REDUCING TRAFFIC CONGESTION BY URBAN PLANNING IMPROVEMENT

Leonardo Silva Vianna

July 18, 2021

1 INTRODUCTION

Carbon dioxide (CO_2) emitted by human activities is one of the most important issues of our era, and the most responsible for Earth's climate changes. Governments, industries and all Earth's inhabitants are making great efforts to reduce CO_2 emissions and vehicle traffic congestion is directly related to this problem.

Greenhouse gas emissions are not the only problem caused by vehicle traffic. The time spent in traffic congestion affects the health of drivers and passengers, causing stress and frustration, and is considered a waste of potentially productive time. It can result in late arrival for employment, meetings, and education, secondarily causing more waste of time because of the need for more time allocation to avoid these delays. Vehicles are subjected to more breakdowns and maintenance due to the intense use of their components during frequent acceleration and braking in congestion.

Problems related to vehicle traffic congestion seems to constantly increase despite the efforts of urban planning stakeholders. Smart cities are being designed to avoid traffic jams by properly distributing their infrastructure, in coordination with the residential population density, providing a better experience for its residents. Unfortunately, inhabitants of older cities still have to live with these problems.

Consequently, the information discovered by correlating vehicle congestions and city venues and facilities can in urban planners to project commercial spots distributions. Avoiding traffic jams should improve the quality of urban environments, decrease CO_2 emissions and enhance community health. The city of Chicago was chosen to test the hypothesis of the capacity to generate useful information to improve its urban planning.

2 METHOD

2.1 DATA GATHERING

The datasets used in this project were gathered from two repositories, the Chicago Data Portal and the venue location service provided by Foursquare. From the Chicago Data Portal, a dataset containing historical congestion estimated by region, obtained through the Chicago traffic tracker service, were downloaded and covered information since 2018. Information about the location of cities venues and facilities was obtained through the Foursquare API.

