Guardian of Streets

“NYC Road Safety Program”

PROJECT BY-

SHUBHAM CHOUDHARI

RITESH RAUT

DEEKSHITH REDDY

JOBIN JOHN

GOPI KRISHNA

CONTENT

01 PROBLEM STATEMENT

02 DATA CLEANING

03 EXPLORATORY DATA ANALYSIS

04 ML MODELS

05 ANALYSIS/CONCLUSION

06 FUTURE ENHANCEMENT

PROBLEM STATEMENT

New York City Council is running a road safety campaign . They would want to ensure the resources are efficiently utilized so that the campaign successfully addresses the major concerns. The council has delegated the responsibility to my analyst team to identify hotspots, major factors causing the accidents . We would also want to find the relationship between accidents and majority of features around it to identify high risk features .

DATA CLEANING

PHASE 1

Identified and handled null values

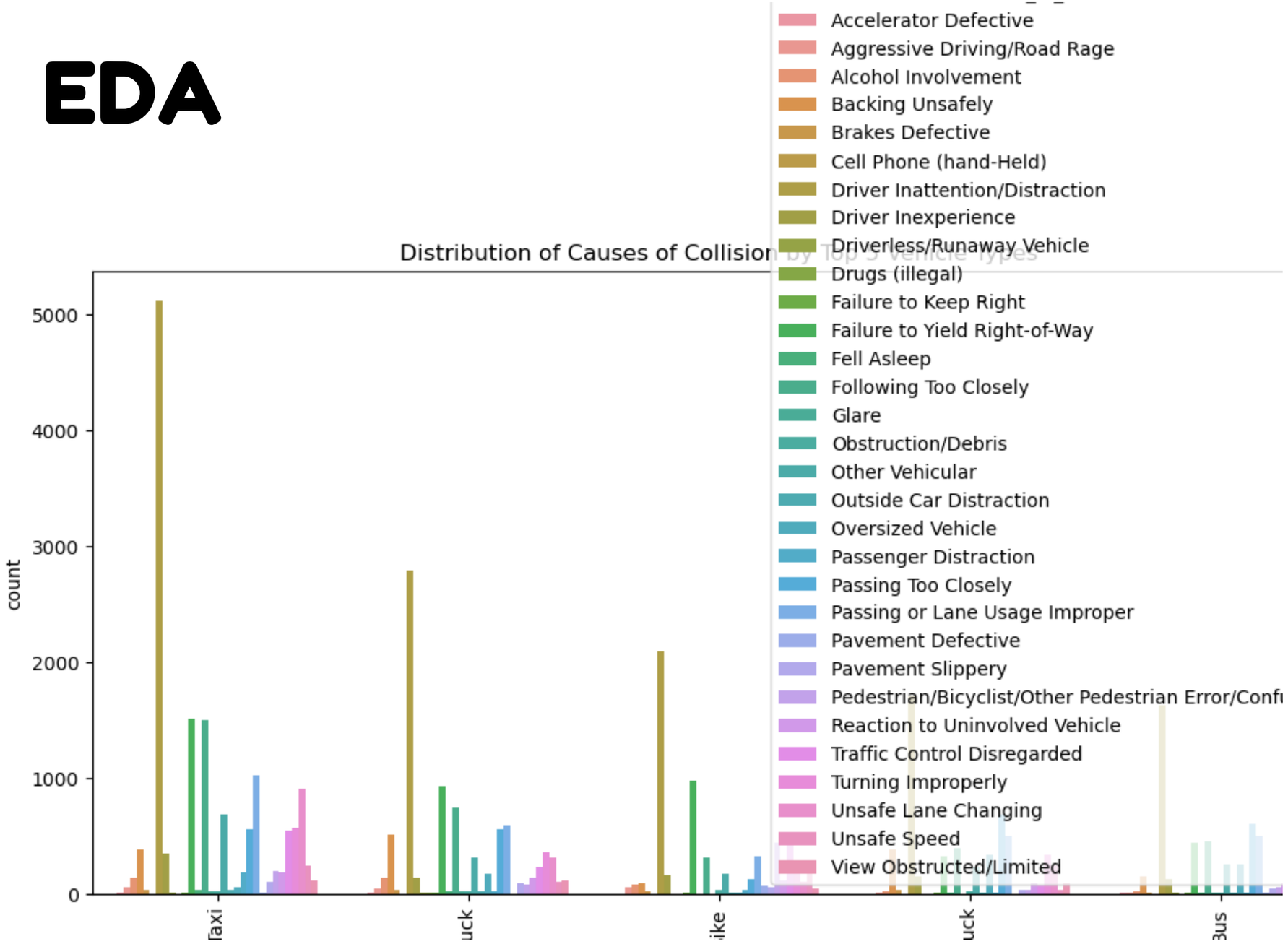
PHASE 2

Identifying not impactful features and dropping them

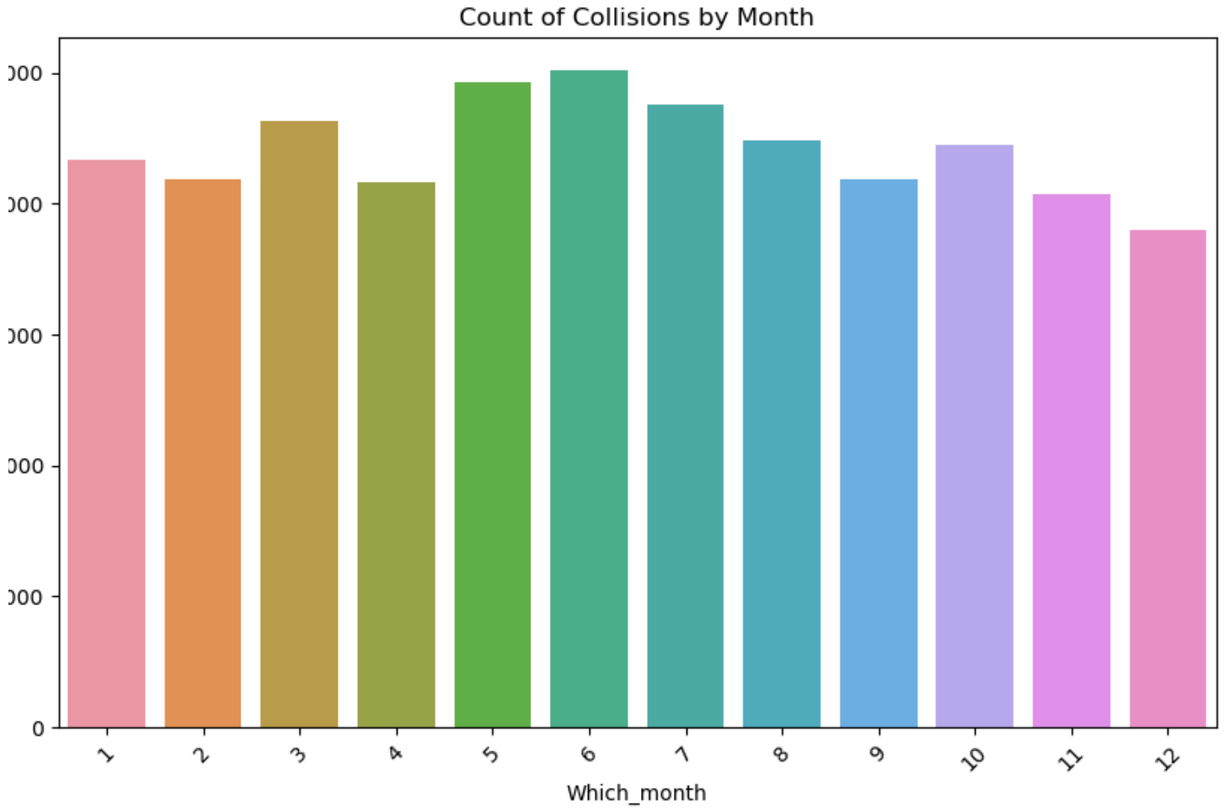
PHASE 3

Converted variables into categorical

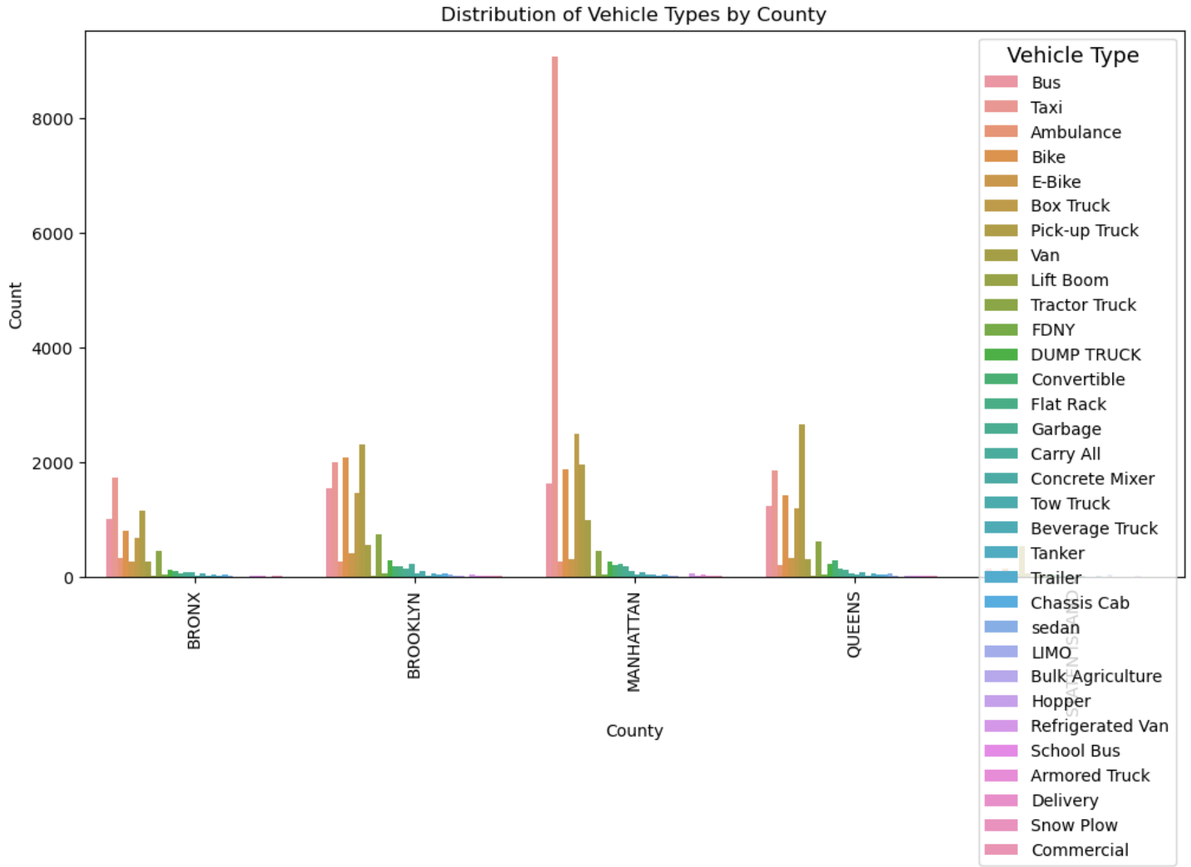
EDA


Distribution of Causes of Collision by Top 5 Vehicle Types
count
5000
4000
3000
2000
1000
0
Accelerator Defective
Aggressive Driving/Road Rage
Alcohol Involvement
Backing Unsafely
Brakes Defective
Cell Phone (hand-Held)
Driver Inattention/Distraction
Driver Inexperience
Driverless/Runaway Vehicle
Drugs (illegal)
Failure to Keep Right
Failure to Yield Right-of-Way
Fell Asleep
Following Too Closely
Glare
Obstruction/Debris
Other Vehicular
Outside Car Distraction
Oversized Vehicle
Passenger Distraction
Passing Too Closely
Passing or Lane Usage Improper
Pavement Defective
Pavement Slippery
Pedestrian/Bicyclist/Other Pedestrian Error/Confu
Reaction to Uninvolved Vehicle
Traffic Control Disregarded
Turning Improperly
Unsafe Lane Changing
Unsafe Speed
View Obstructed/Limited

