A Review of Urban Science Methods: a case study of Mexico City

Contributors:

Bugg, Caleb
Industrial Engineering/Operations Research (IEOR)
The University of California, Berkeley (UCB)

Fighiera, Vincent
Systems Engineering
Civil/Environmental Engineering
UCB

Kho, Jeremy
Systems Engineering
Civil/Environmental Engineering
UCB

Melikov, Pierre
Systems Engineering
Civil/Environmental Engineering
UCB

Contents

Project Background

Mexico City is one of the largest cities in the world, with over 21 million people in the greater metropolitan area. It is also one of the most important cultural and historical centers in the Americas. With such a large amount of people and a high level of vibrancy, mobility in the region can be quite a challenge.

Recently, a major household travel survey was completed for the Metropolitan Zone of the Valley of Mexico called ‘Encuesta Origen-Destino en Hogares de la Zona Metropolitana del Valle de México’. Conducted by INEGI (Instituto Nacional de Estadistica y Geografia) from January to March of 2017, the survey obtained information to facilitate better understanding of mobility of the inhabitants in the metropolitan region. This included data on trip generation, trip attraction, mode choice, trip purpose, trip duration, socio-demographics, and more, which is representative of 34.56 million daily trips occurring in the metropolitan zone.

Project Objective

The objective of the project is to gain deeper understanding about the lifestyles and mobility of the inhabitants of Mexico City to help devise better public policies, and in this way to improve the quality of life and social equity in the region.

This objective will be accomplished through and accompanied by a review of urban science methods and tools, such as principal component analysis, percolation theory, fractals, and the extended radiation model.

Literature Review

Here we present a brief review of the literature, as it relates to urban data science.

Principal component analysis (PCA) and K-Means Clustering
Principal component analysis (PCA) is one of the most valuable results from applied linear algebra. The goal of principal component analysis is to compute the most meaningful basis to re-express a noisy data set. The hope is that this new basis will filter out the noise and reveal hidden structure (Shlens, 2005). In our analysis, a “best expressed” PCA analysis is measured in terms of variance explained by the principal eigenvectors. Therefore, PCA is a great classification tool for feature spaces (independent variables) with large cardinality.

Application to urban data science can be found in a landmark classification paper concerning classification and equity analysis of an urban park system, where the researchers conducted analysis concerning the “multi-dimensional classification of urban parks in Phoenix reflecting their unique physical, spatial, landcover, and built characteristics” (Ibes, 2014). They explain “A Principal Component Analysis (PCA) was then applied… to reduce overlap and redundancy in the independent variable to reveal which unique combinations of factors explain the majority of the variance in park and neighborhood characteristics.” Furthermore, researchers employ K-means clustering, which they describe here “... a cluster analysis was run using the factors identified in the PCA… K-means was then applied to form the clusters, assigning each park a specific group… based on its similarities with respect to the factors.” Then, clustering based on the extracted principal components, e.g. K-means clustering, is a natural additive to PCA theory, and is used in our analysis of the modes of mobility for the surveyed population.

Percolation theory
According to the Makse, “Cities grow in a way that might be expected to resemble the growth of two-dimensional aggregates of particles, and this has led to recent attempts to model urban growth using ideas from the statistical physics of clusters” (Makse et. al, 1995). In this paper, Makse et. al introduce the correlated percolation model which is “motivated by the fact that in urban areas development attracts further development.” Percolation theory allows us to identify areas with a dense presence of Points of Interest (POI), and to then predict city, deman, population, disease, etc. growth based on distance from the centroid location. With 20+ years of development following this publication, percolation theory has become a well-known urban data science tool for growth prediction in physical systems, such as cities, that do not extend by any “regular” pattern. This prediction capability is expressed in the aforementioned paper by noting that the correlated percolation model “offers the possibility of predicting the global properties (such as scaling behavior) of urban morphologies”.

Fractals

Fractal geometry is has been popularly termed the “geometry of chaos” as its use in diverse fields, from physics and biology, to mathematics and geography lend authority to its ability to retrieve underlying structure from seemingly chaotic physical phenomena. Without question, the quintessential resource for a high-level, and in-depth, understanding of fractals in urban data science is the book by Michael Batty and Paul Longley entitled Fractal Cities: A Geometry of Form and Function. In the preface, the authors motivate the material by saying of fractal theory “It burst onto the academic stage… around the idea that the world is chaotic, discontinuous, irregular, in its superficial physical form but that beneath this first impression lies an order which is regular, unyielding and of infinite complexity.” To this end, our paper uses fractal theory to describe the outward growth from centroids of interest based on the “dimensionality of the fractal”.

The extended radiation model

The radiation model was adapted to use in urban data science as an alternative tool to the prevailing gravity law for the prediction of “population movement, cargo shipping volume and inter-city phone calls… ” (Simini et. al, 2012) which relies heavily on parameters dependent on the study area, making generalization difficult and consistent model calibration necessary. In this paper, researchers “introduce a stochastic process capturing local mobility decisions that helps… analytically derive commuting and mobility fluxes that require as input only information on the population distribution.” Thus, this paper introduces the radiation model. Lastly, the paper states “The resulting radiation model predicts mobility and transport patterns observed in a wide range of phenomena… Given its parameter free-nature the model can be applied in areas where we lack previous mobility measurements, significantly improving the predictive accuracy of most of the phenomena affected by mobility and transport processes.”

In our project, the extended radiation model is used origins-destinations (OD) pairs to determine the probability distribution of the trips in our dataset in regards to distance from the origin. For infrastructure purposes, this would help us to advise local municipalities as to the populations visiting their POI’s, as well as provide evidence for the construction of POI’s to better serve their populace.

This concludes our summary of pertinent literature, and our methodology and results follow.

Obtaining Places of Interest (POIs)

While travel surveys provide a wealth of data, they are very expensive and time-consuming. For most major cities, these are conducted once a decade; for smaller cities and towns, it's probably more seldom than that. During that time, a lot can happen that could change the dynamic of the city - new attractions, redevelopment of entire city blocks, changing economic trends, impact of a natural calamity, or just the gradual shift of a city's characteristics. As such, other sources of data could prove to be useful as proxy to the data obtained in conventional surveys.

Collecting Google Places of Interest

One such potential is points of interest registered on Google Places, which is extensive, updated frequently, and relatively accessible for most people. Google Places lists all kinds of establishments, such as restaurants, schools, offices, and hospitals, allowing it to serve as a good indicator of trip attraction or where people are going. However, Google also places a limit on the number of API requests an account is allowed to make in order to separate commercial applications from non-commercial applications. While the conduct of this project is non-commercial, the data that needs to be collected far exceeds Google's limitations. Hence, an efficient algorithm needed to be implemented along with the use of multiple Google accounts.

In order to collect the data efficiently, the Google Places API and Uber's H3 API had to be utilized through a script created by Fahad Alhasoun. Because of the size of Mexico City, the entire area of study was divided into its 194 political districts, which is also the geographic basis of the 2017 travel survey. A snapshot of how this algorithm works can be visualized below:

