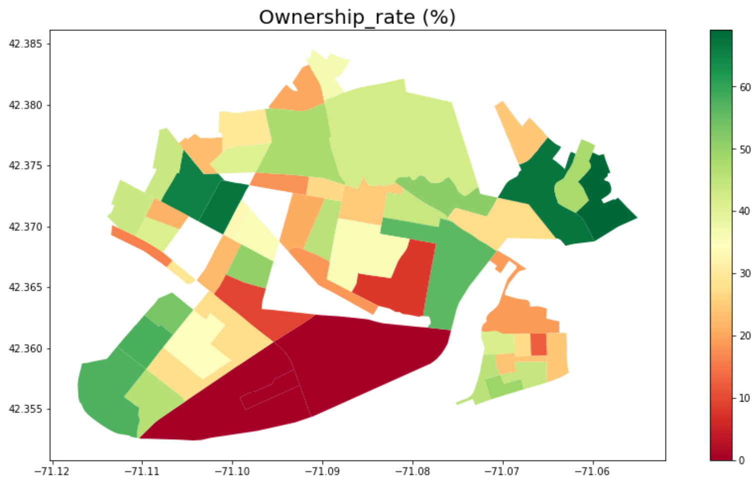

Figure 21. Ownership rate inside KSQ + MIT buffer boundary

Analysis of results

The Kendall Square and MIT community has a great rotation. Even though there are tenured professionals that have been working here for a long time, the working mix is greatly composed of nomads, for which renting options are more valued. In that sense, block groups with high ownership rates are out of the equation for these people.

Beacon Hill, the neighborhood in Boston across the bridge, offers a good amount of bedroom units with a low ownership rate, which makes an appropriate place for renting. Cambridge has a mixture of combinations. Cambridge port counts with a great number of bedroom units, although the ownership rate is very high; Kendall Square has a low housing offer; and in Central Square and its north east the amount of housing available is considerate, with varying ownership rates.

In conclusion, it is believed that the current housing offer is very poor and insufficient to allocate all the workers. It needs to be acknowledged that inside the 51525 workers, its families, students, retirees plus complementary services are not being taken into account, which would highly raise the number of people need to be accommodated.

Additionally, if a solution to this groups is to be proposed a profiling of the population is needed. With that in mind, the population has been divided by age (Figure 22), household income (Figure 23) and family status (Figure 24). The representation is done through a violin plot, useful to represent the distribution of the population. The dotted lines are the 25th and 75th percentiles, while the dashed line is the median.

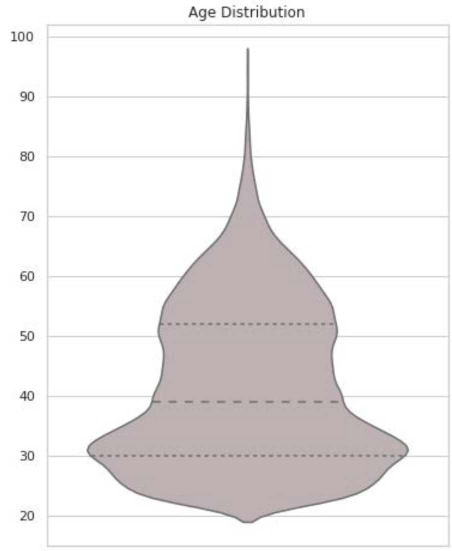

Figure 22. Age distribution for the Kendall Square + MIT buffer boundary

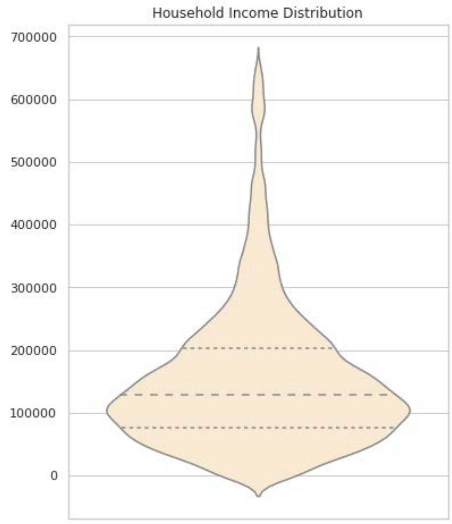

Figure 23. Household income distribution for the Kendall Square + MIT buffer boundary

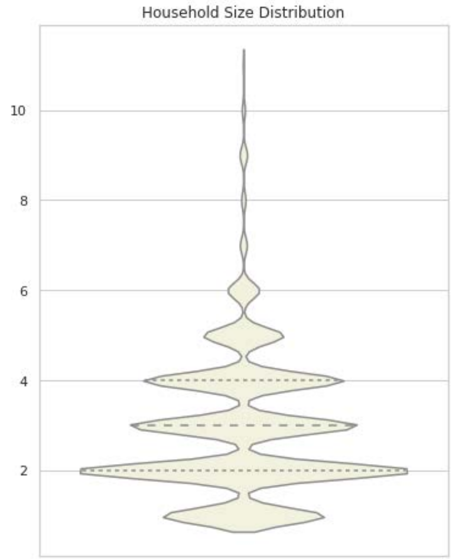

Figure 24. Household Size distribution for the Kendall Square + MIT buffer boundary

It can be concluded, that the area's resident population is highly concentrated around 30 years of age, a household income close to $100,000 that usually share the house with someone else.

6.1.2. Volpe

Given the mentioned problem, MIT is developing a redevelopment program to increase the housing offer for the research and faculty staff. The place selected to do so is Volpe (Figure 25), the last area in Kendall Square that can be built from scratch and help ease the burden of lack of housing. In total, MIT plans to include 10,000 workers along with their families in this area. With a total area of 14 acres, it is intended an approximate 25,000 people/km2 density, similar to that found in Manhattan [79].

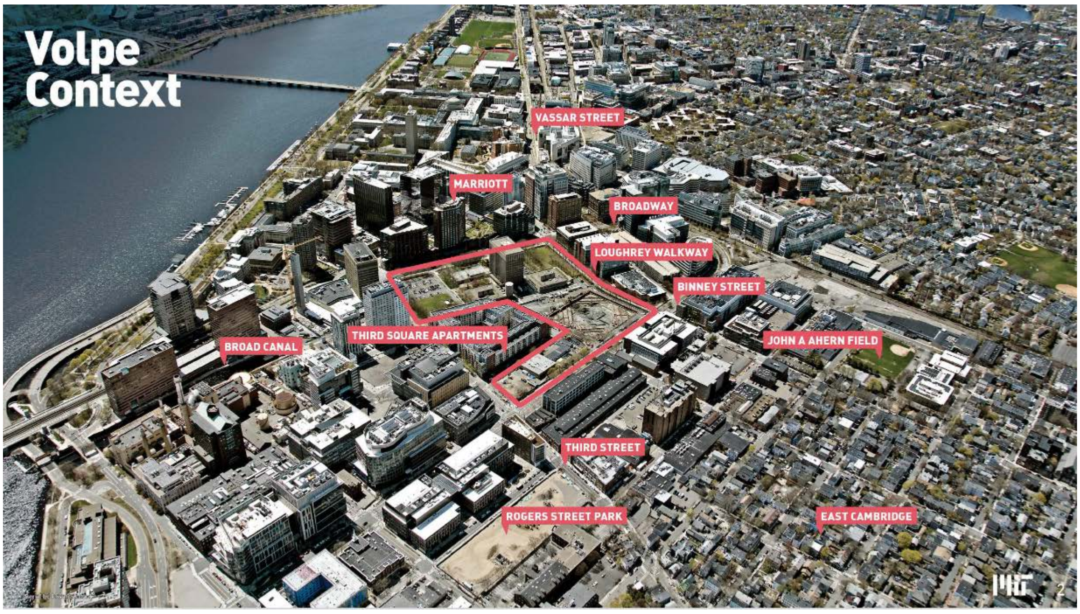

Figure 25. Volpe development area (red square) and the elements of its surroundings

Process

To get to know the housing needs of the total group, it has been proceeded to contact MIT staff regarding statistics of single, married and couples with children. This way, it has been calculated a total population of approximately 14,000 people.

The Volpe Redevelopment Plan started some years ago and much work has been done so far. Meetings with city councils, and community involvement gatherings have been held. Additionally, the development plan has already been approved. The principles revolving around this redevelopment plan focus on revitalizing the area by making it more vibrant. Aspects such as civic life, activation, inclusiveness, comfort and sustainability are the main priorities when thinking of how to approach the issue [80].

However, lately some questions have been raised on whether the plan should be more ambitious and try instead, build a vertical city at a small scale. The concept vertical city is understood as building more vertically at higher heights, while still offering a complete and diverse set of services.

The projected plan, offered a good office space for workers. However, it did not contemplate the housing offer as much, since just 1,400 units of housing and 950 beds for graduate students had been projected. This is why, it is being studied how taller buildings could accommodate the mentioned 14,000 people.

Results

In this exercise, it has been assumed that each worker should get a house key, totaling 10,000. Additionally, based on the development plans, the floor side size is of 100 feet; and the typical gross to net area ratio of 1.2 has been assumed. With that in mind, Table 8 has been obtained, detailing the housing needs of this new population.

At this point, a clarification needs to be made. Even though with the housing mix projected in Table 8 more than 14,000 people could be accommodated, it is believed that the housing needs of young professionals and executives, for example, is totally different, executives demanding larger houses.

Housing Profile occupancy	Net sqf (ORI)	# units	Total net sqf (,000)	Total gross sqf (,000)	Number of floors
Single	200	6000	1200	1440	144
Dual (2 person)	300	1500	4500	540	54
Triple (2 person + child)	500	1000	5000	600	60
Quad (2 person + 2 children)	750	800	6000	720	72
Family (2 person + 3 child)	1000	700	7000	840	84
Total		10000	3450	4140	414

Table 8. Housing profile and total housing square feet needs for Volpe

For the net square feet (sqf) per house profile, ORI systems' reference has been taken as explained in Figure 26. This helps increasing usability of the place, by creating a multifunctional space, and, thus, allowing compact housing. This way, the need for bigger houses, full of monofunctional spaces is decreased. This goes in line with the lab's view of increasing density of people in urban areas, by creating more economically accessible spaces.

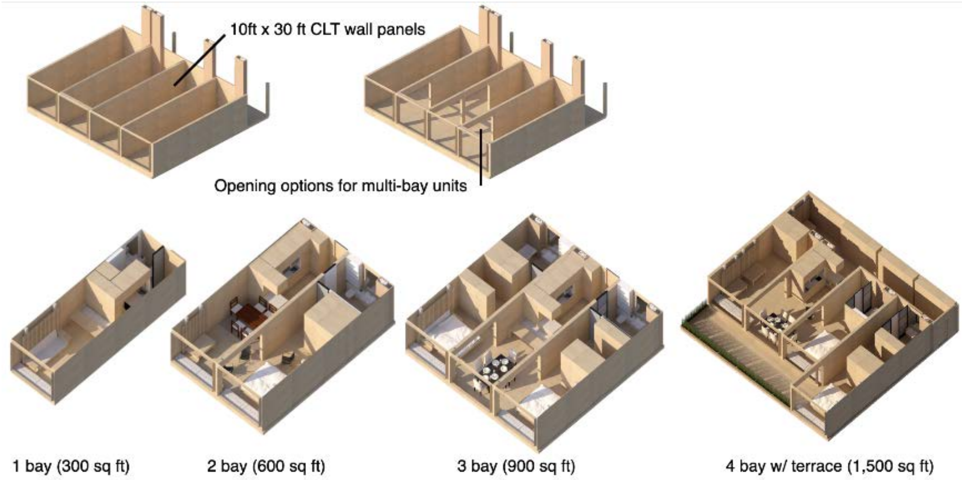

Figure 26. Bay space distribution depending on housing needs and incorporating ORI systems furniture

Moreover, the plan also considered office space for 1.7 M sqf. With a preliminary breakdown of commercial space type, assuming a net area per person validated by industry experts, it is obtained a total space for nearly 5000 workers. The space distribution is summed up in Table 9.

Commercial Type	% of total	Net Area (,000)	Net Area (sqf/person)	Workers
Offices	60%	1,020	250	4080
Wet lab	20%	340	800	425
Research Space	10%	170	400	425
Co-Working Space	10%	170	100	1700
Total Commercial	100%	1,700		4930

Table 9. Accommodated workers based on currently projected office space

<u>Analysis of results</u>

The Total gross square feet for housing to include the whole population is approximately 4 times greater than the devoted in the preliminary plan. Besides that, with the current commercial space allocation, there would just be enough space for half of the workers. All in all, the current Volpe Redevelopment Plan falls short on the ambition of creating a live-work symmetry landscape, which leaves no other choice but to increase buildings height.

This initial plan considered 8 buildings, which leads to project a minimum of 52 residential floors per building. This number of floors would not take into account the additional office space needed. If office space for 10,000 workers was to be built, 43 office floors per building would be needed. In addition to residential and office space, the redevelopment plan also considers open space, accounting to 2.5 acres, and retail and entertainment space of 95,000 sqf. That would require a minimum of 1.2 floors/building. All in all, an average of 86-story buildings would be required to meet with the intended goal.

However, this project might seem overambitious given the surroundings in which this development is englobed, where there are not buildings that high. In that sense, the population to be fitted in Volpe to do live-work could be diminished, although bearing in mind the positive aspects a live-work symmetry district englobes.

6.2. Benchmarks

Benchmarks are always useful when defining something from scratch. Whenever it is wanted to forecast a metric, it is helpful to see how this metric has behaved in comparable situations. This is the way it has been followed in this case. In order to make Volpe/Kendall a vibrant neighborhood, other business districts from the US where there is an innovative and entrepreneurial environment have been taken as reference.

These business district benchmarks have been chosen based on a panel of experts' opinion of well-functioning urban areas. However, these areas have some dissimilarities between them. The area each business district englobes is different, having in some cases an order of magnitude of difference. In addition, some of the areas that have been chosen are purely business offices with little residence. However, it is believed that having dissimilar benchmarks is what gives this research a more holistic view of the situation. All in all, the list is as follows:

City	Area
Atlanta, GA	Downtown
Austin, TX	Downtown
Cambridge, MA	Kendall Square
Chicago, IL	Business District
New York City, NY	Brooklyn - Central Business District
New York City, NY	Manhattan - Union Square Neighborhood
Philadelphia, PA	Downtown
San Francisco, CA	Financial District
Washington, DC	Georgetown

Table 10. Entrepreneurial district benchmarks to assess amenity concentration

Being able to compare Kendall Square with other well-functioning areas definitely gives insight on which indicator is Kendall Square performing well and which one can be improved so as to meet the standards.

6.2.1. Process

For the list above, a script has been written for each, divided in three sections: importing of databases, working on databases and obtainment of output.

