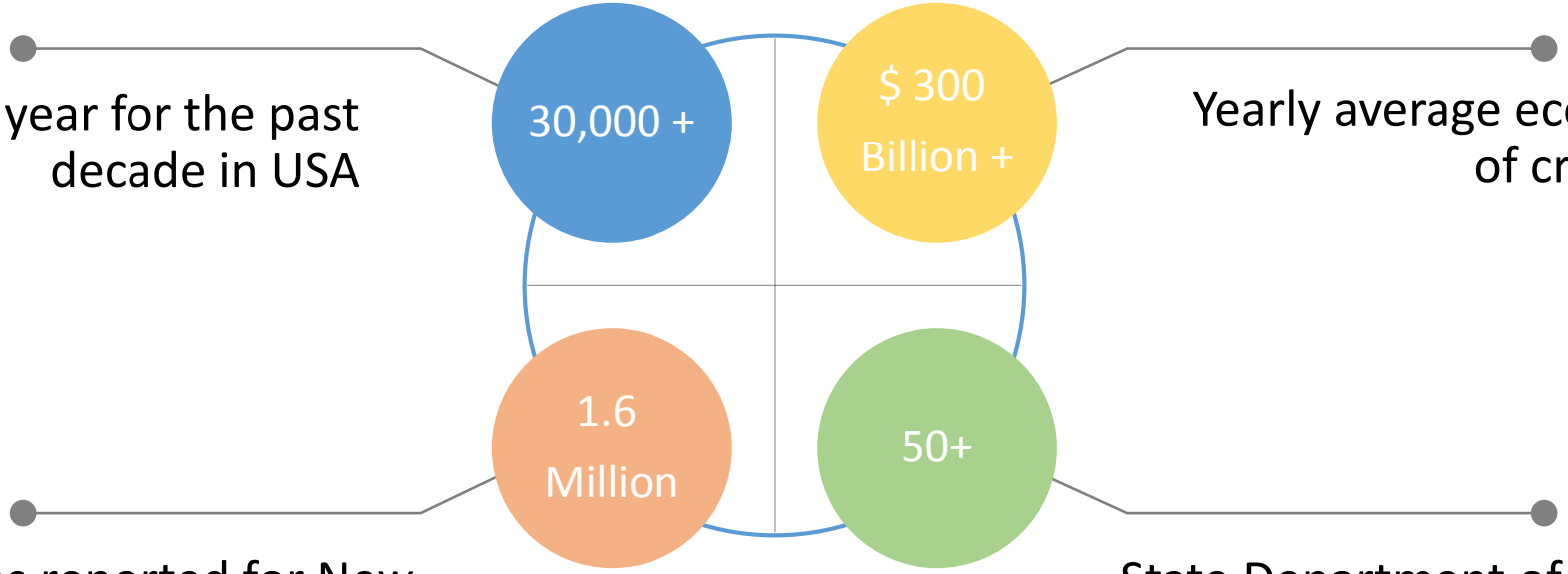
Location Specific Crash Modeling with Machine Learning

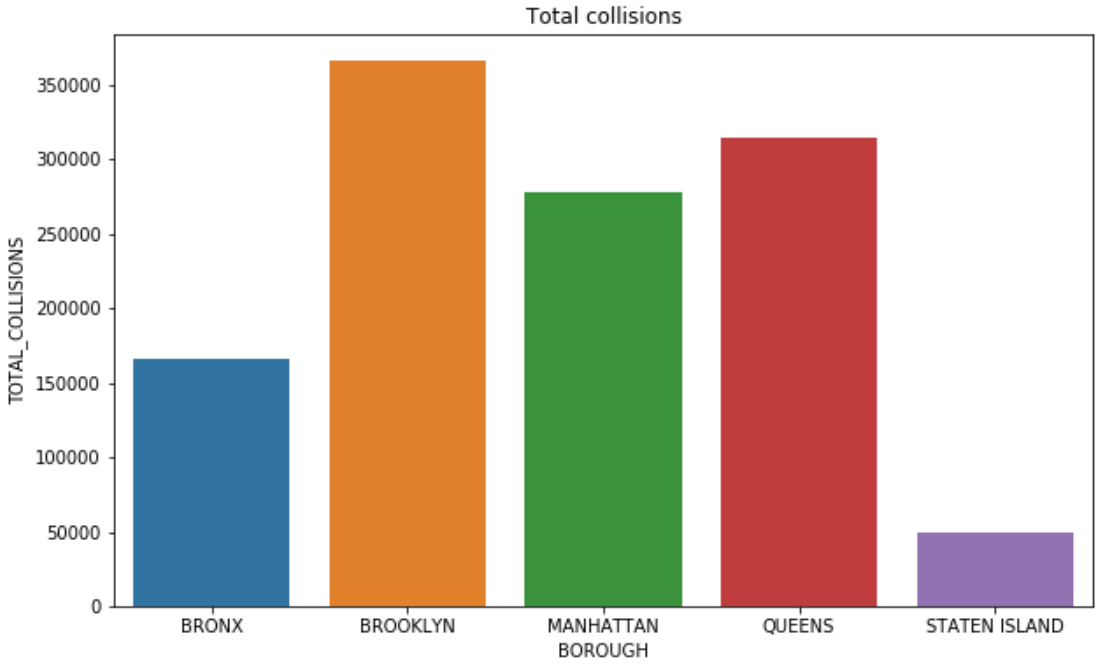
Shrikant Fulari: The Data Incubator Project

