Milestone 1

Data Science Report

CS3753 – Data Analysis

Fall 2020

Aaron Perez

Department of Computer Science

University of Texas at San Antonio

One UTSA Circle, San Antonio, TX 78249

Email: [tle728@my.utsa.edu](mailto:tle728@my.utsa.edu)

Mykel Roa

Department of Computer Science

University of Texas at San Antonio

One UTSA Circle, San Antonio, TX 78249

Email: [bva733@my.utsa.edu](mailto:bva733@my.utsa.edu)

David Levy

Department of Computer Science

University of Texas at San Antonio

One UTSA Circle, San Antonio, TX 78249

Email: [mca880@my.utsa.edu](mailto:mca880@my.utsa.edu)

Idea 1)

Description: The metadata on breaking in cities used by smart cars vs the actual accident locations and what the means for the city’s civil engineering.

Problem addressed: The problem this analysis will attempt to address is that of the city’s infrastructure and engineering of future layouts. Typically, there are high accident areas in cities, but the reality of that vs where people are hard breaking are very different. If there are areas before the accident that can be addressed, thus being proactive vs reactive, I believe this should be the focus of future endeavors.

Background: Major cities can be differentiated by the architecture and the infrastructure layout of the city. What we were thinking as a group is how well each city is structured to handle the amount of people in the city. If there are more accidents, this could mean that there are poorly thought out roads and highways. There could be some cities that are very surprising, given the amount of people in the city. Finding out which cities handle traffic the best can lead to knowing how cities need to adapt and change to fit the amount of people are in them.

As cities grow and develop, the information that we find can help prevent future car accidents and save lives. We can find the similarities of certain city roads between a large group of cities so it can be implemented to other growing cities. Additionally, we can find what specific road structures can create the most amount of accidents.

Skills needed:

* Basic python coding
* Advanced understanding on how to read data
* Advanced understanding on calculating the large amount of data
* Knowledge on the kind of accidents that are mostly affected by city infrastructure and not the drivers

Dataset: <https://www.kaggle.com/c/lyft-motion-prediction-autonomous-vehicles>

Idea Videos: <https://www.youtube.com/watch?v=sHmVMclYGw0&t=283s>