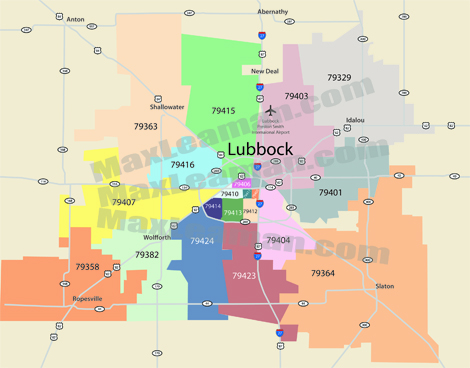
**Battle of Neighborhoods:** Lubbock, Texas



**Project description**:

Lubbock is a city in Texas with approximately 307,500 residents living in the area. As at 2017, according to the US Bureau of Statistics, the total employer establishments are 7,202 which shows an increasing trend. Majority of these businesses include service companies and small businesses. A car salesman is looking to invest in his business and expand to Lubbock. With the increasing number of student population at Texas Tech University, there seems to be more opportunities for a car sales business. Using this project, an analysis scan be done to show the features of the Lubbock neighborhood and where the proposed car business can be established,

The project requires geo-locational information about Lubbock, Texas can provide the salesman data needed to make his decision. The project will help stakeholders including other car salespersons, parents/students/families in Lubbock, other small businesses and the city of Lubbock as a whole.

To illustrate, the project will compare the different zip codes in Lubbock and analyze the neighborhoods with the most common car shops and where there are no car shops. Using the k-means clustering algorithm, this will help give a better understanding of the neighborhood and insights on areas to establish the car shop.

**Data sets and APIs**

The data for Lubbock neighborhood that include the zip codes was found on the [city of Lubbock](https://ci.lubbock.tx.us/departments/gis-data-services/home) website, the [Lubbock area connect](http://lubbock.areaconnect.com/zip2.htm?city=Lubbock&qs=TX&searchtype=bycity) website and the [Geonames](https://www.geonames.org/postal-codes/US/TX/303/lubbock.html) website. The major data scraped from these websites include the zip codes, the location data (longitude and latitude) and the county names.

After data was downloaded from these three websites, the data was joined into one table. There was no missing data based on the zip codes obtained. After cleaning the data, there were a total of 33 samples and 4 features. The features include: zip code, county name, longitude and latitude.

The Foursquare API is a data gathering source that is also used for this project as it has a very large database which gives ability to share business locations and provide information based on location search that will be used to understand businesses in Lubbock. Photos and reviews by users provided by Foursquare API can also be used for gathering insights on car shops in Lubbock.

Python scientific libraries, visualization libraries, packages and dependences will also be utilized in getting information about the city of Lubbock. The k-mean clustering algorithm will be applied on the clusters of categories in the Lubbock neighborhood.

**Methodology**

The GitHub repository was used and contains the data for the main components of this capstone project. The data contains the zip code, county, longitude and latitude. As Lubbock is a smaller Texas city, the neighborhoods are identified by the zip codes of the areas in the county. As seen in figure 1 below:

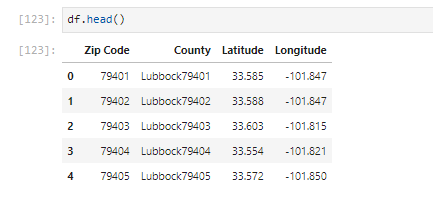


Figure 1

The geopy library in Python was used to get the latitude and longitude values of Lubbock as seen below in figure 2 .

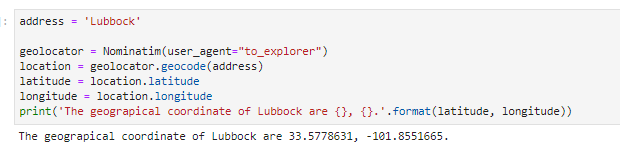


Figure 2

This was done to create the visual geographic details of the city using the python folium library also seen in figure 3. The boundaries of Lubbock city are seen as s Shallowater, Idalou, Wolfforth, New Deal, Slaton and Ransom Canyon.

