Project Proposals

Student Name: - Jitendra Agarwal
Course: - Springboard cohort Jan2
Proposal: - Explanatory Analysis of Traffic pullover pattern for Florida v/s Montana
Data set: https://openpolicing.stanford.edu/data/

Project Proposal 1: - Explanatory Analysis of Traffic pullover pattern for Florida v/s Montana

Data set: https://openpolicing.stanford.edu/data/

1. What is the problem you want to solve?

Explanatory and Exploratory Analysis of Traffic stop data and find patterns and correlation with the violation and final outcome for Florida v/s Montana.

We will explain and assess the manner and extent to which age and race play a role in traffic stops and the final outcome result. It is also interesting to explore if weather, population density plays any role in the traffic stops patterns and frequency.

2. Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn't have otherwise?

More than 20 million Americans are stopped each year for traffic violations, making this one of the most common ways in which the public interacts with the police (Langton and Durose, 2013). This Explanatory analysis can help people understand what are the most common time, type of traffic stops and most common possible outcome. The possible outcome can help build confidence in the police and improve the law and order also.

3. What data are you going to use for this? How will you acquire this data?

Stanford is analyzing a unique dataset detailing more than 60 million state patrol stops conducted in 20 states between 2011 and 2015. This dataset is compiled through a series of public records requests filed with all 50 states, and redistributing these records in a standardized form to facilitate future analysis. The Stanford Open Policing Project data are made available under the Open Data Commons Attribution License for educational project.

Data Source and sample: https://openpolicing.stanford.edu/data/

4. In brief, outline your approach to solving this problem (knowing that this might change later)

- Data volume: - I plan to use 3 heavily populated cities of Florida (Miami, Orlando, Tampa) and compare with the entire state of Montana. I will use one year's data of a state for training, and one year's for validation and one year's data for prediction. Approx. 500K records will be used.

- Data manipulation using Pandas and other Python packages: -
Work with missing or invalid values, Data wrangling steps including filter data by year, clean messy and incomplete data, group and aggregate data

- Data analysis and visualization: - Data Storytelling using mapplotlib python package primarily focusing on the most common type of Bar chart, line chart, Scatterplot, histogram, density plot

5. What are your deliverables?

The proposal will be part of a github repository for my project. All code, data sets and further documentation I write will be added to this repository.
I will also create a Pdf project report for the entire project outcome and observations.
Project report will highlight comparison of heavy density (FL) v/s remotely populated areas (MT)

Exploratory Analysis: -

- What are the most common time of year and time of the day for traffic pullover?
- What is the most common type and cause of traffic pullover?
- What is most common outcome of traffic stop.
- What is the profile of the repeat offenders and where are most common stops?
- What is the role of age, race and time in the traffic stops?

Clustering and analyzing different model:-

- Compare Florida high density populated area with the other less populated state
- For speed related pullover, what is most common age/race.
- What is the most probable time of traffic stop during the day /month /Year?
- How are millennials doing compared to Gen-x and is Florida really senior friendly?
- Who are most likely let go with just a warning based on pattern.
