Seattle Traffic Accident Severity Prediction

Chi Zhang

CONTENTS


01
Introduction

Traffic accidents have risen to the 3rd main reason for mortality among countries by 2020.

Citizens

Economy Society

The purpose of this project is to analyze the collision dataset for the city of Seattle and determinate the key factors to accident severity.

Transportation Governments

Car Rental or Insurance Companies

Vehicle Manufacturers


02
Data & Methodology

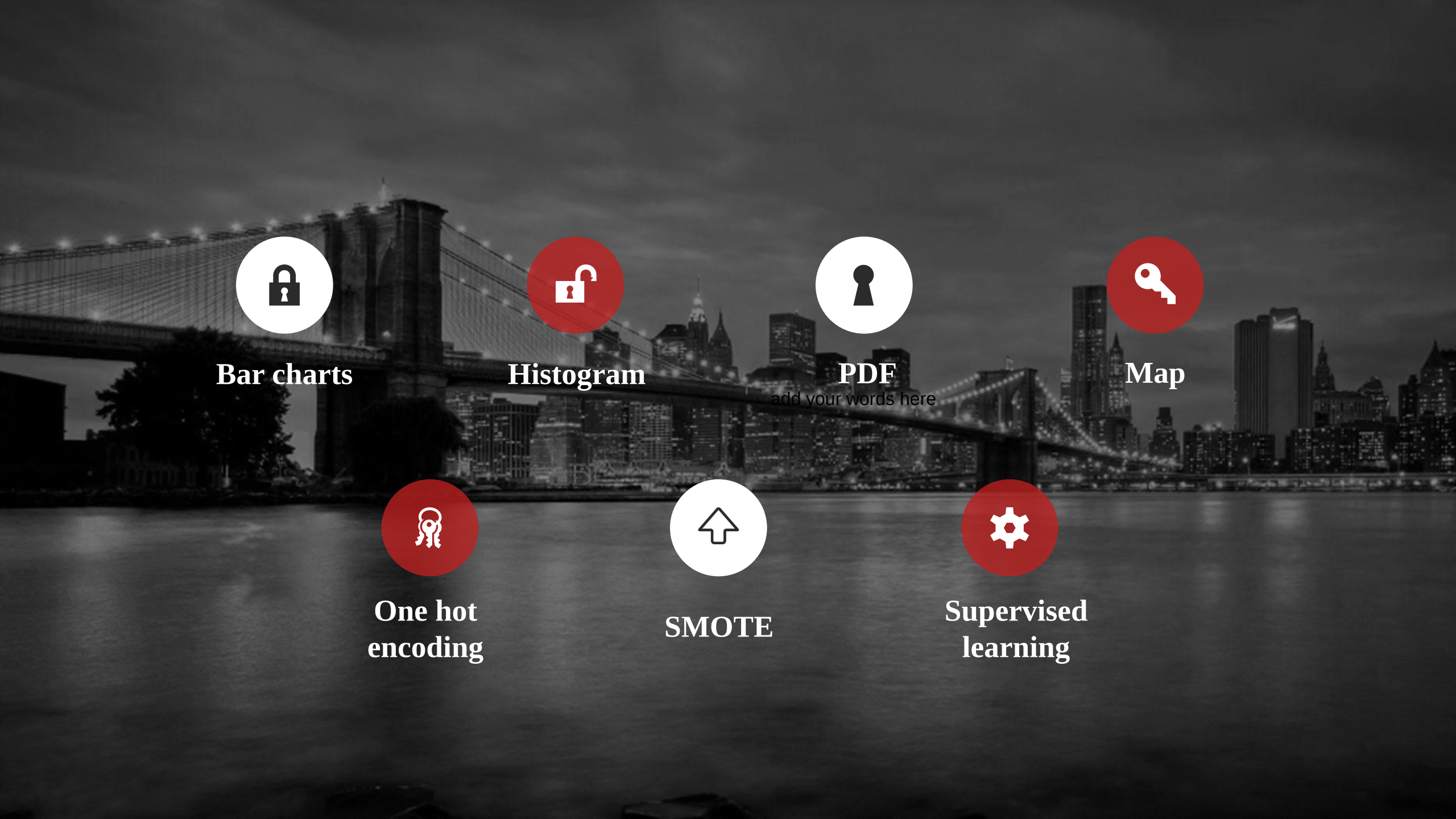
Seattle Department of Transportation provides traffic accident cases from 2004 to 2020.

