US Road Accidents

Early Warning System

Objective: Analyze 'US Road Accidents' and their 'Severity' in terms of impact on traffic delay

Severity is defined 1-4 where 4 is the most severe i.e. longest traffic delay associated with the accident
***Note**: Severity in this project does not include the property or human impact*

Brief Description of the project / approach:
1.) We explored about 50 features to predict the 'Severity' of an accident using a Machine Learning Model. Some of them are:
- Location features such as zipcode, city, state
- Highway vs non Highways
- Climate features such as weather consitions, precipitation etc
2.) The data was stored on AWS Postgres
3.) It was stored and queried using PySpark due to the size of the dataset
4.) **Tableau analysis is for only the more severe accidents, that is 3 & 4 severity**

Description of data source

Source: Kaggle
Datasize: ~3M entries
Time period: February 2016 to December 2019
Scope: 49 US states (No Hawaii and Alaska)

Acknowledgements

https://www.kaggle.com/sobhanmoosavi/us-accidents#US_Accidents_Dec19.csv

- Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, and Rajiv Ramnath. "A Countrywide Traffic Accident Dataset.", 2019.
- Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, Radu Teodorescu, and Rajiv Ramnath. "Accident Risk Prediction based on Heterogeneous Sparse Data: New Dataset and Insights." In proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, 2019.

Questions we hope to answer:

- Is it possible to accurately predict accident severity based on certain features?
- Given specific weather conditions, can we provide an early warning system to drivers?
- Can output be used to develop pre-planning to avoid traffic hotspots during certain conditions (weather, time of day, road features, etc.)?