**Capstone Project Proposal**

**Context**

This project is intended to create a visualization for traffic violation data. Geolocations within the data will be used to create visualizations that consist of pins for each violation, as well as the ability to sort through varies criterial, such as cars make and model, type of violation, race, etc. The visualizations are intended are for the use of public agencies for the purpose of data analysis.

**Real World Use**

An example of a visualization of this data would be vehicle specific which could potentially be used by car insurance agencies in help determine rates for a location.

**Data Source**

The data is sourced from <https://www.kaggle.com/microtang/exploration-the-violations>.

**Tools/Technology**

Developed on python with geo-mapping using Folium, a geospatial data visualization tool which can be used in representing the data.

<https://towardsdatascience.com/creating-a-simple-map-with-folium-and-python-4c083abfff94>

**Structure**

A python program will take the raw data and transform it into usable data for the purpose of this project. This process will filter out unwanted data where the new data will be stored locally. The python program will consist of visualization tools which will produce diagrams and maps that a user may access through a hub like interface for the purpose of analyses.

**GitHub**

The repo for this project is, <https://github.com/kpatel-io/DataEngineering_CapstoneProject>

This repo consists of a slide deck and a readme file with a description of the project

**Deliverables Order**

Readme.txt file for the project uploaded to GitHub

Import data to project

Integrate Folium as mapping tool

Create a few sorting/analyzing tools for data specific criteria