194673 records

37 indicators

9 categorical indicators

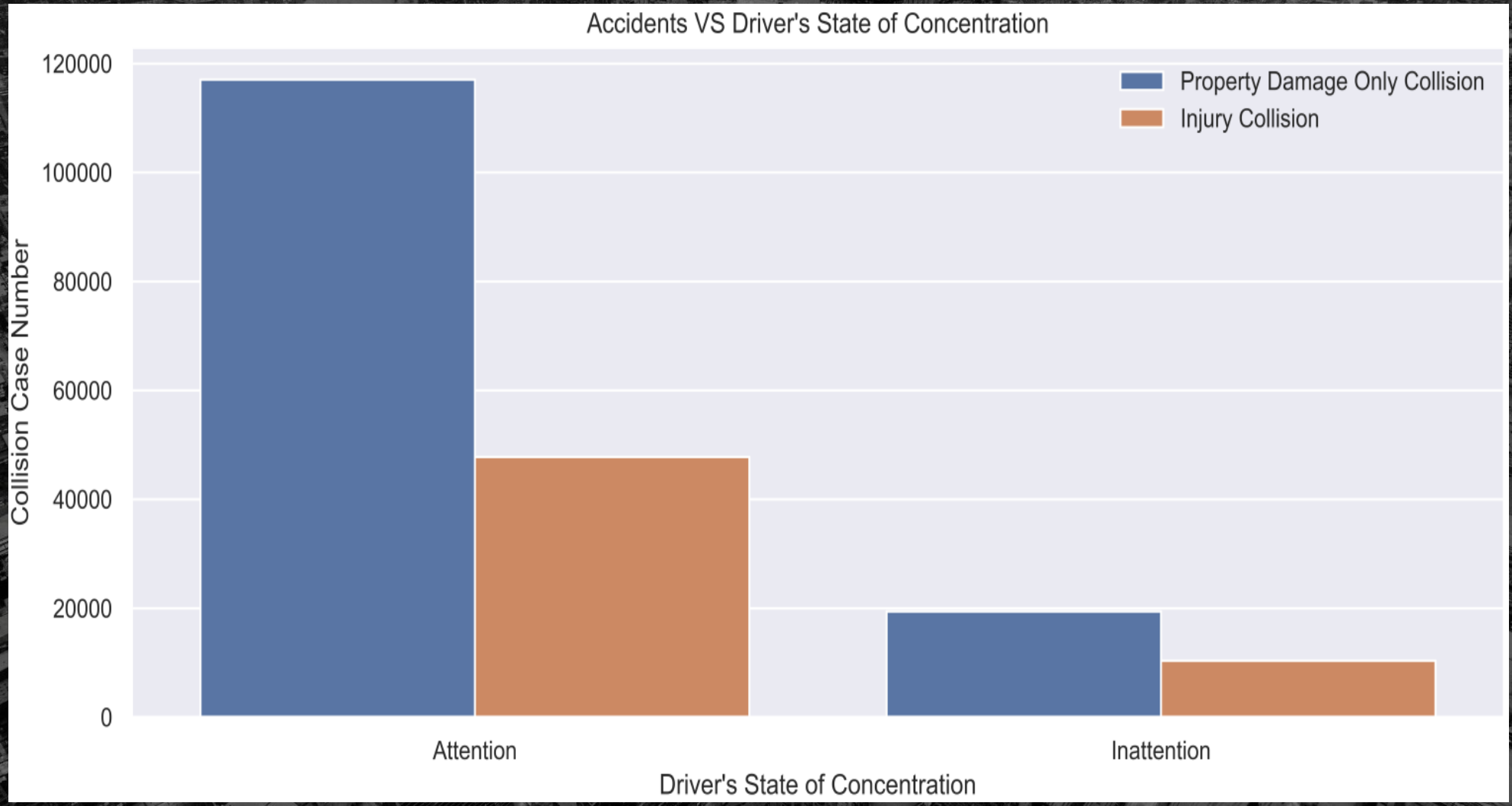
1 output


Bar charts
Histogram
PDF
add your words here
Map
One hot encoding
SMOTE
Supervised learning