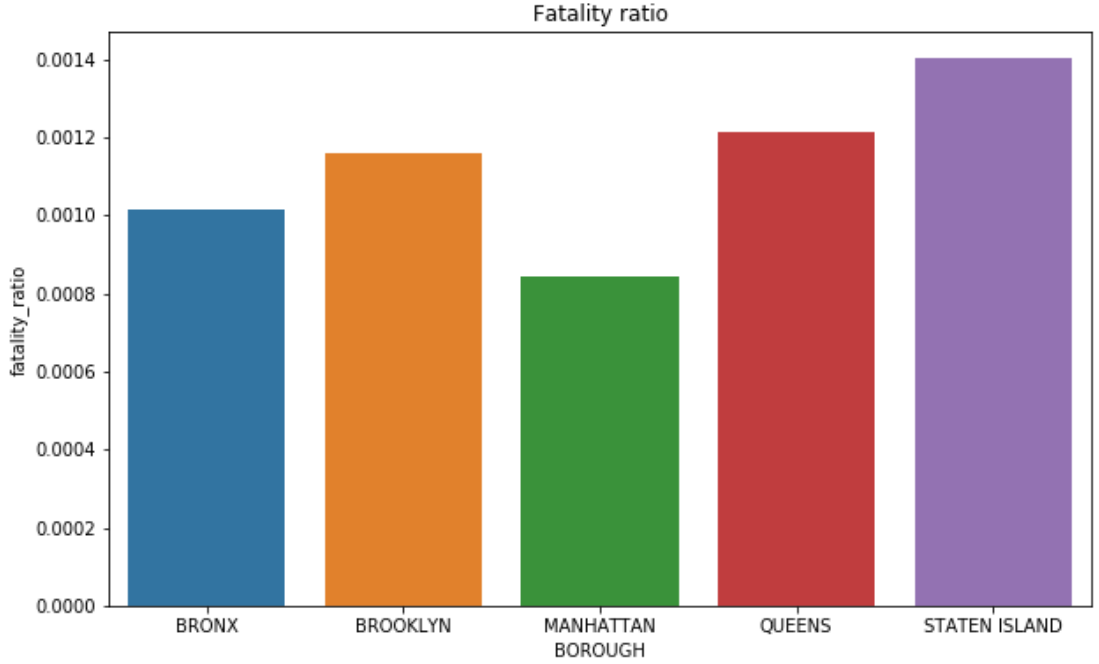
Motivation



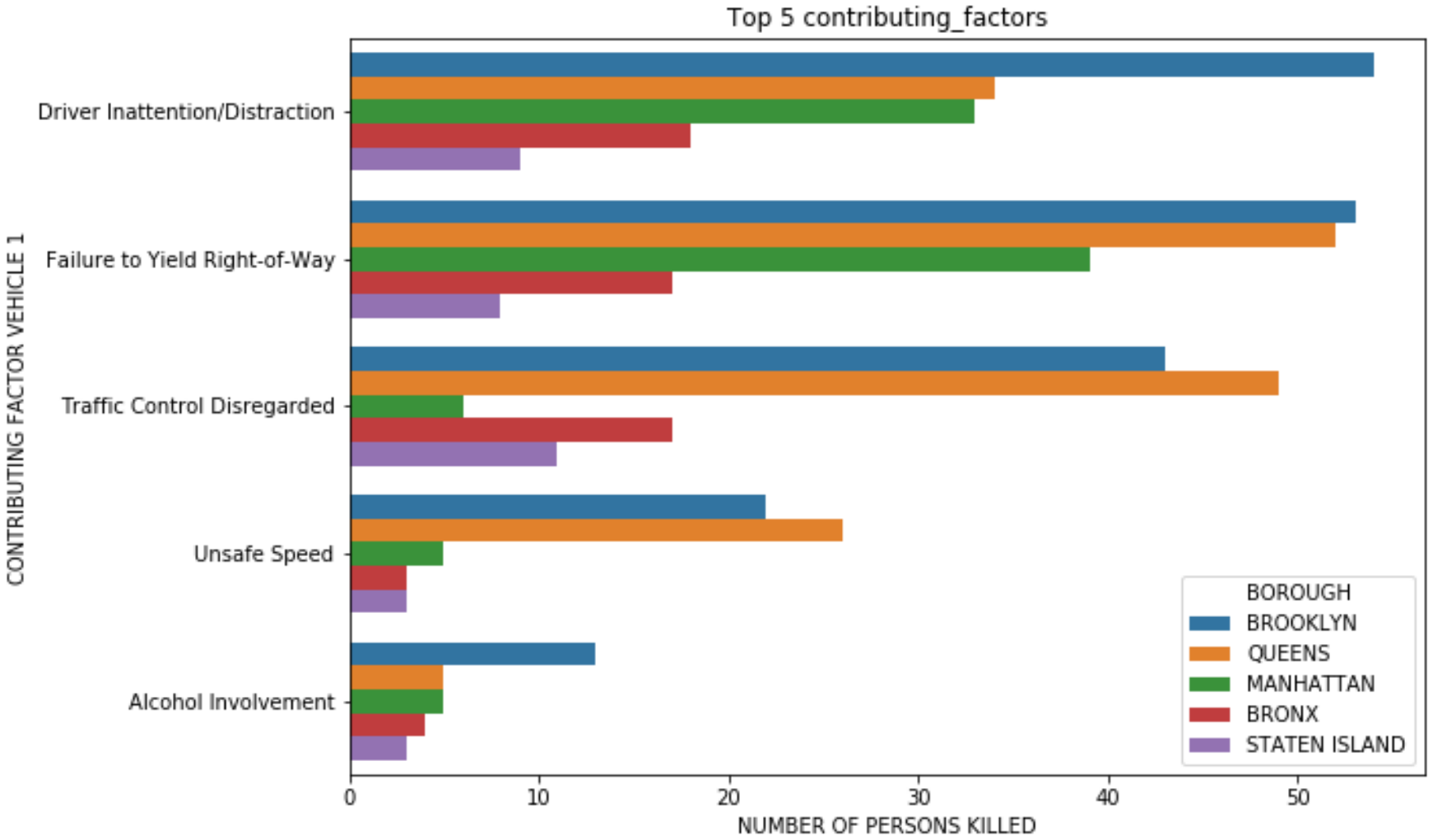
Source: https://www.transportation.gov/, https://www.nhtsa.gov/research

Insights from NYC collision data: Fatality ratio

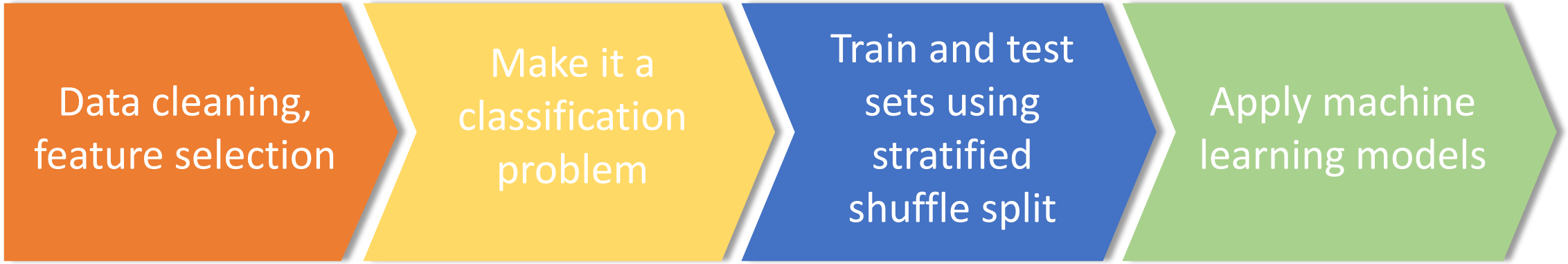




Insights from NYC collision data : Contributing factors



Fatal/non-fatal crash classification model



- Treat null values
- Generate variables such as hour, day of week and month from date time
- Select most helpful features