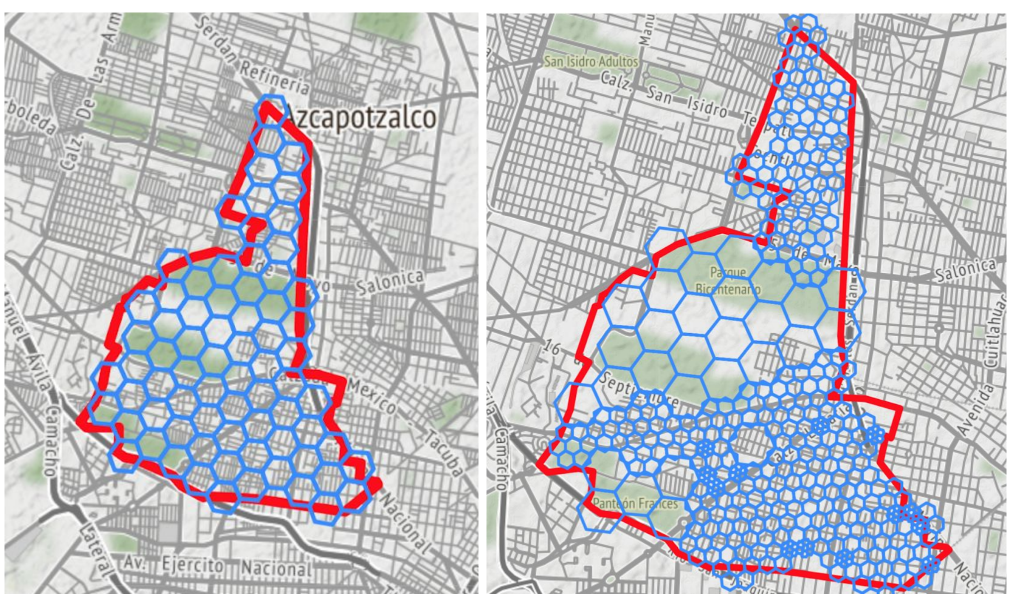


Figure 1. a. Initialized Stage of Algorithm, b. Final Stage of Algorithm

Uber’s H3 system is an application of the concept of fractals. Areas are divided into hexagons, each recursively further divided into smaller hexagons. Hexagons are used because they are good approximations of circles, while minimizing overlap between cells. This is useful as the Google API requires a radius parameter in which the search for places will be made.

The image on left of Figure 1 shows the initialized stage of the algorithm for one particular district. An initial resolution for the size of the hexagons is determined. The larger the initial resolution, the more efficient the script is likely to run, but large resolutions also increase the marginal area unaccounted for near the borders. For each hexagon, a Google Places API request is made. If the request reaches the maximum number of places it can return, the script will divide that hexagon into smaller hexagons and repeat.

As we can see on the image on the right of Figure 1, some areas of the city do not need multiple API requests as there aren’t many establishments to return, such as parks and nature reserves. Downtown city blocks, on the other hand, are recursively splintered until API requests can be made to cover all the establishments. Note: For some areas, such as those with high-rise buildings, the number of establishments on a specific coordinate is larger than the limit that Google Places can return. In this scenario, fragmenting the hexagon will not make a difference and the script eventually crashes. This is a challenge that should be solved moving forward.

Trip Attraction Comparison

In the end, the script was able to return a total of over 733,000 Google Places of Interest across the whole area of study. Plotting these per district, a direct comparison can be made with trip attraction reported in the 2017 travel survey.

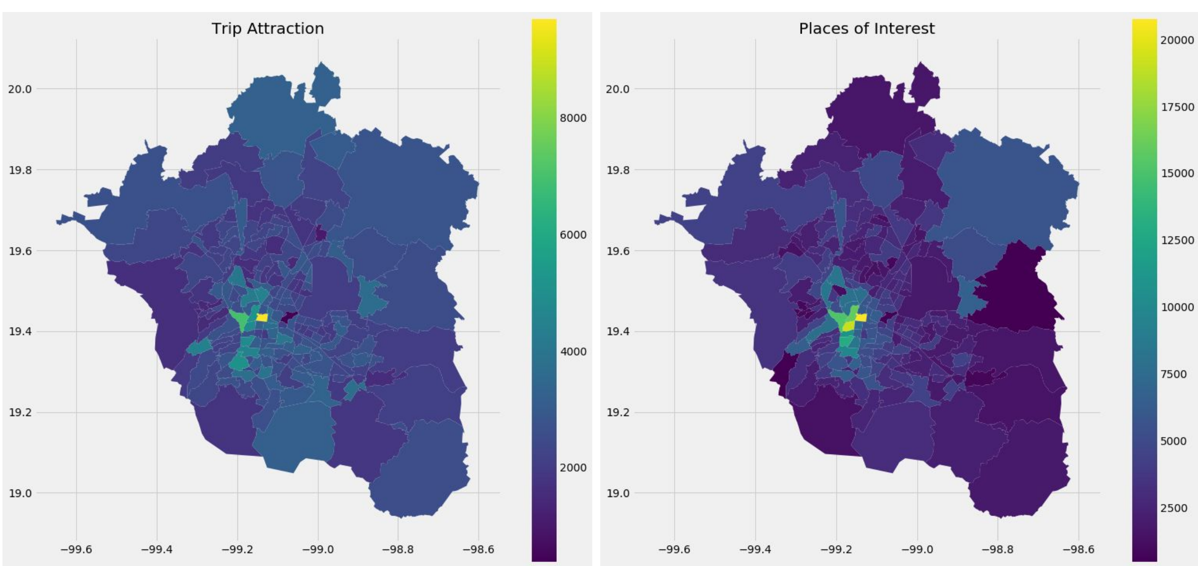


Figure 2. a. Trip Attraction (2017 Travel Survey), b. Places of Interest (Google Places)

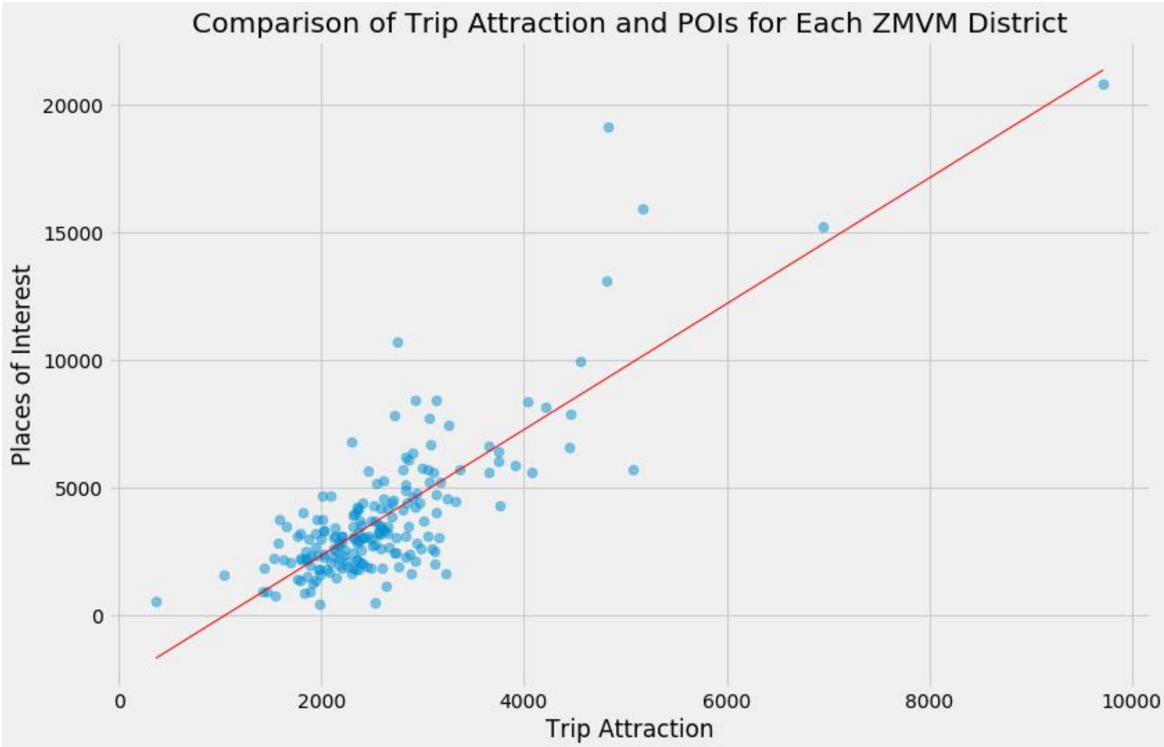


Figure 3. Comparison of Trip Attraction and POIs for Each ZMVM District

While the comparison is not perfect, the distribution of Places of Interest makes a good approximation of the distribution of Trip Attraction. Most notably in Figure 2, the difference between the city center and the rest of the region is similarly stark. After plotting the data in Figure 3, the correlation coefficient of the two variables was determined to be quite high at **0.81**. This comparison will be of great use in a later section, where the Places of Interest are used to model mobility patterns in the city, in place of travel survey data.

