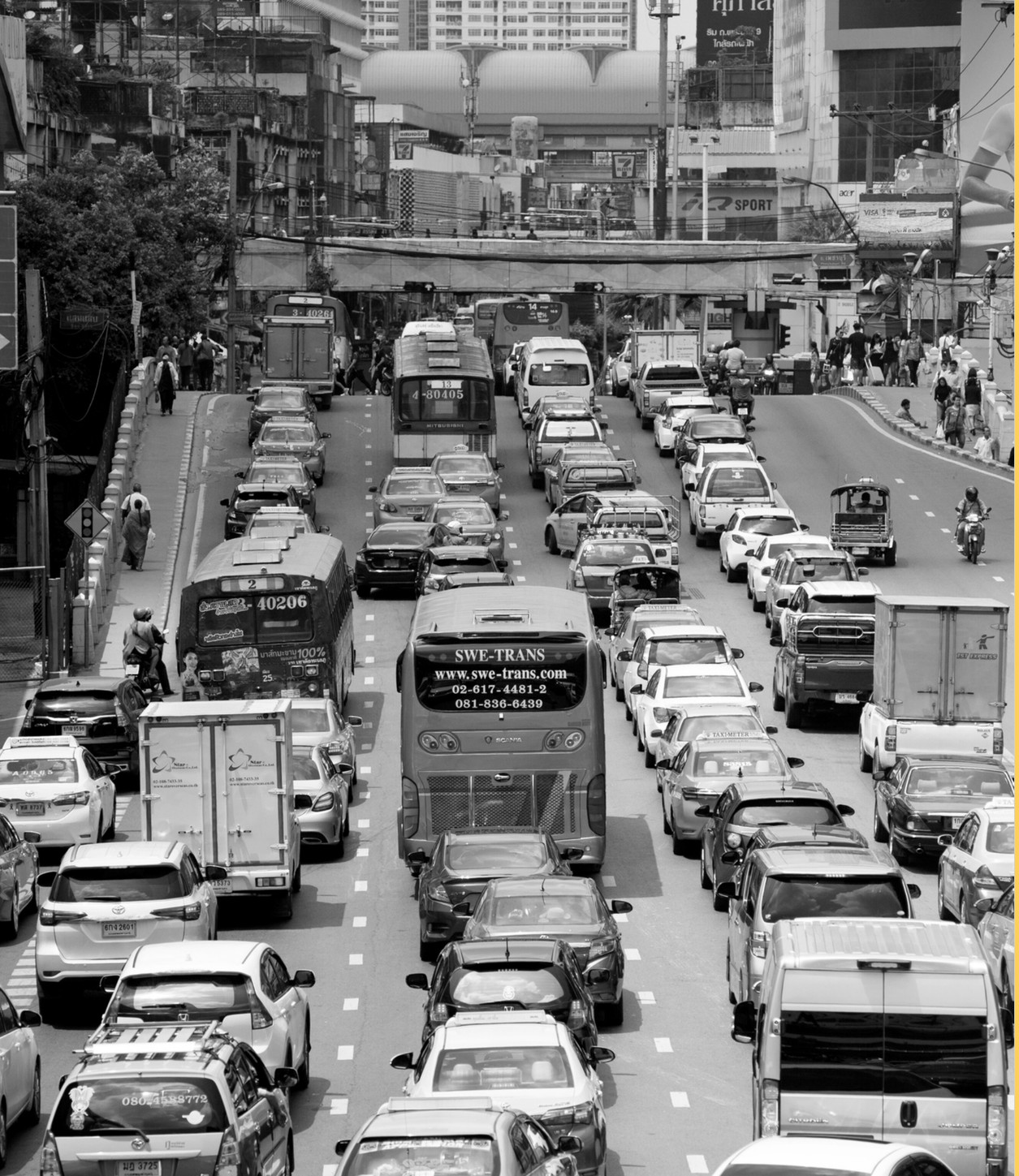
SEVERITY PREDICTION OF VEHICLE ACCIDENTS

By: Julie Leung



BACKGROUND

Driving is an integral part of everyday life for many individuals. However, it's crucial to recognize that motor vehicle accidents remain a leading cause of death in America. Shockingly, there is a 1 in 107 chance of dying in a car crash.