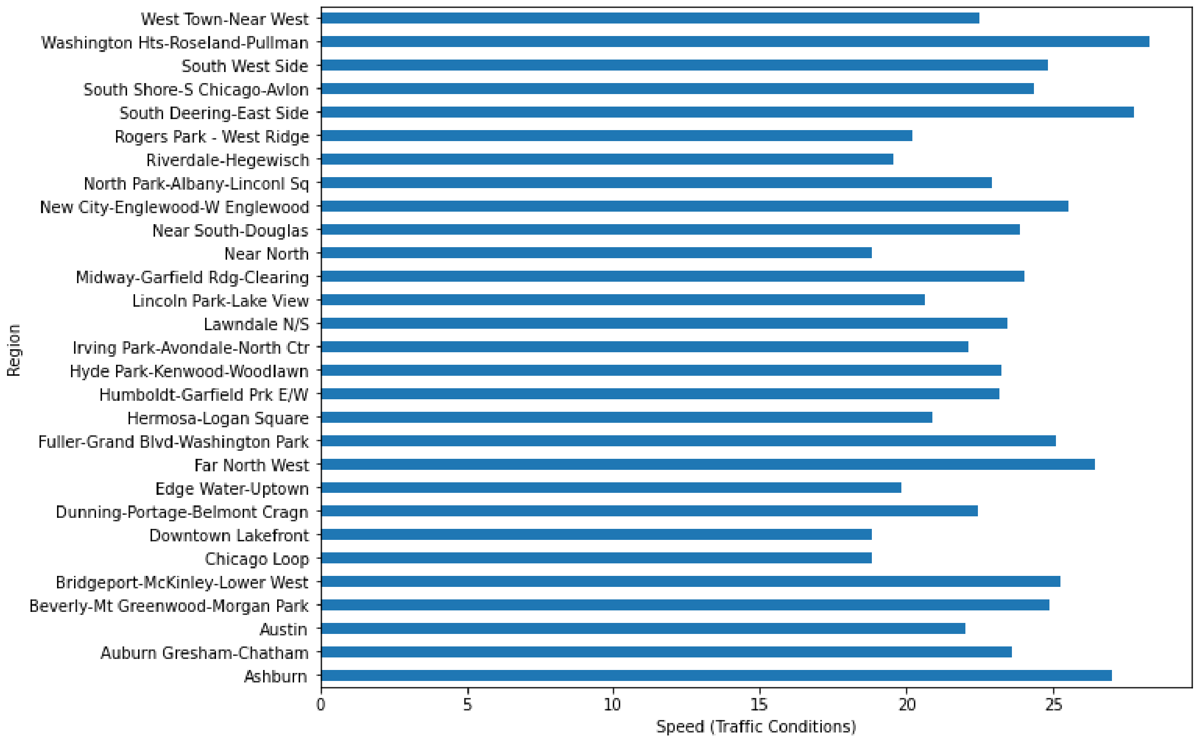
2.2 DATA ANALYSIS

While Foursquare`s data was obtained only by a specific data request using the API, the Chicago Traffic Tracker dataset was previously downloaded, allowing its analysis. This dataset contained information about 29 traffic regions, that was used to estimate traffic congestion by monitoring and analyzing GPS traces received from Chicago Transit Authority buses. The variables were identification and description of the city regions (ID, REGION and DESCRIPTION), the identification, characteristics and timestamp of the readings (TIME, BUS_COUNT, NUM_READS, HOUR, DAY_OF_WEEK, MONTH, RECORD_ID), location coordinates of readings (WEST, EAST, SOUTH, NORTH, NW_LOCATION, SE_LOCATION), and the congestion indicator, the average traffic speed (SPEED). Although it can fluctuate, the speed average for street segments can provide a good understanding of the actual traffic condition. It is important to emphasize that the lower the speed, the greater the congestion in a given region and the worst its traffic condition.

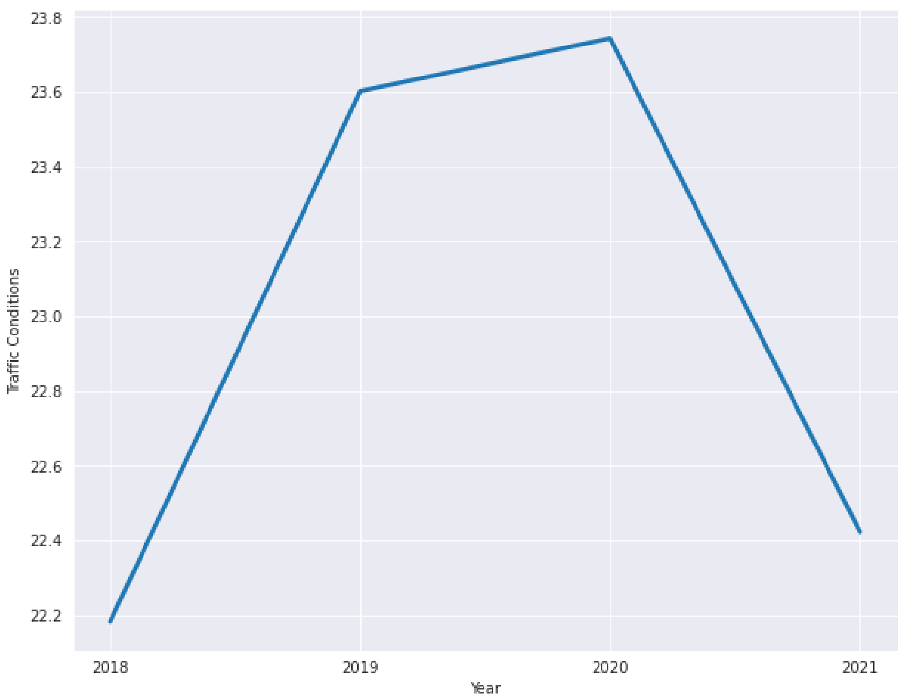
Graphic 1 provides information about the regions existing in the dataset and the average speed (or traffic conditions) from 2018 until nowadays. The Covid-19 pandemic brought outstanding changes in society's behaviour. And Graphic 2 provides a better understanding of the evolution over the years considered the obtained dataset.