Regarding the databases that have been used, first, the geometry file of the state with a census block group capillarity has been obtained from www2.census.gov. The census block group is the smallest geographical unit for which the Census has data. For example, for the case of Georgia (GA), this database looks as follows (Figure 27).

```
# Function to download the geometry from the census data webpage used by the CS
Group
def get_geometry(state_fips, geom_type, year, target_crs="EPSG:4326"):
    print('Getting geometry ({}) for state: {}'.format(geom_type, state_fips))
    geom_type_map={'block_group': ['BG', 'bg'], 'tract': ['TRACT', 'tract'], 'blo
ck': ['TABBLOCK', 'tabblock10']}
    try:
        geom_type_l=geom_type_map[geom_type]
    except:
        print('Unrecognised geometry :'+geom_type)
    geom=gpd.read_file('https://www2.census.gov/geo/tiger/TIGER{}/{}/tl_{}_{}_{}.
zip'.format(
                year, geom_type_l[0], year, state_fips, geom_type_l[1]))
    geoid_col=[col for col in geom.columns if 'GEOID' in col][0]
    geom=geom.rename(columns={geoid_col: 'GEOID'})
    geom=geom.set_index('GEOID')
    geom=geom.to_crs(target_crs)
    geom.index=[int(i) for i in geom.index]
    centroids=geom['geometry'].centroid
    geom['x_centroid']=[c.x for c in centroids]
    geom['y_centroid']=[c.y for c in centroids]
    return geom

# Calling the function
bgs_ga= get_geometry(state_fips=13 , geom_type='block_group', year=2019, target_c
rs="EPSG:4326")
```

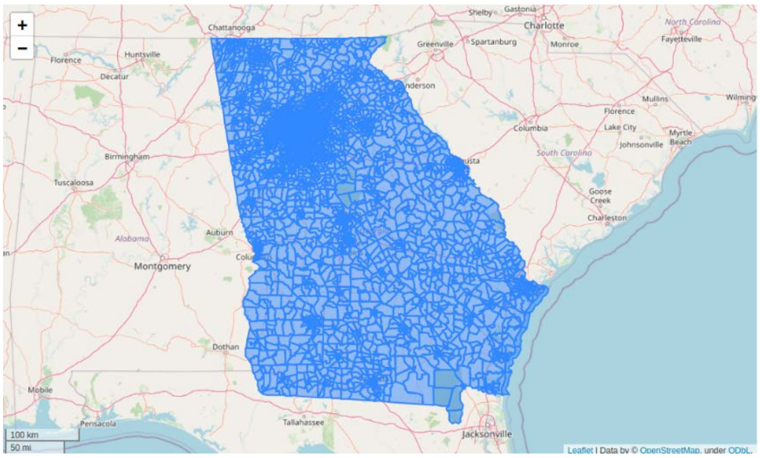

Figure 27. Map of Georgia divided by block group

After that, a shapefile (.shp) with the shape of the wanted-to-measure business district has been uploaded (Figure 28).

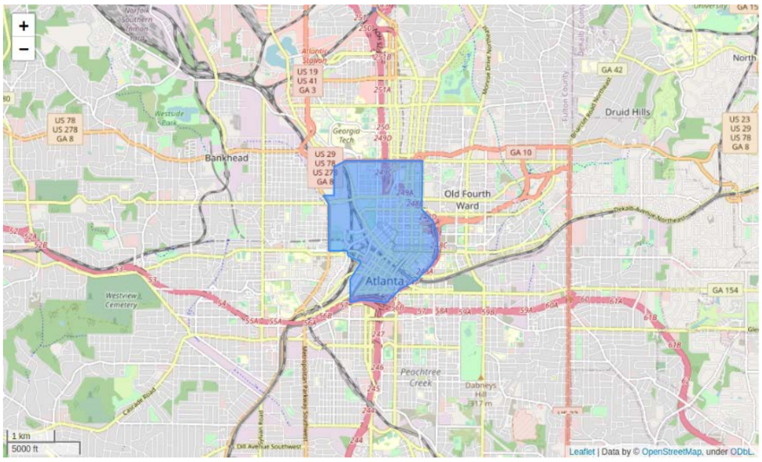

Figure 28. Atlanta Downtown Business District

Following that, the population census data has been extracted from the Census webpage, with a Block Group capillarity. In this case, a small error is being committed, since the data the US government publishes is not 100% accurate. However, this is the most reliable and accessible database source and it has been considered that the margin of error is within an acceptable range. Additionally, the Places and Geometry databases provided by Safegraph have been used.

After having imported all the databases, it has been proceeded to work with them. First, it has been overlapped the business district area with the state block groups database, so as to obtain the block groups belonging to the area objective. Since it is believed that the amenities inside an area do not only serve those living in the exact same area, but also to a greater radius from the boundaries, a 900 meters buffer boundary has been drawn. Then, again, the block groups belonging to this buffer area have been obtained (Figure 29).

```
# Function used by the CS Group to overlap two geometries
# Threshold defines the amount of intersection allowed to be considered a geometr
y inside another one
def get_overlap_by_threshold(geom_1, geom_2, threshold=0.5, plot=True):
    """
    function for subsetting a polygon collection based on overlap with another po
lygon collection
    geom_1 and geom_2 should be geopandas GeoDataFrames
    geom_1 is the geometry to be subsetted
    subsetting is based on overlap with geom_2
    a zone in geom_1 will be included in the output if its area of overlap with g
eom_2 is greater than the threshold
    """
    geom_1['copy_index']=geom_1.index
    geom_1['zone_area']=geom_1.geometry.area
    all_intersect=gpd.overlay(geom_2.to_crs(geom_1.crs), geom_1, 'intersection')
    all_intersect['intersect_area']=all_intersect.geometry.area
    all_intersect=all_intersect[[col for col in all_intersect.columns if not col=
='zone_area']]
    all_intersect=all_intersect.merge(geom_1[['copy_index', 'zone_area']],
        how='left', left_on='copy_index', right_on='copy_index')
    all_intersect['prop_area']=all_intersect['intersect_area']/all_intersect['zon
e_area']
    valid_intersect=all_intersect.loc[all_intersect['prop_area']>threshold]
    final_zone_ids=list(valid_intersect['copy_index'])

    if plot:
        fig, ax = plt.subplots(1, figsize=(10,10))
        geom_1.loc[final_zone_ids].plot(facecolor="none",
            edgecolor='blue', ax=ax)
        geom_2.to_crs(geom_1.crs).plot(facecolor="none",
            edgecolor='red', ax=ax)
    return final_zone_ids

# Calling the function
# bgs_ga is the geodataframe with all the block groups
# Atlanta_cbd is the business district shapefile
bg_overlap_atlanta_cbd = get_overlap_by_threshold(bgs_ga,atlanta_cbd, plot = Fals
e)
```

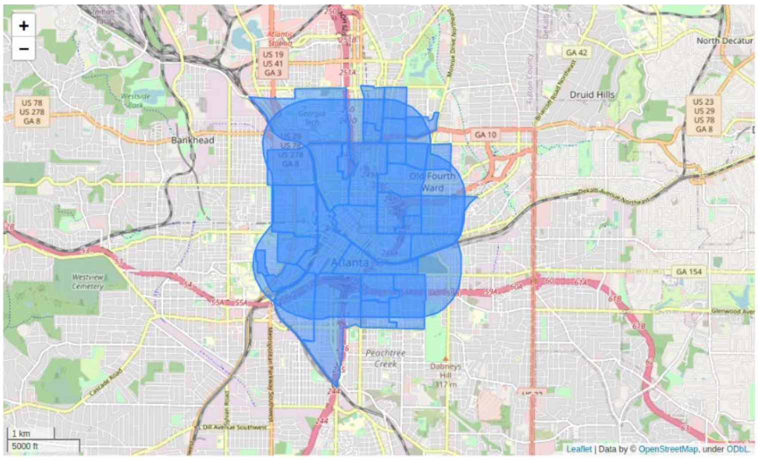

Figure 29. Overlay of Atlanta Business District buffer boundary with the belonging block groups

The criteria for including a block group inside the buffer area has been as follows: If more than 50% block group's area belonged to the buffer boundary it has been included; otherwise, it has been rejected. Having identified the block groups, it has been proceeded to calculate the total population of the block groups belonging to the buffer boundary making use of the census data.

Later on, the amenities inside the business district have been calculated as well as its area. Having all this information, a .csv file with the following appearance has been created (Table 11). The metrics that have been thought to be important are the following (Table 12):

Amenity	Count	Amenities per 1000	Total sqf	Sqf per amenity	Sqf per person	Amenity sqf by total area (%)
Restaurants and Other Eating Places	268	16.06	2119295	7908	127	3.97
Offices of Physicians	87	5.21	822989	9460	49.32	1.54
Other Personal Services	76	4.55	387877	5104	23.24	0.73
Traveler Accommodation	53	3.18	189273	3571	11.34	0.35
Personal Care Services	53	3.18	152745	2882	9.15	0.29
Religious Organizations	42	2.52	95133	2265	5.7	0.18
Lessors of Real Estate	39	2.34	193813	4970	11.61	0.36
Jewelry, Luggage, and Leather Goods Stores	39	2.34	161715	4147	9.69	0.3
Health and Personal Care Stores	33	1.98	122186	3703	7.32	0.23
Museums, Historical Sites, and Similar Institutions	31	1.86	425180	13716	25.48	0.8
Clothing Stores	29	1.74	85280	2941	5.11	0.16
Advertising, Public Relations, and Related Services	29	1.74	139460	4809	8.36	0.26
Management, Scientific, and Technical Consulting Services	27	1.62	154068	5706	9.23	0.29
Justice, Public Order, and Safety Activities	27	1.62	144492	5352	8.66	0.27
Depository Credit Intermediation	27	1.62	66893	2478	4.01	0.13
Automotive Repair and Maintenance	24	1.44	483066	20128	28.95	0.91
Drinking Places (Alcoholic Beverages)	24	1.44	124717	5197	7.47	0.23
Offices of Other Health Practitioners	24	1.44	53936	2247	3.23	0.1
Other Amusement and Recreation Industries	22	1.32	540569	24571	32.39	1.01

Offices of Real Estate Agents and Brokers	21	1.26	93485	4451	5.6	0.18

Table 11. Top 20 amenity categories and related metrics for Atlanta Business District

Metric	Description
Count	# of amenities of each category inside the business district
Amenities per 1000	# of amenities of each category per 1000 people living in the buffer area
Total sqf	Total square feet of the amenities of the same category
Sqf per amenity	Avg square feet per amenity
Sqf per person	Avg square feet per person living in the buffer area
Amenity sqf by total area (%)	Percentage of square feet that represents the category over the total area of the business district

Table 12. Amenity metrics obtained for each district

This process has been repeated over and over for the list of business districts mentioned above (Table 10). Then, all the .csv files have been grouped inside a .xlsx file where the number of amenities for each business district was easily comparable.

6.2.2. Results

The ratio thought to be more useful for amenity planning purposes is the amenity per 1000 people as this gives a sense of the number of amenities needed to have in an area depending on the size of the population. As explained, these ratios have been obtained for more than 150 categories that are present in those districts.

Amenities per 1000	Kendall Square	Atlanta	Austin	Brooklyn	Chicago	George town	Philly	San Francisco	Union Square
Restaurants and Other Eating Places	0.59	4.56	1.13	1.66	6.63	2.58	6.2	3.43	2.65
Lessors of Real Estate	0.54	0.66	0.7	0.86	5.03	2.23	6.84	0.18	0.77
Other Amusement and Recreation Industries	0.15	0.37	0.26	0.31	1.05	0.48	0.81	0.4	0.84
Offices of Other Health Practitioners	0.15	0.41	0.52	0.39	2.72	0.46	1.74	1.57	2.6
Management of Companies and Enterprises	0.14	0.12	0.04	NA	0.37	NA	0.16	0.31	0.07
Other Miscellaneous Manufacturing	0.12	0.22	0.09	0.07	0.72	0.08	0.44	0.19	0.9
Offices of Physicians	0.08	1.48	1.66	0.23	2.99	0.63	2.17	0.54	1.09
Management, Scientific, and Technical Consulting Services	0.08	0.46	0.09	0.05	0.93	0.06	0.49	0.64	0.33
Depository Credit Intermediation	0.07	0.46	0.35	0.07	0.66	0.32	0.59	0.58	0.24
Other Financial Investment Activities	0.07	0.34	0.17	0.07	1.32	0.04	0.4	1.5	0.29
Other Personal Services	0.05	1.29	0.39	0.36	1.34	0.19	0.71	0.46	0.36

Colleges, Universities, and Professional Schools	0.05	0.12	0.22	0.08	0.23	0.13	0.22	0.06	0.07
Other Schools and Instruction	0.05	0.29	0.09	0.16	0.71	0.19	0.36	0.4	0.7
Child Day Care Services	0.05	0.15	0.13	0.19	0.14	0.15	0.25	0.05	0.06
Activities Related to Real Estate	0.05	0.2	NA	0.03	0.42	0.04	0.28	0.21	0.12
Religious Organizations	0.05	0.72	0.26	0.17	0.72	0.57	0.87	0.13	0.44
Automotive Repair and Maintenance	0.05	0.41	0.04	NA	0.11	0.04	0.07	0.1	0.05
Grocery Stores	0.05	0.14	0.13	0.1	0.37	0.13	0.48	0.08	0.1
Personal Care Services	0.03	0.9	0.79	0.58	2.58	1.53	2.08	0.66	1.93
Building Equipment Contractors	0.03	0.09	0.09	0.02	0.17	0.1	0.06	0.17	0.15
Offices of Dentists	0.03	0.34	0.09	0.13	1.13	NA	0.63	0.55	0.27
Clothing Stores	0.03	0.49	NA	0.34	1.21	1.2	0.83	0.13	1.35

Table 13. Top Amenities per 1000 residents in 9 US Business Districts

However, some of them are considered to be more important and representative. For example, Restaurants and Other Eating Places, which considers not only full-service restaurants but also limited-service ones and snack bars (nonalcoholic), shows varying numbers across the considered business districts. Kendall Square, with 0.59 Restaurants per 1000 residents, is the one with the least number, while other districts such as San Francisco's Financial District are on 3.43. There are some other business districts that show a great number of restaurants per a thousand inhabitants as Chicago's Central Business District with 6.63.

The results obtained have not been entirely as expected. It is acknowledged that the obtained ratios do offer a guideline on which to base the urban design. However, the working hypothesis was that all the business districts would show similar ratios regarding amenities, with slight variance between them. Nonetheless, a big difference has been encountered.

One of the reasons of this difference might be found in the sampling area. Smaller business districts are penalized in comparison to bigger ones, when considering the buffer area. In other words, the area added when buffering 900 meters, represents a larger share of the total area for smaller business districts. In consequence the population added to the ratios also vary, thus, disturbing the metrics. Figure 30 and formulas (5)(6)(7)(8) help better understand this issue.

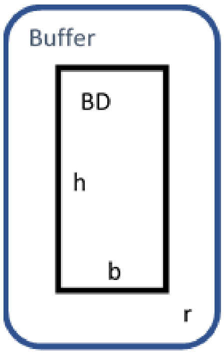

Figure 30. Representation of a Business district (BD) and its Buffer boundary

$$A_{BD} = b \cdot h \tag{5}$$

$$A_{Buffer} = b \cdot h + 2 \cdot (b + h) \cdot r + 4 \cdot (\frac{1}{4} \cdot \pi \cdot r^2) \tag{6}$$

$$Small\ BD\ (b = 2\ mu, h = 4\ mu, r = 1\ mu) \rightarrow\ A_{BD} = 8\ mu^2; A_{Buffer} = 20 + \pi\ mu^2 \rightarrow \frac{A_{Buffer}}{A_{BD}} \cong 3 \tag{7}$$

$$Big\ BD\ (b = 4\ mu, h = 8\ mu, r = 1\ mu) \rightarrow\ A_{BD} = 32\ mu^2; A_{Buffer} = 56 + \pi\ mu^2 \rightarrow \frac{A_{Buffer}}{A_{BD}} < 2 \tag{8}$$

A_{BD} : *Area Business District*
A_{Buffer} : *Area Buffer Boundary*
b: *base*
h: *height*

This might help to partially explain the obtained results although it is not the whole picture. The own characteristics of each business district might have a bigger saying. In order to subtract the influence of the area of the business district, spatial metrics have been also considered, as the "Amenity sqf by total area (%)". Surprisingly, the results have also been very volatile, with no clear outcome from this analysis. Additionally, this metric loses importance when shifting towards buildings at higher heights. In addition, a normalization of the metrics could have been carried out. However, this approach was thought to be time consuming and would not offer a much better view than the actual one.

