**Safe Neighborhood – Team 4 - (06/22/2019)**

**(Gael G / Lori B / Sisay T / Sujita K / Van S)**

**Introduction:**

Team 4 (Team Safe Neighborhood) picked a project to visualize crime pattern in the Dallas County.

The first purpose of this project is to show the concentration of crimes in Dallas county. The second purpose of this project is to show the median home prices in Dallas county.

By connecting these two data sets and layering the map, we want to show if there is any relationship between median home prices and crime.

The project started out as a group discussion on the topic. Once crime was decided as the topic, we focused on DFW area. There are 13 counties in DFW area. The initial idea was to visualize crime for the top 5 counties based upon population count. Due to limitation of data, we narrowed our project to looking at crime and home values in Dallas County only.

There were 593 different types of offenses listed. For this project we included only five types of offenses that are more related to crimes in a residential area. This is because the data we obtained from Zillow had residential home prices only. The five types of crimes that we looked into are Theft, Burglary, Robbery; Assault; Children; Death/Suicide, and Sexual assault.

**Description of data used:**

The following data sets were used throughout the project.

* Median home prices were obtained from Zillow.
* Citydata.com website was used to obtain the profile data for all counties. For example: population, median income and median home prices.
* Dallasopendata.com was used to obtain the crime data for all of DFW area.
* Geocod.io was used to obtain the latitude and longitude for the offence streets.

**Description of technology used:**

* API call to the above-mentioned websites were made in order to collect the data.
* Jupiter notebook was used to clean and manipulate the data.
* MongoDB and plotly.js were used for website rendering.
* Git Hub was used as a repository to store analysis and reports.
* Zillow API did not have json option. Data had to be converted from XML to json.

**Limitations of data and analysis:**

The crime data obtained had some limitations. There were no latitude and longitude. The cities were spelled differently. Once the city names were cleaned up, 27,000 rows of data remained for Dallas county.

The other challenge was the street address. Although the data pulled was for DFW locations, the latitude and longitude showed out of state addresses. This could have been because of numerous reasons. Also, the vast number of rows of data slowed the loading process. In order to exclude the out of state data points, we took only 5970 sample data points for this exercise.

The other challenge we encountered was in getting data for the same year. Since different data sources were used for different data sets, it was not possible to get data sets for the same period.

Crime data used was for the year 2014 whereas the home prices are the most recent as of April 2019.The county data used were for 2016 and 2018.

The Zillow website did not have json data. XML data had to be converted to json which was very challenging.

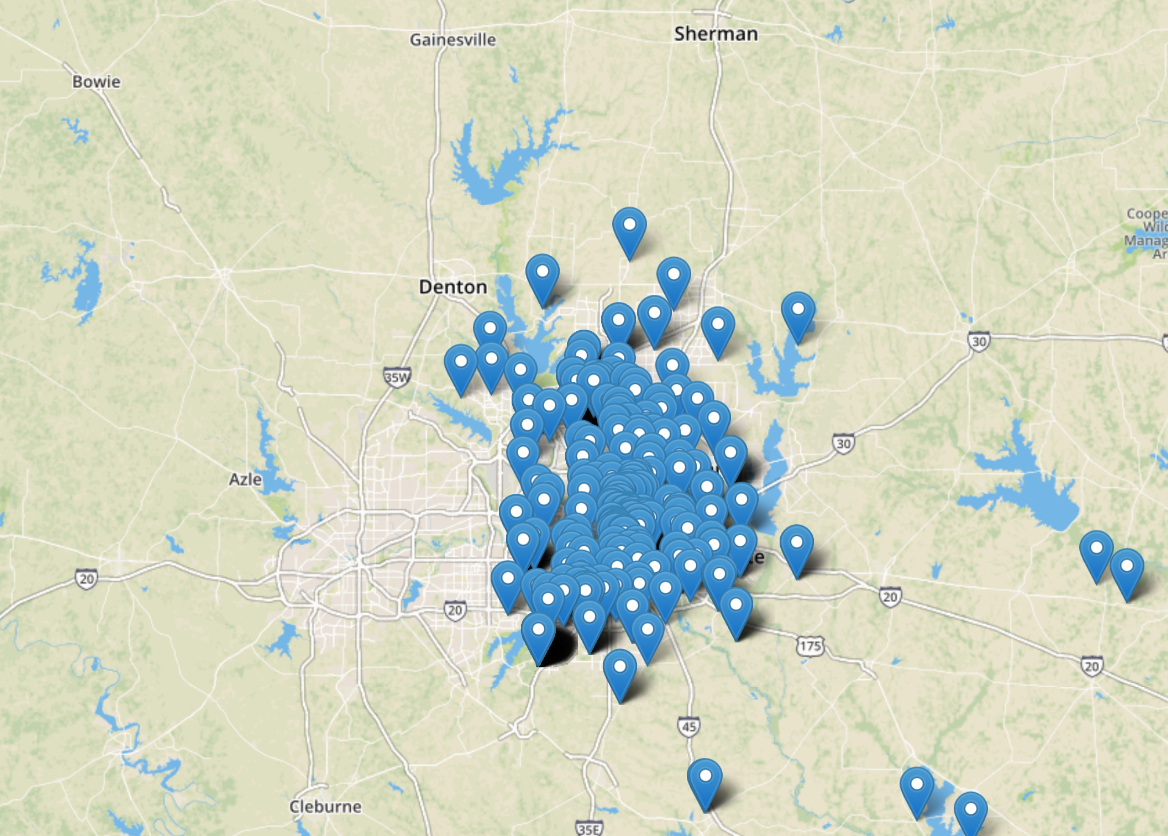
We also had to change data sources multiple times due to unavailability of required data from a particular data source.

**What could we have done better:**

* We could have added all indexes
* We could have had all the visuals in one file.
* We could have added additional details to the crime chart.

**Appendix of All Vizzes:**

1. Static chart – Chart showing county descriptions like population, average home value and average household income.
2. Interactive chart – A bar chart showing a trend of the above information when a county is selected.
3. Map – A map that shows the crime locations and household income.



1. A map that shows median home values.

