FLATIRON SCHOOL

COLLISION STATS

CALIFORNIA 2016 - 2020

2021

ERIC AN

AGENDA

Our content today is
divided into four parts.
Each part will be described in detail.

01 | Background

The purpose of this data analysis.

02 | Methodology

How the data was examined.

03 | Analysis

Presentation of findings.

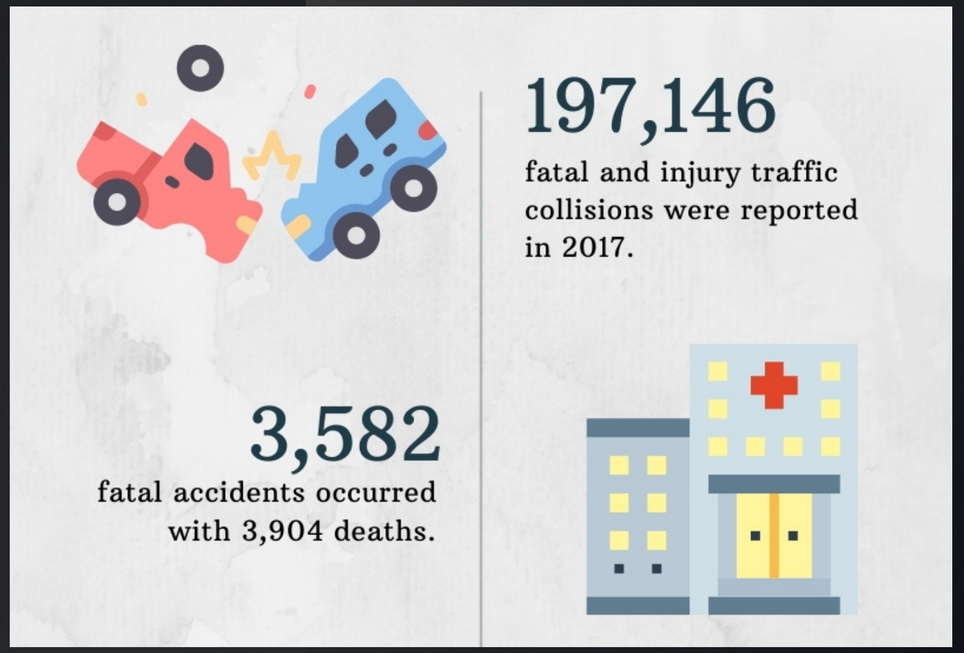
04 | Summary

Future work and other considerations.

COLLISION STATS

Why does it matter?

- A better understanding of how collisions occur
- Physical toll of lives lost, lives ruined, and lives interrupted
- Emotional toll of losing loved ones
- Secondary costs of property damage, traffic jams, increased enforcement



Source: https://westcoasttriallawyers.com/ca/los-angeles-car-accidents-lawyer/california-car-accident-statistics/

METHODOLOGY

examine the automobile traffic collision history from years 2016 - 2020 in California

Utilize visualization techniques to analyze collision data and use machine learning models to predict collision severity.

Data Breakdown

Broken down by:
Time-related
Conditions-related
Party Stats
Victim Stats
Collision Nature
Alcohol-Related
Pedestrian/Bicycle/Motorcycle

Visualizations

heatmaps, countplots, boxplots, pareto diagram, stripplots, violinplots, and more

Prediction Modeling

logistic regression, knn, random forests, and other algorithms using a gridsearch pipeline

Time Statistics

1. Collision Severity By Year *(heatmap)*
2. Collisions By Month *(barplot - shaded by magnitude)*
3. Collisions Severity By Day of Week *(countplot)*
4. Type of Collision and Number of People Injured/Killed By Day of Week *(boxplot)*
5. Violation Type By Day of Week *(countplot)*
6. Collisions by Time of Day *(barplot/lineplot - pareto diagram)*
7. Collision Severity Based on Time Window *(countplot)*
8. Type of Vehicle Collision Based on Time Window *(countplot)*
9. Other Party Vechicle Type Based on Time Window *(countplot)*
10. Collisions Based on Holidays and Time Window *(countplot)*
11. Hit and Run by Time Window *(countplot)*
12. Similarly, Hit and Run and Collision Severity *(displot: histogram)*
13. PCF Violation Category by Time Window *(barplot - stacked percentage)*