BUSINESS PROPOSITION

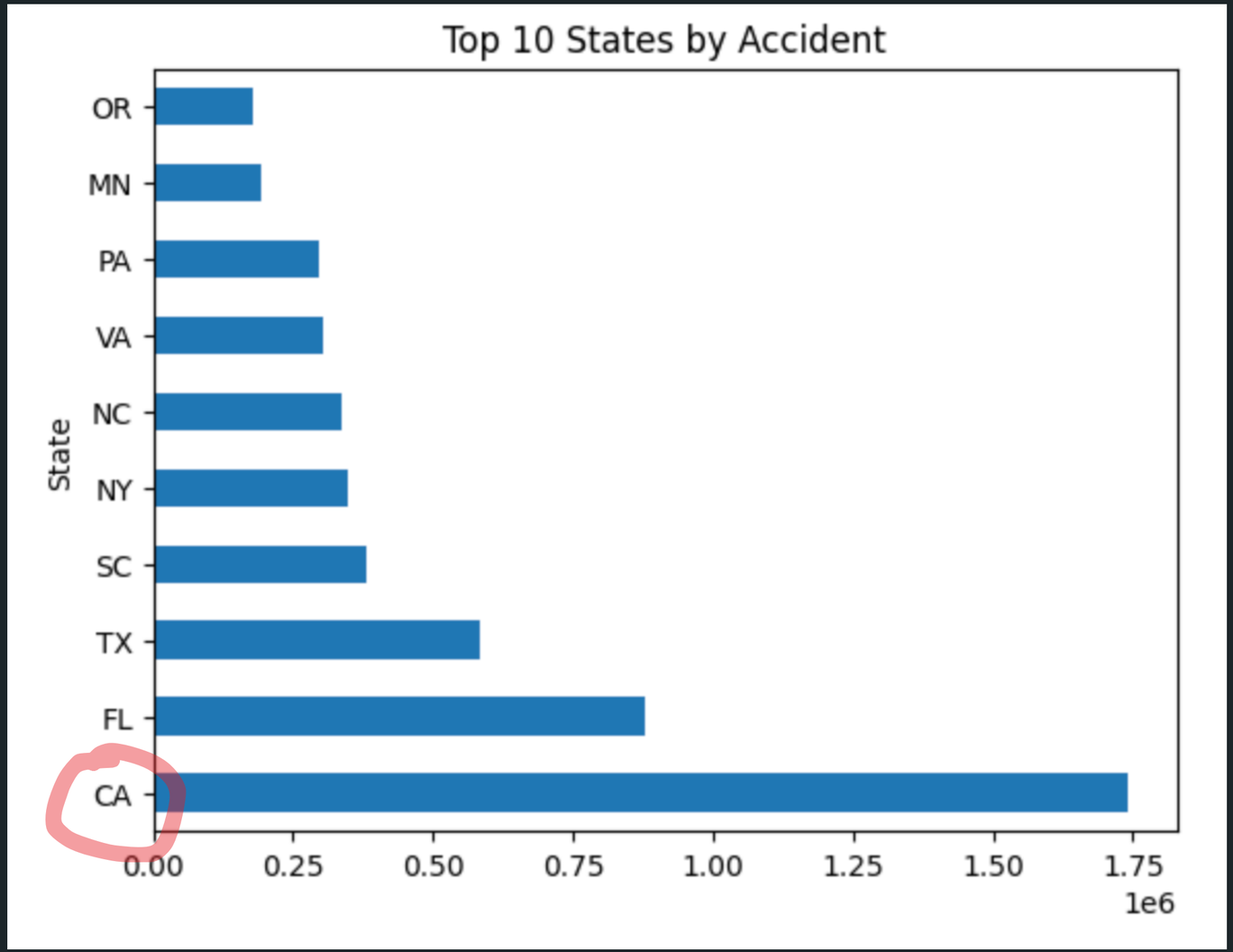
In an effort to kickstart their comprehensive safety action plans aimed at mitigating the severity of vehicle accidents, the Department of Transportation (DOT) is seeking to identify the top three cities with the highest rate of severe crashes.

DEFINING THE METRICS

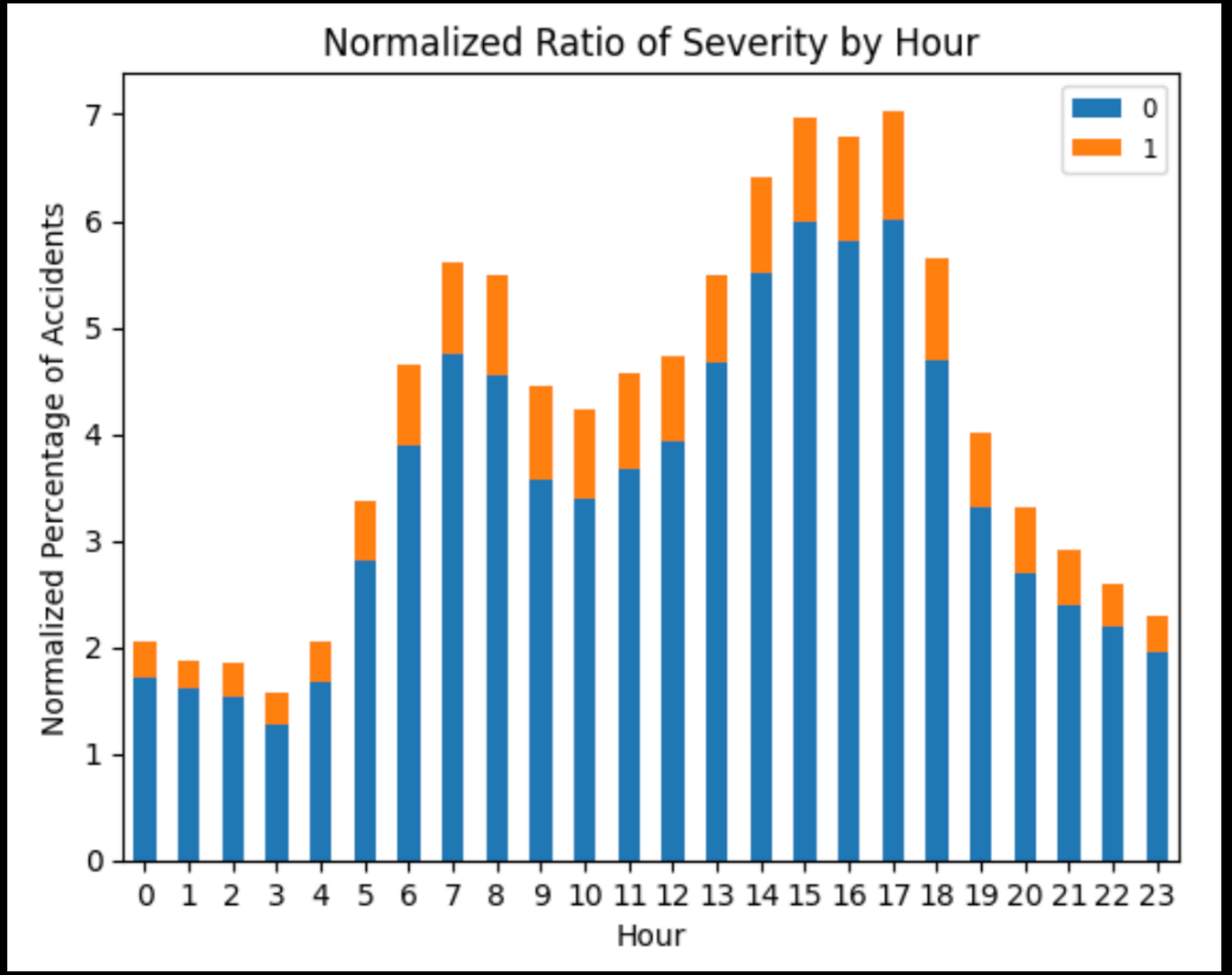
- Severity = 0: defined as least impact on traffic (i.e., short delay as a result of the accident)
- Severity = 1: defined as significant impact on traffic (i.e., long delay)