Moving forward, one potential path to explore is the study of which districts are underrepresented by the Google Places of Interest compared to the actual trip attraction. What is the cause of such patterns? What are the socio-demographic profiles of the people who live or work in these districts? The data from Google Places can even be filtered to show the distribution of essential services and facilities, such as schools and hospitals, to be able to identify communities that need additional resources allocated to them.

Binning of Places of Interest to a Grid

Counting the number of Places of Interest per district is necessary for direct comparison with the 2017 travel survey data. However, for other purposes, it would be useful for the Places of Interest to be on a coordinate system. Instead of storing the specific coordinates of each establishment, they were assigned to bins, such that the map is divided into a grid. This will be useful later on for percolation and clustering of Places of Interest.

To keep it simple, binning was implemented by rounding coordinates to 3 decimal places. For Mexico City, this approximates to 100 meters per direction. Hence, each grid cell is approximately 1 Hectare or about 1 city block. Sorting the grid cells by the number of Places of Interest, we can easily ascertain the city blocks with the most activity, shown in the table below.

	lng	lat	Count
47405	-99.135	19.433	278
47408	-99.135	19.436	241
45316	-99.141	19.433	222
45314	-99.141	19.431	197
47041	-99.136	19.433	188

Table 1. City Blocks with the Most Places of Interest

In Table 1, we can numerically observe that the city blocks with the highest amounts of Places of Interest seem to be pretty near each other, with their longitudes and latitudes just a few thousandths different.

Looking into Google Maps, we can quickly see where these places are. Unsurprisingly, these blocks are found in the district called Centro Historico, which was the district highlighted in yellow in Figure 2. Looking at the map, numerous prominent establishments seem to be in this area, including museums, government offices, metro stations, plazas, malls, and more.

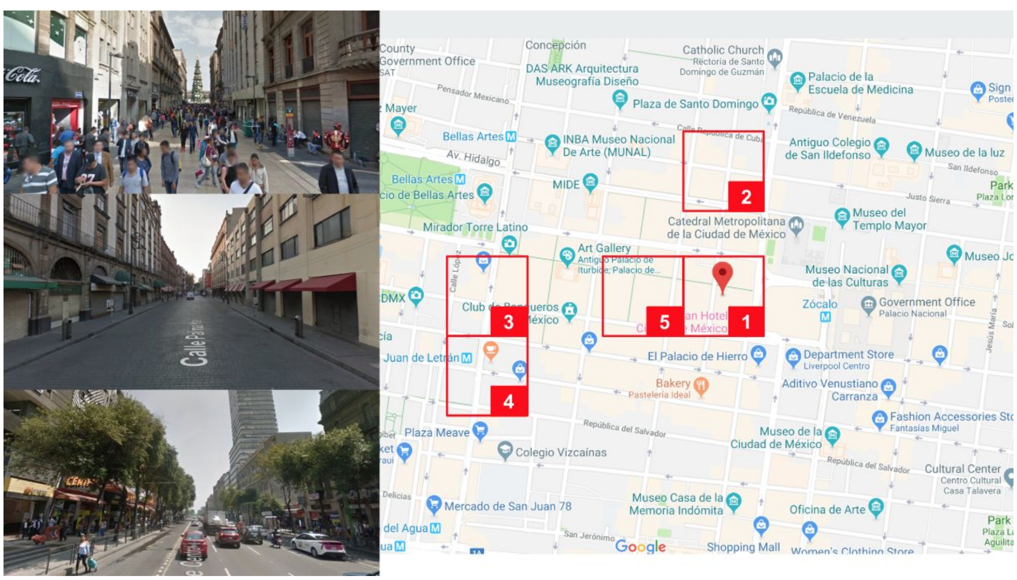


Figure 4. City Blocks with the Most Places of Interest (Mapped)

Evidently, this is the city center of Mexico City. Let us now look at the frequency distribution of the number of Places of Interest per grid cell, plotted below.

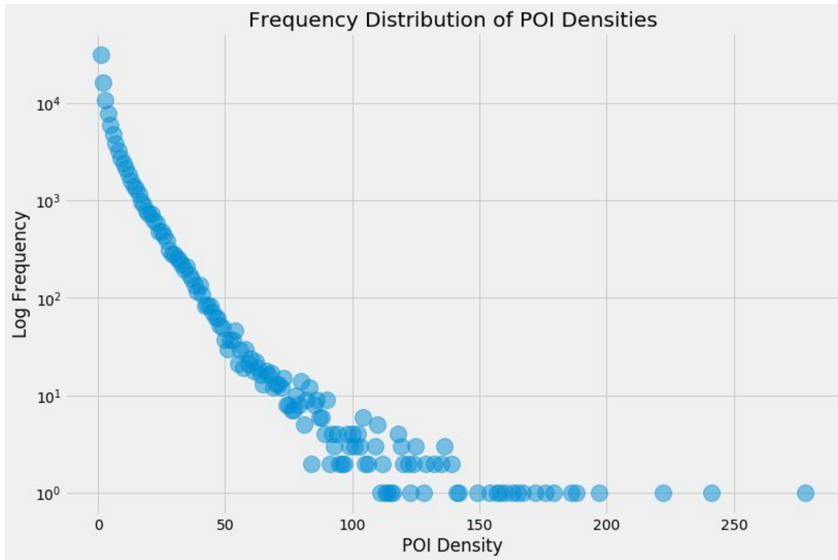


Figure 5. Frequency Distribution of Places of Interest Per Grid Cell

As expected, the frequency distribution shows an exponential decay. Very few blocks have high densities, while numerous blocks have fewer Places of Interest. Figure 6 shows the distribution of densities over distance, where an exponential decay from the city center is also observed. Generally, the further from the center (r=0), the lower the density of Places of Interest per cell.

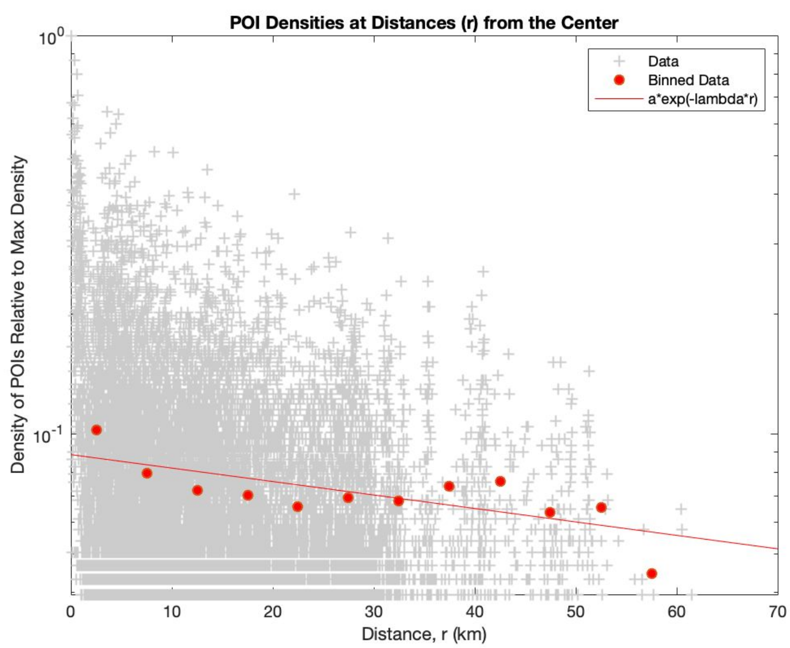


Figure 6. Distribution of the Density of Places of Interest over Distance

Determining Trip Attraction Clusters

One of the limitations of the 2017 travel survey is that the spatial granularity of the data is limited to the political districts of Mexico City. In actuality, trips assigned to a district are not distributed equally. For instance, less urbanized areas would have a town center, where most economic activities occur. Furthermore, political boundaries are generally arbitrary and may be transcended by functional city clusters.