Conditions Statistics

14. Collision Severity Based on Lighting *(countplot)*
15. Violation Category Based on Lighting *(heatmap)*
16. Collision Severity Based on Weather *(countplot)*
17. PCF Violation Category Based on Weather *(barplot - stacked percentage/heatmap)*
18. Collision Severity Based on Road Conditions *(jointplot - histogram/countplot)*
19. Collision Severity Based on Intersection/State Hwy *(countplot/scatterplot)*

Party Statistics

20. Party By Gender and Age and Race *(stripplot)*
21. Gender/Age/Race of Parties at Fault *(swarmplot & violinplot)*
22. Vehicle Type and Pre-Collision Movement By Age for Parties at Fault *(histogram)*
23. Violation Category By Age for Parties at Fault *(boxenplot)*

Victim Statistics

24. Victim Role By Age *(scatterplot)*
25. Victim Injury Level Based on Insufficient Safety Equipment *(FacetGrid: countplot)*

Collision Statistics

26. Injuries or Death Based on Violation Category *(pointplot)*
27. Type of Collision and Severity Level For Collisions with Injuries or Death *(countplot)*
28. Type of Collision and Violation Category/What the Vehicle was Involved With *(countplot)*
29. Vehicle Towed Away Based on Collision Severity *(heatmap)*
30. Pre-Collision Direction and Collision Severity *(stripplot & pointplot & violinplot)*