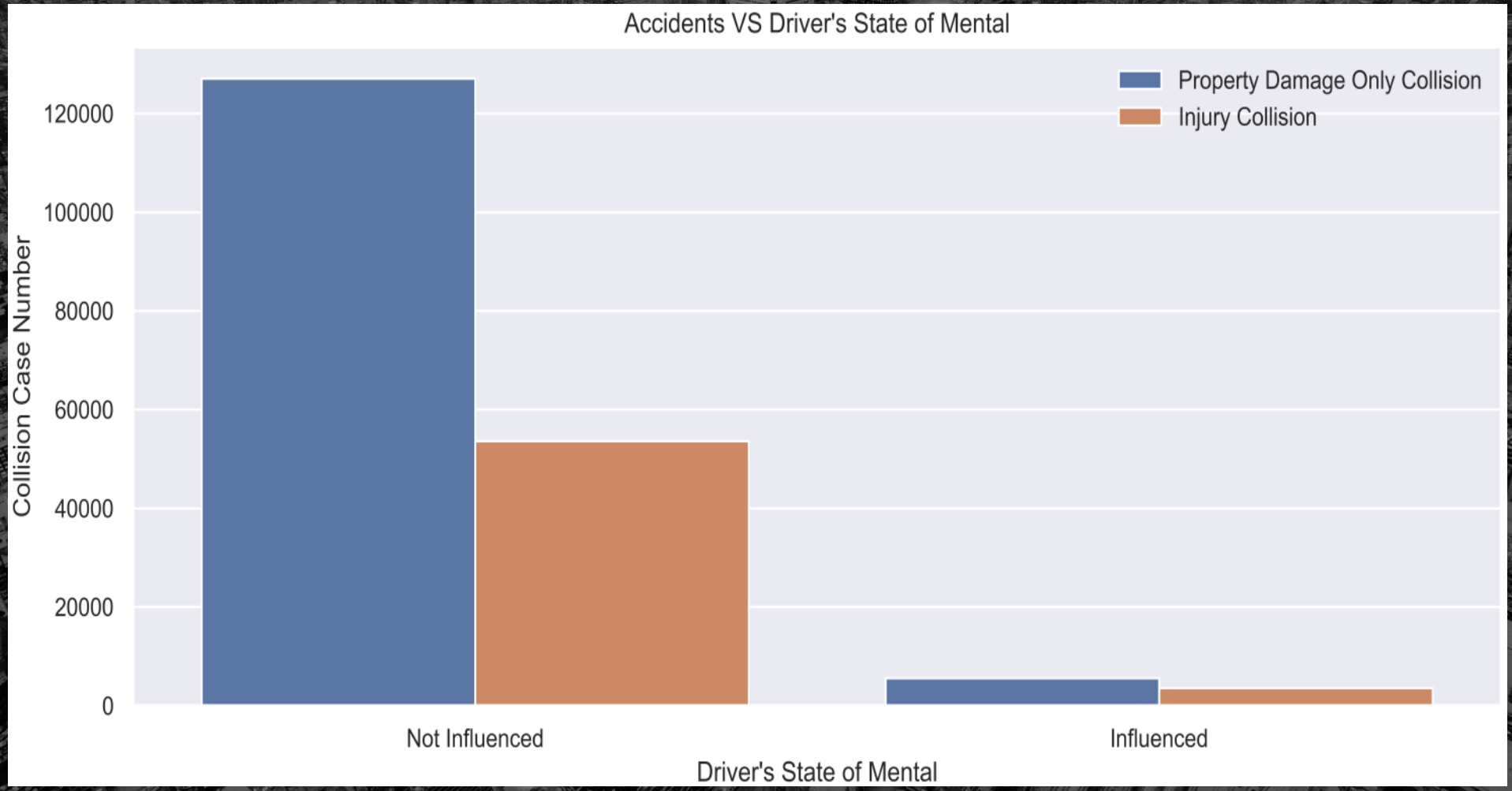
03

Results & Discussion

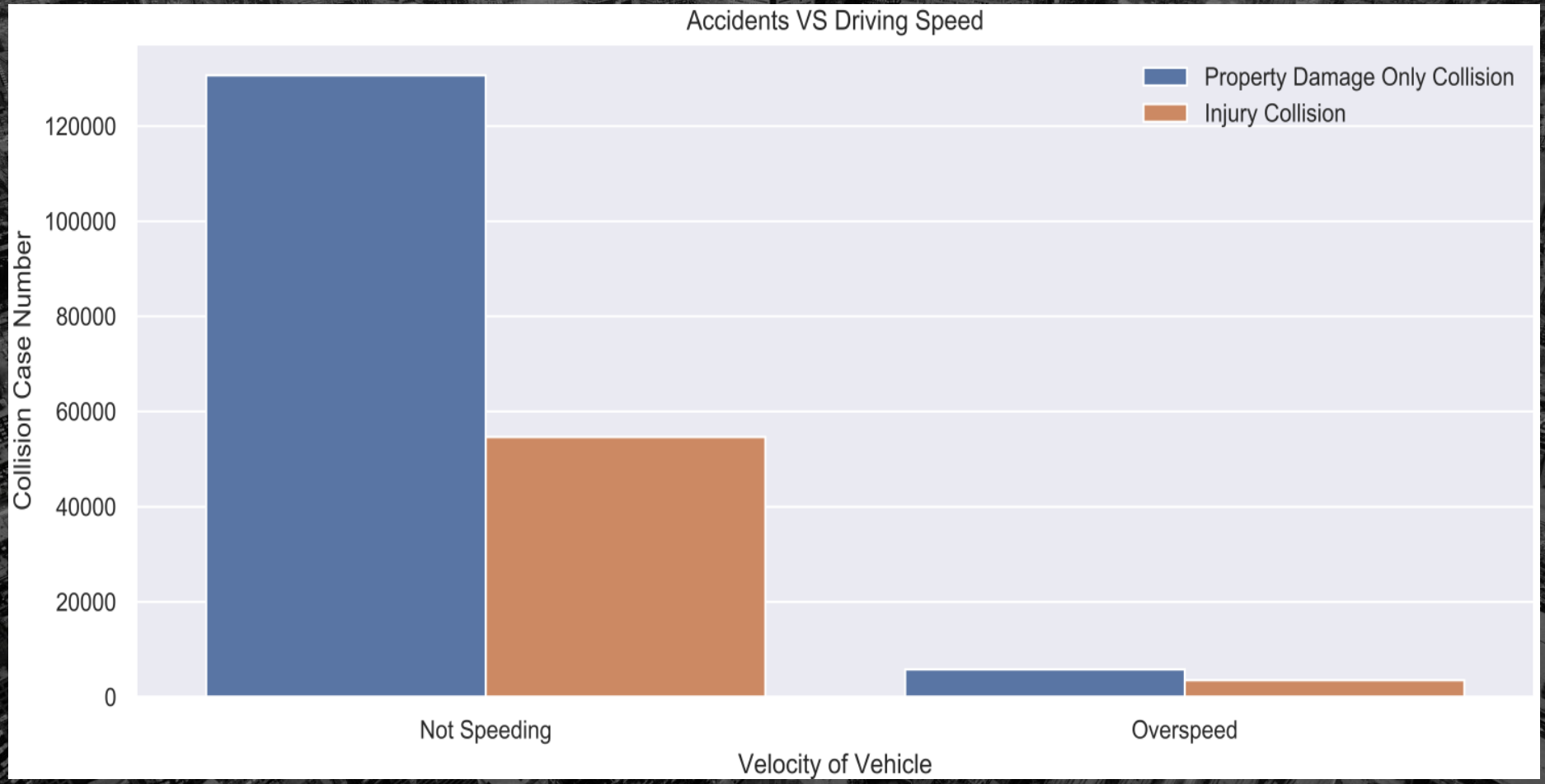
Inattention drivers were 5.9 percentile much likely to come across serious collisions than ordinary ones .



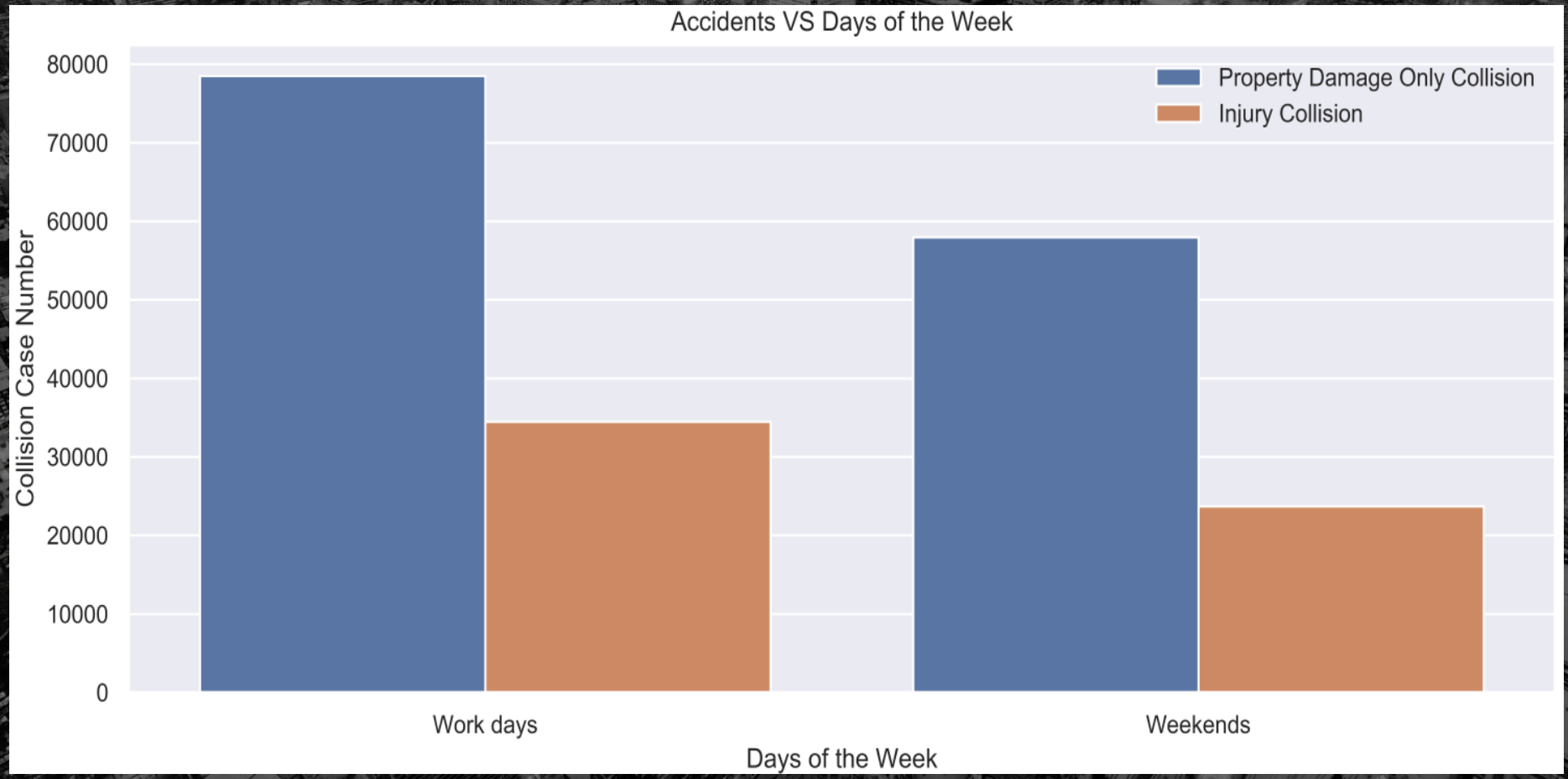
The possibility of serious car accidents caused by drunk or drug drivers is 9.4 percentage points higher than self-disciplined people generally.