Setting the Percolation Threshold

In order to form such functional clusters, the binned Places of Interest obtained in the previous section was plotted on the shapefile of the area of study. The percolation threshold was then adjusted until distinguishable clusters were formed. Although this process may seem subjective and arbitrary, the percolation threshold decided on is dependent on the objective of the study. For this project, the threshold selected was at 40 Places of Interest per cell block. With only 1.44% of blocks remaining after this threshold selection, it is a fairly strict cut-off. However, lower values for the threshold would make the central section of Mexico City indistinguishable with just one large cluster. Figure 7 below provides visualization of the cells with Places of Interest and the corresponding cells after the threshold was applied.

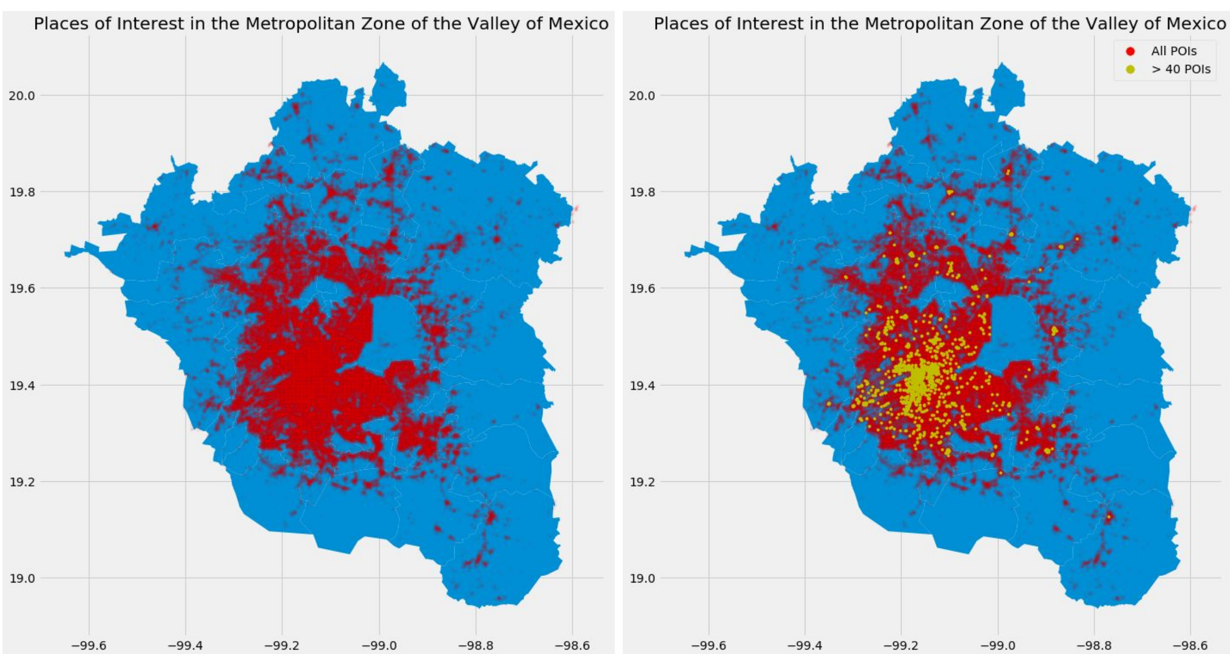


Figure 7. a. Binned Places of Interest, b. Places of Interest with Threshold of > 40 POIs

Fractal Properties

Based on this threshold, the fractal properties are shown in Table 2 and Figure 8.

Scaling Factor	No. of Black Boxes
2	4
4	10
8	27
16	70
32	161
64	333
128	544

Table 2. Box Count of the Resulting Percolation Matrix

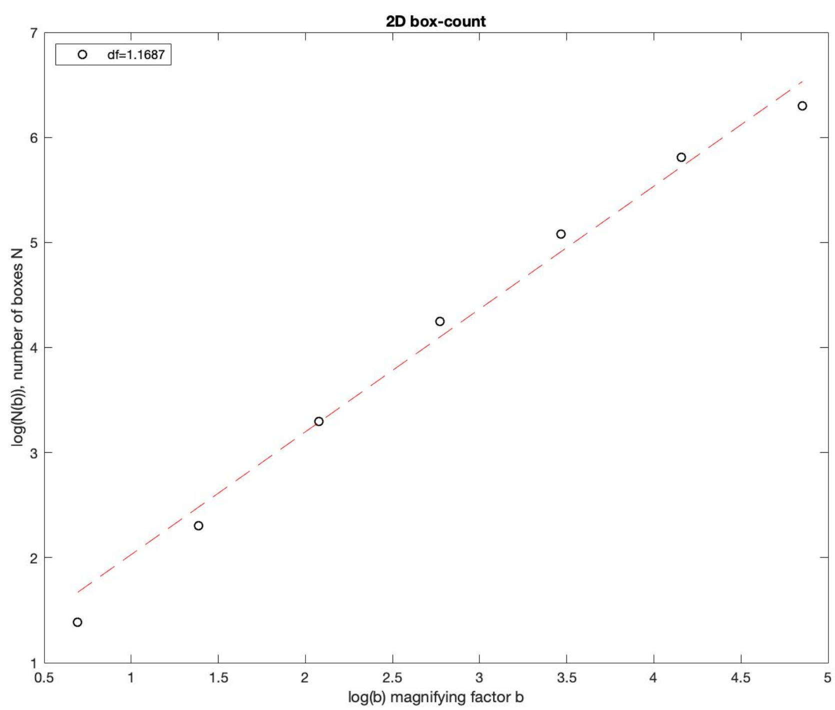


Figure 8. Determination of Fractal Dimension and Plot of Box Count

Because of the strict percolation threshold imposed, the resulting box count is fairly low, also causing the resulting fractal dimension to be low.

Cluster Identification

After forming the clusters, the clusters are labeled to make them distinguishable from each other. To make the most use out of these identified clusters, they were then plotted on a map, as shown in Figure 9 below.