6.2.3. Analysis of results

After having studied the points of interest (POI) data, it has been concluded that, even though this analysis offers valuable insight, it is not definite to define the amenities that would be needed in a newly developed area. Ratios of amenities per 1000 people or sqf per capita are valuable as references. However, these ratios do not take the individual characteristics of each area into account. Additionally, it has been assumed that the coverage of the amenities were confined to the 900 meter buffer boundary of the business district. Nonetheless, this is an assumption that does not take into consideration the numerous people that commute every day to work in and out of that zone. Therefore, it has been thought that the next approach should not be focused in areas (e.g. Volpe-like areas) but rather in people (e.g. prospect residents of Volpe). If it is possible to characterize the visits of different groups of people to each amenity, it can be later defined the number of amenities needed depending on the population of that area.

6.3. Massachusetts

In this step, based on the Safegraph patterns data, it has been analyzed the amenities that people visit every month regardless of the location of these amenities. Safegraph has data for more than 65k POIs in Massachusetts, with its patterns database containing monthly count of visits to each POI disaggregated by block group. With this data at hand two are the intended outputs:

- Visits per month per person to each different amenity at a block group level
- More reliable count of visits to each POI

The former offers valuable insight on the frequency of visit to any kind of amenity, while the latter gives an approximate view of the capacity of a POI. Both combined, help answer quantity and proximity of each amenity that should be placed in an area. All in all, the process followed has been as follows.

6.3.1. Process

First of all, the databases have been imported. Both Census and Safegraph's data have been combinedly used. Regarding Census Data, the greatest available breakdown is census block groups. In this sense, age and income distribution data have been imported. As to Safegraph, both places and patterns databases have been used. First step, has been cleaning the data, erasing the less useful columns and those rows that included null data. Later on, the information has been reordered inside a dictionary in which the keys were the amenities categories; and the values a pandas data frame containing just each category data. This step, although not necessary, simplified the process to come.

Then, two main data frames have been created, in which each row represents the visits of people from a block group to an amenity. The difference between both data frames is what the block group represents. While in the first data frame the block group represents the block group where the visitors live, the second one represents the workplace.

```
# safegraph_ma contains the Safegraph POIs information of MA
list_categories = safegraph_ma['top_category'].unique()
category_dict = dict()
for i in range(len(list_categories)):
    category_dict[list_categories[i]] = safegraph_ma[safegraph_ma['top_category']
 == list_categories[i]]

# Get the column indexes
category_aux = category_dict[list_categories[0]]
col_tc = category_aux.columns.get_loc('top_category')
col_ln = category_aux.columns.get_loc('location_name')
col_pk = category_aux.columns.get_loc('placekey') # unique id
col_nc = category_aux.columns.get_loc('naics_code')
col_rv = category_aux.columns.get_loc('raw_visit_counts') # registered visits
col_nv = category_aux.columns.get_loc('normalized_visits_by_state_scaling') # sta
te normalized visits


all_df = []
for i in range(len(list_categories)):
    category = category_dict[list_categories[i]]
    category_placekey = list(category['placekey'])
    if i%10 == 0:
        print(i, "/", len(list_categories))

    for j in range(len(category)):
        cat_ind = category.index[category['placekey']==category_placekey[j]].to_l
ist()
        visits = category.at[cat_ind[0],'visitor_home_cbgs'] # Change by 'visitor
_home_cbgs' if home visits are to be analysed
        visits = json.loads(visits)
        visits_df_aux = pd.DataFrame.from_dict(visits, orient = 'index', columns
=['Visits_bg'])
        #Add columns
        visits_df_aux['Placekey'] = category.iloc[j, col_pk]
        visits_df_aux['Top_category'] = category.iloc[j, col_tc]
        all_df.append(visits_df_aux)

visits_live_df = pd.concat(all_df)
```

At this stage of the study, it is not of high relevance which exact amenity they visited, but instead, the category of the amenity. Following this guideline, the number of visits per block group to each different amenity type has been obtained. The output resulting from this script looks as follows (Table 14):

Geoid_home	Visits_bg	Placekey	Top_category
250259817001	705	22d-222@62j-shz-4n5	Restaurants and Other Eating Places
250259817001	330	272-223@62j-shz-4n5	Restaurants and Other Eating Places
…	…	…	Restaurants and Other Eating Places
250250503001	40	227-222@62j-sj7-2tv	Beer, Wine, and Liquor Stores
250158204001	36	225-222@62k-ckb-q2k	Beer, Wine, and Liquor Stores
…	…	…	Beer, Wine, and Liquor Stores
250092532022	80	zzw-222@62j-sx7-mp9	Clothing Stores
250092151012	45	zzw-222@62j-ptw-2c5	Clothing Stores
…	…	…	Clothing Stores

Table 14. Sample of visits per block group to selected categories

In Table 14, Geoid_home represents the block group from which the visits are coming. Visits_bg stands for the total monthly visits to a placekey from a given block group. Placekey is a Safegraph unique id associated to one POI; and Top_category is the amenity category of the given placekey.

Before jumping into any conclusions, some flags need to be raised. The Safegraph data has some voids that need to be taken into consideration so as to know the limitations of the model that has been developed. Regarding the visits, there are 3 layers to be aware of before driving any conclusion. Figure 31 helps better understand this rationale.

- First, the normalized layer. Safegraph does not monitor everyone in the state. For example, in Massachusetts they monitor just around a 5% of the people. Therefore, any visits result obtained needs to be normalized to acknowledge for this monitoring percentage. In this case, the state normalization multiplier would be 20.
- Second, the raw visits layer. This number Safegraph offers is the total amount of visits that have been registered to each amenity. The raw visits multiplied by the state normalization multiplier accounts for the normalized visits.
- Third, the visits by block group layer. In this layer, Safegraph offers the number of the registered visits that have come from each block group. However, the sum of this third layer does not amount to the raw visits due to two reasons.
    - First reason is due to Safegraph sometimes not being able to categorize the block group where the visitor either lives or work.
    - Second reason is due to Data Privacy Act, which blocks them from publishing accurate collected data from block groups with less than 4 visitors to an amenity. Block groups with 2 to 4 visitors are represented as if there were 4 visits, while unique visits are not reported. In order to mitigate this phenomenon, monthly data has been taken instead of weekly one, even though the difference is not entirely amended.

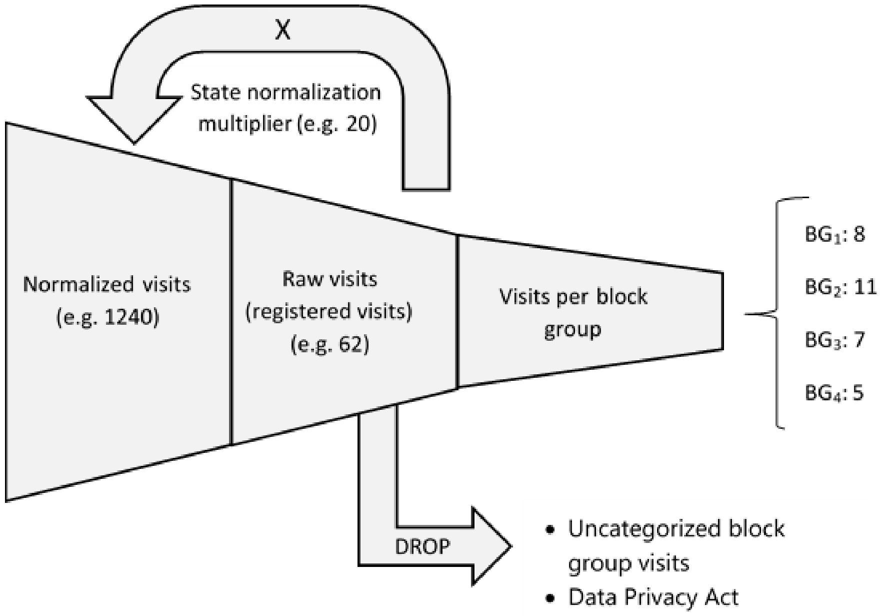

Figure 31. Safegraph's visits layers funnel

Additionally, as it has been stated before in the chapter 5.1.2 there are two additional limitations.

- Safegraph's accuracy when assigning a block group, either where a person lives or work, is not optimal. In this case, some misassignment can happen which can slightly lead to false conclusions.
- Safegraph does not monitor a person throughout the whole day, so the number obtained both at a person and POI level, just represent a fraction of the total visits. This fraction has been assumed to be the same for both cases as it has been explained in formula (3).

Additional columns to the outcome table have been added so as to get comparable ratios between block groups.

- A normalization of the visits has been carried out based on the monitoring percentage of each block group. At first, this normalization was done equivalently across all the block groups, normalizing all the visits by the state normalization multiplier. However, this led to false conclusions since the devices tracked per block group are not constant throughout the state (some block groups might be overrepresented while other are underrepresented). That is why an individualized multiplier for each block group has been calculated based on the number of devices tracked and the population of each block group as soon as this information was made available. Safegraph delivered an additional database where the monitoring percentage of each block group was detailed. Figure 32 shows the new way of normalizing the visits.

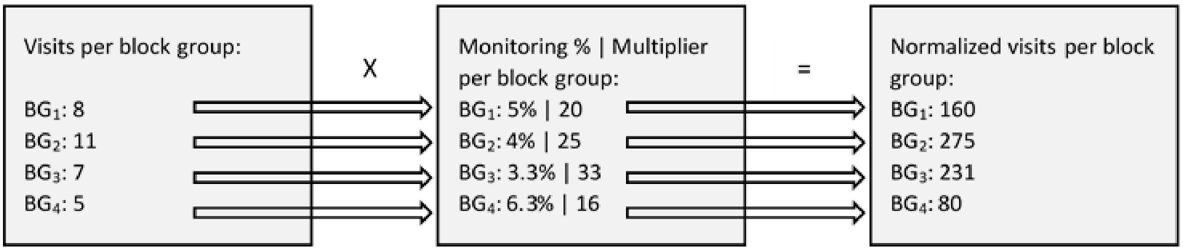

Figure 32. New approach to calculate normalized visits, based on an individualized block group multiplier

- Also, an additional column with the unique POI visited of each category, has been calculated in order to see block groups' prevalence for distinct or repeating POIs. This can give a sense on whether residents/workers of a block group attend diverse destinations or, their visits are more concentrated in some specific block groups
- Besides that, it has been proceeded to obtain some metrics. With the Census Data block group's population, the visits per person have been obtained. Additionally, the capacity of each type of amenity has been calculated as the mean visits to each amenity inside a category. With the visits per person and the number of visits an amenity typically takes, the number of people an amenity serves can be obtained. This last calculation is based on the assumption that the mean visits to an amenity is a good reference of its capacity.

Not all block groups are equally important for the sake of this project. Since Volpe's redevelopment plan is to be located in a highly-dense area, in other words, an urban area, it makes sense that the analysis is focused just on the block groups that meet this condition. Additionally, block groups with low sampling are hardly extrapolated. Since this sampling bias can induce some errors (some visits might go unregistered due to low sampling or privacy issues), a minimum threshold has been established so as to discard those block groups that might not have the enough tracking. In that sense, block groups that have a density below 1500 habitants/km^2, threshold above which an area is catalogued as urban [81], have been filtered out, leaving 2388 block groups in MA out of the 4984 originals. Additionally, a monitoring percentage threshold has been established in 2%. In total, 1793 block groups are being considered.

6.3.2. Preliminary results

Having the previous constrains as a starting point Table 15 has been obtained. Table 15 shows the most present categories in urban, Safegraph minimally monitored MA block groups. The POI per 1000 metric gives a rough sense of the number of POIs that would be required for each category in urban regions. This sets the first milestone for this project. If compared with Table 2, where Cesar Hidalgo et al. Work is presented [65], some disparity between the numbers can be observed, even though not all categories are comparable one to one. The present approach, oftentimes, projects a greater number of POIs per 1000 residents.

Top_category	Count	Visits bg norm mon	Visits pPOI	Visits pp	POI per1000	1 POI per X
Restaurants and Other Eating Places	8369	8668812	1035.82	3.72	3.59	278
Other Amusement and Recreation Industries	2632	2262933	859.78	0.97	1.13	885
Museums, Historical Sites, and Similar Institutions	2541	2548819	1003.08	1.09	1.09	917

Gasoline Stations	1608	1638824	1019.17	0.70	0.69	1449
Grocery Stores	1422	1741390	1224.61	0.75	0.61	1639
Religious Organizations	1421	483087	339.96	0.21	0.61	1640
Elementary and Secondary Schools	1346	1033540	767.86	0.44	0.58	1731
Health and Personal Care Stores	1139	1139967	1000.85	0.49	0.49	2046
Automotive Repair and Maintenance	907	241120	265.84	0.10	0.39	2569
Sporting Goods, Hobby, and Musical Instrument Stores	783	690844	882.30	0.30	0.34	2976
Other Miscellaneous Store Retailers	769	602442	783.41	0.26	0.33	3031
Offices of Dentists	712	180810	253.95	0.08	0.31	3273
Offices of Other Health Practitioners	697	187866	269.54	0.08	0.30	3344
Offices of Physicians	677	267549	395.20	0.11	0.29	3443
Personal Care Services	676	181216	268.07	0.08	0.29	3448
Child Day Care Services	675	267861	396.83	0.11	0.29	3453
Clothing Stores	582	393532	676.17	0.17	0.25	4005
Drinking Places (Alcoholic Beverages)	580	453361	781.66	0.19	0.25	4018
Automobile Dealers	579	299936	518.02	0.13	0.25	4025
Automotive Parts, Accessories, and Tire Stores	546	281497	515.56	0.12	0.23	4269
Beer, Wine, and Liquor Stores	538	334164	621.12	0.14	0.23	4332

Table 15. Number of amenities for most visited categories based on visitation patterns for Massachusetts urban and minimally monitored by Safegraph block groups

In Table 15 each category represents the following:

- Top_category: category of the POI
- Count: Number of POIs belonging to each category in the considered block groups
- Visits bg norm mon: Monthly visits per block group to each top category after normalizing by the block group monitoring percentage
- Visits pPOI: ratio of 'visits bg norm mon' by the count of POIs of each category; it represents the capacity of a POI based on Safegraph standards
- Visits pp: ratio of 'visits bg norm mon' by the population of the considered block groups
- POI per1000: ratio of 'Visits pp' by 'Visits pPOI' normalized by 1000 residents
- 1 POI per X: number of residents per POI

On one side, Table 15 shows the state view, where all the visits have been summed up by category and general metric at a state level have been obtained. On the other side, the block group breakdown can be studied. However, the results, once again, offer a high variability between block groups with no clear outcome that can be extracted from this.

Taking the most visited categories as an example, a state heat map has been plotted where the represented variable is the visits per person. Online tool Kepler.gl has been used to do so. Being a normalized metric by the population, one would expect to obtain similar values across the block groups.

