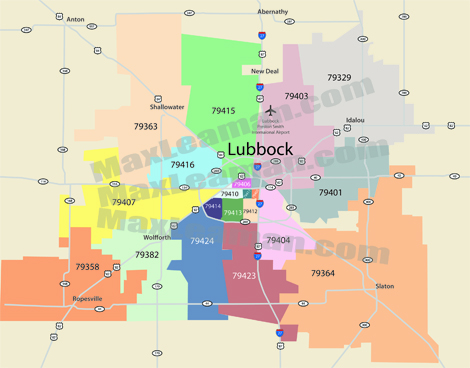
**Battle of Neighborhoods:** Lubbock, Texas



**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.