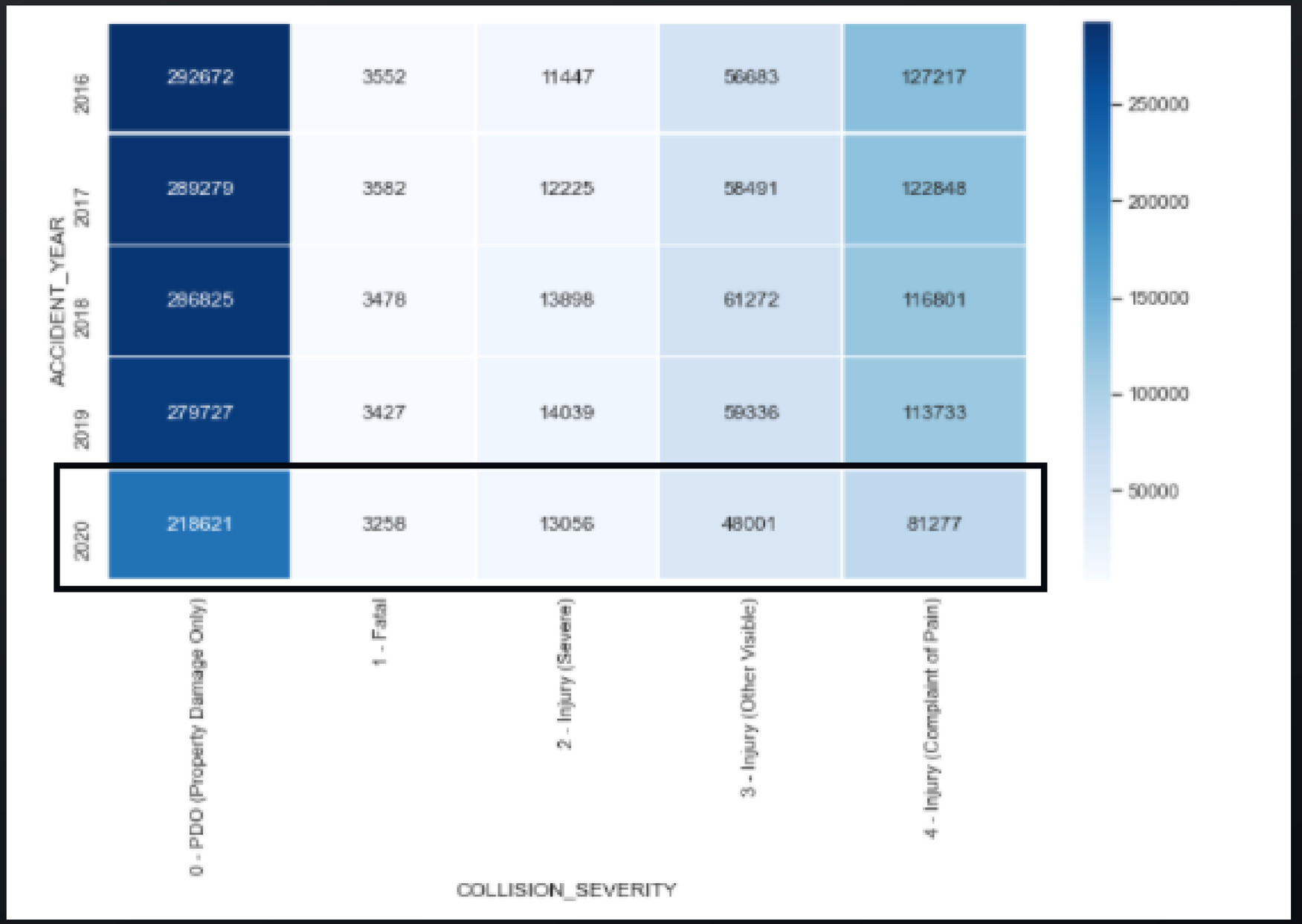
Alcohol-Involved Statistics

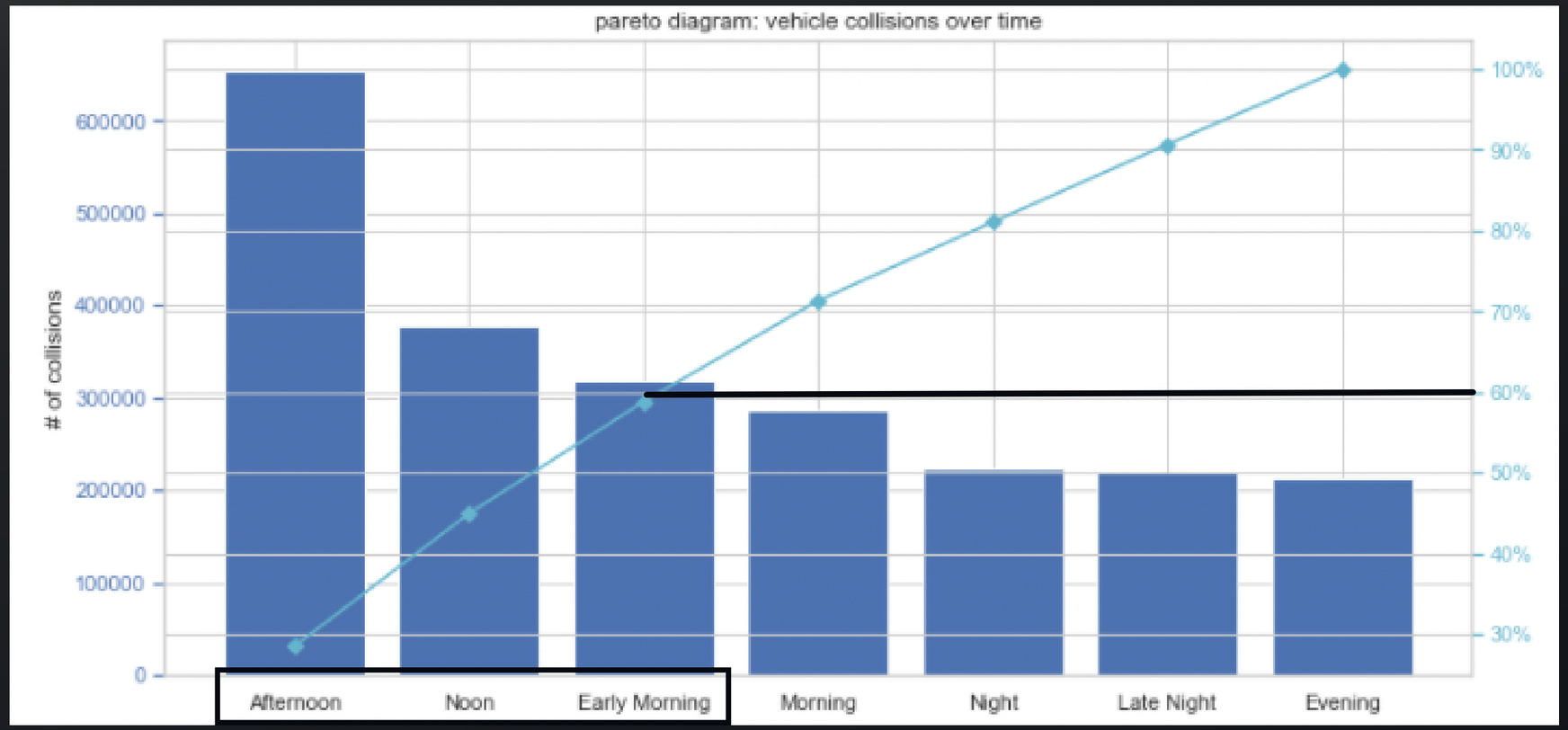
31. Age/Gender and Sobriety of Party at Fault *(swarmplot)*
32. Alcohol-Involved Collisions Over Time *(overlapping barplots)*
33. Alcohol-Involved Collisions by Vehicle Types and Injuries/Death *(countplot)*
34. Alcohol-Involved Collisions by Date *(barplot)*
35. Alcohol-Involved Collisions by County (Top 20) *(barplot)*