Figure 33 represents the heat map of Restaurants and Other Eating Places, with numbers varying greatly between zero visits and 153 per block group. Figure 34, on the contrary represents Bars where the min and max are zero and 42 visits respectively.

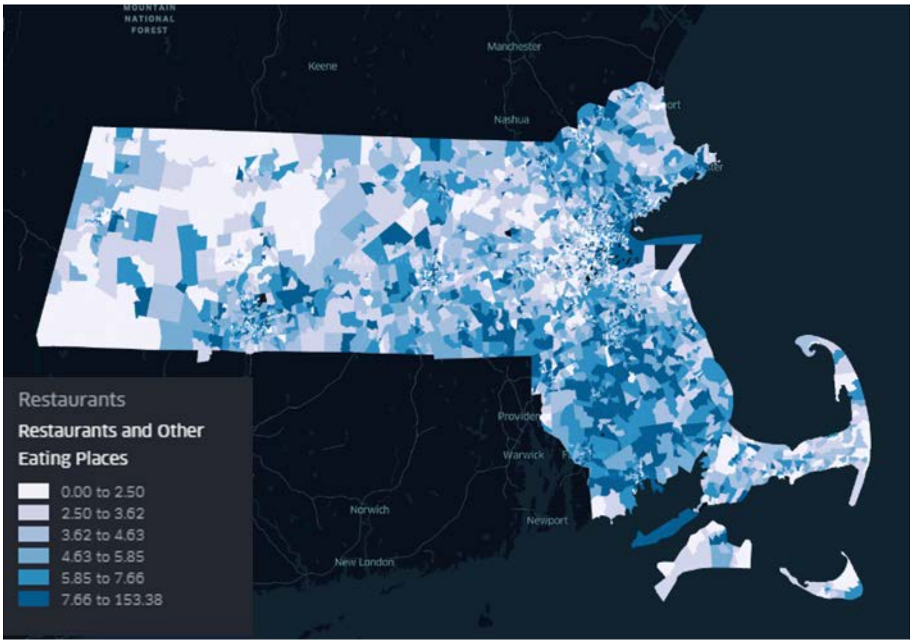

Figure 33. Massachusetts heat map of visits per person to Restaurants and Other Eating Places

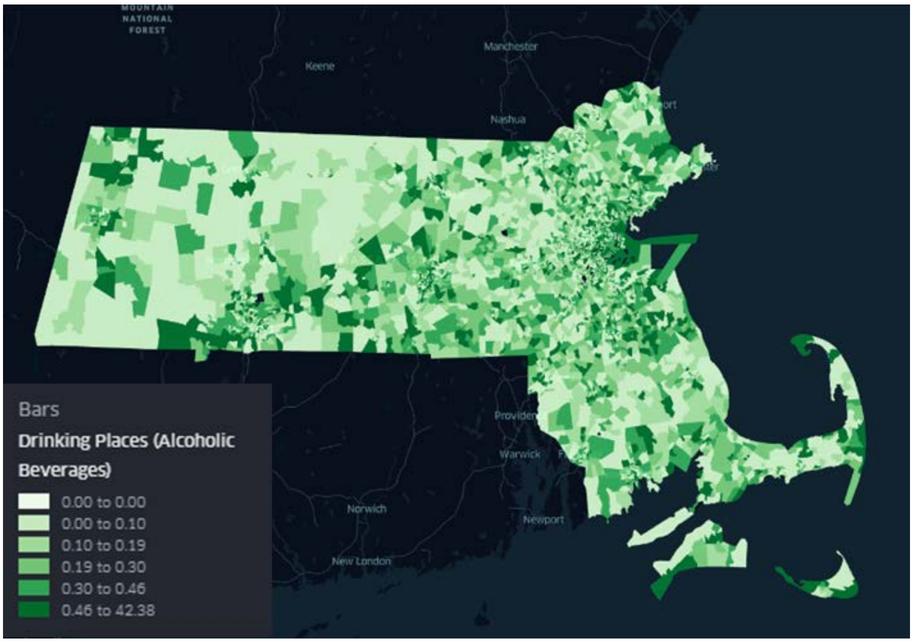

Figure 34. Massachusetts heat map of visits per person to Bars

The high disparity found makes it difficult to extract conclusions just from this approach. It is then proceeded to analyze whether the particularities of each block group have a saying in the variability found in the data.

6.3.3. Linear regression model

The first step has been focused on calculating the table for the whole state of Massachusetts, which englobes a highly diverse population. However, ending the analysis here makes little to no sense since the differences between different groups of people are not visible.

It has been proceeded then to build a linear regression model, where the input variables of the model were the income and age of the population of a specific block group and the output was the total visits of the block group to an amenity type. As a first example, it has been performed this model just for Restaurants. Table 16 is a sample of the input table passed along to the Linear Regression model, where the first column is the block group id, the second column (Visits bg norm mon) represents the individually normalized block group visits and it is the output variable, while the rest of the columns are the age and income distribution of each block group (e.g. u25 u10k means under 25 years old, under 10k household income)

Geoid home	Visits bg norm mon	u25 u10k	u25 u15k	...	o65 u100k	...	o65 u200k	o65 o200k
250010101005	260	0	0	...	19	...	0	7
250039001001	1479	0	13	...	0	...	0	0
250039001004	2703	21	0	...	0	...	0	0

Table 16. Sample of the input table for the Linear Regression model

To begin with, the outliers have been filtered out through the Z-score method. The Z-score is calculated as stated in formula (9). Later a threshold value is established so as to filter out extreme odd values. This threshold value has been set at 3, as it is usually done, which keeps 99% of the input entries [82].

$$Z_{score} = x - \frac{\mu}{\sigma} \quad (9)$$

Since the total visits can be a function of the working population of the area, it would not be suitable to calculate the Z-scores based on the total visits (some block groups might have greater values just because the living population is bigger). It is then, sensible to do it for the visits per person column.

Following that, it has been proceeded to divide the data frame in the training (80%) and validation (20%) set. The training set accomplishes the data with which the model will be built, while the validation set permits checking if the model is accurate enough. Prior to this step, the dataset has been shuffled so as to randomly select the rows included in each set.

Later on, it has been proceeded to calculate the linear model. At first, a simple Linear Regression was performed, where the results obtained were not very appealing, with values of R^2 below 0.4. There were a large number of variables trying to explain the model and some of them were counteracting the others. Then, a new linear model was performed called Lasso Regression in which the variables that do not help explain the model are penalized, and the ones that are more representative are kept.

All in all, a Lasso regression for each amenity category has been performed by running the following code.

```
# df_regression_res_pop is the input table
# Shuffling so as to randomly select the training and validation sets
df_regression_res_pop = shuffle(df_regression_res_pop)
# Selecting the input and output columns
list_X = df_regression_res_pop.columns.difference(['Geoid_home', 'Visits_bg_norm_
mon'])
X_pop = df_regression_res_pop[list_X]
y_pop = df_regression_res_pop['Visits_bg_norm_mon']
# Separating the training and validation sets
train_perc = 0.8 # 80% training, 20% validation
X_pop_train = X_pop.iloc[:int(train_perc*len(X_pop)), :]
y_pop_train = y_pop.iloc[:int(train_perc*len(X_pop))]
X_pop_val = X_pop.iloc[int(train_perc*len(X_pop)):, :]
y_pop_val = y_pop.iloc[int(train_perc*len(X_pop)):]
# Training the algo
regr_pop = linear_model.Lasso(alpha = 1.0, fit_intercept = False, positive = True
) # All coefficients shoudl be true, right?
regr_pop.fit(X_pop_train, y_pop_train)
# Validating the algo
score = regr_pop.score(X_pop_val,y_pop_val)
```

6.3.4. Results

As an example, Figure 35 shows the fit for restaurants, where the x-axis represents the actual normalized visits while the y-axis represents the predicted visits, done so by the profiling of each block group.

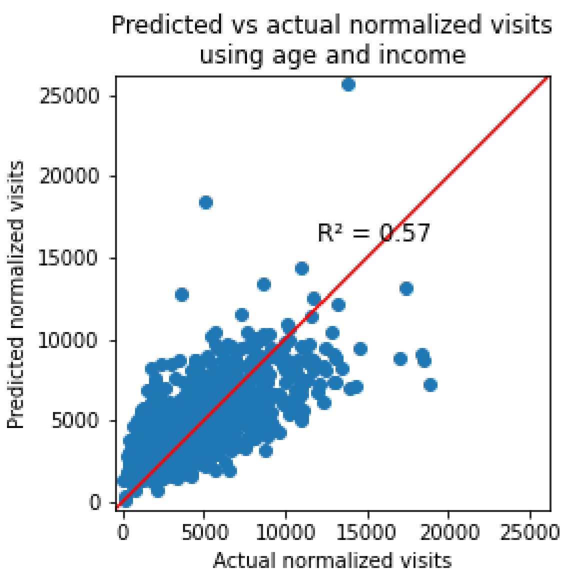

Figure 35. Predicted vs normalized visits to Restaurants and Other Eating Places, using age and income as input variables

Additionally, it is included the table with the coefficients of each age and income group (Table 17). The interpretation of these coefficients is very clear, with each of them representing the marginal contribution of each group; in other words, how many more visits would there be if an additional person from that age-income group was included in the population.

Group	Coefficient
u25 u30k	29.15
u45 u25k	23.18
u25 u15k	21.67
u25 u10k	21.15
u25 u50k	20.69
u65 u200k	18.25
u45 u10k	17.11
u65 u150k	16.83
u25 u125k	16.61
u65 u125k	16.46

Table 17. Marginal coefficients of each age-income group when predicting visits to Restaurants and other Eating Places

At this point, it is important to highlight a concept often repeated in the Data Analytics environment: "Correlation does not mean Causation". Even if the model might show a correlation, there might be additional variables that are falling outside the model that are not being taken into account.

6.3.5. Analysis of results

As a conclusion, in Figure 35 it can be seen that the performed linear regression do show a trend although the fit is not accurate enough. This trend is similar in other amenity categories. These graphs are useful to a make a fuzzy sense of amenity needs depending on a block group's population characteristics, although not as a truthful tool.

6.4. Accessibility analysis

In the following chapter it has been tried to answer whether accessibility to amenities influences the visitations patterns of each block group. It has been done so by leveraging the already mentioned Safegraph results. Even though these were obtained for the whole state of Massachusetts, they have been reduced to a smaller area for computational purposes. This area englobes Cambridge, part of Sommerville and Boston's Fenway, Back Bay, North End and Downtown regions.

6.4.1. Process

To assess the proximity analysis, graph & node theory-based pandana library has been used which easily helps building networks. The network type has been established as walk, as it is trying to be assessed the walking accessibility. There are two main queries that can be run for this purpose.

- Nth closest amenity query: enables to obtain a predefined number of close amenities to the node being considered
- Aggregation query: it will output the number of amenities within a given walking radius.

For the purpose of this analysis, the aggregation query has been used as it is believed to better represent the accessibility to multiple diverse amenities. The walking radius has been established in 500 m, focusing on hyper-accessibility. The analysis has been performed for multiple amenity categories, prioritizing those of which that are most visited by the residents of the abovementioned area. As an example, "Restaurants and Other Eating Places" (Figure 36), and "Clothing stores" (Figure 37) are showed.

Figure 36. Accessible Restaurants and Other Eating Places in a 500 meter walkable radius

Figure 37. Accessible Clothing Stores in a 500 meter walkable radius

6.4.2. Results

The results are very representative. Looking at the "Restaurants and Other Eating Places" network (Figure 36), it can be seen how North End has a big restaurant offer, with some nodes having more than 175 accessible restaurants. Other Boston areas, such as Back Bay, still have a good restaurant accessibility. Cambridge, in contrast, does not stand out. Harvard Yard and Massachusetts Avenue, the zones with the highest restaurant offer, still show a poor offer when comparing with Boston; not to talk about Kendall Square's offer, which is even poorer.

Regarding Clothing stores (Figure 37), Back Bay is undoubtedly the distinguished commercial venue, with Downtonwn having also some remarkable presence. Newbury Street, in Back Bay, is the place where all clothing firms want to have a store. Cambridge's clothing store offer is very deficient. North of Kendall Square there is a commercial mall which helps slightly mitigate this stark comparison.

After analyzing these maps, it surges the question on whether Northend residents are more prone to visit restaurants or people living in Back Bay, same so with clothing stores. Not to limit the analysis to just these regions, the whole area has been considered. The working hypothesis is that accessibility does affect visitation patterns although this conclusion needs to be backed up by data.

Safegraph's data is organized by block groups, while the pandana network is organized by nodes. To be able to correlate both databases, an accessibility metric per block group has been calculated as the mean of all nodes' accessibility, those of which belonging to the same block group.

Once done so, a scatter plot displaying the visit per person to an amenity against the accessible amenities of that type has been drawn (Figure 38). In the graph, it is displayed this relationship for restaurants.

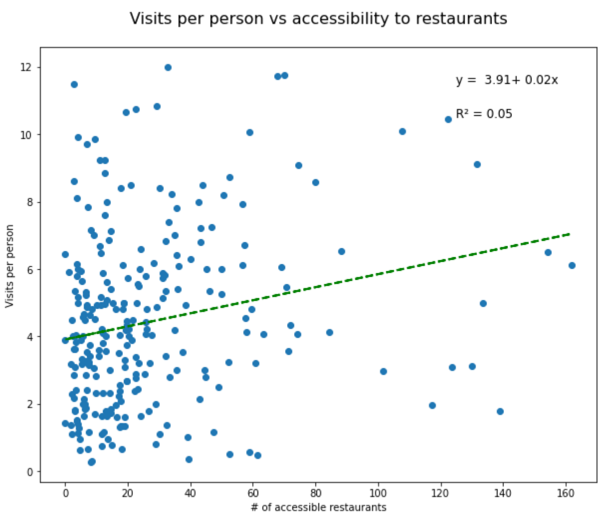

Figure 38. Visits per person vs accessibility to restaurants in Kendall Square and its surroundings for Restaurants

6.4.3. Analysis of results

The results obtained in Figure 38 for Restaurants do not prove a clear relationship from which insights could be extracted; however, the sample is small enough not to drive major conclusions.

Up until now, the POI analysis carried out has been done so leveraging Safegraph data which it is believed it has given valuable results and a rule of thumb method, presented in Table 15, to approach amenity planning in urban areas

6.5. Clustering analysis

The following chapter revolves around a clustering analysis performed with Replica's demographic data. As it has already been discussed, Replica's databases have some limitations. They do not offer a POI granularity, which is indeed needed to discuss visitation patterns of amenity categories. However, in the following chapter, it is explored whether they can shed light on behavioral patterns. In that sense, it is believed that a clustering analysis could help further analyze the data at hand. This step is totally exploratory as there is no clear output to be obtained from the process. To do so, Replica's Kendall Square's people data has been used and it's been tried to cluster people based on three variables: age, household income, and household size.