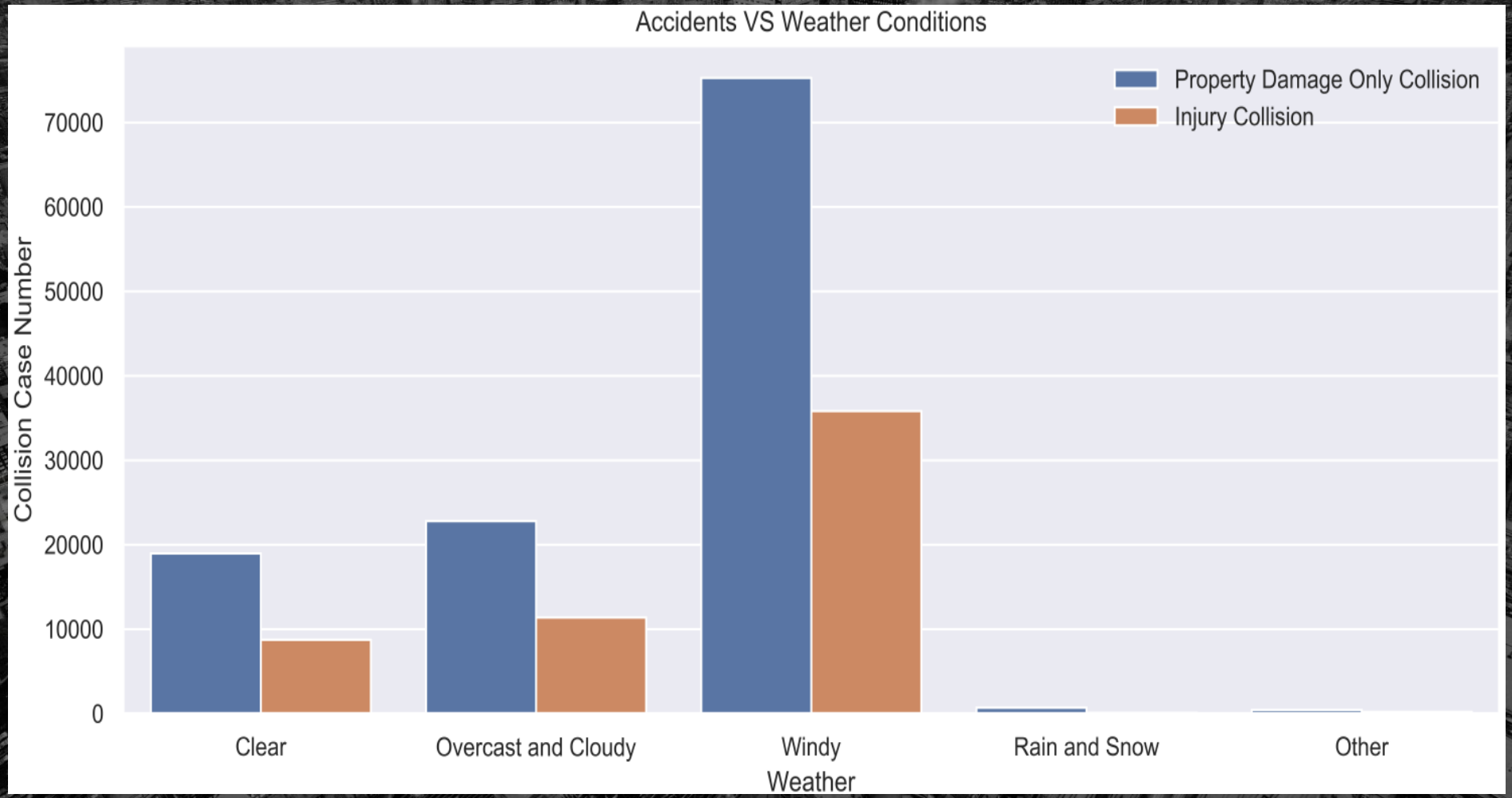
Overspeeding also led to an 8.3 percentage point higher proportion of harmful collisions compared to traveling at normal speeds.



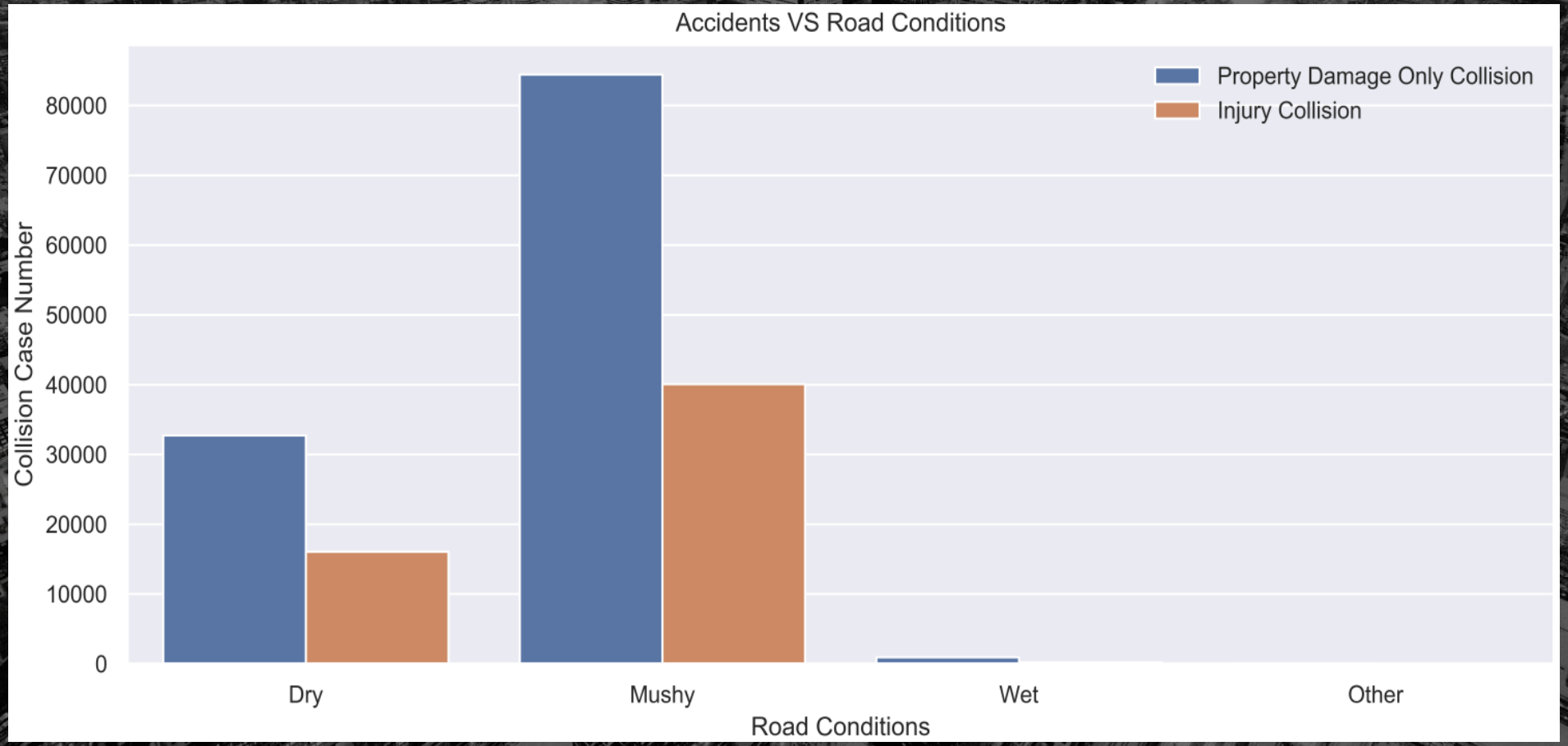
Whether the day was among weekends, from Friday to Sunday, had nuances to the severity of crashes, but they were nearly the same at the 30% level.