Note that traffic conditions improved in 2020 (the average speed increased), probably caused by some circulation restrictions imposed by the health authority. The current year of 2021 is incomplete for analysis and modelling purposes. Apparently, 2019`s average traffic conditions could be an outlier. Although more information is necessary to affirm this statement, 2020 is certainly an outlier and 2019’s average is highly similar to its one. Consequently, all project analyses were performed in 2018.

Graphic 1 - Traffic conditions in each Chicago`s region, from 2018 until nowadays.

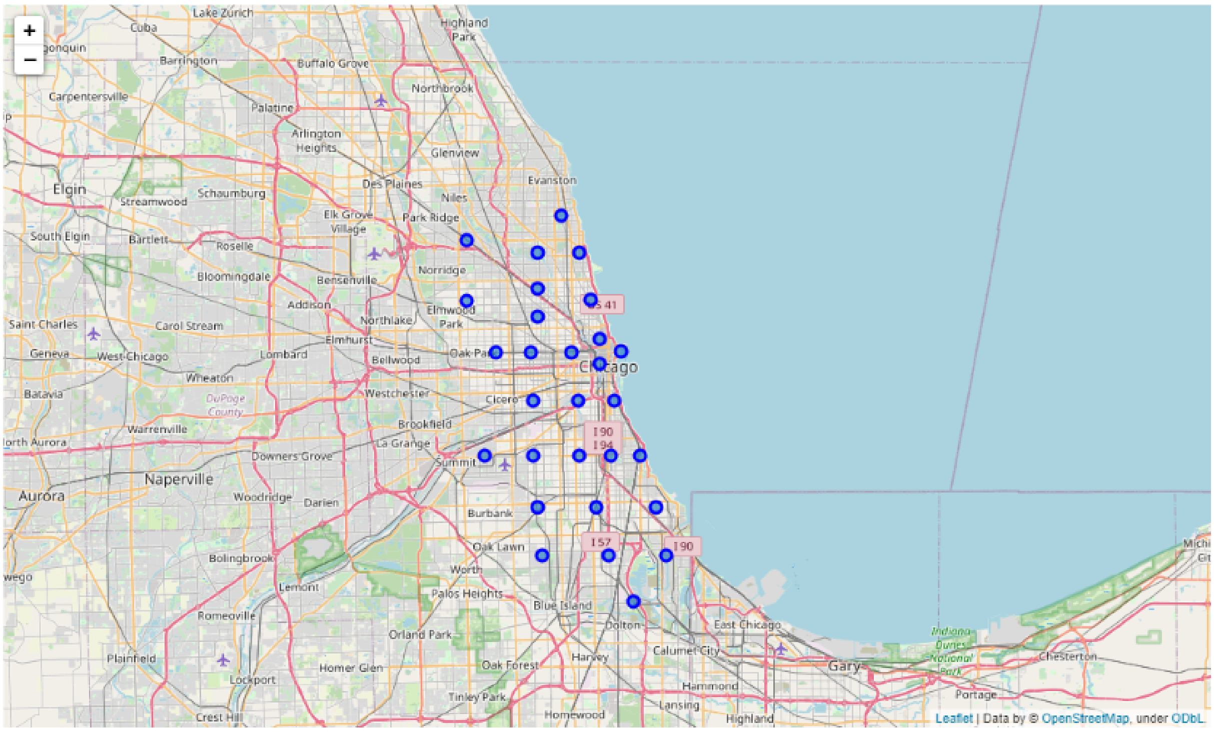


Graphic 2 - Evolution of traffic conditions.



Map 1 shows the location of all Chicago's regions analyzed in this project, with information about the traffic conditions contained in the Chicago Traffic Tracker dataset.

Map1 - Chicago’s regions analyzed in the project.



2.3 DATA PROCESSING

The Chicago Traffic Tracker dataset timestamp attributes did not include a specific attribute with the year of each record. Therefore, the year was extracted from the full timestamp, later dropped from the dataset together with other attributes time-related. In sequence, the records were also filtered considering only the 2018 year.

The location of each Chicago`s region was informed through the latitude and longitude of your cardinal points (north, south, east and west). Consequently, the central point of each region was determined and added to the dataset, as well as its range. The range was calculated applying the Haversine formula between the central point and the external point of the smallest radius of each region’s area. This information was determined because it is needed to request information for the Foursquare API.