6.5.1. Process

First of all, the data has been filtered for people older than 21 years old and it has been also cleaned so as not to work with misleading data. The underlying reason to do so is to just focus on target people for Volpe, setting students aside. Later, outliers have been filtered out by the z scores algorithm, as the outliers do not represent the larger population, and introduce noise to the model. In addition, since the selected variables have all different scales, they have been standardized, so as to have a 0 mean and a 1 standard deviation. This process is deemed essential so that the algorithm does not mistakenly assume that some variables have greater importance than others.

```
# new_df -> original df with age, household income and household size as columns

# Dropping outliers
z_scores = stats.zscore(new_df['age']) # Age
new_df = new_df[(np.abs(z_scores) < 3)].reset_index().drop('index', axis = 1)
z_scores = stats.zscore(new_df['household_income']) # Household income
new_df = new_df[(np.abs(z_scores) < 3)].reset_index().drop('index', axis = 1)
z_scores = stats.zscore(new_df['household_size']) # Household size
new_df = new_df[(np.abs(z_scores) < 3)].reset_index().drop('index', axis = 1)

# Standardization
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
data_scaled = scaler.fit_transform(new_df[new_df.columns.difference(['person_id']
)])
data_scaled = pd.DataFrame(data_scaled)
new_df_stand = new_df.copy()
new_df_stand [new_df_stand.columns.difference(['person_id'])] = data_scaled
```

The clustering analysis decided to be performed has been k-means given the ease to implement it and the highly acceptable results the algorithm converges to. This algorithm needs the number of clusters as an input. In consequence, to better decide the number of clusters to divide the population, the elbow test was carried out (Figure 39). For every value of k (# of clusters), it is calculated the WCSS (Within-Cluster Sum of Square), which is the sum of the squared distance between any point and the centroid of the cluster. The number of clusters, k, is optimal when the change of slope is most pronounced. In this case, 4 clusters are the appropriate number.

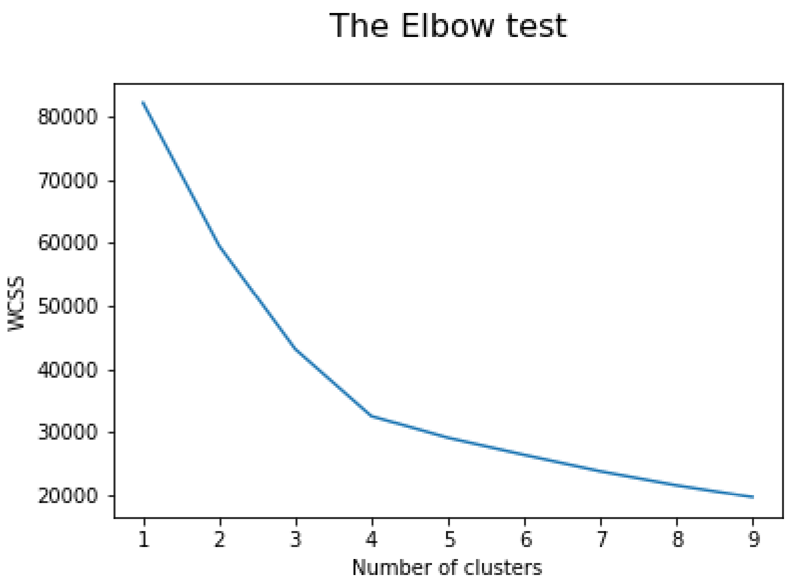

Figure 39. Elbow test to decide optimal number of clusters in k-means clustering method

Once defined the number of clusters, it has been proceeded to perform the k-means clustering methods, stating multiple initialization points so as to avoid the random initialization bias this algorithm englobes.

```
kmeans = KMeans(
    init="random",
    n_clusters=kl.elbow,
    n_init=10,
    max_iter=300,
    random_state=42
)
data_rescaled1 = scaler.inverse_transform(input_cluster)
data_rescaled1 = pd.DataFrame(data_rescaled1)
new_df_stand_resc = new_df_stand.copy()
new_df_stand_resc [new_df_stand.columns.difference(['person_id'])] = data_rescale
d1
new_df_stand_resc ['Cluster_KMeans'] = kmeans.labels_
new_df_stand_resc.head(3)
```

For clustering analysis, visualization is sometimes very helpful. However, it is better performed in 2D and when there are fewer points. For this case, box plots have been thought to be the most accurate representation. The lower line represents the minimum of the cluster and the upper one the maximum. Then, the body represents the distribution of each variable all along each cluster.

These are the results obtained for age (Figure 40), household income (Figure 41) and household size (Figure 42):

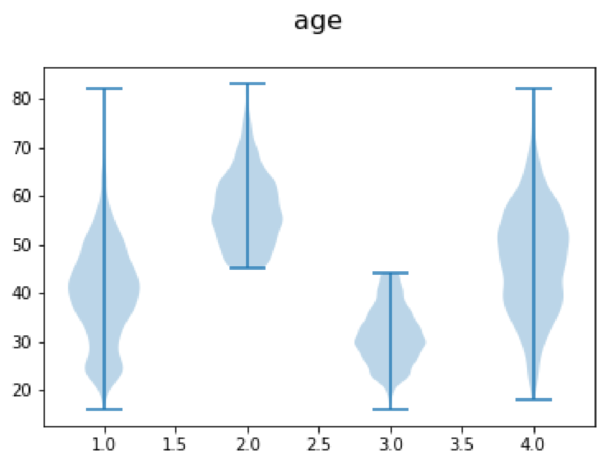

Figure 40. Violin plot of the age distribution by cluster

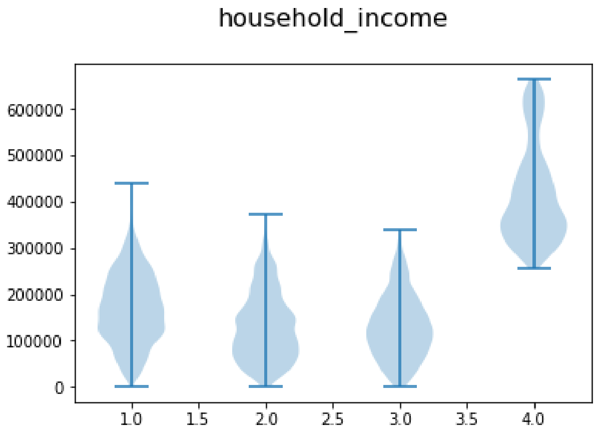

Figure 41. Violin plot of the household income by cluster

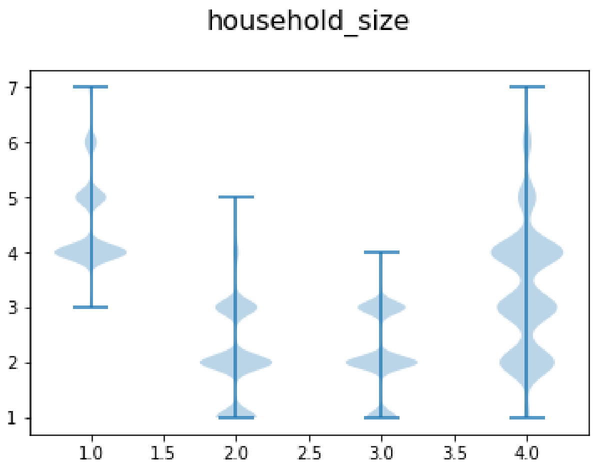

Figure 42. Violin plot of the household size by cluster

Based on the graphs above it can be inferred the way the algorithm has clustered the different people groups:

- Cluster 1 corresponds to middle-aged, middle-income families – Middle-aged Professionals
- Cluster 2 corresponds to old middle to low-income couples – Empty Nesters
- Cluster 3 corresponds to middle-income young couples – Young professionals
- Cluster 4 corresponds to diverse-age, high-income and family cluster – Executives

Once the clusters have been formed, it has been proceeded to correlate this with the travel patterns of the people belonging to each cluster. Given the limitations of the lack of POI breakdown mentioned before, it has been intended to prove the clusters correlation with the travel purpose 'Eat', which is in the end restaurants and other eating places. As a working hypothesis, it was intended to prove that either younger or higher-income people would have more visits to restaurants.

6.5.2. Results

However, the results were somewhat unexpected. The clusters formed by the algorithm did not show any difference when looking at the Eat travel purposes (Figure 43).

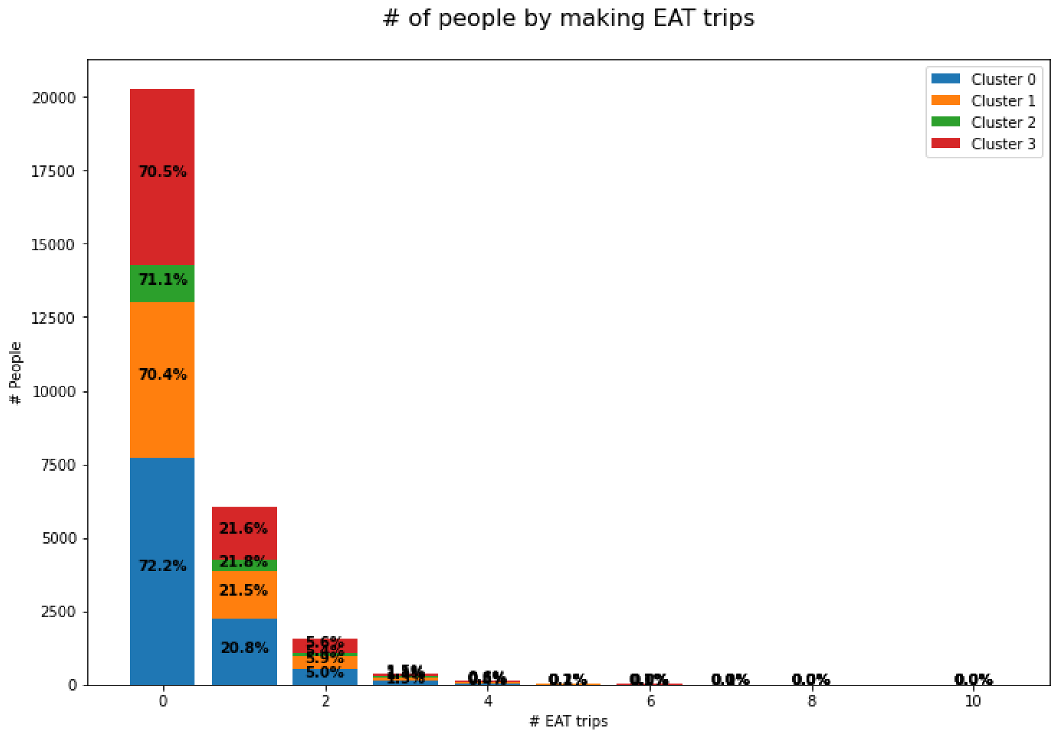

Figure 43. Number of people vs number of Eat trips carried out in a day by cluster

Intending to assess the differences regarding the travel patterns, a count per person of each travel purpose has been carried out (Table 18). Some insights can be extracted. Young professionals show a greater work travels per person that its comparables, which can be due to having a small share of retired-age and young people, the latter group with a higher unemployment rate. However, the results, even if some differences between clusters is shown, are not clear enough to jump into any major conclusions.

Cluster	Home	Recreation	Work	Shop	social	Eat	Maintenance	School	Other
0	1.177	0.172	0.473	0.570	0.365	0.377	0.210	0.001	0.083
1	1.314	0.193	0.499	0.680	0.438	0.412	0.262	0.001	0.102
2	1.180	0.191	0.428	0.593	0.381	0.388	0.238	0.001	0.114
3	1.310	0.195	0.547	0.676	0.438	0.408	0.257	0.000	0.106

Table 18. Count of travel purposes per person for each cluster

Continuing down this road was not thought to be the wisest decision. It would have required a high amount of work to try to get something clear from this analysis. Many other clustering algorithms could have been tried but the results would have been most certainly similar. Consequently, it was decided to give a step back and get simpler, as well as helpful, metrics. In that sense, for age (Table 19) and income (Table 20) variables 4 groups were formed and the metrics per person for each travel purpose were obtained.

Age group	Home	Recreation	Work	Shop	Social	Eat	Maintenance	School	Other
(0, 25]	1.174	0.182	0.465	0.580	0.382	0.379	0.207	0.006	0.086
(25, 45]	1.225	0.174	0.48	0.601	0.383	0.394	0.228	0.000	0.088
(45, 65]	1.323	0.202	0.529	0.691	0.449	0.408	0.262	0.000	0.113
(65, inf)	1.304	0.188	0.537	0.609	0.382	0.368	0.259	0.000	0.089

Table 19. Count of travel purposes per person for each age group

Income group	Home	Recreation	Work	Shop	Social	Eat	Maintenance	School	Other
(0, 75000]	1.21	0.172	0.6	0.629	0.392	0.402	0.229	0.001	0.088
(75000, 150000]	1.246	0.185	0.521	0.63	0.398	0.396	0.239	0.001	0.096
(150000, 250000]	1.28	0.187	0.462	0.639	0.423	0.4	0.241	0	0.1
(250000, inf)	1.293	0.196	0.432	0.628	0.411	0.39	0.246	0	0.101

Table 20. Count of travel purposes per person for each age group

6.5.3. Analysis of results

There are some insights that can be drawn from this summary tables:

- Regarding Eat travel purposes, the visitation patterns are shown to be similar regardless of the age or income groups.
- Shopping is the most frequent travel carried out in one day. Middle-aged people are more prone to do so with no clear differences when looking into income.
- Same happens with Social activities, which refer to people visiting friend's houses
- With Work travels an interesting insight can be extracted. As people get older, they travel to work more frequently, which can lead to believe that younger people are more prone to work from home. Also, lower-income people travel due to work more often, concluding that their jobs might require a more on-site presence.

6.6. Safegraph & Replica

The analysis carried out so far intended to extract insights from the both databases available separately; POI conclusions based on Safegraph's data and trip and people information based on Replica's.

In this chapter, it is intended to merge both Replica and Safegraph databases so as to obtain a practical tool that can associate each trip carried out by a given person to a POI. This chapter aims to bring together all the knowledge acquired during the application of the previous ones. This tool can later aid in the decision-making of defining the number of amenities needed in a given area based on the age and income profile of the residents.

6.6.1. Process

To start with, a more ambitious region has been considered to analyze the data. Instead of just focusing on Kendall Square or similar places, the Boston Metropolitan Area has been chosen as a greater diversity of people is achieved (Figure 44). To be more precise, the people living in that area and the trips they make have been analyzed. Previous analysis were intentionally constrained to some places, which might have led to a sampling bias, in which working age, high-income people were either predominant or over sampled.

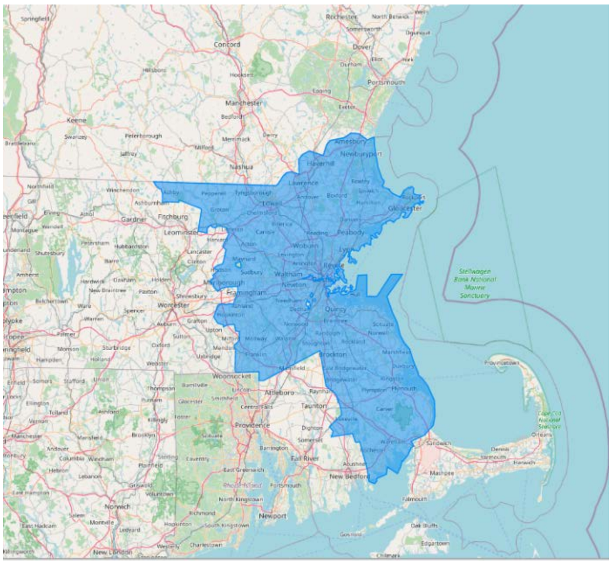

Figure 44. Map of the Boston Metropolitan Area

Before jumping into any analysis, Replica's data scientists have been contacted in order to increase the quality of their data. The databases at hand would just offer the travel purpose and the destination block group of the trip as a means of knowing the POI people were visiting. In that sense, it was demanded a North American Industry Classification System (NAICS) code to be included. The NAICS code is a six-digit number that refers to every industry. Replica did not have this level of accuracy in their data, so, instead, they just added the NAICS category, which is the first two digits. Latitude and longitude coordinates of the destination were also a request, although the accuracy of these was not ideal.