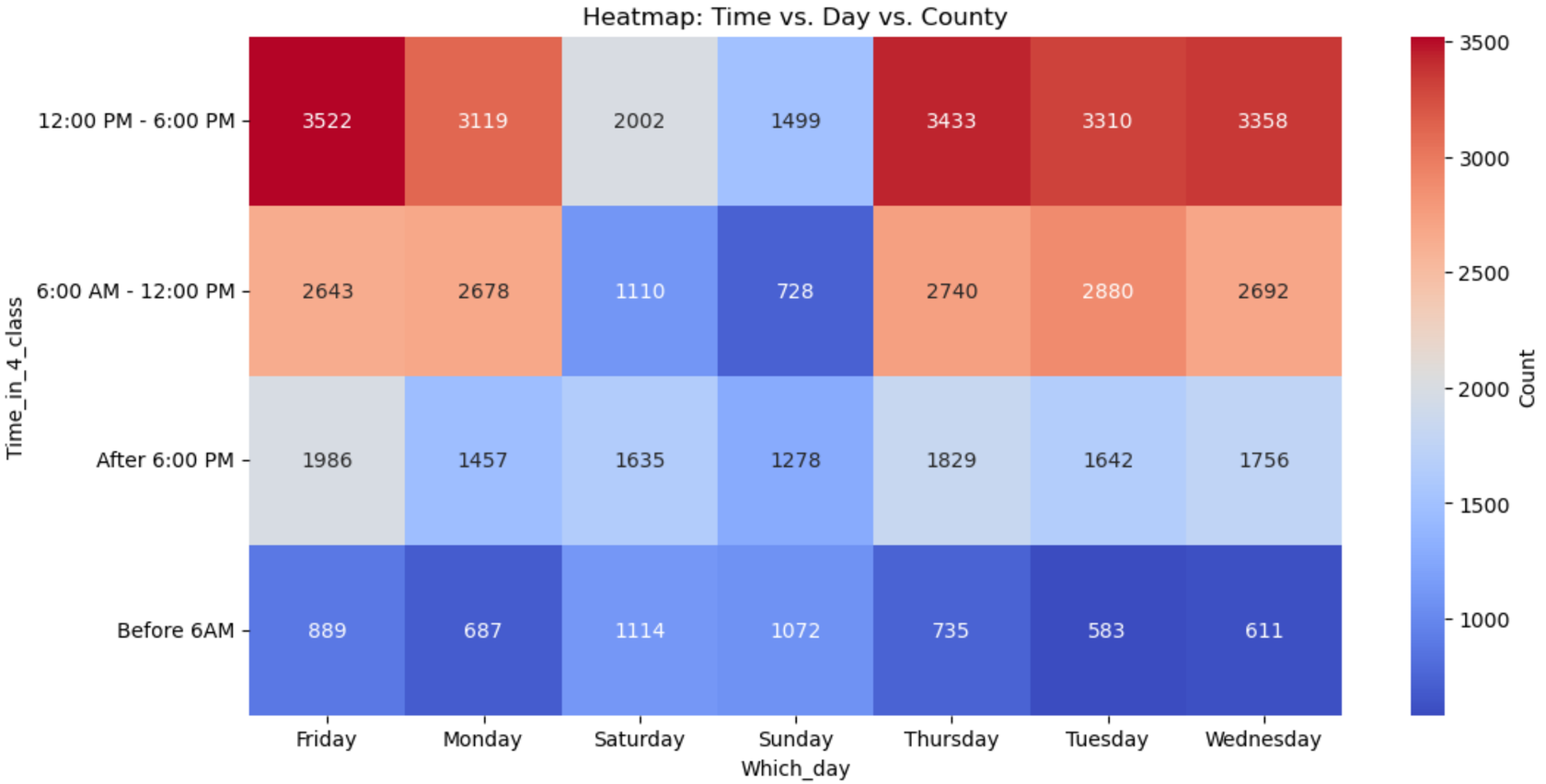
EDA

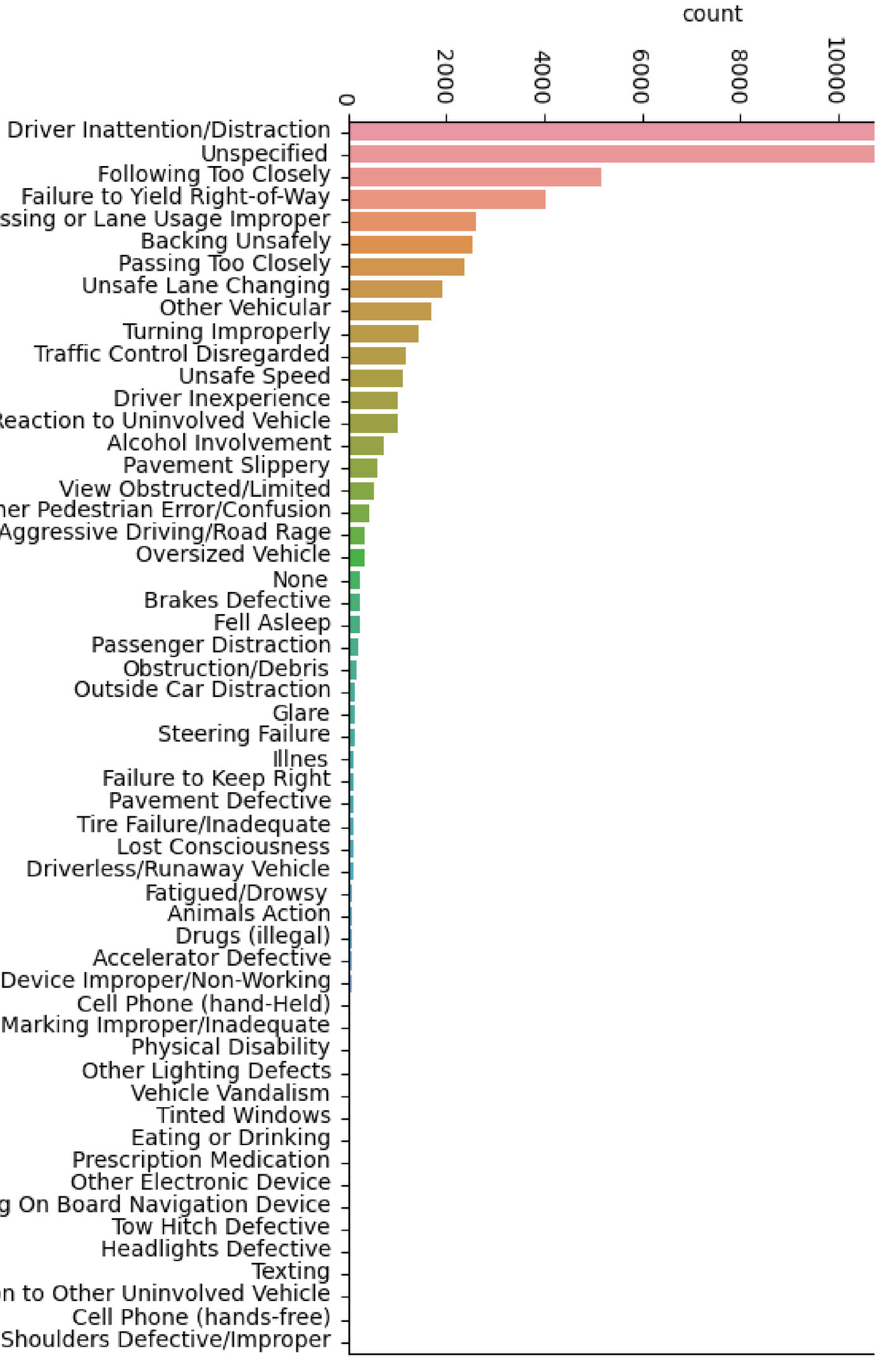

Count of Collisions by Month
000
000
000
000
0
1
2
3
4
5
6
7
8
9
10
11
12
Which_month