- Output variable: If Fatalities >0 then Fatal, else non-fatal event

- Used stratified shuffle sampling to get reliable accuracy score

- AdaBoost
- Logistic Regression
- Support vector classifier
- K-nearest neighbor
- Decision Tree
- Random Forest

Pack the entire application in Tkinter using Python

Welcome to New York City Micro-Crash modeling

Model development inputs

Select if you would like a model based on Zip Codes or Borough

Select

Select the corresponding Zip Code or Borough

Select

Click here to generate model and results

Model performance charts

Kindly click here for project summary

Zone summary charts

Welcome to New York City Micro-Crash modeling

Model development inputs

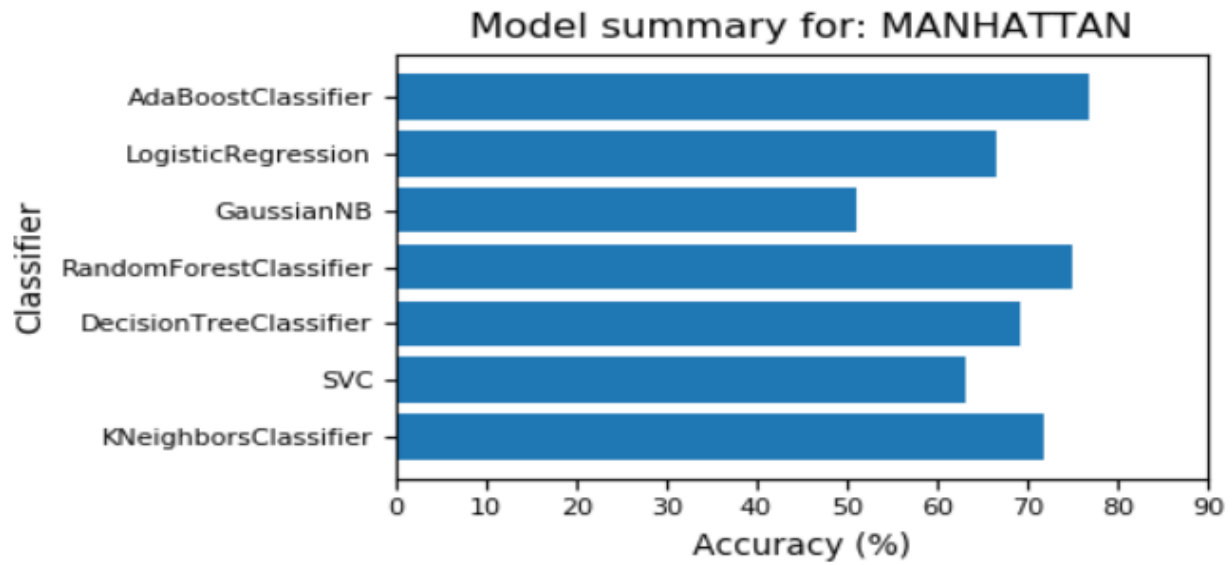
Select if you would like a model based on Zip Codes or Borough