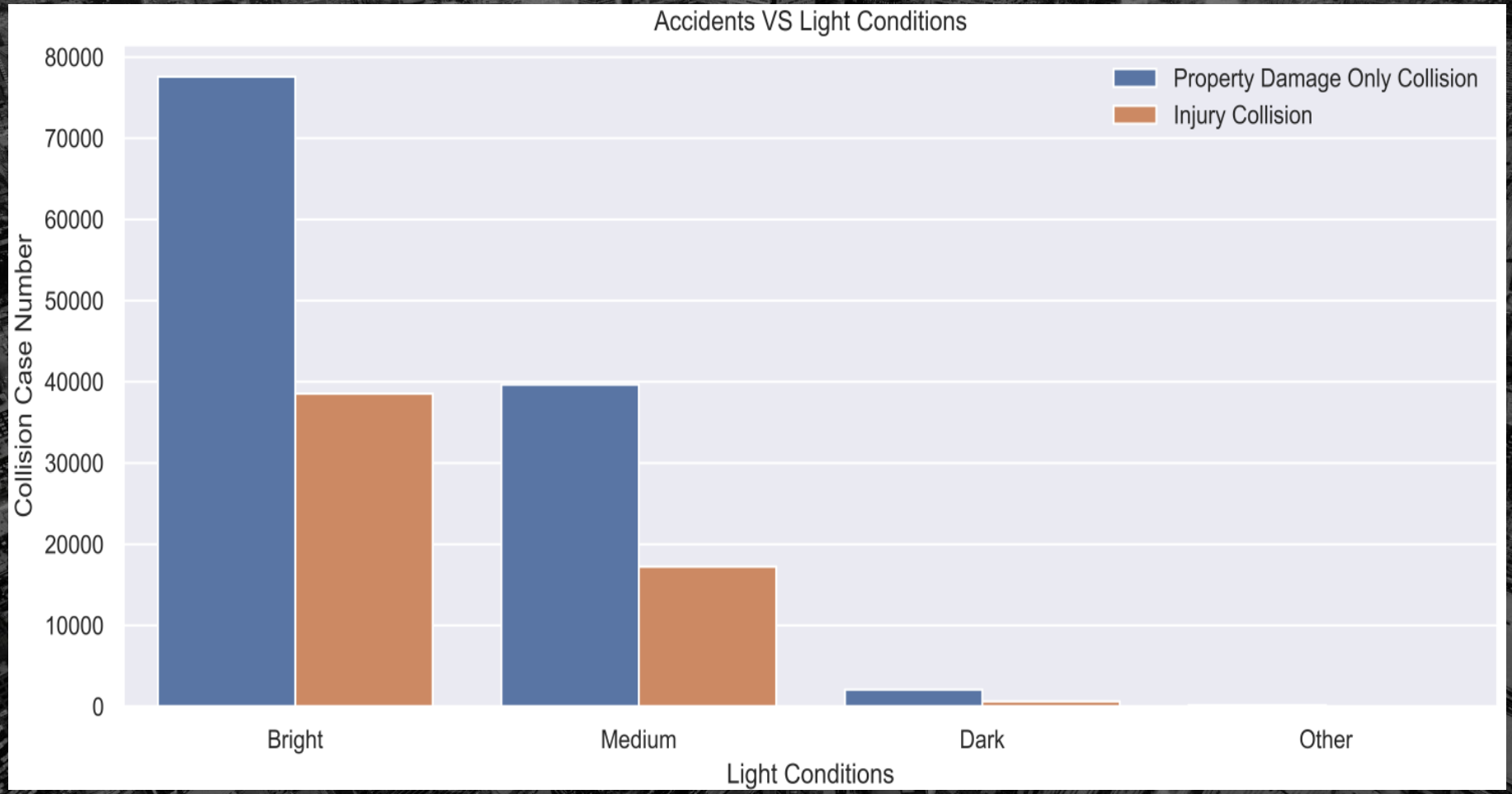
The weather types have dramatically different influences on the severity of car accidents.



