**README**

This repository contains all data processing steps and preliminary raw data sets used for my data 115 final group project on California highway Collision data. this repository contains multiple files and their contents are to be disclosed as according below. the initial data set used for this project was gathered from kaggle and the link to the dataset can be found here:

**https://www.kaggle.com/alexgude/california-traffic-collision-data-from-switrs**

This dataset included all Collision data reported to the California Highway Patrol from 2001 to

2020 it was instrumental in the formulation of this final project and served as the main data set

for my analysis, within this set was included many descriptors of every collision and both

qualitative and quantitative data surrounding facets of each accident all as reported to the CHP

**REPOSITORY CONTENTS:**

**RAW:** Contains the initial CSV used for this project. (250k randomly sampled entries from the kaggle set.)

**PROCESSED:** Contains all spreadsheets and CSVs processed for this final project.

**VISUALIZATIONS:** Contains all visualizations used for this final project

**DOCUMENTATION:** Contains this readme and the link to the video describing the documentation steps taken to process and analyze this data. As well as my jupyter notebooks and the final paper written about this subject.

*DOCUMENTATION VIDEO: https://youtu.be/ck59mJYV3fU*