BOROUGH

Select the corresponding Zip Code or Borough

MANHATTAN

Click here to generate model and results

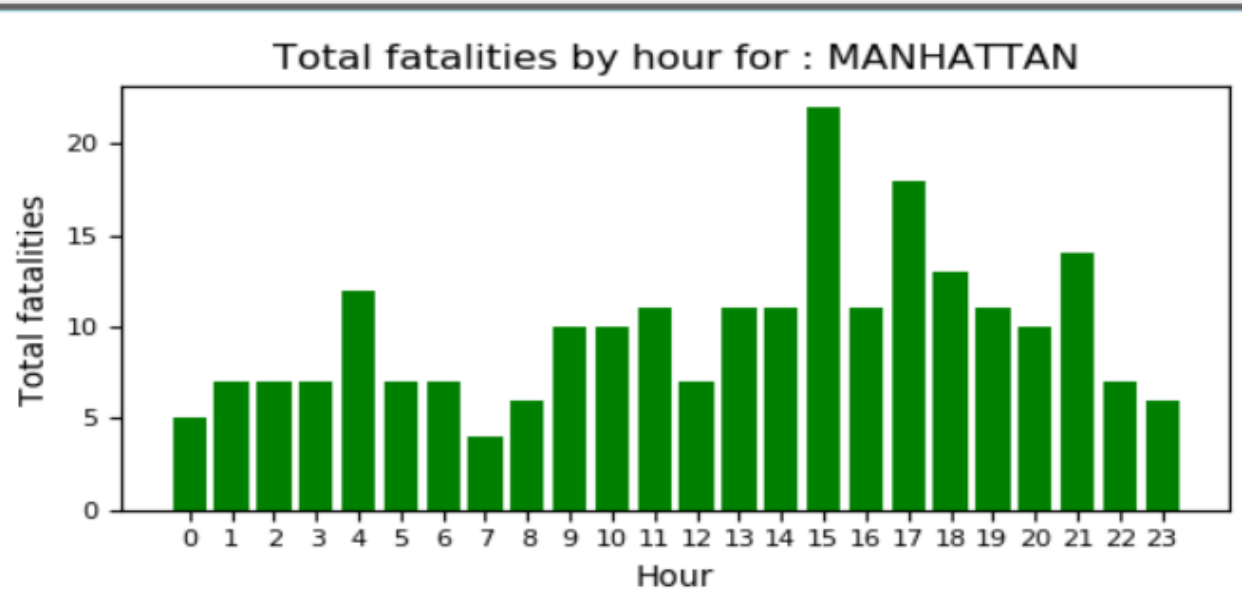


Kindly click here for project summary

Motivation:

1. Vehicle crashes occur all over the world, with varying severity, traffic rules, vehicle characteristics, driving behavior, weather etc., each of which play a significant role in the outcome of the crash.

2. While most of the data science applications for crash analysis have been only in advanced visualization, there is limited study of developing advanced machine learning algorithms that can help predict future crashes and its severities based on several parameters.



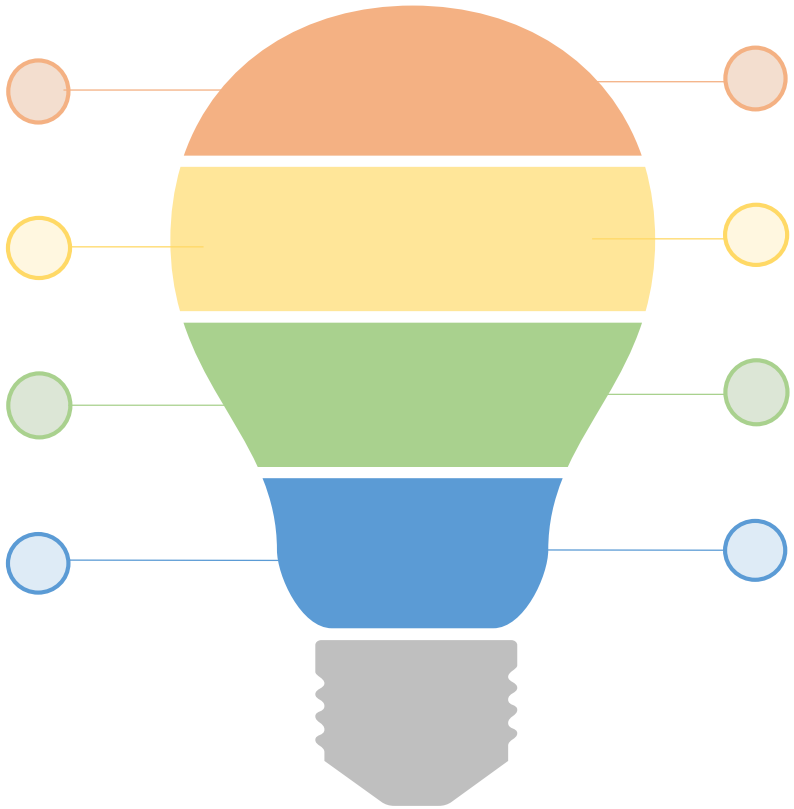
Scalability and future work

Provide diverse end user applications

Enhance accuracy of models

Expand database for wide coverage

Develop a centralized application



Benefits

Instant control measures to prevent crashes

More reliable crash predictions

Better understanding of crash data

One stop for all interacting agencies

THANK YOU!!