DATA SOURCE

- The dataset is from Kaggle.
- 7.7 million recorded accidents
- The accidents are from February 2016 to March 2023.
- Includes all states and DC, excluding Alaska and Hawaii

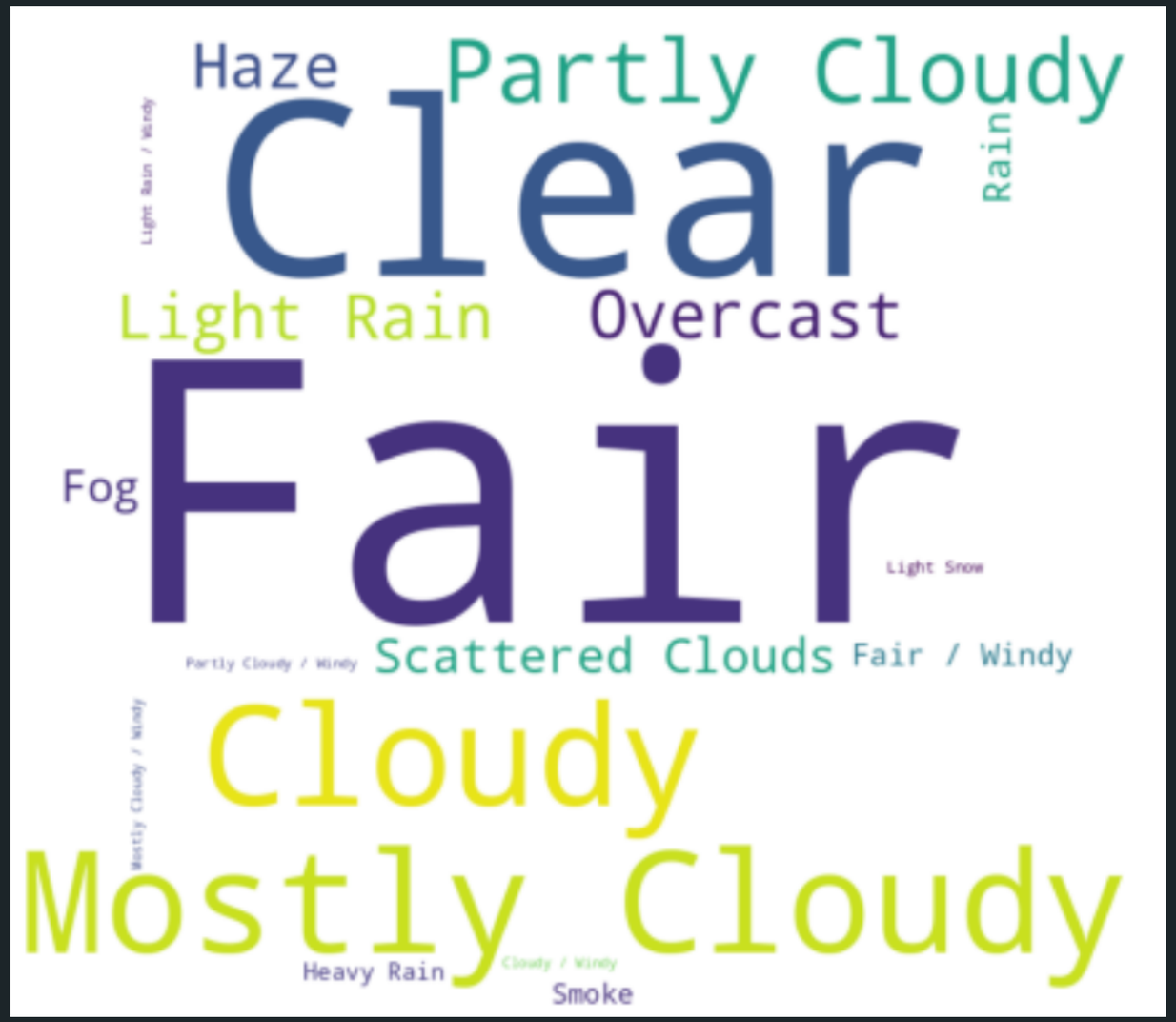


ANALYSIS

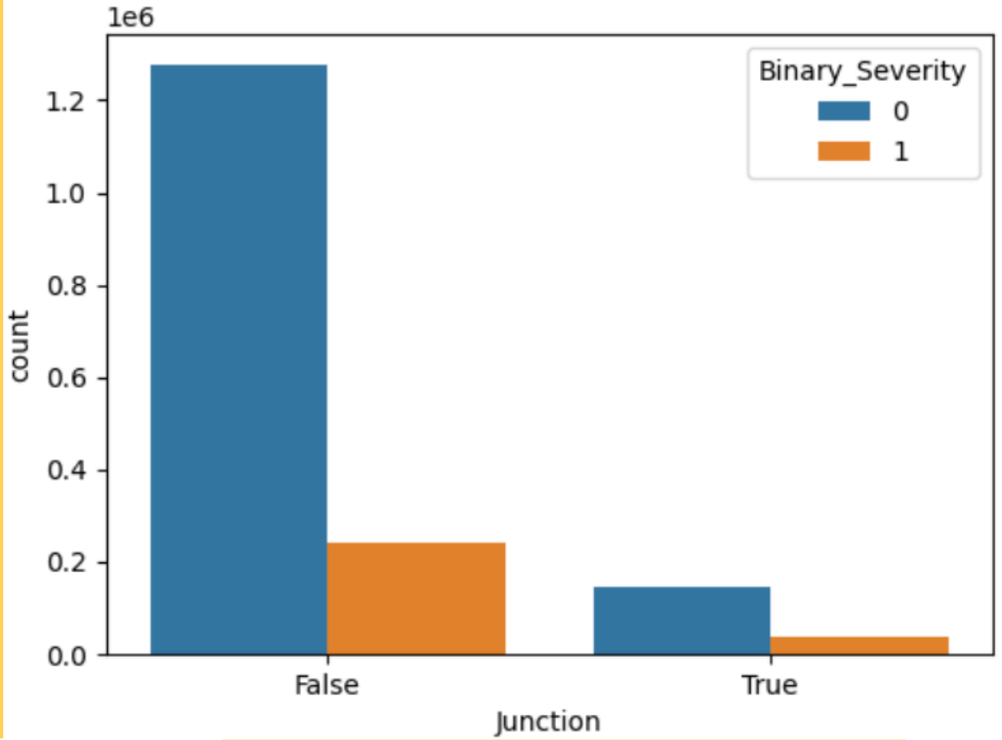

Normalized Ratio of Severity by Hour
Normalized Percentage of Accidents
Hour
0
1
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
0 1 2 3 4 5 6 7

WEATHER CONDITION

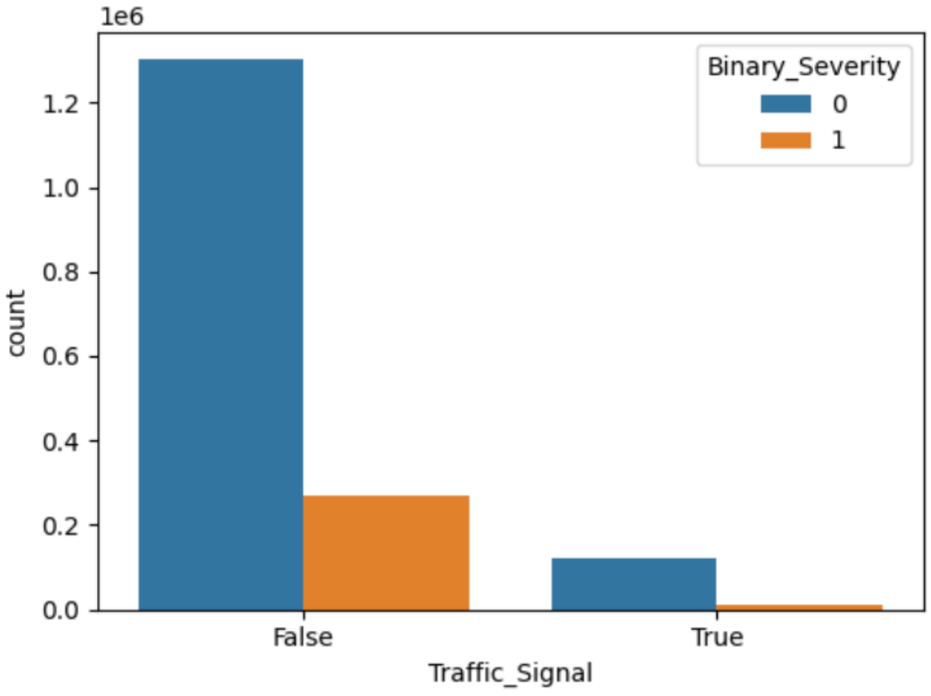
Majority of accidents occur under fair/clear or cloudy weather conditions rather than during snow or rain, which aligns with the typical weather conditions in California.