Hereafter, the Foursquare API was accessed through a GET request to return all venues localized in the range of the latitude and longitude from the central point of Chicago’s region. The venues were received in a JSON format, that was filtered by categories and organized in a dataset indexed by each region. A one-hot encoder was applied in the venues and used to build a dataset for the modelling phase. The version defined for the API was ‘20181231’, considering the 2018 year analyzed in this project.

2.4 MODELLING PHASE

For the modelling phase, the scikit-learn library was used for a clustering task. The venue categories obtained from the Foursquare API were clustered after the encoder procedure, using each Chicago’s region as a referenced instance (or index). The clustering task was executed with the KMeans algorithm, with the number of clusters defined as 4.

3 RESULTS AND DISCUSSION

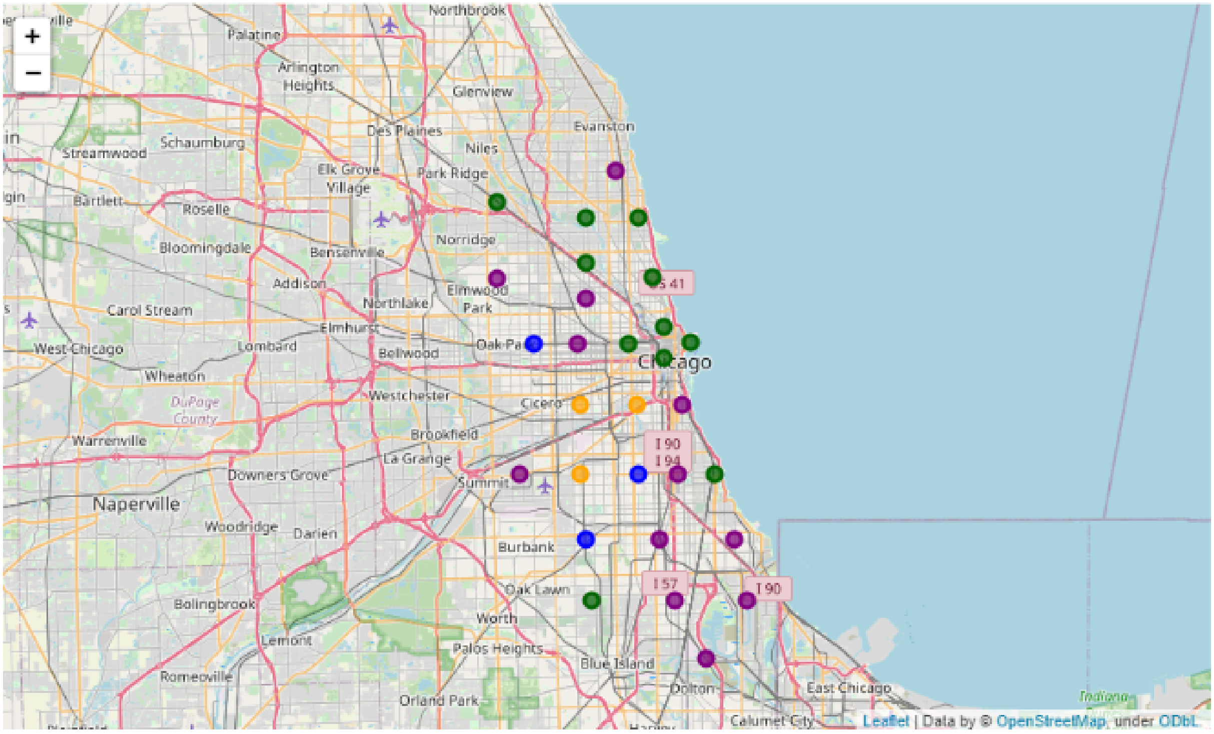
The modelling phase resulted in regions labelled with one of the 4 clusters defined. The labelled cluster information was joined with the first dataset containing the traffic conditions (speed). The average speed of each cluster is presented in Table 1 and Graphic 3 demonstrate the cluster distributions for region. Cluster 1 presented the lowest average speed representing the worst traffic congestion scenario, followed by cluster 2, cluster 3 and cluster 0 (see Table 1).

