**dataRockStars**

Team: Rod Skoglund

Lucy Hohler

Giovanni Rigters

Objective:

The KCPD reached out to our company, dataRockStars, for help. They have a ton of crime data they want to make use of and need some help making sense of it. They have their hands full keeping up with all the incidents that happen in the surrounding areas as is, so they would like for us to come up with an easy way for them to be able to look at the data and make important decisions on which troublesome neighborhoods need more attention.

Not only would they like to know which areas are the **most crime** ridden, they would also like to know if the data can give them any answers on how to prevent some of these crimes from happening in the first place.

For example, if the data shows that men between the ages of 24-30 commit an extensive amount of crime, we might be able to take active measures in outreach programs for this demographic.

Data set:

The data we have access to is 2018 crime data in the surrounding Kansas City areas on the Missouri side only. We use this data to create interactive visualizations for the police department.

We started with more than 120K pieces of data. After the initial cleaning, we dropped to ~93K pieces of data. After major cleaning (removing zip codes = 99999, Tulsa-OK, Clearwater-FL, etc.) and filtering to just look at the suspect, we dropped to a little more than 18K pieces of data. This was the basis for our website.

We used ETL to extract the data and store it into our mySQL database. We are then loading this data into our Flask file.

Project work

We are doing the work of cleaning the data, writing the code and using both an interactive chart and map to visualize this data.

**Cleaning the data**

We had to filter through a ton of data that was not useful. We had to get rid of empty cells (no sex), delete wrong data (ages above 100) and convert data where we had similar cities (Kansas City and KC).

**Writing the code**

* HTML/CSS/Bootstrap was used to create our website
* Jupyter Notebook, Python and Flask where used for writing the code and loading it into our database.
* MySQL was used to store our data
* D3 and Leaflet was used to create a visually interactive chart and map

**Visualizing the data**

Our data page uses advanced tables to make filtering data not only easy and fast, but also efficient.

Through flask we connect with the database and pull the required data into our JS pages. For the map we also have to throw our errors as we noticed that we were getting lat, long errors depending on how many rows of data we wanted to pull.

For the scatter plot we had to create additional columns to calculate the percentages, which we then pull in to make our plot interactive.

**Conclusions**

The KCPD was very satisfied with our work and wants to work with us full-time and not on a contract basis anymore. We were able to give them an easily navigable solution which is also interactive.

The police staff on their end needs to do a more thorough job of entering the correct data for the crime reports. Frequently we observed data entries missing for sex and ages and locations that did not make any sense.