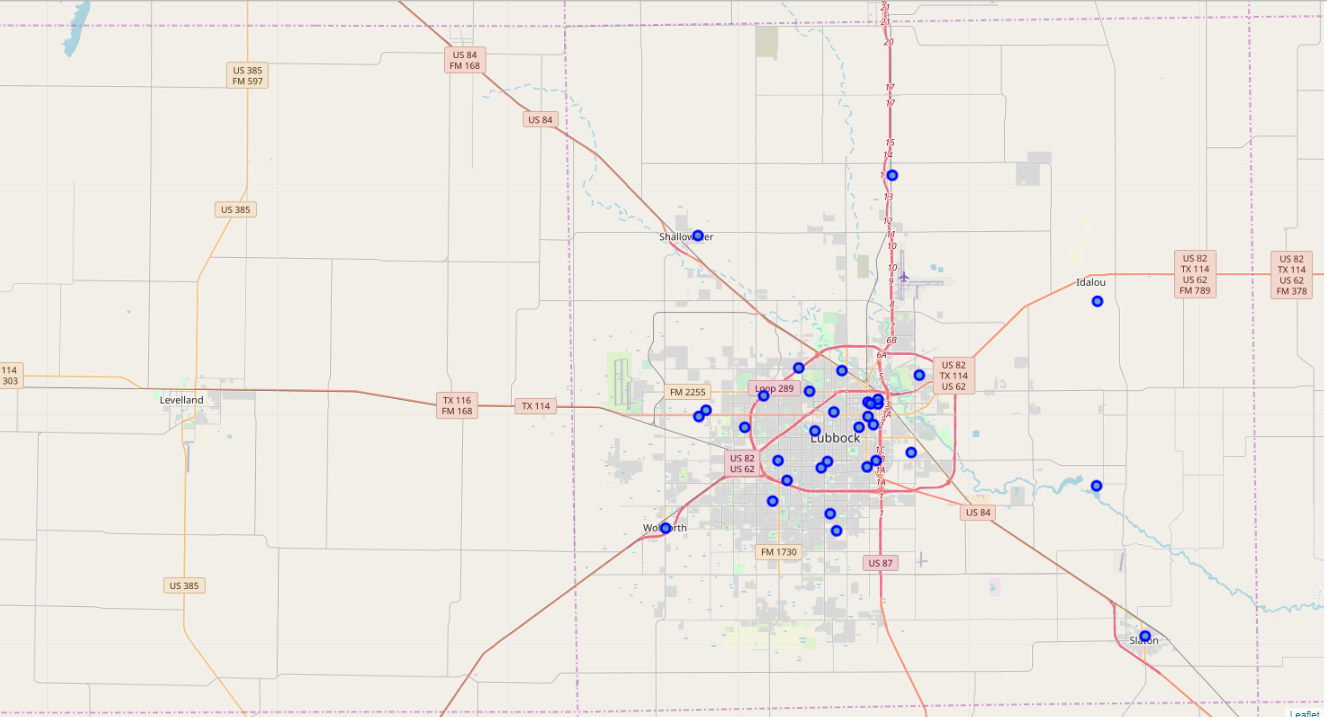


Figure 3

Using the Foursquare API, the neighborhoods and clusters were explored and analyzed, The limit used was 100 and the radius defined for each neighborhood was 500 based on the longitude and latitude information. After defining the functions and getting the venues names, Figure 4 shows an example of nearby venues in a particular neighborhood based on the json files downloaded from Foursquare.

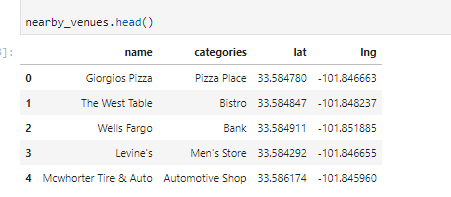


Figure 4

In total, there were 242 venues downloaded for the city of Lubbock. The Figure 5 below shows the information on how many venues grouped by the zip codes provided.

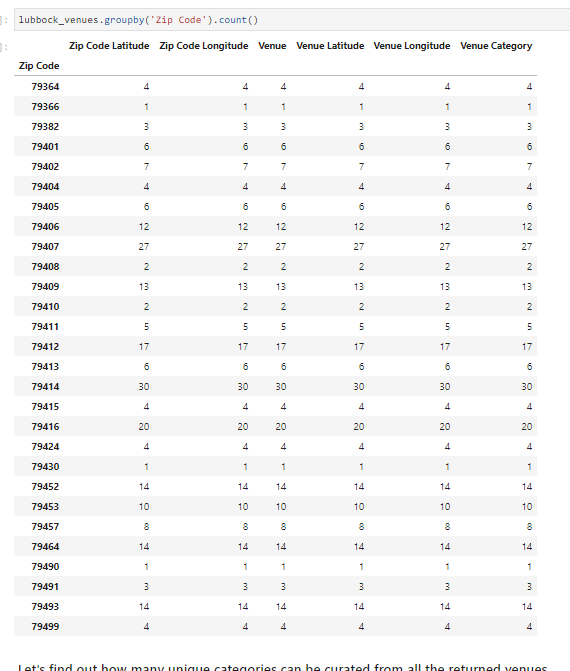


Figure 5

From the information shown, it can be seen that Neighborhood with zip code 79414 has the highest number of venues. Although there were 102 unique venue categories based on the information on Lubbock city.

After getting this information, we also grouped the zip codes according to the top 10 common venues in each neighborhood using grouping and sort functions and have figure 6 as seen below. Please note that figure 6 only shows top 5 venues. The top ten can be found in the Python Notebook.