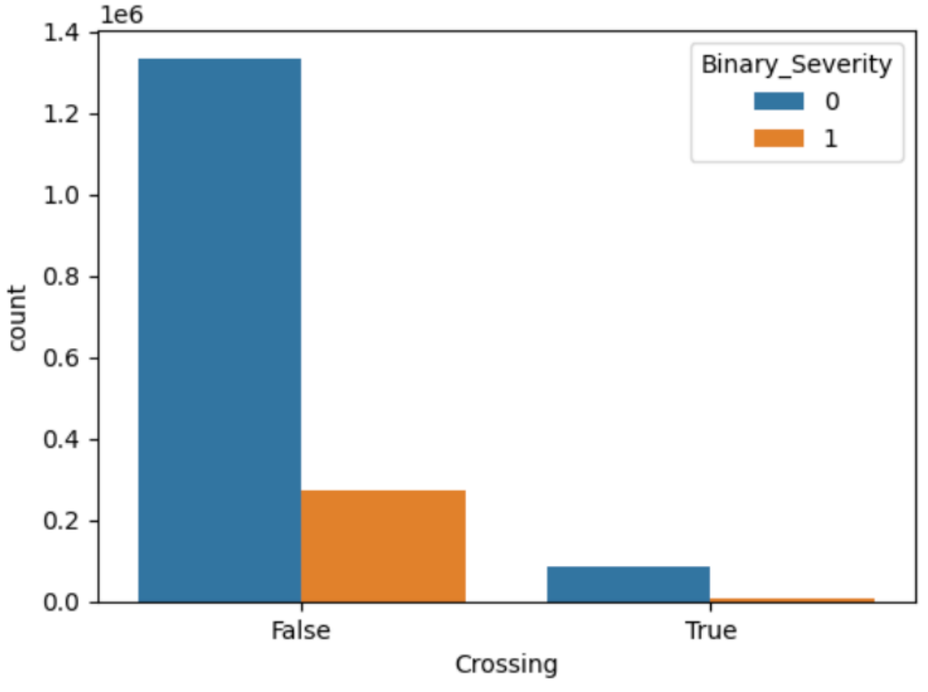
PRESENCE



ROAD FEATURES



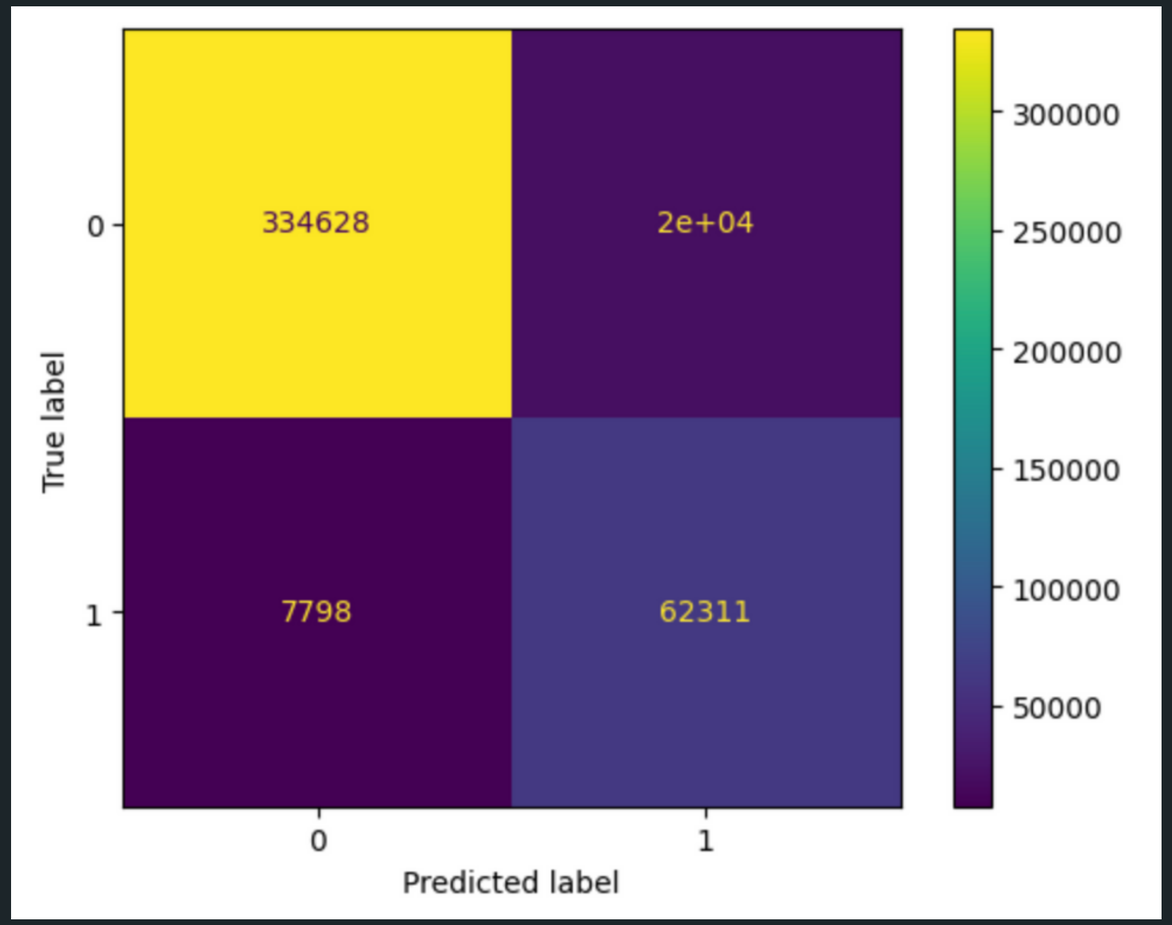
OF