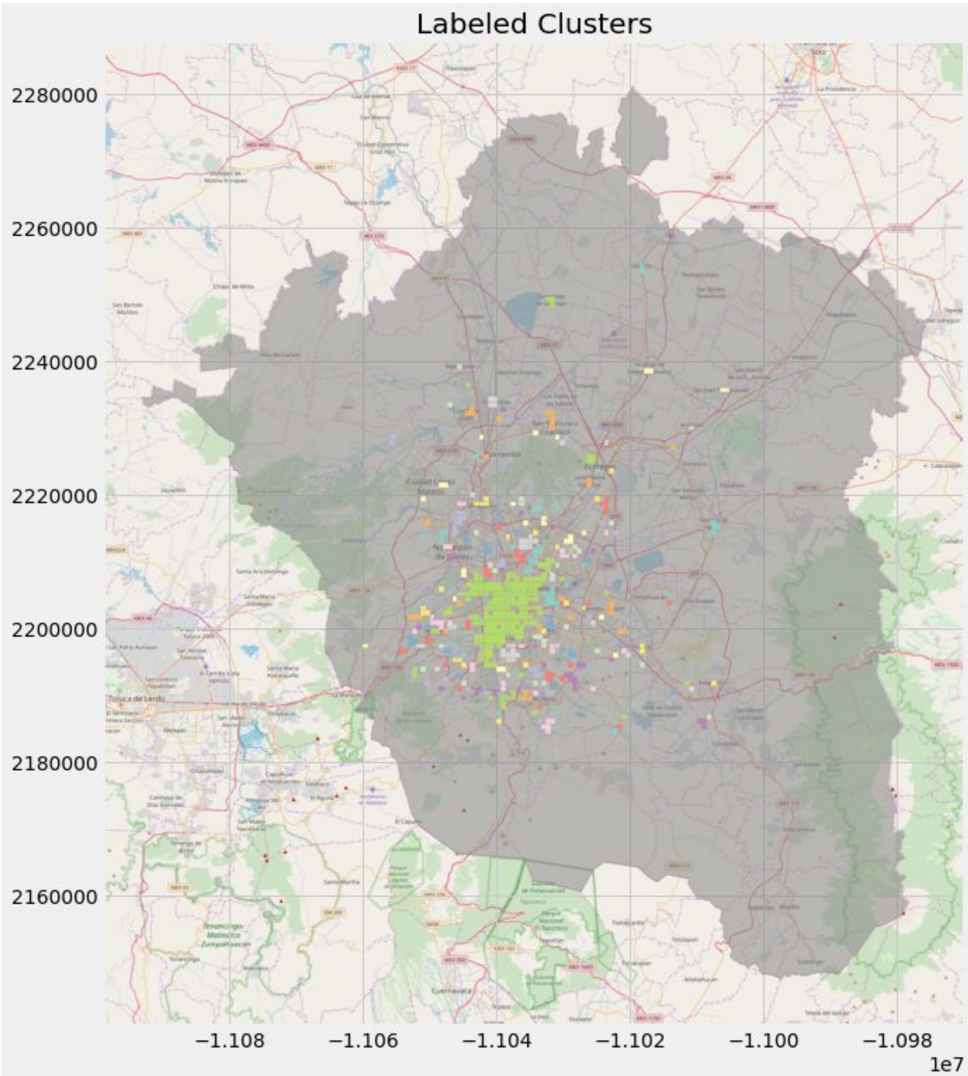


Figure 9. Labeled Clusters Overlayed on Map

Based on this, it is determined that there is a large primary cluster colored green. On the map, this coincides with, but exceeds, the Centro Historico district, which was previously identified as the city center. After the primary cluster there are many others of varying size across the area of study. The frequency distribution of cluster sizes is shown on Figure 10.

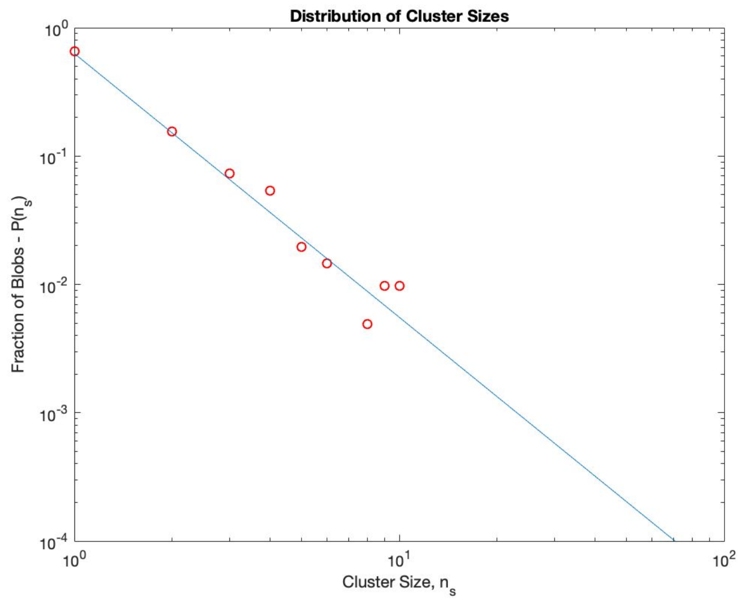


Figure 9. Distribution of Cluster Size

Here, we observe the distribution on a logarithmic scale, such that there are very few large clusters, but numerous smaller sized clusters.

These results are based on a percolation threshold of 40 POIs, which was subjectively chosen. However, there are many other potential thresholds. The critical threshold, p_c, was determined at 0.5927, where the most number of clusters would emerge. Figure 11 shows the cluster size distribution for the critical threshold and the size of the largest cluster for different thresholds.

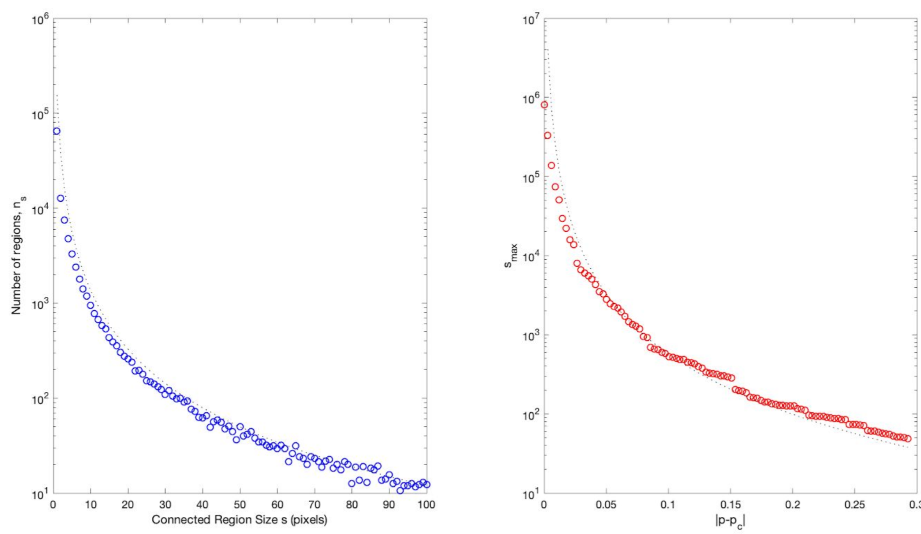


Figure 11. a. Cluster Size Distribution for the Critical Threshold, b. Size of the Largest Cluster for Different Threshold Values

Applications

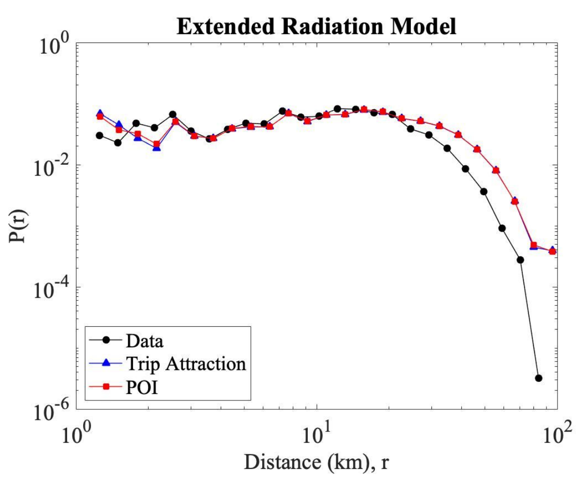
In order to gain deeper insight into the functions and characteristics of these clusters, further study and local expertise would be necessary. However, there are many potential ways for these clusters to be utilized. For instance, the clusters can help in transportation planning to determine bus terminal locations where people tend to congregate, or to identify mass transit routes and stations with an understanding of which city blocks functionally interact with each other.

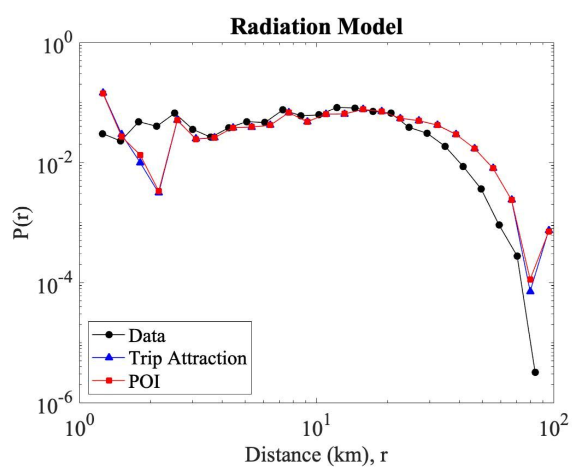
Furthermore, the inverse of the methodology applied here can be used to identify clusters with low numbers of Places of Interest. Paired with granular population distribution data, this can help with planning for housing infrastructure. Clusters with high population and low Places of Interest could be the subject of further study to analyze what services and facilities are lacking in the area, such as schools, hospitals, and other government services.