The association between replica and Safegraph databases has had two iterations.

At first, since latitude and longitude coordinates of the trip destination were made available, these were used to do the association. Making use of scipy's library, spatial's kdtree function was used to obtain the nearest neighbors of any point. In that sense, the 10 closest amenities to the trip's endpoint were found. In amenity-dense areas, these 10 amenities would be enclosed within a 250 meter radius from the endpoint; that was not the case for lower amenity dense-areas. That is why, a 250 meter radius was defined, in which the closest amenities had to be enclosed. This way of approaching the problem would have been appropriate if the endpoint data was reliable. However, later on, the advice from Replica's data scientists was not to make use of the latitude and longitude coordinates data and use the block group data instead, which was less precise but much more reliable.

As such, in order to associate the POIs that could have been the destination of a trip, the POIs enclosed within a block group were selected. Alongside this geospatial requirement, POIs that matched the NAICS category were filtered. For a given trip, many POIs could comply with the geospatial and NAICS category conditions. In that sense, the way to associate a specific POI to a trip was a weighted randomization. The options to choose from were the POIs that complied with the requirements, while the weights were the Safegraph's recorded visits. This way, POIs that had more registered visits by Safegraph were more prone to be the destination of a Replica trip.

The following script shows the process mentioned above to do the association between both databases. It is done for a given Thursday, but the process for a given Saturday is analogous.

```
# trips_bma_th: dataframe containing the trips to amenities carried out by Boston
 Metropolitan Area population (~6M rows)
# safegraph_ma: dataframe containing POI information of MA

# Initializing the placekey column
trips_bma_th['placekey'] = ''
# For loop running through the trips_bga_th dataframe
for i in range(len(trips_bma_th)):
  # Extracting the block group and NAICS category of each trip
  destination_bg = trips_bma_th.at[i, 'destination_bgrp']
  naics_cat = trips_bma_th.at[i, 'naics_category']
  # Filtering POIs inside the block group with that NAICS category
  pois_in_bg = safegraph_ma[(safegraph_ma['naics_category'] == naics_cat) & (safe
graph_ma['poi_cbg'] == destination_bg)]
  # If the dataframe where the POIs that meet the conditions is empty then, enter
  if (pois_in_bg.empty == False):
  # Weighted randomization association based on Safegraph's recorded visits
    trips_bma_th.at[i, 'placekey'] = random.choices(list(pois_in_bg['placekey']),
 list(pois_in_bg['raw_visit_counts']))[0]
  if i%100000 == 0:
    print(i)
trips_bma_th
```

Table 21 shows a sample of the Replica and Safegraph association database. The first columns correspond to the Replica database while the last one (*placekey*) is the result of the association with the Safegraph database. Each *placekey* is unique for each POI and has a linked POI name, category etc.

activity_id	destination_bgrp	person_id	naics_category	placekey
3249348448254220000	250010104002	5801932163825710000	62	224-223@62j-y9f-pgk
240410798909245000	250010120012	6187529526728990000	23	222-223@62j-ymn-dy9
2992220528788860000	250010127002	16279219265392000000	53	222-222@62k-3q3-m8v
13028753685376400000	250010137001	6110800023112910000	54	zzw-222@62k-3yf-k75
8878259724321330000	250010137001	13152676800029600000	54	zzw-222@62k-3yf-k75

Table 21. Sample of Safegraph and Replica association database for a given Thursday

6.6.2. Results

Having obtained Table 21, this opens a window of opportunities. On one hand, the number of total visits each POI receives on a given weekday or weekend day can be obtained (Table 22).

placekey	location_name	top_category	count_visits
zzw-222@62j-sj7-jd9	Boston Logan International Airport Terminal A	Support Activities for Air Transportation	12302
zzw-222@62j-sz7-dgk	Gateway Center	Lessors of Real Estate	11929
222-223@62j-ptf-tsq	Costco	General Merchandise Stores, including Warehouses	7613
zzw-222@62k-phk-x89	Chick-fil-A	Restaurants and Other Eating Places	6291
222-225@62j-ptq-kmk	Sephora	Health and Personal Care Stores	6133

Table 22. Sample of count of visits to each POI based on Safegraph and Replica association for a given saturday

Additionally, the profile of the people visiting these POIs can be extracted. Figure 45 shows the age and income distribution for a Costco Store. The blue surface represents the distribution of the Costco visitors while the orange surface represents the Boston Metropolitan Area distribution. Horizontal lines represent extreme values and the median. The comparison with the considered population is deemed necessary to be able to extract valuable conclusions. For this particular case, it can be concluded that Costco tends to have older aged visitors while the population of the area considered is more concentrated in younger residents. Regarding the income, it does not seem to be a great difference between the BMA population and the Cotsco visitors, although Costco visitors might tend to have a slightly higher individual income.

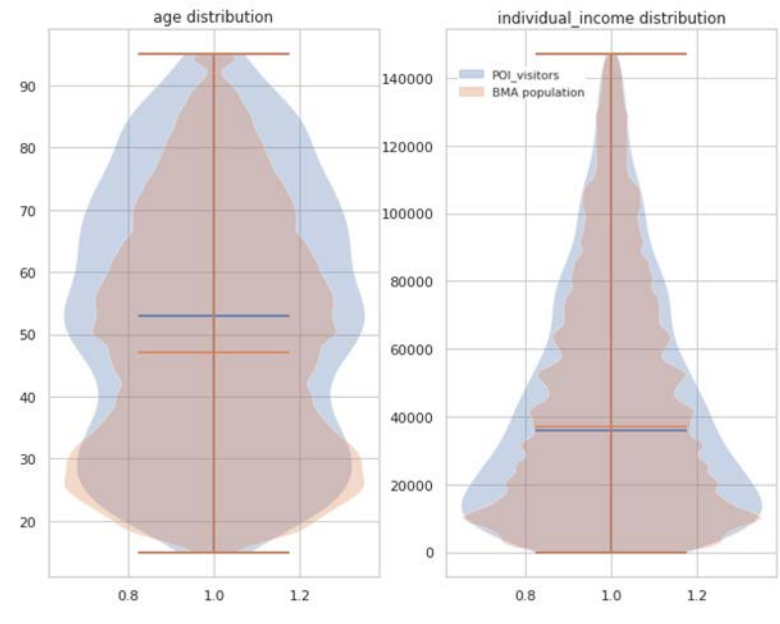

Figure 45. Costco's and BMA age and income distribution

The performed analysis could be similarly applied to POI categories instead of just a POI itself.

Having an extended version of Table 21 for both a given Thursday and Saturday, these have been concatenated to obtain monthly visit numbers. In that sense, the Thursday database has been taken as a reference of a typical weekday; same procedure has been followed for Saturdays and weekend days. This way, the weekday visits have been multiplied by 5 and the weekend day ones by 2 to obtain the weekly visits. In order to get the monthly visits, a multiplier of 4.33 has been applied, resultant of dividing the 52 weeks of the year by the 12 months.

Having the total visits, it has been proceeded to divide the population in age and income groups. The age bins have been established in 25, 45 and 65 years; and the income ones in 10k, 40k, 75k and 150k USD. Having created this age-income groups, it has been proceeded to obtain an aggregated view for each category. In other words, the number of visits each age-income group does to each category has been summed up. Following that, a similar approach to the one carried out in chapter 6.3.1, has been carried out, where an average of visits to a POI has been taken as a reference of the POI visit capacity. In addition, not only the number of POIs, but the area of these was also obtained. Making use of Safegraph's Geometry database, the median square meters of Cambridge and Boston POIs was obtained for each category. With that, Table 23 was obtained as an output which sets another milestone in this project; obtaining metrics for each POI category based on the socioeconomic profile of the people that will make use of them.

Age group	Income group	Top category	Visits pp	N POIs p1000	N sqm p1000
(0.0, 25.0]	(0, 10000]	Book Stores and News Dealers	0.163	0.041	10.886
(45.0, 65.0]	(0, 10000]	Book Stores and News Dealers	0.213	0.053	14.296
(65.0, inf]	(10000, 40000]	Beer, Wine, and Liquor Stores	0.503	0.184	40.024
(25.0, 45.0]	(10000, 40000]	Beer, Wine, and Liquor Stores	0.467	0.170	37.135
(25.0, 45.0]	(40000, 75000]	Restaurants and Other Eating Places	9.015	2.192	320.006
(45.0, 65.0]	(40000, 75000]	Restaurants and Other Eating Places	8.921	2.169	316.686
(45.0, 65.0]	(75000, 150000]	Grocery Stores	2.105	0.391	93.358
(65.0, inf]	(75000, 150000]	Grocery Stores	2.069	0.384	91.759
(0.0, 25.0]	(150000, inf]	Clothing Stores	0.802	0.287	63.375
(25.0, 45.0]	(150000, inf]	Clothing Stores	0.681	0.243	53.755

Table 23. Sample of output table regarding number of POIs and square feet per amenity category depending on age-income group

6.6.3. Analysis of results

In short, a powerful amenity planning tool has been developed. The reach and impact of it is yet to be discovered, although it does set the path for the next researcher to explore on how to drive more important conclusions from the information at hand. At this stage, the analysis performed is just confined for the BMA, although with the required data at hand, it could be extended to other places where new insights could be obtained.

As a final remark, it is believed that the contribution of this tool is of tremendous value, since each anonymized visit has an associated socioeconomic profile and can be linked with a destination POI. The accuracy of the tool goes in line with that offered by Replica, whose data is highly reliable; and the algorithm developed in this project, with this data at its root, is also well-grounded.

6.7. Application in Volpe

Once the tool has been obtained, Volpe in Kendall Square has been used as the use case. The population that wants to be fitted inside this area is mainly concentrated, although not limited to postgraduate students. The age and income groups that of the residents of this area, including MIT staff plus their families, would look something like Table 24. This table has been obtained based on experts' opinion and MIT confidential surveys.

Age group	Income group	Number of people
(0, 25]	(0, 10000]	1000
(25, 45]	(75000, 150000]	8000
(25, 45]	(150000, inf)	2000
(45, 65]	(75000, 150000]	2000
(45, 65]	(150000, inf)	1000

Table 24. Age and Income group distribution of Volpe prospect residents

With this distribution (Table 24) and the metrics obtained in Table 23, the amenity needs of this population are obtained (Table 25).

Top category	N_POIs	N_sqm
Restaurants and Other Eating Places	23.70	3460.70
Museums, Historical Sites, and Similar Institutions	9.56	33325.84
Other Amusement and Recreation Industries	8.29	9772.56
Religious Organizations	5.78	2042.01
Automotive Repair and Maintenance	4.61	1557.67
Grocery Stores	4.04	965.91
Personal Care Services	3.86	597.73
Offices of Dentists	3.71	821.74
Health and Personal Care Stores	3.69	729.75
Elementary and Secondary Schools	3.64	7143.22
Offices of Other Health Practitioners	3.64	697.39
Offices of Physicians	3.06	1787.37
Child Day Care Services	2.74	1756.13

Table 25. Top amenity categories needed by prospect Volpe residents based on the socio-economic profile

The total square meters for this population accounts for 96,300, following this procedure. It needs to bear in mind that the approved plan just projected 95,000 square feet (approximately 9000 square meters), which falls short on the ambition of creating a vibrant, walkable and sustainable district. If the intention is for residents to be able to live and work in Volpe, while also thrive in that same place, an additional 12 floors per building would be needed to be included.

6.8. Comparison with starting point

Back in section 4.2.2 it was stated that César Hidalgo et al.'s work [65] was taken as a starting point. It was considered a good reference from which to expand the knowledge. It is then reasonable to compare the results obtained with the own methodology with their results.

However, some flags need to be raised beforehand. The POI database used is not the same one in both projects. While in this project Safegraph Places has been used, they obtained the information through the Google Places product. Even though it is believed that both data bases contain most of the POIs, some minor differences could be found. Additionally, the POI categories are not the same ones across both databases, which makes it more difficult the one by one comparison.

Having said that, it is proceeded to compare Table 2 retrieved from César Hidalgo et al.'s work [65], with the two milestones that have been accomplished throughout this project. On one hand, the obtained table based on Safegraph visiting patterns (Table 15) and, on the other hand the one obtained with Replica's trip data (Table 23). The result of that comparison is Table 26.

Amenity Category	Safegraph	Replica	Amenity Mix of Urban Neighborhoods
Restaurants and Other Eating Places	3.59	1.31 - 2.66	1.84
Gasoline Stations	0.69	0.23 - 0.49	0.04
Grocery Stores	0.61	0.23 - 0.48	0.25
Religious Organizations	0.61	0.35 - 0.75	0.96
Elementary and Secondary Schools	0.58	0.23 - 0.48	0.76
Clothing Stores	0.25	0.13 - 0.29	0.49
Drinking Places (Alcoholic Beverages)	0.25	0.09 - 0.2	0.35
Furniture Stores	0.10	0.05 - 0.1	0.20

Table 26. Sample of the comparison of POIs per 1000 residents between Safegraph-based, Replica-based methods & Cesar Hidalgo et al.'s work [65]

To assess the differences between the three methods, there are some particularities that need to be acknowledged:

- The way to obtain the results varies from method to method. While in Safegraph-and Replica-based methods, visiting patterns to POI are taken as a mean of calculating the required POIs, in the *Amenity Mix of Urban Neighborhoods* [65], a ratio of the number of POIS per 1000 residents is calculated.
- The population considered for each method are different. For Safegraph-based method, the population of Massachusetts minimally monitored urban block groups are considered. For the Replica-based method, on the contrary, the population of Boston Metropolitan Area is taken into account. Lastly, the ratio for the *Amenity Mix of Urban Neighborhoods* [65] has been calculated based on the population of 47 US cities.
- In the *Amenity Mix of Urban Neighborhoods* [65], the authors analyze the POI in various cities, considering the city as a whole. On the other hand, in the Safegraph-based method, a minimum population density is applied to each block group. In that sense, it can happen, that some block groups, even though they are in a city, they do not fall inside Safegraph's methodology since they do not meet that requirement.
- Additionally, the use of different amenity categories across the methods makes it even more difficult to compare the results, since some Safegraph categories might englobe more than one category adopted by Cesar Hidalgo et al. [65]
- Replica-based method is based on age and income groups, so the range between the minimum and maximum is offered.

Taking all of this into consideration, it can be concluded that there are some categories for which Safegraph-and Replica-based methods oversize the POI needs, whereas for others they are undersized, in comparison with Cesar Hidalgo et al.'s research [65] as it can be observed in Table 26. For example Restaurants and Other Eating Places Safegraph-based procedure claims that 3.59 restaurants per 1000 residents are needed in stark comparison with Replica, between 1.31 and 2.66, and the reference, 1.84, methods. However, this might make sense since for the Safegraph-based procedure, uniquely urban block groups are being considered where the agglomeration of restaurants is more present (The *Amenity Mix of Urban Neighborhoods* and replica-based methods also consider suburban areas). On the contrary, Cesar Hidalgo et al. describes a bigger need for Clothing stores, reasonable given the presence of retail stores in suburban areas.

6.9. Energy Tool

The CS group is developing a tool to quantify the energy consumption and the CO2 impact derived from our day-to-day actions, to later propose policies to change our behavior, thus, reducing the carbon footprint. The geographical boundary set for this study is Kendall Square's political boundary plus an additional block group to which Kendall Square is being extended. The population being considered for this study accounts to 66.169 people, englobing people living in the area and commuting into the area.