MODELING

MODELS

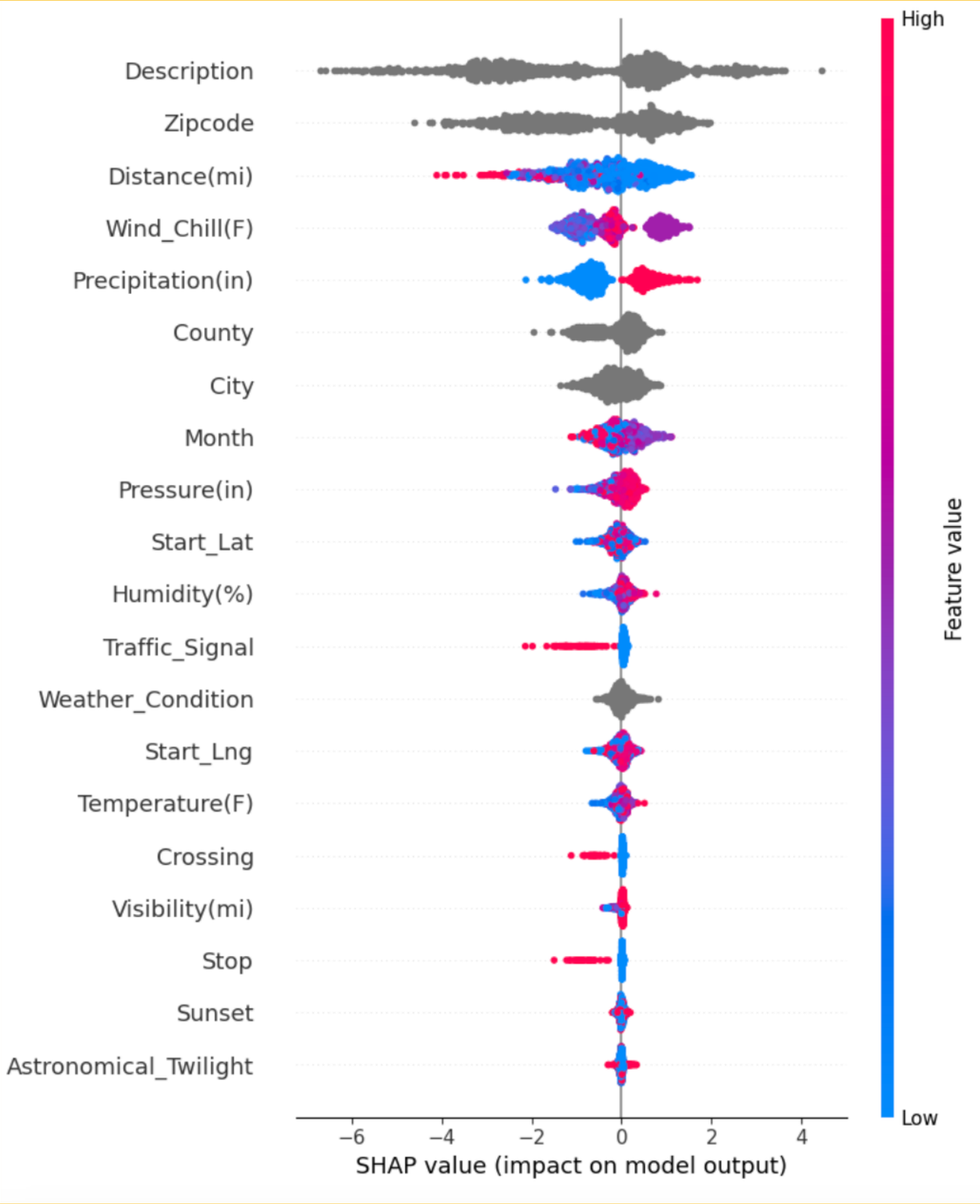
- Logistic Regression
- Random Forest
- **Catboost**