Table 1 - Average speed (traffic conditions) by cluster.

Cluster (colour in Map 2)	**Average Speed (mph)**
0 (blue)	24.04
1 (green)	20.63
2 (purple)	22.80
3 (yellow)	23.55

Likewise, the distribution of the labelled region in Chicago can be visualized in Map 2. The central regions of Chicago (Near North, West Town - Near West, Downtown Lakefront and Chicago Loop) concentrate the worst traffic conditions, labelled with cluster 1, besides other regions. It is important to emphasize that the clusters were not defined with the regions speed but with their venues and facilities. According to this, the clusters represent the agroupment of regions with venues and facilities with the same characteristics, which also have similar traffic conditions expected. This state is supported by the different average speeds demonstrated by each cluster.

Map 2 - Clusters distribution (cluster 0: blue, cluster 1: green, cluster 2: purple, and cluster 3: yellow).

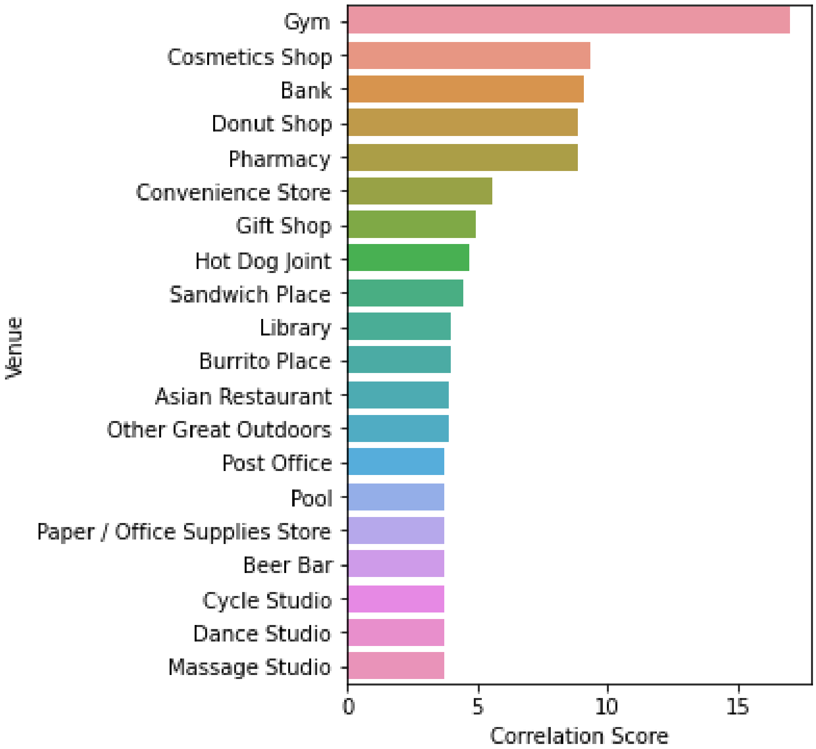


Finally, the venue and facility classifications obtained from Foursquare were used in a feature selection algorithm to determine the correlation with the traffic conditions. A linear regression function was built in each attribute - the venue and facility classification - and used to measure its correlation with the target value - traffic condition (or speed). Graphic 3 shows the 20 most correlated venue categories. Gyms presented the highest correlation result, followed by cosmetic shops, banks, donut shops, and pharmacies.

3.1 STUDY LIMITATIONS

The coordinates used to acquire the list of venues and facilities used a generalization of Chicago's regions and a better geographic delimitation could improve the correlation with the traffic conditions of each spot. Likewise, although it was not used, the application of statistical tests would confirm the statements made in this research.

Graphic 3 - Most correlated venue categories.



4. CONCLUSION

The correlation between venue classifications and traffic conditions obtained with this project can help urban planning stakeholders to better distribute the city's venues and facilities through the city's master plan. This information provides an instrument to mitigate traffic jams in the streets and avenues. This better flow of vehicles can enhance the life experience in the city, improve the health conditions of its population and reduce greenhouse gases emissions.