There are different groups, each responsible for a set of policies. The energy team focuses on local energy production and CO2 capture. The Diet Team studies the impact of the current food consumption to later propose alternatives to reduce it. The Mobility Team is in charge of measuring the impact of transitioning towards a lightweight micro mobility, electrification of the fleet etc. Lastly, the Land Use team assesses the impact of live-work symmetry, walkable amenities, compact housing, building retrofitting and hybrid work. Additionally, there is the user interface team, which is in control of merging all the modules and creating the interactive tool. Even though the teams are focused on one topic, there are some interventions that fall in the conjunction of more than one team, requiring collaboration.

The author of this thesis has been involved with the Land Use Team. In addition to being the coordinator for this team, developing live-work symmetry and walkable amenities policies, more precisely the impact they have in mobility, have been the main tasks.

The ultimate goal of this project is to determine what it would take to reduce the per capita CO2 impact to meet the Paris Agreement recommendations of keeping it below 2 t CO_2 if an increase of temperature above 2 °C is to be avoided [83].

6.9.1. Process

As a general overview, live-work symmetry and walkable amenities interventions are to be analyzed from the mobility side. To assess the live-work symmetry intervention, it has been assumed that the people that transition to live-work adopt the mobility characteristics of the people already doing live-work in Kendall Square. However, the trips that are affected are just those related to commuting to and from work. Trips to amenities are to be analyzed separately as a means of maintaining both interventions independent. In that sense, walkable amenities intervention assesses how mobility patterns would change if accessibility to amenities was increased. These interventions are implemented as sliders from 0 to 100, and the user details the percentage of each intervention that wants to include in the model.

In the following lines, the step-by-step process followed to develop the live-work symmetry and walkable amenities sliders is explained.

First, the databases have been imported. The trips table included the following columns:

- Activity_id: unique id for each trip
- Person_id: unique id for each person doing the trip
- Tour_type: type of trip
    - Commute: trip from home to work and vice versa
    - Other home based: trip with origin or destination at home
    - Work based: trip with origin or destination at work
- Mode: mobility mode used for the trip
- Distance_meters: distance traveled during the trip

Also, a population table has been imported where information regarding the block group of residence, work or school was provided. With this information, it has been defined the commute status of every person considered for this study. The distribution of people is as follows (Table 27):

Commute Status	Live	Work	Number of people
Commute_in	Outside KSQ	KSQ	56744
Commute_out	KSQ	Outside KSQ	4257
Live_work	KSQ	KSQ	2130
Live	KSQ	N/A	3038

Table 27. Population distribution by commute status

The people commuting in account for the people that live outside Kendall Square but work there. The exact opposite goes for the people commuting out. Live-work people are understood as the people that live and work in Kendall Square, while live people are those that live in Kendall Square but there is either no information about their workplace, or they do not work.

With that in mind, intending to work with more manageable databases, trips have been classified into the following distance categories in kilometers: (0, 1], (1, 2.5], (2.5, 5], (5, 10], (10, inf). Following that, the trips database has been divided in three, one for each tour type. The reason behind this decision is that the live-work symmetry and walkable amenities policies do not affect equally to each trip type. Later, all the trips have been grouped by the commute status of the people making the trip, the mode of transport used and the distance group. Each of these, the unique combination of the three columns, is referred as trip group from now on.

Regarding the live-work symmetry intervention, people would adopt the mobility behavior of the people living and working in Kendall Square. But they do not do so for every single trip. The trips affected by the live-work symmetry intervention are those related to commuting to-and-from work. Work-based and other home-based trips are not altered. This implies that people would still maintain their travel patterns to other non-work/home destinations, even though they have adopted the live-work symmetry. The application of the slider works as a percentage. Depending on the percentage of live-work symmetry intended to be achieved, the same amount of people that are not currently doing live-work symmetry would transition to that status. It is not the scope of the author's work to determine the increasing needs of housing that would be required, although that demand is being considered by other members of the Land Use Team.

The walkable amenities intervention is just confined to the selected geographical boundary. For that reason, people living in Kendall (live_work, commute_out, and live) see their other home-based trips changed by this intervention. People working in Kendall (live_work, and commute_in) would alter their work-based trips. Having the 15-minute city in mind, it is proposed that with a 100% of walkable amenities residents and workers in Kendall Square would find everything they need within a 1km distance. Therefore, the percentage of this ideal scenario determines the percentage of trips reduced to a distance below 1 km.

Table 28 helps better understand the impact each intervention has on each tour type trips of each commute status. (C: commute, W: work-based, H: other home-based).

Intervention by tour type		Commute status			
Intervention	Tour type	Commute_in	Commute_out	Live_work	Live
Live-work symmetry	C	Yes	Yes	N/A	Yes
	W	No	No	No	No
	H	No	No	No	No
Walkable amenities	C	No	No	No	No
	W	Yes	No	Yes	No
	H	No	Yes	Yes	Yes

Table 28. Application of the interventions depending on tour type and commute status

To finish with, some useful metrics have been calculated (Table 29):

Metric	Desctiption
Avg_dist (m)	Average distance of the trips belonging to each trip group, in meters
Count_trips	number of trips of each trip group
Kms_traveled	total kilometers traveled
N_trips_S	number of trips of each commute status, equal across all the groups that share commute status
N_people_S	number of people of each commute status, equal across all the groups that share commute status
%_mode_S	percentage of trips that each mode of transport represents of the total trips carried out by a commute status
%_distGroup_S	percentage of trips that each distance group represents of the total trips carried out by a commute status
%_trips_S	percentage of trips each trip group represents of the total trips carried out by a commute status
%_trips_T	percentage of trips each trip group represents of the total trips carried out
%_kms_S	percentage of kms each trip group represents of the total trips carried out by a commute status
%_kms_T	percentage of kms each trip group represents of the total trips carried out

Table 29. Metrics calculated for the trips after the live-work symmetry and walkable amenities interventions have been applied

As things stand, the live-work symmetry and walkable amenities sliders only incise on the distance of the trip, preserving the mode of transport originally chosen. For instance, if a car a car is used for a 10 km trip; now that the distance gets below 1 km, the person still decides to drive the car. It is therefore acknowledged that the module is incomplete and that is why it is passed along to the mobility team. They are responsible of applying some other modules, like the decision of the mode of transport based on the distance, policies to promote lightweight mobility, and electrification of the fleet.

6.9.2. Results

As it has been mentioned before, the literature supports the use of the average distance travelled for commute as a proxy for sustainability of the trip [71]. With that in mind, a sensitivity analysis of the average kilometers per person has been carried out (Figure 46), assessing the impact of implementing each intervention.

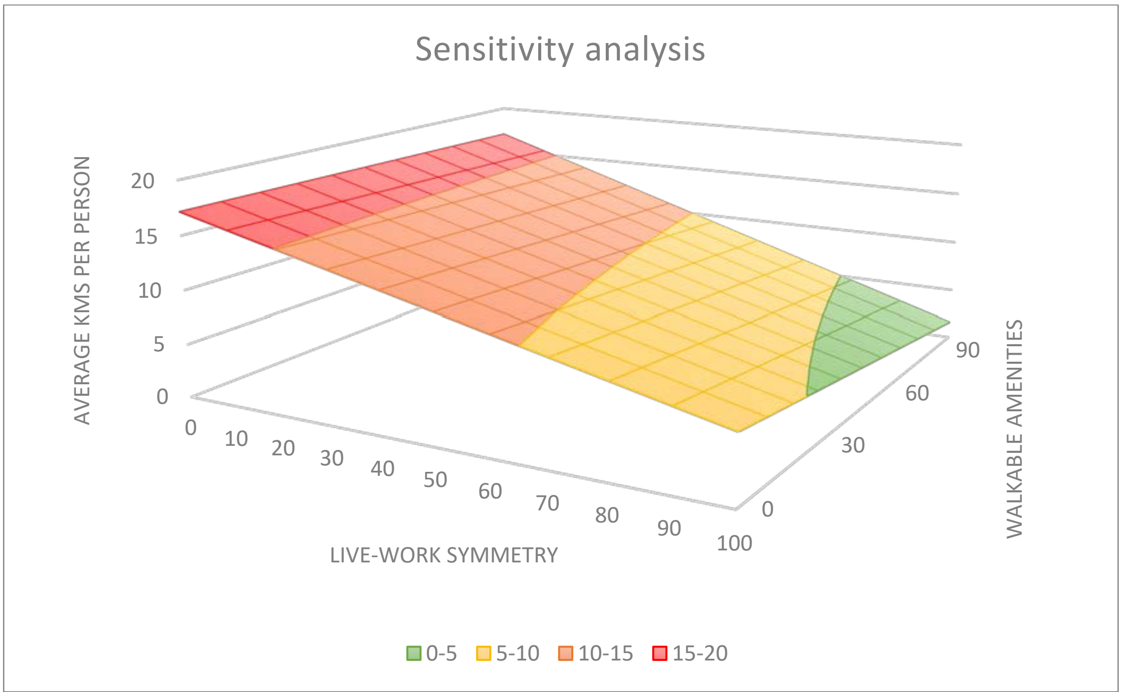

Figure 46. Sensitivity analysis of the average kms per person travelled when applying the interventions

6.9.3. Analysis of results

If the interventions were to be analyzed independently, it is concluded that pleading only for walkable amenities would have a very small impact. Even if 100% walkable amenities are reached, the average distance travelled per person still is in the red area, meaning an average above 15 kilometers per person. On the other hand, live-work symmetry on its own does have a significant impact going from the 15-20 kilometers (red) to 5-10 kilometers (yellow). However, the greatest impact is achieved when both interventions work jointly. If a 100% live-work symmetry was to be applied, meaning 100% of the people would live and work in Kendall Square, and a 100% walkable amenities, finding in a 15-minute walking radius everything needed in the daily life, the average distance travelled per person would fall below the 2 kilometers per day.

If distances were to be translated to CO2 impact, a reduction could be seen coming from two different sources:

- If travel distances are reduced the total CO2 impact of the trips is therefore reduced (e.g. a 10 km car trip becomes a 1km car trip).
- When distances are reduced, people opt for other modes of transport that originally were not considered (e.g. a 10 km car trip becomes a 1km bike trip)

The first of the sources is understood as the direct impact of the live-work symmetry and walkable amenities interventions. That is the CO2 reduction that would happen if people still maintained their behavior regarding the mode of transport. The second source, though, would be the consequence of applying micro-mobility solutions.

It is not the scope of the Land Use Team to assess the CO_2 impact of the mobility patterns of the populations, since that falls under the Mobility Team filed. However, by applying a proxy, the average kilometers traveled per capita, it has been proved the improvement on sustainability these sliders would have.

7. CONCLUSIONS

The methodology presented in this document enables to expand the knowledge on amenities planning in order to achieve the widely beneficial walkable amenities. Reduction of commuting, vibrant neighborhoods, livable districts and more sustainable areas are the direct consequences of achieving the desired walkable communities.

Leveraging on big data analysis, a contribution to this field has been accomplished. As the first milestone of this thesis, a rule-of-thumb tool was developed to parametrize the amenity needs for urban areas.

Not settling for this, a more case-based approach was developed. The second milestone of this project, englobed a tool detailing the amenity needs for age and income groups. This tool enables a more precise urban planning in newly-developed areas as it has been shown with Volpe.

Finally, as part of a CS lab-wide project a quantification of the live-work symmetry and walkable amenities policies has been done for Kendall Square. The learnings extracted from this exercise prove that the collective application of these interventions is what maximizes the desired result towards a more walkable and sustainable district.

8. DISCUSSION

This thesis is part of the CS Group intention to rethink how cities are built. While this thesis is focused on walkable amenities, other researchers' studies contribute to the vision of building vibrant and livable communities from other points of view. New mobility options, better mobility infrastructure, sensorization of the city etc. are just but examples of the CS work which fall under the same umbrella.

It is the author's belief that this thesis has contributed to better understand amenity needs in communities. Already developed areas can assess whether they lack of any basic amenities, whereas newly developed areas can trust this tool for having an estimate of the demand of different amenities and plan accordingly.

The work developed here does not come to an end with the writing of this thesis. The output and insights obtained from this thesis, will go on to the next step. They will serve as the input for an optimization tool being developed right now at the CS by research scientist Ronan Doorley. His intention is to apply this insights into a generative model which can build a business district from scratch based on some predefined constraints. The inputs that would determine the amenity offer would be the age and income of the prospect residents, just as it has been described in chapter 6.7 with Volpe.

Additionally, it is believed that this tool has no geographical boundaries. The same approach for other cities and countries could be carried out should the data be available. In contrast, if there is no data and an amenity offer based on US standards is to be built, a cost-of-living based normalization could be done to Table 23 in order to have a good proxy on how to approach the walkable amenities issue.

In the same vein, there are two important events worth mentioning: City Science Summit and EmTech. These two events happening in October will have CS researchers as speakers. The contribution of this thesis will be portrayed in two of the talks. On one hand, the already mentioned amenity optimization tool and on the other hand the Energy tool. Following this event, the writing of two research papers is intended, in which oneself would be a contributor.

Looking back there are many things that could have been done differently but which it is believed have helped the learning process. Being too ambitious at first, and wanting to solve many problems at once, lead to ambiguous solutions with no clear application. However, when breaking down the problem into smaller pieces, it is acknowledged the progress made and the final contribution which has helped better understand a long waited issue.

Looking into the future, this is nothing but the first stone into which further research can be sustained. Analyzing not only the categories but also the type of amenities within the category each age-income group visits, can be an interesting approach from which valuable lessons can be extracted. This can lead to defining, for instance, whether discount-sale or luxury brand stores are more appropriate for a district. Moreover, if spending data for the POIs is made available, by correlating visiting and spending patterns, a range of opportunities would get opened.

9. COST ANALYSIS

In the following chapter the costs incurred during the development of this project are detailed.

9.1. Accounting items

The following section breaks down the accounting items one by one.

9.1.1. Software

The software item is composed by the tools and the databases used.

Software tools

Most of the software tools used for this project are open access. The coding and visualization tools are open access, whereas Microsoft 365 is a monthly paid subscription. The software tools used and associated costs are summed up in Table 30.

Name	Developer	Cost [$/mo]	Quantity [mo]	Total [$]
Visual Studio Code	Microsoft	Open access	NA	NA
Jupyter notebook	Ipython Project	Open access	NA	NA
Google Colab	Google	Open access	NA	NA
Kepler.gl	Uber	Open access	NA	NA
CityScope	ML CS Group	Open access	NA	NA
Microsoft 365	Microsoft	2.99	7	20.93
Total				20.9

Table 30. Software tools costs

Databases

The cost of the databases used across the project fall under the Software item. Safegraph (Table 31) has granted an Academia discount thanks to which their data could be used with no cost; whereas for Replica (Table 32) a lab-wide license has been purchased.

Name	Cost [$/row]	Quantity [# rows]	Total [$]	Academia discount [$]	Total after [$]
Safegraph	0.1	838526	83852.6	83852.6	0

Table 31. Safegraph database cost

Name	License fee [$/year]	# of users	License fee [$/year/user]	Monthly fee [$/mo]	Quantity [mo]	Total [$]
Replica	16000	4	4000	333.3	4	1333.3

Table 32. Replica database cost

9.1.2. Labor

The labor costs (Table 33) correspond to the dedication of Pablo Barrenechea as a Visiting Student, Ronan Doorley as Research Scientist 1, Markus ElKatsha as Research Scientist 2, Luis Alonso as Research Scientist 3, and Kent Larson as Principal Investigator.

