Capstone Three Project Proposal
Edward Seymour

Millions of people are on the road everyday. In fact, there are more than seven million cars registered in LA county alone. With that many cars on the road, and pedestrians and cyclists, it can be quite dangerous to commute to everyday activities like work and grocery shopping, However, some routes are more dangerous than others.

I will create a model to return the likelihood of an accident occurring on the different routes returned by the Google Maps API when prompted by a user, such that the user can accurately choose the safest journey to take more than 75 percent of the time. For this project, I will be solely focusing on the city of Los Angeles.

Success criteria for this project are:
1. Speed - The application needs to run quickly in the background so that the time lost is negligible.
2. Accuracy - The application needs to correctly identify the safest route, with an success rate of 75%

While weather and time are certainly factors in accidents, there is an obvious human component to traffic collisions, which we cannot accurately grasp with this dataset. The set contains a column on whether a phone was being used during the accident, however there is a chance that some people would lie so this isn't that useful. As well, our final model can't know if any of the other drivers on the road are currently using their phones, or if they're angry, or tired, or any number of other factors that could cause an accident.

Stakeholders for this project include drivers in Los Angeles, as well as city planners. Drivers will be well informed about the safety of each route provided by the Google Maps API, which they can use in their everyday lives. City planners will be able to identify roads that pose a greater risk to the public. Through further investigation of these dangerous routes, city planners could determine the cause of the increased risk.

The data that will be used is listed below,
1. SWITRS dataset on traffic collisions in California from 2001 to 2022. Sqlite files include info on collisions, parties involved, injuries, and a few weather factors. This is the dataset that will be used for modeling. The set contains over 300,000 collisions in the city of Los Angeles from 2007 to 2021.
2. OpenWeather API provides weather data used for training the model, as this set has more useful weather factors, mainly visibility, than the California traffic collision dataset.
3. Google Maps API to provide routes on demand to users, who will then be given the likelihood that each route will result in an accident. API