From:
Gengkui Liu

Subject:
Trip Data Visualization Tool for Driver Behavior Exploration

Date:
September 23, 2022

Audience

Insurance Companies

Intention

To create a geodata visualization tool that allows data scientists in insurance companies to explore driver's dangerous behaviors in a trip. The data scientists will be able to identify actions like speeding/hard brake/sharp turn/frequent changing lanes using this tool. As a result, they can assess each trip and assist the auto insurance price estimation for the driver. If the vehicle in a trip did speeding, then the price of the auto-insurance policy might increase. If there was no risky actions in a trip, the price would be lower. The data scientists can even integrate this trajectory data exploratory tool into their larger machine learning workflow if they want.

Geographic Extent

All states in the USA

Deliverables

The final product will be a Python-based, Jupyter Notebook-hosted tool that contains codes and visualizations. The data scientists can easily integrate this notebook into their cloud platforms such as Google Cloud Platform. Then they can plug in their own data source using Big Query and run this visualization tool. It serves as a tool to explore the trajectory data, and a documentation that describe the process of cleaning, processing, and visualizing the trip data.

Budget

I plan to spend 10 hours each week on this project. These hours include collecting dataset, communicating with my mentor/instructor, researching articles to implement the features in this tool, optimizing my code and fixing bugs, etc. I will specify where I find the data, what each line of code does through comments. In terms of the visualization component in the notebook, I will use a basemap with 19 levels of detail; I will also visualize acceleration for detecting speeding, velocity for detecting sharp turn/changing lanes, etc.

Technology & Extraneous Costs

I will use free, open source trajectory dataset. This dataset at least needs to have attributes like longitude, latitude, altitude, heading. If it is a tabular dataset like csv, then each row should represent the attributes of the vehicle in one second of time.

I will use Python as the programming language and Jupyter Notebook as the development environment. Both are free.

Product Development & Delivery

The product will be created on GitHub as a repository. I will also create slides to introduce this tool to people from the insurance company and let them know the benefits of using it.

Timeline

Week 1

- Decide a project idea that I am interested in.
- Communicate with the instructor and get his feedback
- Adjust some details according to his advice

Week 2

- Communicate with my mentor Katie and receive her feedback
- Take a deeper look at the data source she provided to me and see if they will be useful

Week 3

- Determine a dataset.
- Research on what Python libraries will be necessary for this project
- Research on what attributes are necessary for visualizing risky actions
- Update with my instructor and mentor

Week 4

- Coding
- Fixing bugs
- Update with my instructor and mentor

Week 5

- Coding
- Fixing bugs
- Update with my instructor and mentor

Week 6

- User testing
- Optimize my code

Week 7

- Submit my project
- Make slides

Last Updated: 9/23/2022