EDA


Distribution of Vehicle Types by County
Count
8000
6000
4000
2000
0
BRONX
BROOKLYN
MANHATTAN
QUEENS
County
Vehicle Type
Bus
Taxi
Ambulance
Bike
E-Bike
Box Truck
Pick-up Truck
Van
Lift Boom
Tractor Truck
FDNY
DUMP TRUCK
Convertible
Flat Rack
Garbage
Carry All
Concrete Mixer
Tow Truck
Beverage Truck
Tanker
Trailer
Chassis Cab
sedan
LIMO
Bulk Agriculture
Hopper
Refrigerated Van
School Bus
Armored Truck
Delivery
Snow Plow
Commercial

EDA


Heatmap: Time vs. Day vs. County
12:00 PM - 6:00 PM
6:00 AM - 12:00 PM
After 6:00 PM
Before 6AM
Time_in_4_class
3522 3119 2002 1499 3433 3310 3358
2643 2678 1110 728 2740 2880 2692
1986 1457 1635 1278 1829 1642 1756
889 687 1114 1072 735 583 611
Friday
Monday
Saturday
Sunday
Thursday
Tuesday
Wednesday
Which_day
Count
3500
3000
2500
2000
1500
1000


count
0
2000
4000
6000
8000
10000
Driver Inattention/Distraction
Unspecified
Following Too Closely
Failure to Yield Right-of-Way
ssing or Lane Usage Improper
Backing Unsafely
Passing Too Closely
Unsafe Lane Changing
Other Vehicular
Turning Improperly
Traffic Control Disregarded
Unsafe Speed
Driver Inexperience
Reaction to Uninvolved Vehicle
Alcohol Involvement
Pavement Slippery
View Obstructed/Limited
her Pedestrian Error/Confusion
Aggressive Driving/Road Rage
Oversized Vehicle
None
Brakes Defective
Fell Asleep
Passenger Distraction
Obstruction/Debris
Outside Car Distraction
Glare
Steering Failure
Illnes
Failure to Keep Right
Pavement Defective
Tire Failure/Inadequate
Lost Consciousness
Driverless/Runaway Vehicle
Fatigued/Drowsy
Animals Action
Drugs (illegal)
Accelerator Defective
Device Improper/Non-Working
Cell Phone (hand-Held)
Marking Improper/Inadequate
Physical Disability
Other Lighting Defects
Vehicle Vandalism
Tinted Windows
Eating or Drinking
Prescription Medication
Other Electronic Device
g On Board Navigation Device
Tow Hitch Defective
Headlights Defective
Texting
n to Other Uninvolved Vehicle
Cell Phone (hands-free)
Shoulders Defective/Improper

BUILDING OUR PREDICTIVE MODEL

Target variable - Persons Injured

Independent variables - Month,Vehicle type,time,street name and county

POSSIBLE MODEL CONSIDERED

- **RANDOM FOREST**
  R2 : 86%
- **GRADIENT BOOSTING**
  R2 : 13%
- **DECISION TREE**
  R2 : 71%

BASED ON THE ABOVE DATA WE ARE USING A RANDOM FOREST REGRESSOR FOR THE PREDICTION

IMPORTANT FEATURES AFFECTING OUR TARGET VARIABLES

features	importance
street_name	0.306358
VEHICLE TYPE	0.137222
Which_month	0.135169
Cause_of_collision	0.106185

CONCLUSION

- Based on the feature importance calculated by the model, we can see that the location (street), and the type of vehicle an important contributors to the accident probability.
- Hence, our analysis suggests these two features should be the primary focus of resource allocation.
- We have also identified primary factors that cause collisions, which can be used as a target for the safety campaign to address.
- Our predictive model can help the council gauge the overall fatalities that can occur given a particular location, vehicle type

FUTURE ENHANCEMENT

We have successfully provided descriptive analysis based on the data we had and also built a predictive model that can predict a number of fatalities. We plan to improve this model by removing its concise feature dependency so that it can predict the general number of fatalities even in the absence of some features. We also want to focus on developing a probable crash score that will calculate the likelihood of a crash based on the driver and vehicle history.

THANK YOU