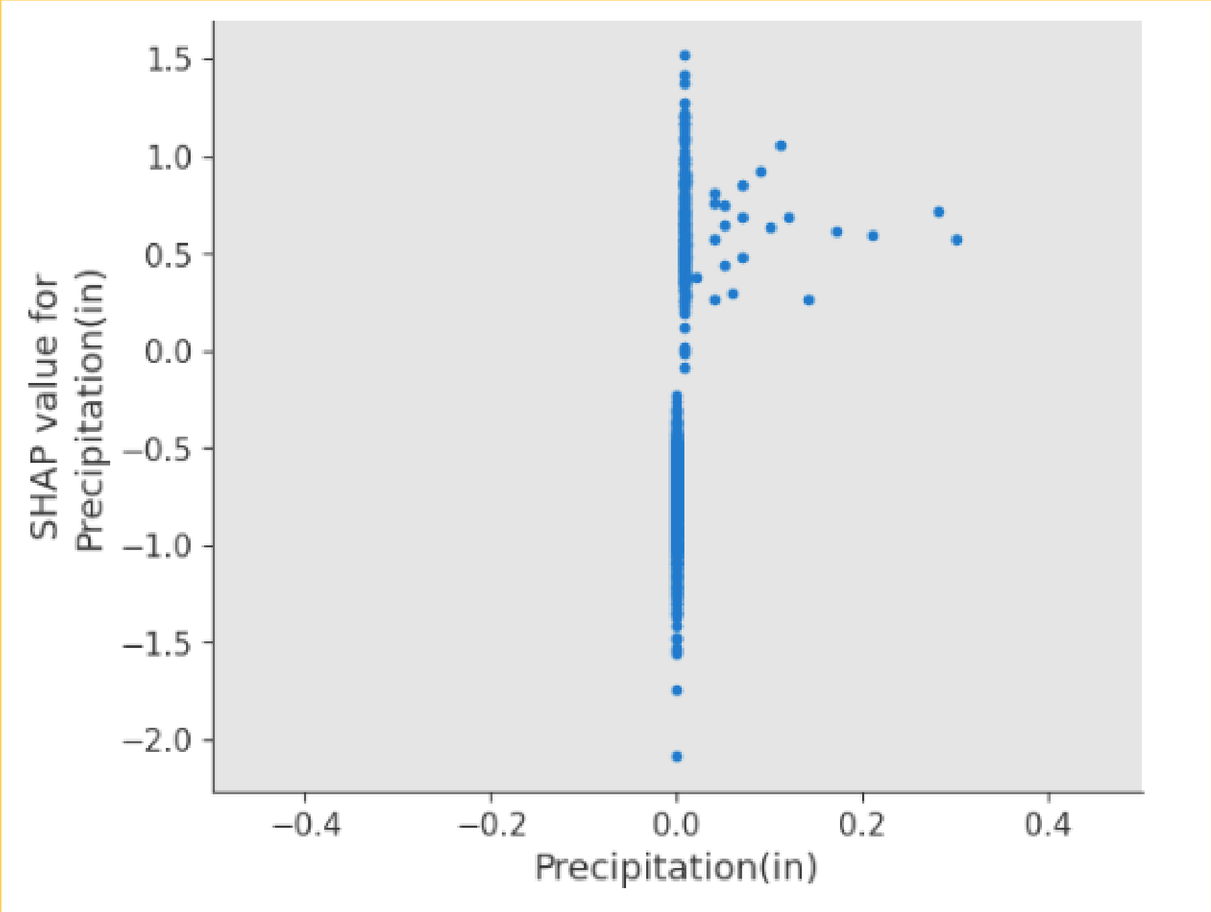


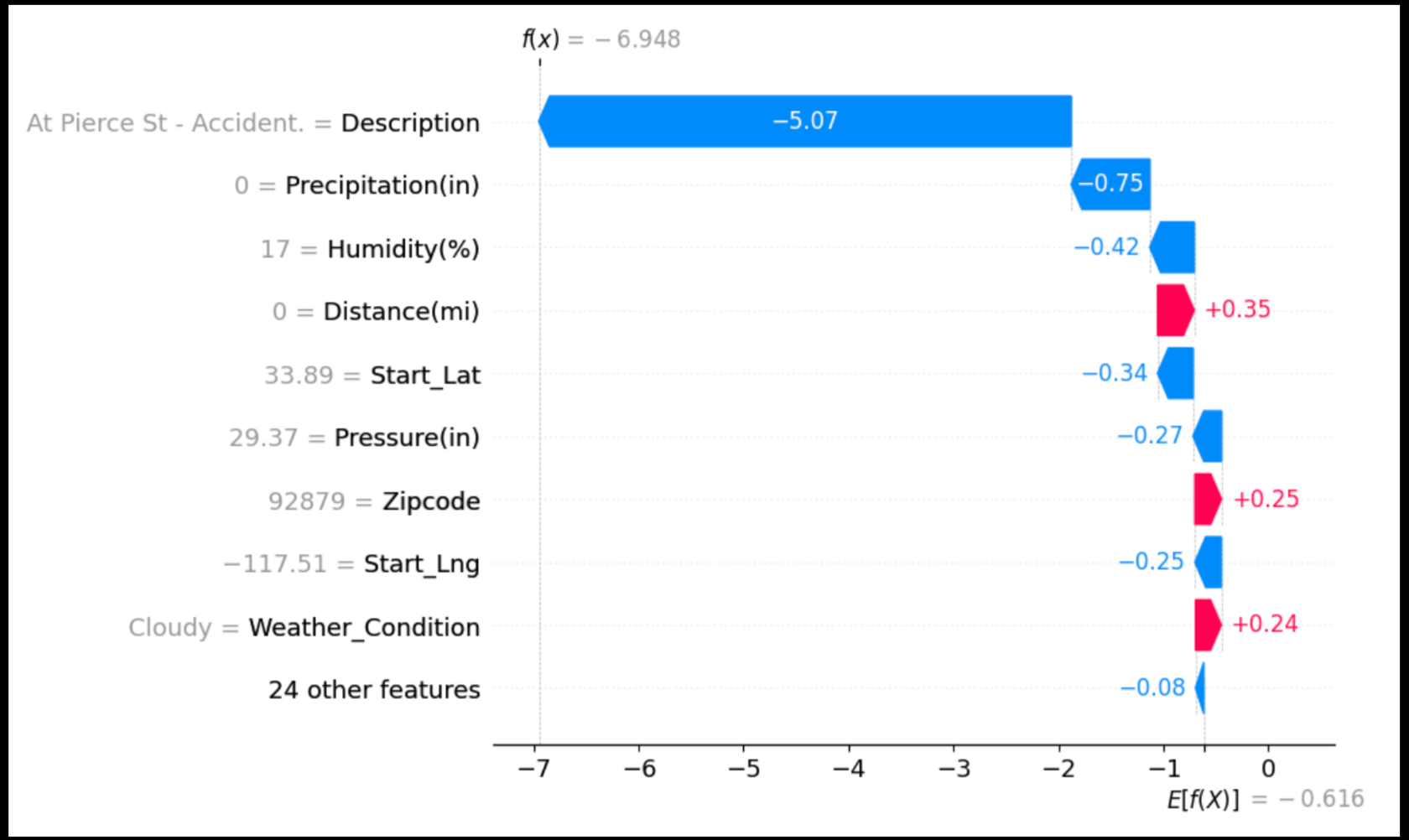
Catboost Confusion Matrix

	precision	recall	f1-score	support
0	0.98	0.94	0.96	355104
1	0.75	0.89	0.82	70109
accuracy			0.93	425213
macro avg	0.86	0.92	0.89	425213
weighted avg	0.94	0.93	0.94	425213