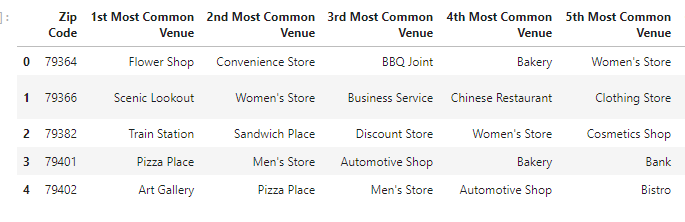


Figure 6

As we can see, there are common venues in the neighborhood, therefore using the unsupervised k-means machine learning algorithm, the neighborhoods are clustered based on their similarities. The k-means algorithm will create and optimize the clusters based on the positions of centroids and the defined iterations. Based on the frequency for an “automobile or motorcycle shop”, the neighborhood was clustered into 5. This lets us determine which neighborhoods have a higher number of car shops in the neighborhood. It will also help with the analysis based on the project description for the car salesman looking to expand to Lubbock city. The figure below shows the cluster label for each neighborhood.



Figure 7

**Results**

The results from the k-means clustering algorithm suggests that the neighborhood can be categorized into 5 clusters based on the occurrence of the automotive shops.

It should be noted that Neighborhood with zip code 79403 shows “Nan” as this area is for ranching and farming.

Cluster 0: Neighborhoods with a high occurrence of automotive shops

Cluster 1: Neighborhoods with some occurrence of automotive shops

Cluster 2: Neighborhoods with little occurrence of automotive shops

Cluster 3: Neighborhoods with very little occurrence of automotive shops

Cluster 4: Neighborhoods with no occurrence of automotive shops

Also, we can view the cluster in a folium map as seen below

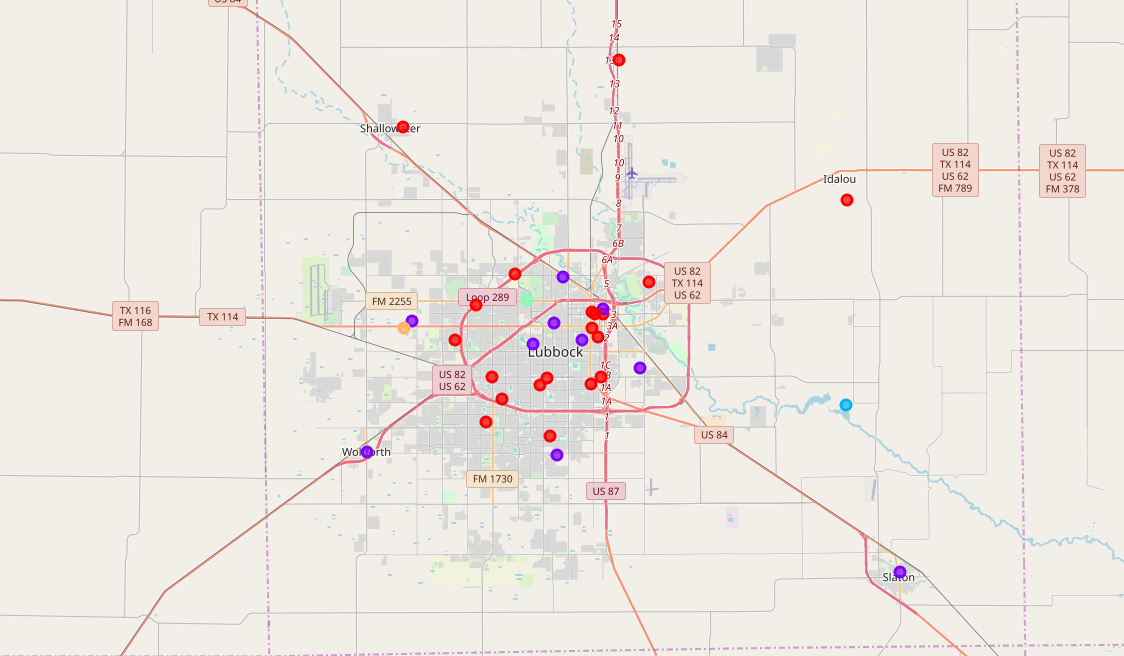


Figure 8

**Discussion**

Lubbock is a smaller Texas city than Austin, Dallas and Houston. But as can be seen from the analysis have very concentrated small businesses that are frequented. The analysis shows that the most frequented area is neighborhood with zip code 79414. This may be because this is closer to the South Plains Mall. Lubbock can be accessed from each location in 20 minutes but the concentration of the mall and shops in this area may be the reason for this classification. Although 79414 is most frequented, there is no automotive shop in the top 10 venues for this area. Instead the areas 79401 and 79402 closest to the University has the automotive shop has the 3rd and 4th most common venues respectively. Therefore, this project recommends that the car salesman can capitalize on these findings for car shops closer to the University which offers a unique selling proposition to market themselves to the students, parents and their families. The car salesman should avoid neighborhood 79403 which is majorly used for ranching and farming. Also, neighborhoods 79490, 79430 and Ransom Canyon are not populated with many people and it will not be advisable to set up car shops here.

**Conclusion**

As a result of Lubbock being a small city, there are highly concentrated densely and sparsely populated areas. The densely populated areas are closed to Texas Tech University and sparsely populated are usually for ranching, farming or city purposes. Data analysis shows the need for car shops but salesmen who wish to invest their business in the city are to be aware and ensure that they set up car shop in areas where they are highly visible and marketable. The neighborhoods in cluster 0 are the most preferred to set up a car shop has it has high potential for business opportunities.