Predicting Mobility Patterns: The Radiation Model

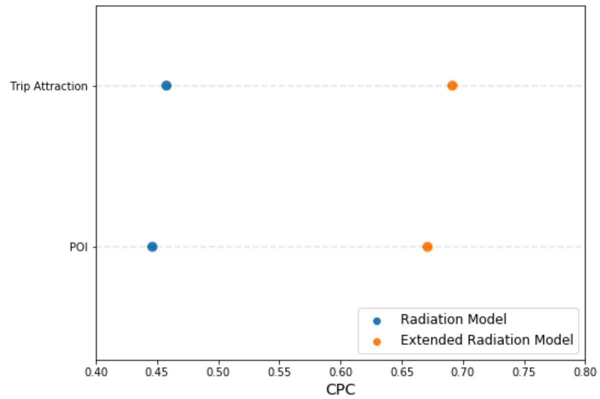
In this section, we use the radiation model to predict mobility patterns (*Simini, F., Gonzalez, M., Maritan, A. & Barabasi, A. A universal model for mobility and migration patterns. Nature 484, 96–100 (2012), Simini, F., Maritan, A. & Neda, Z. Human mobility in a continuum approach. PloSone 8, e60069 (2013))*. As mentioned in our review of the literature, the model can be applied in areas where we lack previous mobility measurements, significantly improving the predictive accuracy of most of the phenomena affected by mobility and transport processes. Here we fit this model in the Valley of Mexico City to evaluate its accuracy. We also compare two different input data set to draw conclusion on which data makes the model perform better. Two different radiation models were used: the parameter free radiation model and the calibrated extended radiation model (Yang, Y.X., Herrera, C., Eagle, N. & Gonzalez, M.C. Limits of Predictability in Commuting Flows in the Absence of Data for Calibration. Sci. Rep. 4, 5662; DOI:10.1038/srep05662 (2014))

In both models, we used origins-destinations pairs, districts - summarized by to their centroid -, and two different proxies for the attractiveness of each district. Firstly, we used the number of incoming trips since they were directly available in the survey. Secondly, we used the number of POIs in each district, a much easier variable to access to when such a rich survey is not available. The following figure shows the probability density as a function of the distance of travel.





We use the common part of commuters (CPC), to quantitatively measure the goodness of flow estimation.



We observe two main behaviors. First, the calibrated extended radiation model gives a much better estimation of trips distribution providing an accuracy around 70% against 45% for the parameter free radiation model. Then, the use of POIs gives predictions slightly worse than when we use the trip attraction. Thus, it is completely justified to use that feature in places where no census data is available.

Analysis of Behaviours Related to Means of transport

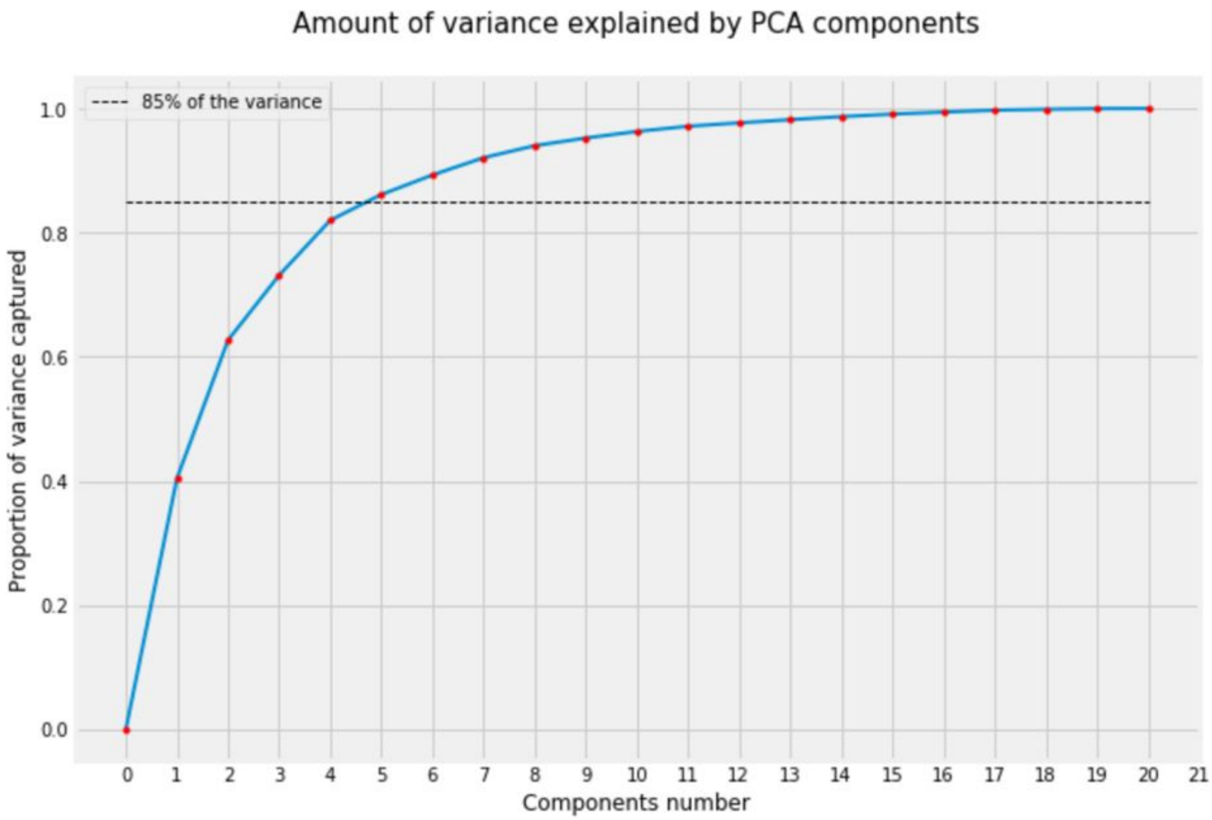
In this section, we look at the different types of transport used by city dwellers in Mexico City. The idea is to obtain the main types of behaviour that prevail in the city: how and to what extent transport is used, associated and combined.

To do this, we have a large database in which more than half a million urban dwellers respond to their means of transport. Each person is surveyed on each of the twenty different types of transportation available in the city. The result is a table measuring 531594 x 20 filled with zeros and ones depending on whether or not the person is using the means of transportation in question.

The purpose of the study is to cluster behaviour to produce different major trends representative of transport in Mexico City. The large number of variables (twenty in total) and respondents (half a million) invites us to first reduce the size of the problem.

a) Dimension reduction

We chose the *Principal Component Analysis* (PCA) method to do this. This aims to capture as much of the total variance of the data as possible with a reduced number of variables - called *Principal Components* (PC). The PCA method is already implemented in the *sklearn* library of *Python* and the formatting of the data having previously done, it can be applied directly.
However, the new size of the projected database, which is an adjustable parameter, remains to be determined. The criterion is the proportion of the initial variance captured by the PCs of the method: the sum of the variances captured by each PC must reach a certain threshold. The latter will not be the subject of a study here, and we will set it at 85%, a widely used and widespread value.
The proportion of the initial variance captured by the variables is then plotted against the number of variables used by the PCA method. We then obtain the following figure:

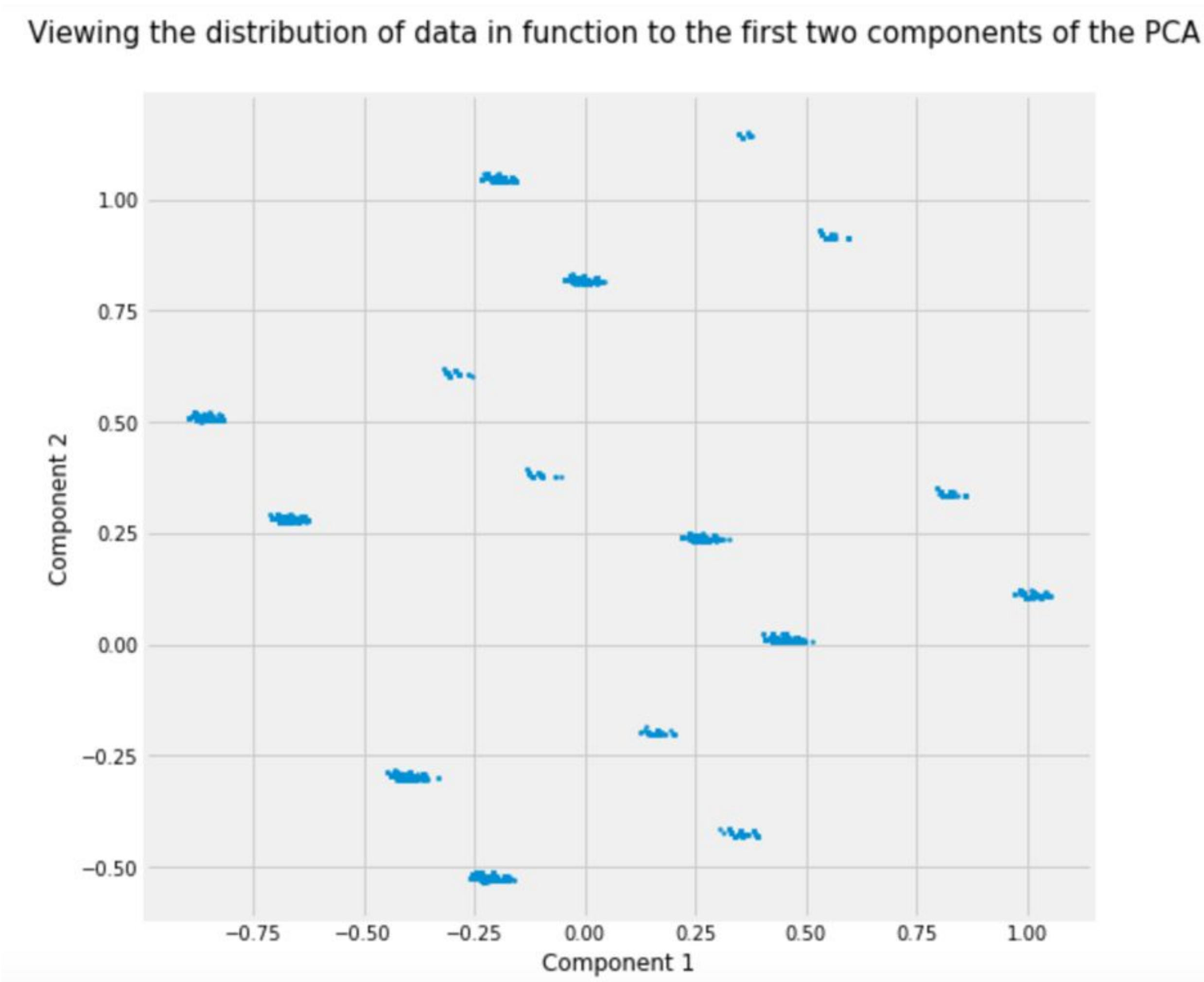


It can be seen that the first five PCs contain 86% of the total initial variance. The number of PCs is therefore set at five for the rest of the study.

We therefore apply the PCA method to our database by reducing the dimension of twenty to five. We then obtain a table of size 531594 x 5: five columns corresponding to the PCs and the 531594 rows being the data projections on these five new variables.

b) Determination of the relevant number of clusters

We want to determine here the ideal number of clusters. Some tools, based on the calculation of a score or distances, can be used to quantitatively determine this number. Although we are aware of the limitations of this approach, we have opted here for a visual approach, the other more quantitative methods could be applied during a more in-depth study.
To visually identify the ideal number of clusters, we project our initial data according to the first two PCs (these are ranked in order of importance, the criterion being the proportion of variance captured). We then obtain the following graph:



We immediately notice the existence of a pattern, repeated four times, itself composed of four "tasks". On the basis of this graph, we can therefore choose a number of clusters equal to four or sixteen. We choose four of them for the rest of the study.

c) Clustering & results

The *Kmean* clustering algorithm is applied to data projected in five dimensions. This algorithm works iteratively: it initializes by randomly selecting centroids among the data at the beginning, then calculates the distances from each point to each centroid, assigns each point to the nearest centroid and finally recalculates a new centroid per group. It repeats the operation until the centroids converge. The optimal centroids are finally obtained, which reflect an "average" behaviour of each group. Each centroid is therefore re-projected into the initial base - the one composed of twenty variables - in order to be able to interpret them. We obtain the following vectors:

Variables	Cluster 1	Cluster 2	Cluster 3	Cluster 4
walked_on_the_street	0.026380	-0.012214	0.997744	0.006218
automobile	0.003146	0.993784	0.002990	0.517063
Collective_Micro	0.005278	-0.005600	0.001889	0.021485
Metro	0.031710	-0.000342	-0.001386	0.194710
Taxi_site	0.091201	0.278000	0.003482	0.144019
Bicycle	0.014088	0.015931	0.000066	0.013323
Metrobus_Mexibus	0.026488	-0.051834	-0.008853	0.138506
Bus	0.035548	0.026322	-0.001944	0.039952
Moto	0.010898	-0.017344	0.002072	0.054280
RTP_Bus_M1	0.005306	0.004397	0.000178	0.003767
Suburban_train	0.035120	0.042777	-0.002042	0.039285
Taxi_Internet_App	0.001988	0.007419	0.000010	0.002817
School_transport	0.001776	0.011646	0.000674	0.006803
Light_rail	1.006250	0.987626	-0.001611	0.004526
Mototaxi	0.000356	0.000379	-0.000014	0.000129
Trolleybuses	0.002469	-0.000185	0.000761	0.009751
Bicitaxi	0.008154	-0.003172	0.001429	0.029743
Other_means_of_transport	0.005178	-0.004643	0.001682	0.019474
Mexicable	0.001077	0.000207	0.000299	0.002597
Transportation_of_personnel	0.001437	-0.000850	0.000491	0.004582

For each cluster, the variables are classified by importance of the absolute value. This operation aims to isolate the significant modes of transport for each of the centroids. We can see above each centroid with the four most important variables:

Cluster 1	
Variables	
walked_on_the_street	0.997744
Metrobus_Mexibus	-0.008853
Taxi_site	0.003482
automobile	0.002990

Cluster 2	
Variables	
automobile	0.993784
Light_rail	0.987626
Taxi_site	0.278000
Metrobus_Mexibus	-0.051834

Cluster 3	
Variables	
Light_rail	1.006250
Taxi_site	0.091201
Bus	0.035548
Suburban_train	0.035120

Cluster 4	
Variables	
automobile	0.517063
Metro	0.194710
Taxi_site	0.144019
Metrobus_Mexibus	0.138506

The first cluster brought together 20% of the respondents, the second 23%, the third 37% and the last 19%. The population of respondents is therefore distributed relatively evenly among the four clusters.

We can therefore establish the four main types of behaviour in terms of transport:
- The first - which corresponds to the first cluster - includes people walking mainly on foot, and almost not using the metrobus or the mexibus.
- The second group includes those who use their cars and light rail on a massive scale.
- The third group includes those using almost exclusively light rail.
- Finally, the last one includes those using their cars and moderately metro, taxi and metrobus or mexibus.