Catboost: Global Explanation of Feature Importance

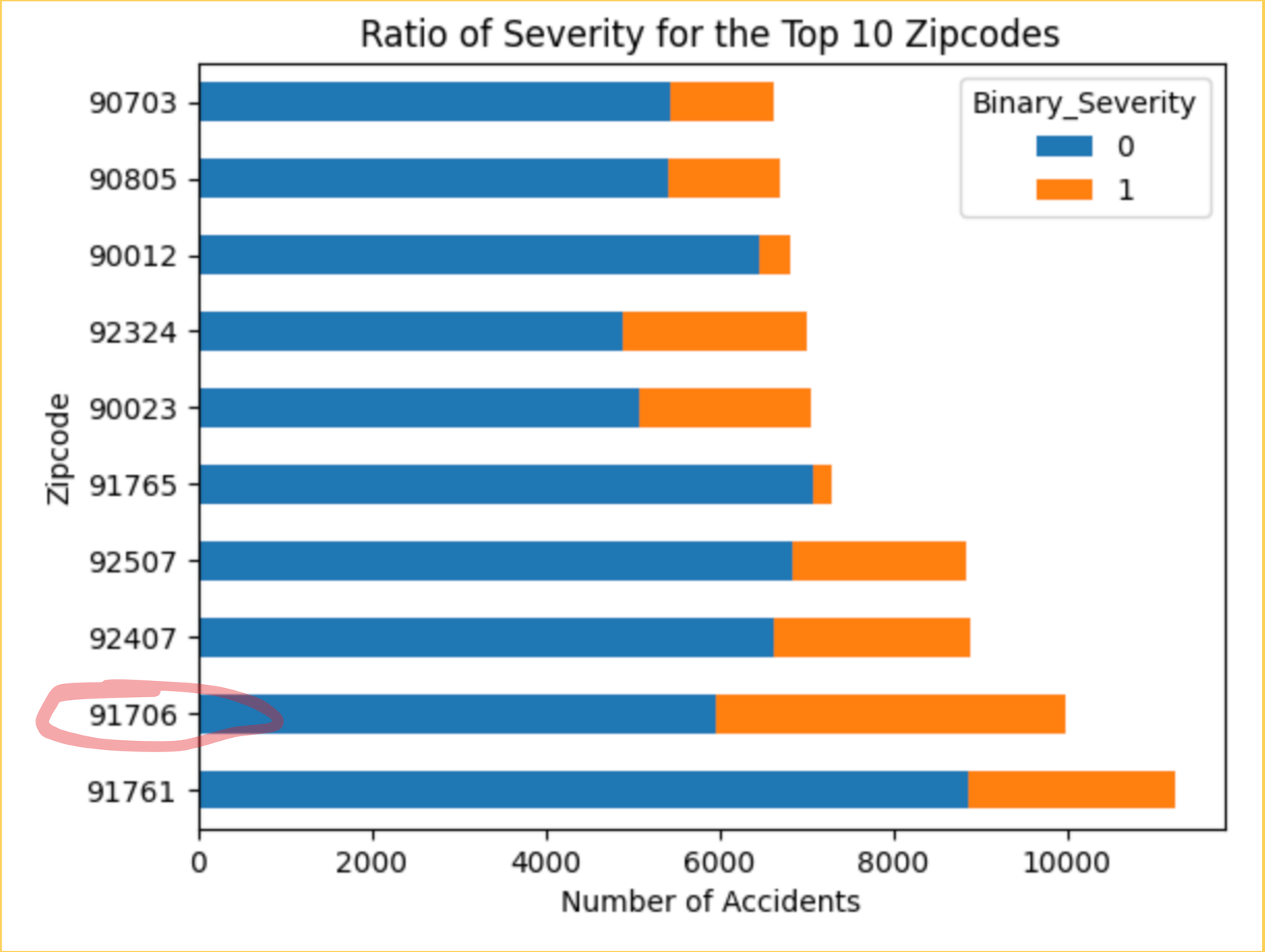


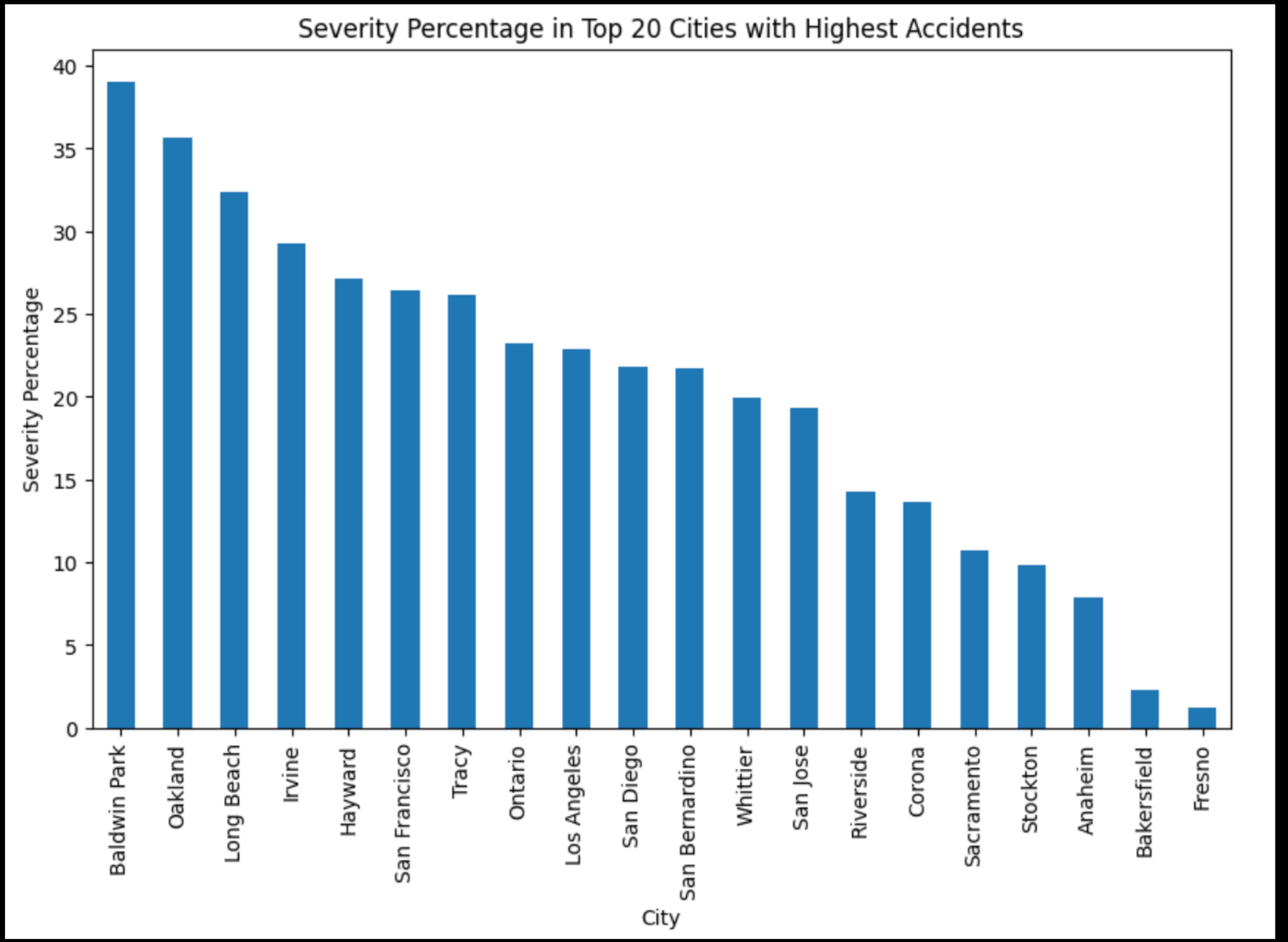




*A waterfall plot of the 888th observation

Catboost: Local Explanation of Feature Importance


Ratio of Severity for the Top 10 Zipcodes
Binary_Severity
0
1
Zipcode
90703
90805
90012
92324
90023
91765
92507
92407
91706
91761
0
2000
4000
6000
8000
10000
Number of Accidents


Severity Percentage in Top 20 Cities with Highest Accidents
Severity Percentage
0
5
10
15
20
25
30
35
40
Baldwin Park
Oakland
Long Beach
Irvine
Hayward
San Francisco
Tracy
Ontario
Los Angeles
San Diego
San Bernardino
Whittier
San Jose
Riverside
Corona
Sacramento
Stockton
Anaheim
Bakersfield
Fresno
City

CONCLUSION

I recommend the DOT begin at the following cities as they have the highest severity percentages:

- Baldwin Park
- Oakland
- Long Beach

FUTURE WORK

- Include more features:
  - Driver’s behaviors: Reckless/Distracted Driving, Under the Influence
  - Vehicle Conditions: Mechanical Error, Model Year
  - Population: are denser areas more prone to accidents?
  - Age: Which age group is involved in the most accidents?
  - Survival: any fatalities?
- Time-Series Analysis on predicting the number of future accidents

Further investigation into these aspects could provide valuable insights for accident prevention measures.

Thank You!

Questions?