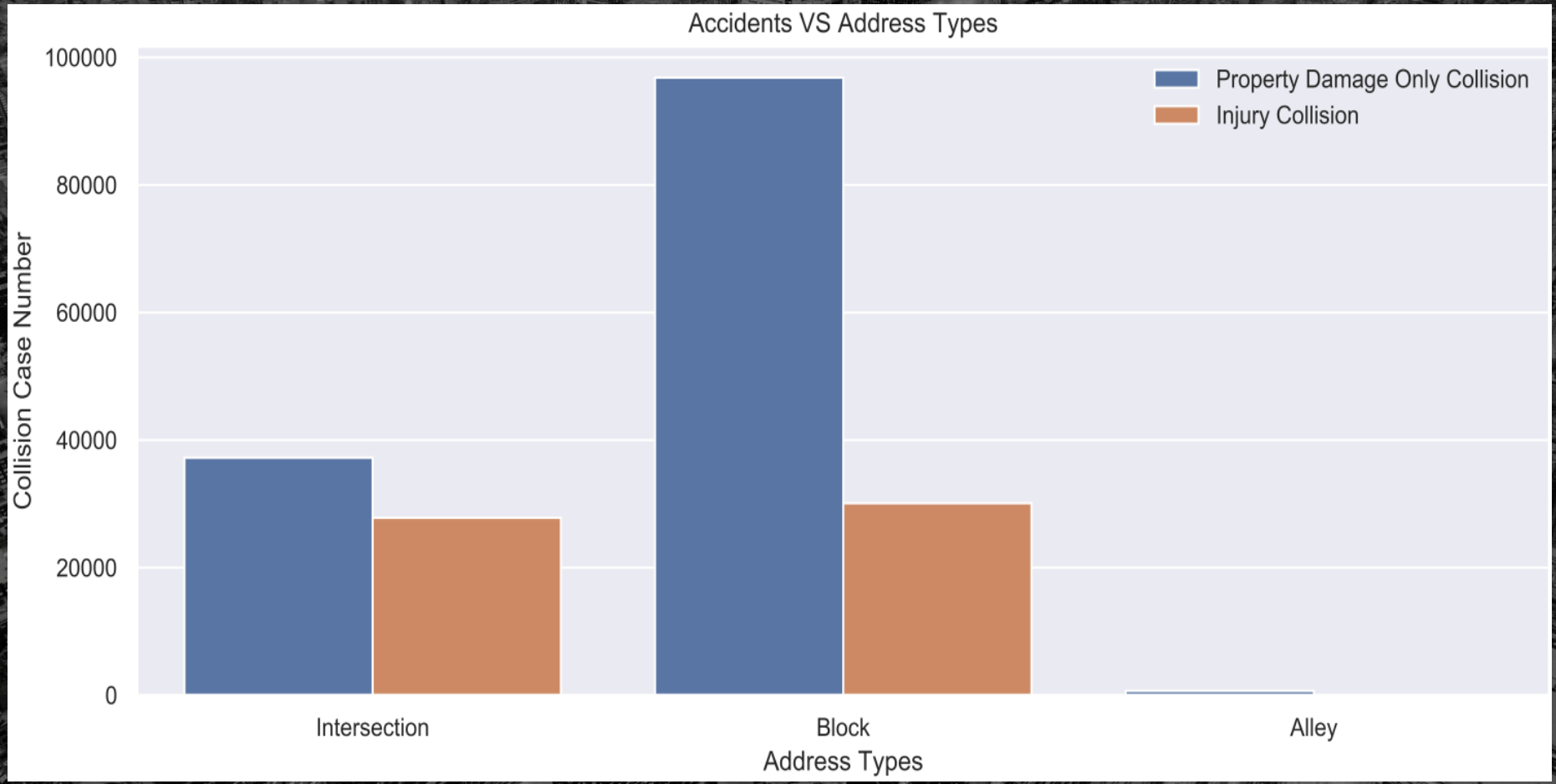
The serious accident rate was less than 20% under mushy road condition, which was 14.2 percentile below other kinds of roads.



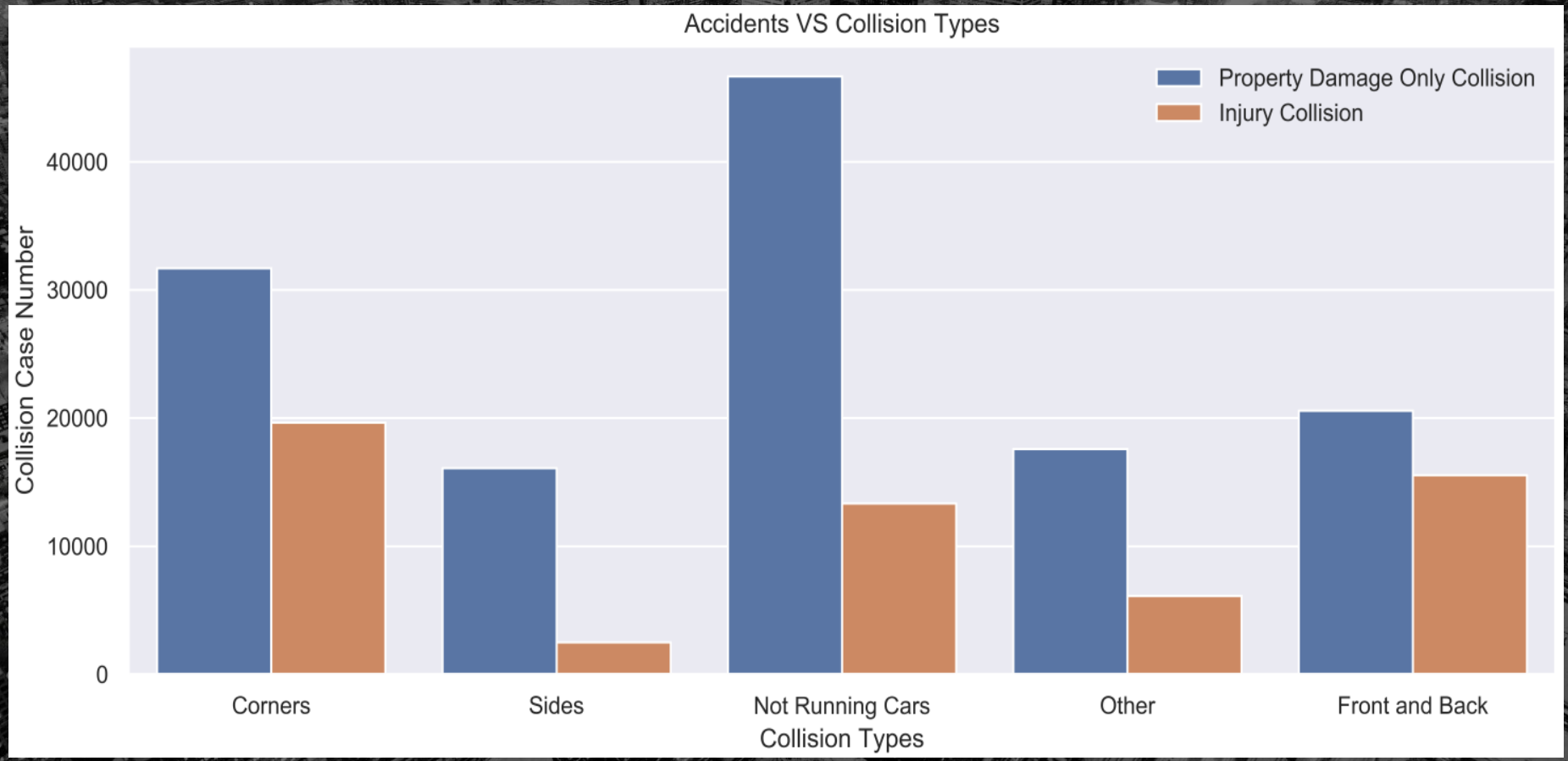
The serious accident rates among light conditions sound beyond expectation .



Three address types had a remarkable influence on crash severity.



Face to face collisions led to a high proportion of injury accidents at 43.0% .