Name	Cost [$/hr]	Quantity [hr]	Total [$]
Visiting student	10.10	1040	10504
Research Scientist 1	54.7	13	711
Research Scientist 2	55.8	26	1451
Research Scientist 3	57.4	39	2239
Principal Investigator	105	10	1050
Total			15954.5

Table 33. Labor costs

9.1.3. Equipment

Table 34 shows the costs derived from the amortization of the equipment used.

Name	Acquisition fee [$]	Years of amortization [yr]	Monthly amortization rate [$/mo]	Quantity [mo]	Total[$]
Computer	1200	6	16.7	7	116.7
Monitor	200	4	4.2	6	25.0
Desk	800	5	13.3	6	80.0
Chair	300	5	5.0	6	30.0
Total					251.7

Table 34. Equipment amortization costs

9.2. Summary

Table 35 shows the total costs incurred during the project which account to 20345 $.

Item	Partial [$]	Accumulated [$]
Labor	15954.5	15954.5
Software	1354.3	17308.8
Equipment	251.7	17560.4
Indirect Costs (25%)	4390.1075	20344.6
Total		20344.6

Table 35. Total costs of the project

REFERENCES

[1] City Science Media Lab, "City Science," [Online]. Available: https://www.media.mit.edu/groups/city-science/overview/.

[2] B. K. a. J. Wagner, "The rise of innovation districts: A new geography of innovation in America," *Brookings Institution,* 2014.

[3] A. H. Hawley., "Population density and the city," *Demography,* vol. 9, no. 4, pp. 521-529, 1972.

[4] T. Kemeny., "Immigrant Diversity and Economic Development in Cities: A critical Review," 2013.

[5] L. A. P. T. B. G. T. N.-H. W. G. a. K. L. Arnaud Grignard, "The impact of new mobility modes on a city: A generic approach using ABM," *Springer,* pp. 272-280, 2018.

[6] H. M. K. W. A. D. J. Reck, "Mode choice, substitution patterns and environmental impacts of shared and personal micro-mobility," in *: 21st Swiss Transport Research Conference (STRC 2021)*, 2021.

[7] LSE Cities / EIFER, "Cities and Energy: Urban Morphology and Heat Energy Demand, Final report," London, 2014.

[8] E. Glaeser, Triumph of the City: How our greatest invention makes us richer, smarter, greener, healthier and happier, 2012.

[9] G. Sjoberg, "The Origin and Evolution of Cities," *Scientific American,* pp. 54-63, 1965.

[10] TED, "Urbanization and the evolution of cities across 10,000 years".

[11] P. M. Hohenberg, "The historical geography of European Cities: An interpretative essay," 2004.

[12] Investopedia, "What is the American Dream," 01 08 2022. [Online]. Available: https://www.investopedia.com/terms/a/american-dream.asp. [Accessed 31 08 2022].

[13] R. S.-G. María Sánchez-Vidal, "Las ciudades y su rol en el comercio y los sistemas de transporte," *Panorama Social,* no. 32, pp. 49-64, 2020.

[14] S. M. J.-S. P. a. F. R. Jan David Bakker, "Of Mice and Merchants: Connectedness and the location of economic activity in the iron age," *The Review of Economics and Statistics,* vol. 103, no. 4, pp. 652-665, 2021.

[15] United Nations, "World Cities Report 2020," 2020.

[16] Center for Sustainable Systems, University of Michigan, "U.S. Cities Factsheet," 2021.

[17] European Commission, "Reclaiming city streets for people: Chaos or quality of life?," 2004.

[18] C. Moreno, Z. Allam, D. Chabaud, C. Gall and F. Pratlong, "Introducing the "15-Minute City": Sustainability, Resilience and Place Identity in Future Post-Pandemic Cities," *Smart Cities,* vol. 4, pp. 93-111, 2021.

[19] C. Moreno, "The 15-minute city | Carlos Moreno".

[20] R. W. Caves, Enciclopedya of the City, 2005.

[21] L. Mumford, "What is a city?," *Architectural Record,* pp. 92-96, 1937.

[22] McKinsey, "Urban world: Mapping the economic power of cities," 2011.

[23] A. G. Luis Alonso, "Data-driven, evidence-based simulation on urban dynamics. Use case Volpe," 2019.

[24] P. J.-F. C. P. S. e. a. Balland, "Complex economic activities concentrate in large cities," *Nature,* vol. 4, pp. 248-254, 2020.

[25] United Nations, "Sustainable development goals," [Online]. Available: https://www.un.org/sustainabledevelopment/cities/. [Accessed 21 08 2022].

[26] C. M. Guy, "Classifications of retail stores and shopping centers: some methodological issues," *GeoJournal,* vol. 45, no. 4, pp. 255-264, 1998.

[27] F. J. C. G. E. P. G. J. R. S. T. R. S. Ramón López de Lucio, "Centros urbanos frente a nuevas centralidades comerciales. Un análisis del sur metropolitano de Madrid," *Cuadernos de Investigación urbanística,* vol. 14, 1996.

[28] R. L. d. Lucio, "Espacio publico e implantación comercial en la ciudad de Madrid: calles vs grandes superficies," *Red de cuadernos de investigación urbanística,* 2006.

[29] J. Carpio-Pinedo, "Localización y evolución del comercio y servicios a pie de calle en los entornos urbanos. Factores locales frente a multi-accesibilidad. El caso de la ciudad de Madrid," 2014.

[30] M. B. H. C. S. B. B. B. A. P. Shi Kai Chong, "Economic Outcomes Predicted by Diversity in Cities," *EPJ Data Science,* vol. 9, no. 17, 2020.

[31] Quay Communications Inc, "Public Bike System Feasibility Study," 2008.

[32] Government of South Australia, MIT Human Dynamics Lab, Bank SA, Optus and DSpark, MIT Living Lab, "Atlas of Opportunity," [Online]. Available: https://opportunity.mit.edu/.

[33] H. Zhang, "The Impact of Urban Sprawl on Environmental Pollution: Empirical Analysis from Large and Medium-Sized Cities of china," *International Journal of Environmental Research and Public Health,* vol. 18, 2021.

[34] J. K. Peter Newman, The End of Automobile Dependance: How cities are moving beyond car-based planning, Island Press, 2015.

[35] R. S. S. J. A. B. L. D. P. Daniel A. Cox, "Public Places and Commercial Spaces: How Neighborhood Amenities Foster Trust and Connection in American Communities," 20 10 2021. [Online]. Available: https://www.americansurveycenter.org/research/public-places-and-commercial-spaces-how-neighborhood-amenities-foster-trust-and-connection-in-american-communities/. [Accessed 31 08 2022].

[36] A. V. N. Claudia Yamu, "Fractals: a multiscale approach in regional and urban planning strategies," *Handbook on Planning and Complexity,* 2020.

[37] R. L. K. K. W. P. J. Terry Nichols Clark, "Amenities Drive Urban Growth," *Journal of Urban Affairs,* vol. 24, no. 5, pp. 493-515, 2002.

[38] US Bureau of Labor Statistics, "American Time Use Survey," 2019. [Online]. Available: https://www.bls.gov/tus/a2-2019.pdf. [Accessed 15 08 2022].

[39] B. S. F. Alois Stutzer, "Stress that Doesn't Pay: The Commuting Paradox*," *The Scandinavian Journal of Economics,* vol. 110, no. 2, pp. 339-366, 2008.

[40] R. Buderi, "The Where Futures Converge: Kendall Square and the Making of a Global Innovation Hub," MIT Press, 2022, pp. 1-7.

[41] QS, "QS World University Rankings 2023: Top global universities," 08 06 2022. [Online]. Available: https://www.topuniversities.com/university-rankings/world-university-rankings/2023. [Accessed 31 08 2022].

[42] F. B. L. Jeffrey R. Kenworthy, "Automobile dependence in cites: An international comparison of urban transport and land use patterns with implications for sustainability," *Environmental Impact Assessment review,* vol. 16, no. 4-6, pp. 279-308, 1996.

[43] M. Millias, "A Critical Appraisal ff the Design, Construciton and Influence of he Unité d'Habitation, Marseilles, France," *Journal of Architecture and Urbanism,* vol. 39, no. 2, pp. 103-115, 2015.

[44] B. Marchand, "La Ville Radieuse de Le Corbusier: Les Paradoxes d'une Utopie de la Société Machiniste," *Les Cahiers du Développement Urbain Durable,* pp. 63-78, 2016.

[45] J. Jacobs, The Death and Life of Great American Cities, New York, 1961.

[46] TED, "Brilliant designs to fit more people in every city," Boston, 2012.

[47] A. H. S. M. Mikhail Chester, "Parking infrastructure: energy, emissions, and automobile life-cycle environmental acounting," *Environmental Research Letters,* 2010.

[48] R. C. Christopher T. Boyko, "Clarifying and re-conceptualising density," *Progress in Planning,* vol. 76, no. 1, pp. 1-61, 2011.

[49] A. Bendix, "Amazon has opened a homeless shelter in a high-rise on its Seattle campus — here's what it looks like," *Business Insider,* 24 05 2020.

[50] Victoria State Government, "20-minute neighbourhoods," 18 05 2022. [Online]. Available: https://www.planning.vic.gov.au/policy-and-strategy/planning-for-melbourne/plan-melbourne/20-minute-neighbourhoods. [Accessed 22 08 2022].

[51] e-architect, "Project H1 Masterplan in Seoul," *e-architect,* 29 10 2021.

[52] Bloomberg, "Make Way for the 'One-Minute City'," 05 01 2021. [Online]. Available: https://www.bloomberg.com/news/features/2021-01-05/a-tiny-twist-on-street-design-the-one-minute-city. [Accessed 22 08 2022].

[53] The Mayor EU, "Ljubljana - the European capital with no cars in its center," 13 11 2018. [Online]. Available: https://www.themayor.eu/en/a/view/ljubljana-the-european-capital-with-no-cars-in-its-center-1700#:~:text=The%20Slovenian%20capital%20of%20Ljubljana,pedestrians%20rather%20than%20personal%20vehichles.. [Accessed 22 08 2022].

[54] World Economic Forum, "A Look At The Car-Free Zone In The Heart Of Ljubljana, Slovenia," 31 03 2022. [Online]. Available: https://www.weforum.org/videos/a-look-at-the-car-free-zone-in-the-heart-of-ljubljana-slovenia. [Accessed 22 08 2022].

[55] Bloomberg, "Driving Through Paris? That's Going to Get a Lot Tougher in 2024," 18 02 2022. [Online]. Available: https://www.bloomberg.com/news/articles/2022-02-18/paris-is-banning-traffic-crossing-the-city-center-from-2024. [Accessed 22 08 2022].

[56] BBC, "Five of the most radical car-free city experiments," 17 03 2014. [Online]. Available: https://www.bbc.com/future/article/20140318-five-car-free-city-experiments. [Accessed 22 08 2022].

[57] Lloyds Banking Group, "Homes close to a supermarket can boost house prices by more than £21,000," 2018.

[58] Fast Sale Today, "Living near a McDonald's lowers your house price by 24%," 2016.

[59] National Association of Realtors, "Public Transportation Boosts Property Values," 2014.

[60] National Bureau of Economic Research, "School Spending Raises Property Values," 2003.

[61] W. M. F. R. Sebastian Brandt, "Do places of worship affect housing prices? Evidence from Germany," *Hamburg Contemporary Economic Discussions,* no. 48, 2013.

[62] A. S. Gerald A. Calino, "Beautiful city: Leisure amenities and urban growth," *Journal of Regional Science,* vol. 59, pp. 369-408, 2019.

[63] R. Oldenburg, Celebrating the Third Place, 2001.

[64] R. S. S. J. A. B. L. D. P. Daniel A. Cos, "American Survey Center: Public Places and Commercial Spaces: How Neighborhood Amenities Foster Trust and Connection in American Communities," 20 10 2021. [Online]. Available: https://www.americansurveycenter.org/research/public-places-and-commercial-spaces-how-neighborhood-amenities-foster-trust-and-connection-in-american-communities/. [Accessed 22 08 2022].

[65] E. C. A. S. César A. Hidalgo, "The Amenity Mix of Urban Neighborhoods," *Habitat International,* vol. 106, 2020.

[66] K. H. B. V. Erik Elldér, "When local access matters: A detailed analysis of place, neighbourhood amenities and travel choice," *Urban Studies,* vol. 59, no. 1, pp. 120-139, 2022.

[67] A. S. F. W. H. S. Petter Næss, "Residential location, commuting and non-work travel in two urban areas of different size and with different center structures," *Progress in Planning,* vol. 128, pp. 1-36, 2019.

[68] D. C. X. D. A. P. Esteban Moro, "Mobility patterns are associated with experienced income segregation in large US cities," *Nature communications,* vol. 12, 2021.

[69] MIT Media Lab Human Dynamics, Universidad Carlos III de Madrid, "Atlas of inequality," [Online].

[70] B. G. B. B. X. D. A. '. P. E. M. Takahiro Yabe, "Behavioral changes during the pandemic worsened income diversity of urban encounters," 2022.

[71] T. N. F. W. Kobe Boussauw, "Relationship between Spatial Proximity and Travel-to-Work Distance: The Effect of the Compact City," *Regional Studies,* vol. 46, no. 6, pp. 687-706, 2012.

[72] J. K. A. S. Edward L. Glaeser, "Consumer City," *Journal of Economic Geography,* vol. 1, no. 1, pp. 27-50, 2001.

[73] Energy Star, "What is Energy Use Intensity (EUI)?," [Online]. Available: https://www.energystar.gov/buildings/benchmark/understand_metrics/what_eui. [Accessed 28 08 2022].

[74] S. C. K. S. M. W. A. M. S. S. R. C. C. K. L. Agnis Stibe, "Empowering Cities for Sustainable Wellbeing," in *11th International Conference on Persuasive Technology*, Salzburg, 2016.

[75] K. L. Agnis Stibe, "Persuasive Cities for Sustainable Wellbeing: Quantified Communities," *Springer International*, 2016.

[76] J. M. B. Pavón, "Dynamic Urban Planning: Scale to Measure Urban Resilience," 2021.

[77] H. J. Mihai Tivadar, "Endogenous generation of amenities and the dynamics of city structure," *Regional Science and Urban Economics*, vol. 47, pp. 49-56, 2019.

[78] V. H. Duncan Balck, "Urban evolution in the USA," *Journal of Economic Geography*, vol. 3, pp. 343-3722, 2003.

[79] US Census, "2020 Census Demographic data Map Viewer," [Online]. Available: https://mtgis-portal.geo.census.gov/arcgis/apps/MapSeries/index.html?appid=2566121a73de463995ed2b2fd7ff6eb7. [Accessed 30 08 2022].

[80] MIT, "MIT Volpe Redevelopment Project: Planning Board 07/20/21," Cambridge, 2021.

[81] US Census, "Places".

[82] Six Sigma Study Guide, "Z Scores (Z Value) & Z Table," [Online]. Available: https://sixsigmastudyguide.com/z-scores-z-table-z-transformations/#:~:text=A%20Z%20score%20of%203%20refers,covered%20by%20the%20z%20score.. [Accessed 09 09 2022].

[83] United Nations, "Nationally determined contributions under the Paris Agreement," Glasgow, 2021.