Trip Analysis by Sociodemographic (SD) Stratum

The largest dataset, tviaje, analyzed the individual trips of our surveyed population. The researchers extracted four (overlapping) groups of commuters based on their destination types. These groups are: (1) Religious Movement, whose destinations were religious precincts, (2) the Working Class, whose destinations were work offices or a factory/workshop, (3) Recreational, whose destinations were cultural centers, recreational areas, sports arenas, or the gym, and (4) Commercial, whose destinations included places of commerce, the market, stores, or shopping centers, along with restaurants, bars, or cafeterias. Descriptive statistics are given in Table 1.

	Entries (%)	Unique Persons	Average Travel Time
Cumulative	531595 (100%)	160,095	0.72 Hours
Religious	3091 (0.6%)	2810	0.55
Work	54439 (10.24%)	39945	1.03
Recreational	15771 (2.97%)	13863	0.74
Commercial	89020 (16.75)	63,749	0.57

Regression analysis

Regression analysis allows the researchers to analyze travel time metrics as an indicator of willingness to travel by SD stratum. This metric gives decision makers a starting point for countermeasures to the inaccessibility of present in the world's major cities. Results are shown in Figure (i). We notice the bell-shaped curve of the distribution of time, which has a reasonable social interpretation, i.e members of the lowest (1) and highest (4) SD stratum do not travel far to POI's, given the fact that members of the lowest stratum usually lack transportation (or social) access to POI's outside if their immediate vicinity, and members of the highest stratum construct their neighborhoods to be self-containing (to live where the majority of POI's are located). In contrast, we usually interpret stratums 2 and 3 as suburban populations, who have the capability to travel to POI's, and the simultaneous need to reside outside the most expensive parts of the city, which contains the densest concentration of POI's, and therefore need to travel to them. More quantitative analysis can be run on this data set, overlaid with this SD information, to support this notion, and gather evidence for specific accessibility countermeasures that ensure fairness of travel time and capability across all SD stratums.

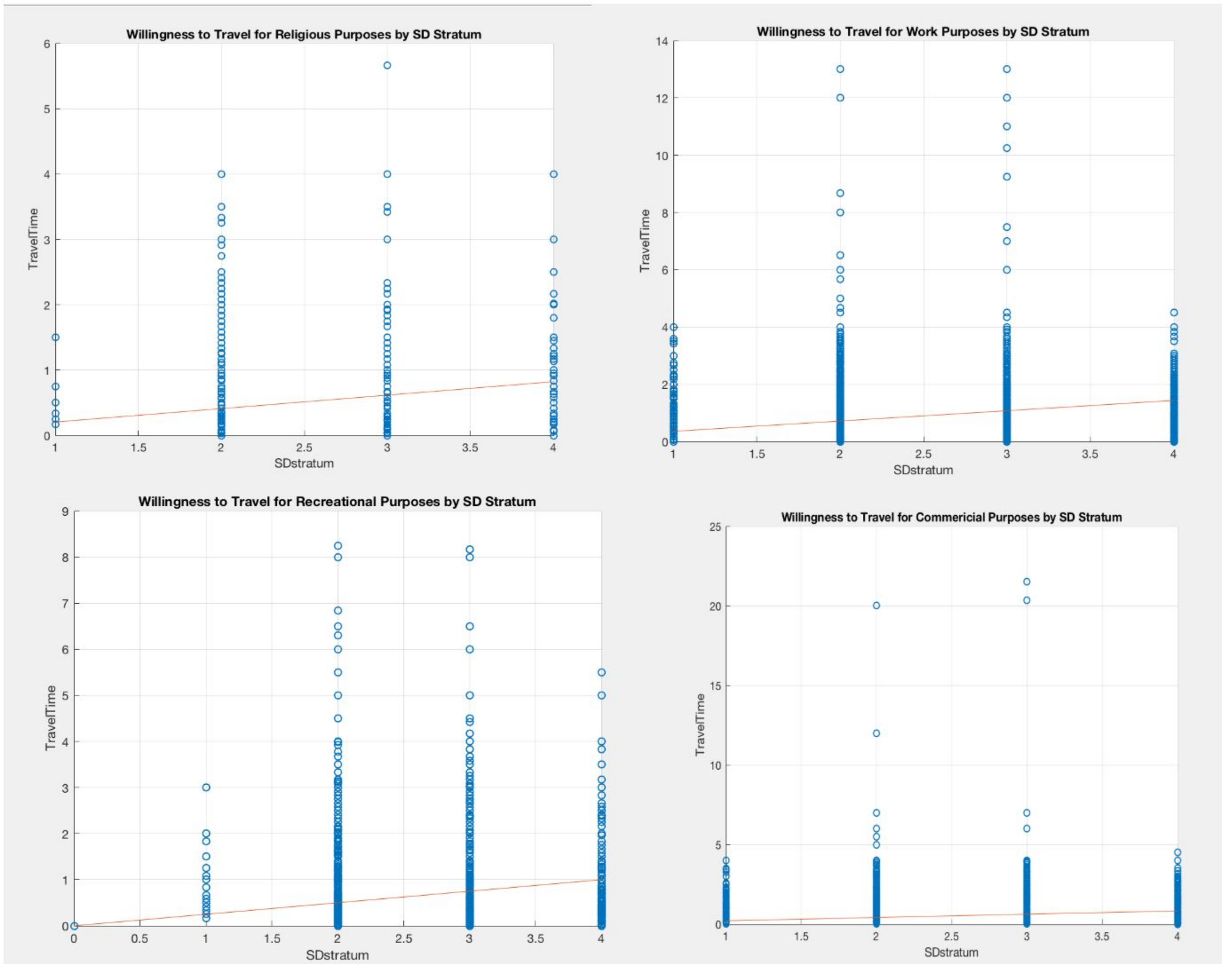


Figure (i): Willingness/Ability to Travel by Sociodemographic Stratum based on trip dataset. The red line represents a linear regression function, which would allow us to impute data for trips based on SD stratum.

This analysis will become more granular in two directions, by the analysis techniques, and further utilization of the data. We expect many factors related to the social hierarchy to appear in our data, and would like to make a well-developed argument to incite the investment of public resources into the increased accessibility of POI's.

Conclusion

The well-defined analysis of complex human socio-technical systems has become the interest of interdisciplinary groups around the World, as urban planners seek to analytically amend current city and country infrastructure to better accommodate the continued expansion of the World's major cities and metropolises. The purpose of this study was broad, that is, the work of the researchers demonstrate initial steps that city and urban planners may take to better characterize their cities based on current data techniques drawn from various disciplines. Many of the methods used in this paper reduce the complexity of the dataset, while simultaneously extracting useful information. To this end, the exponential growth of measured and recorded data lends import to the understanding and implementation of these and other methods for use in city and urban planning.

Furthermore, techniques which extract values from the features of the dataset, and produce equations and models that determine variables of interest (e.g. travel time) can be used in the absence of data for both model calibration and prediction. As indicated in our final section, the researchers are highly interested in objective measurements for the accessibility of POIs by Sociodemographic stratum, as we hope to utilize these methods as a tool to produce social equity and accessibility in the world's major cities.

Acknowledgments

We thank our research professor, Dr. Marta Gonzalez, for her expertise and introduction to the proprietors of the datasets at The National Institute of Statistics and Geography in Mexico. We thank these collaborators for the provision of the data and access to their information.

References

Batty, Michael, and Paul A. Longley. Fractal cities: a geometry of form and function. Academic press, 1994.

Ibes, Dorothy C. "A multi-dimensional classification and equity analysis of an urban park system: A novel methodology and case study application." Landscape and Urban Planning 137 (2015): 122-137.

Makse, Hernán A., Shlomo Havlin, and H. Eugene Stanley. "Modelling urban growth patterns." Nature 377.6550 (1995): 608.

Shlens, Jonathon. "A tutorial on principal component analysis." arXiv preprint arXiv:1404.1100 (2014).

Simini, Filippo, et al. "A universal model for mobility and migration patterns." Nature 484.7392 (2012): 96.

*Codes used in the cleaning and analysis of data is available upon request