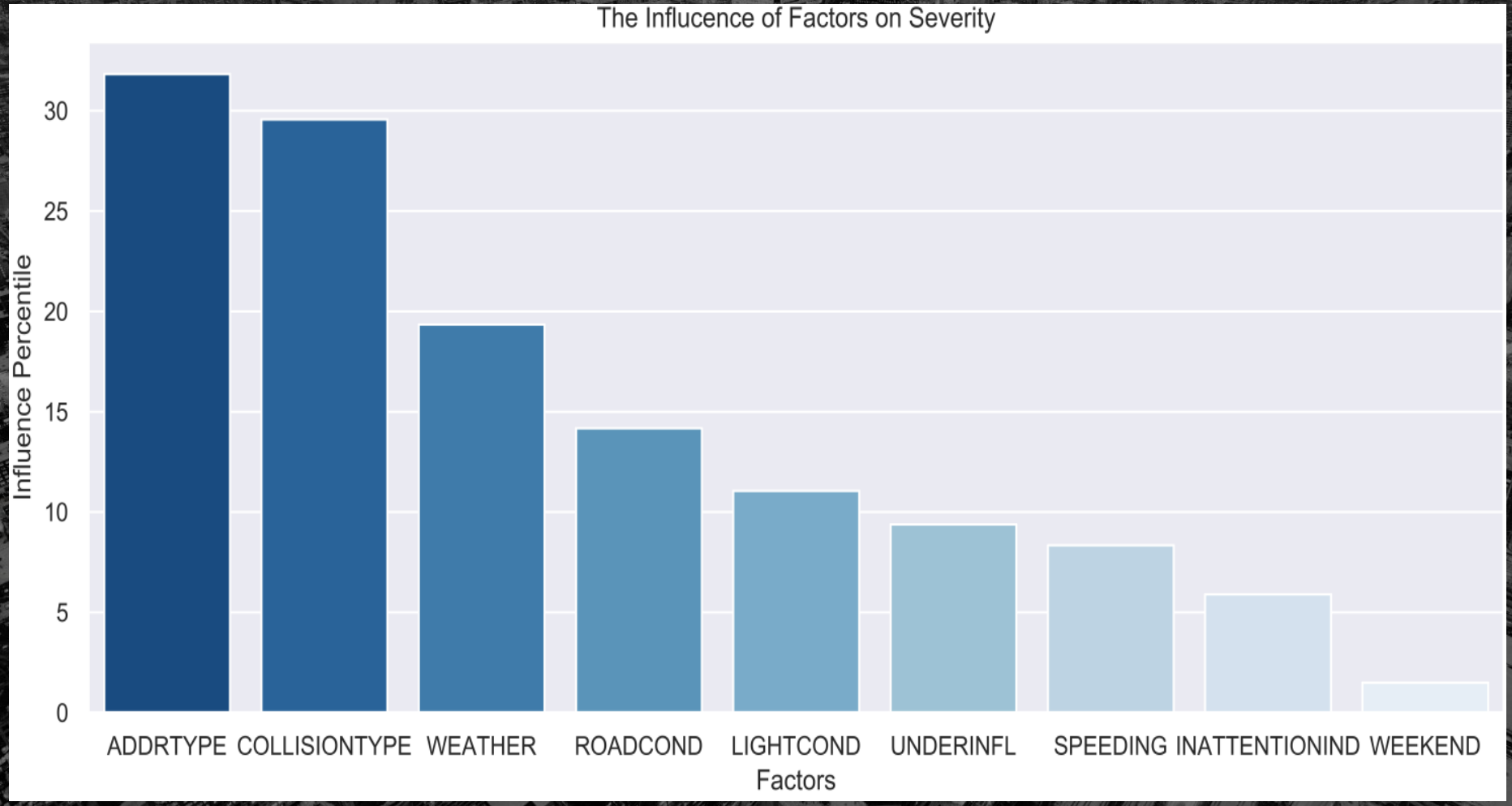


Natural factors surpassed human factors on the impacts of car accidents.

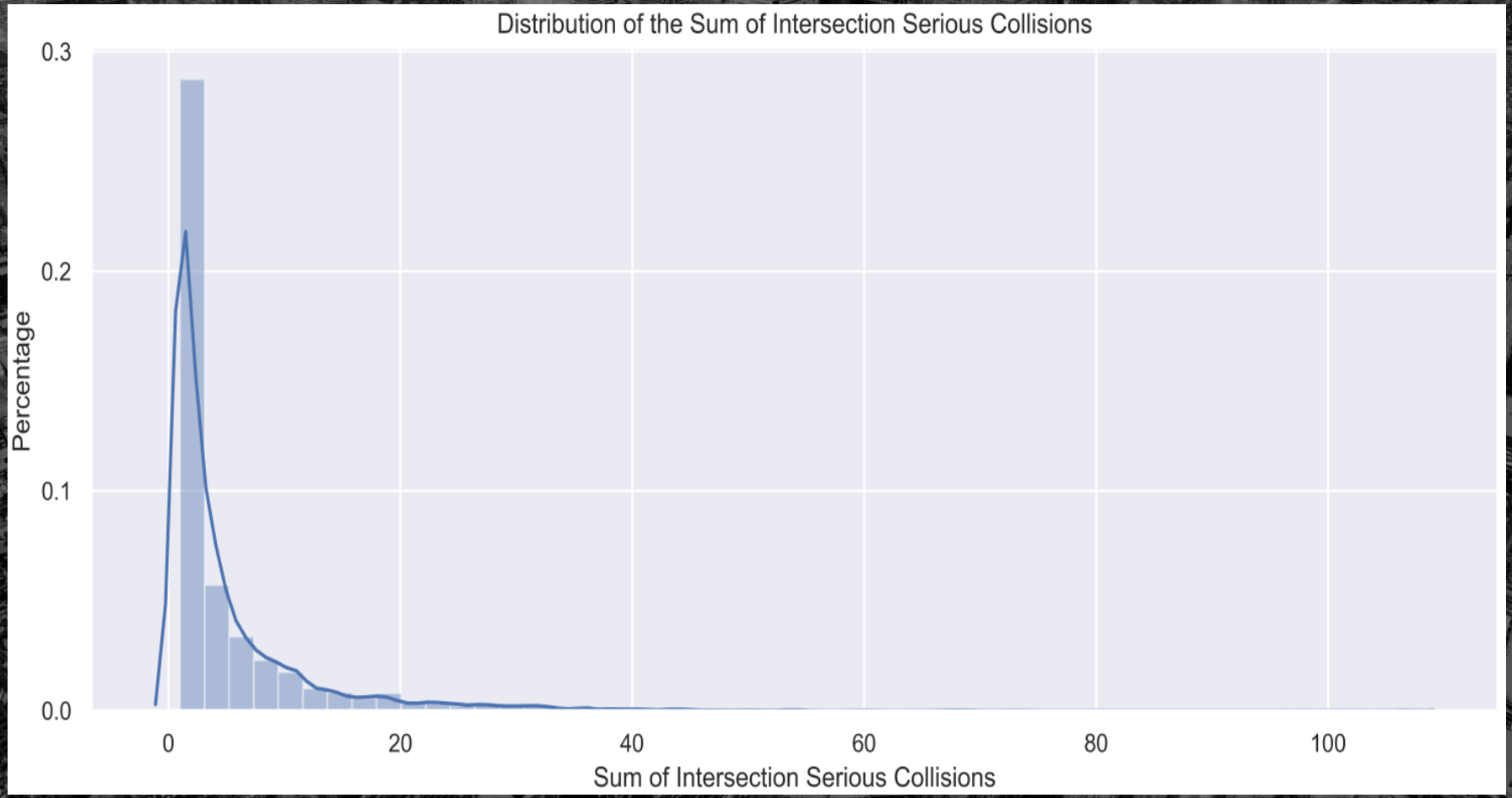
The address was the most important reason among 9 features.

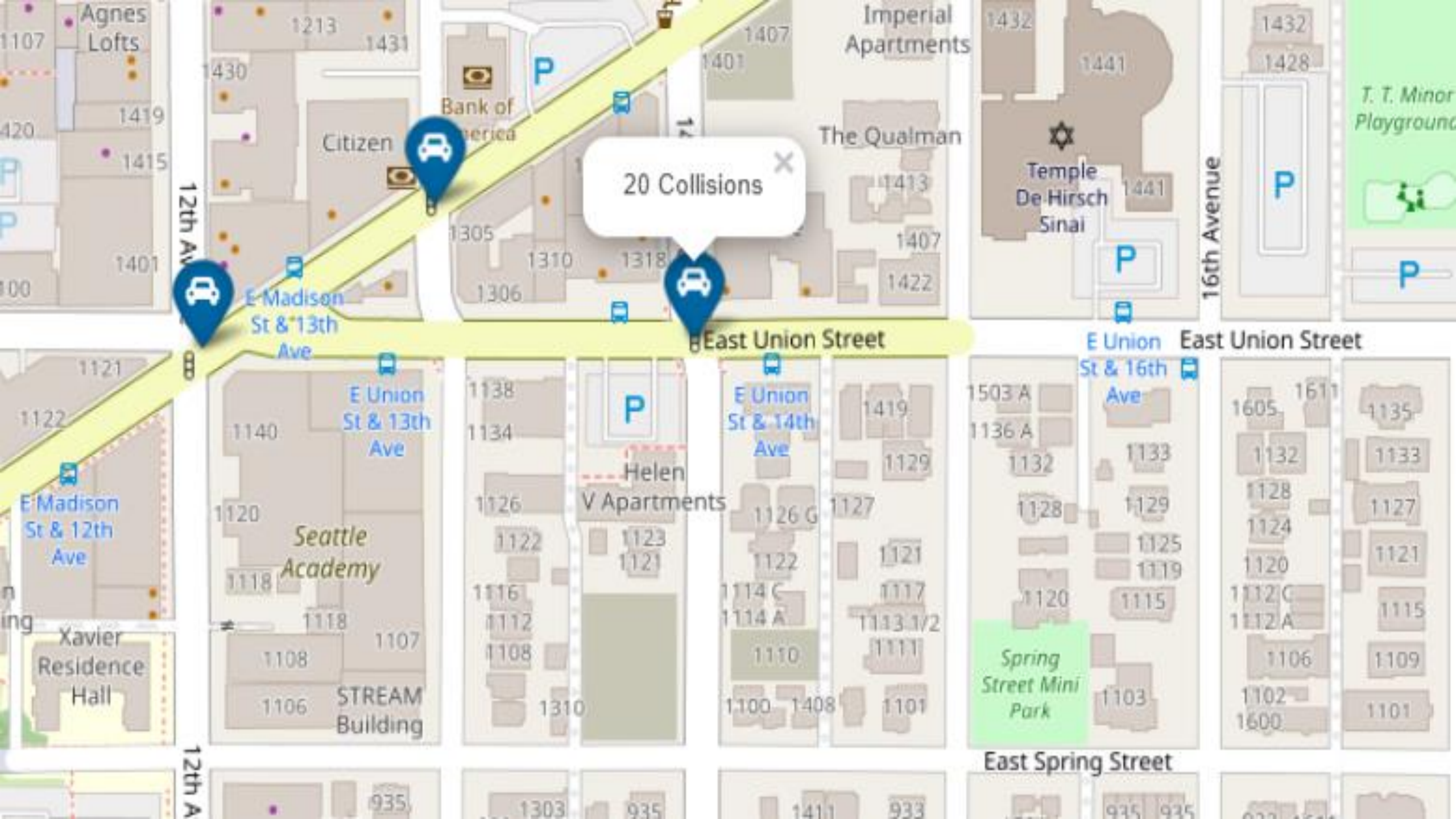
All human factors had an influence on the severity of less than 10 percentile fluctuation.

Bad circumstances properly contribute to casualties in road accidents.



26.7% of serious intersection collisions happened at where the frequency of serious accidents' occurrence was more than 20 times.



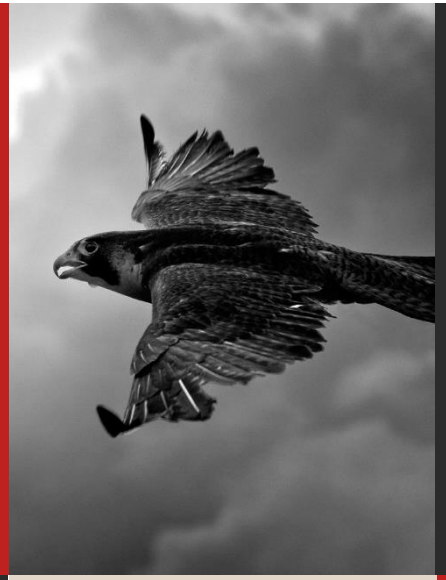

20 Collisions
East Union Street
E Madison St & 13th Ave
E Union St & 13th Ave
E Union St & 14th Ave
E Union St & 16th Ave
E Madison St & 12th Ave
12th Ave
16th Avenue
East Spring Street
Agnes Lofts
Citizen
Bank of America
Imperial Apartments
The Qualman
Temple De Hirsch Sinai
T. T. Minor Playground
Helen V Apartments
Seattle Academy
STREAM Building
Xavier Residence Hall
Spring Street Mini Park

Model Name	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
KNN	67.00	50.00	21.00	30.00
Decision Tree	62.89	45.53	70.32	55.28
SVM	63.00	45.68	71.18	55.65
Logistic Regression	62.35	45.14	71.80	55.44
Random Forest	62.57	45.48	70.73	55.36

I choose the Random Forest model (n=100) finally because it had the least processing time.

The models could use more improvements to reduce false positive predictions.

Adjust the hyper-parameter of existed models or apply new models



Combine some numerical variables into predictive models

04

Conclusion

Governments should update facilities to warn and protect drivers and walkers.

Car rental and insurance companies can customize service.

Vehicle manufacturers should organize some research about improving the mechanical structure of cars.

Thank you!