Pedestrian/Bicycle/Motorcycle Statistics

36. Total Injuries and Deaths By Time of Day *(barplot)*
37. Non-Vehicular Injuries/Deaths and Roads Where It Occured *(no plot)*

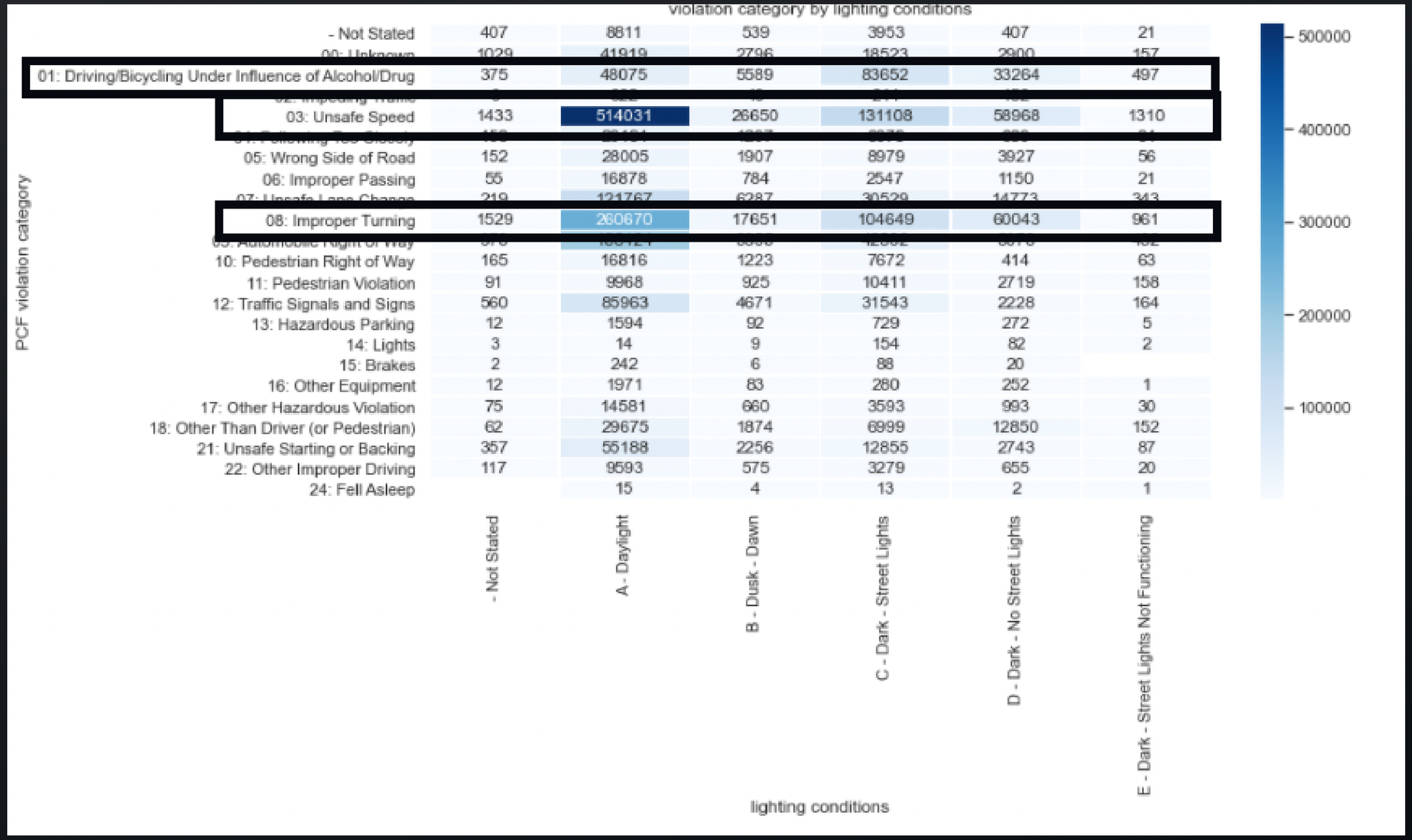
COLLISION SEVERITY BY YEAR

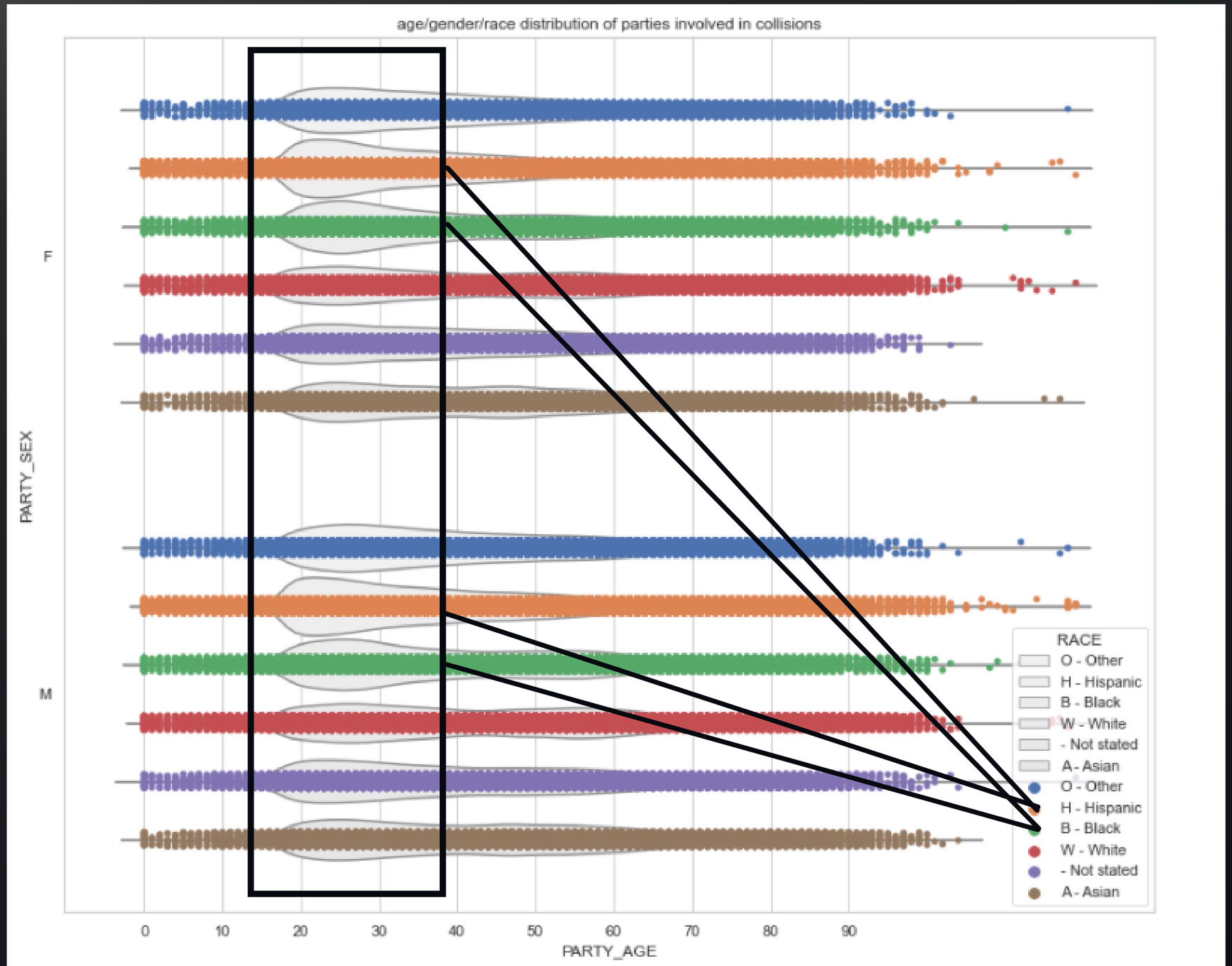




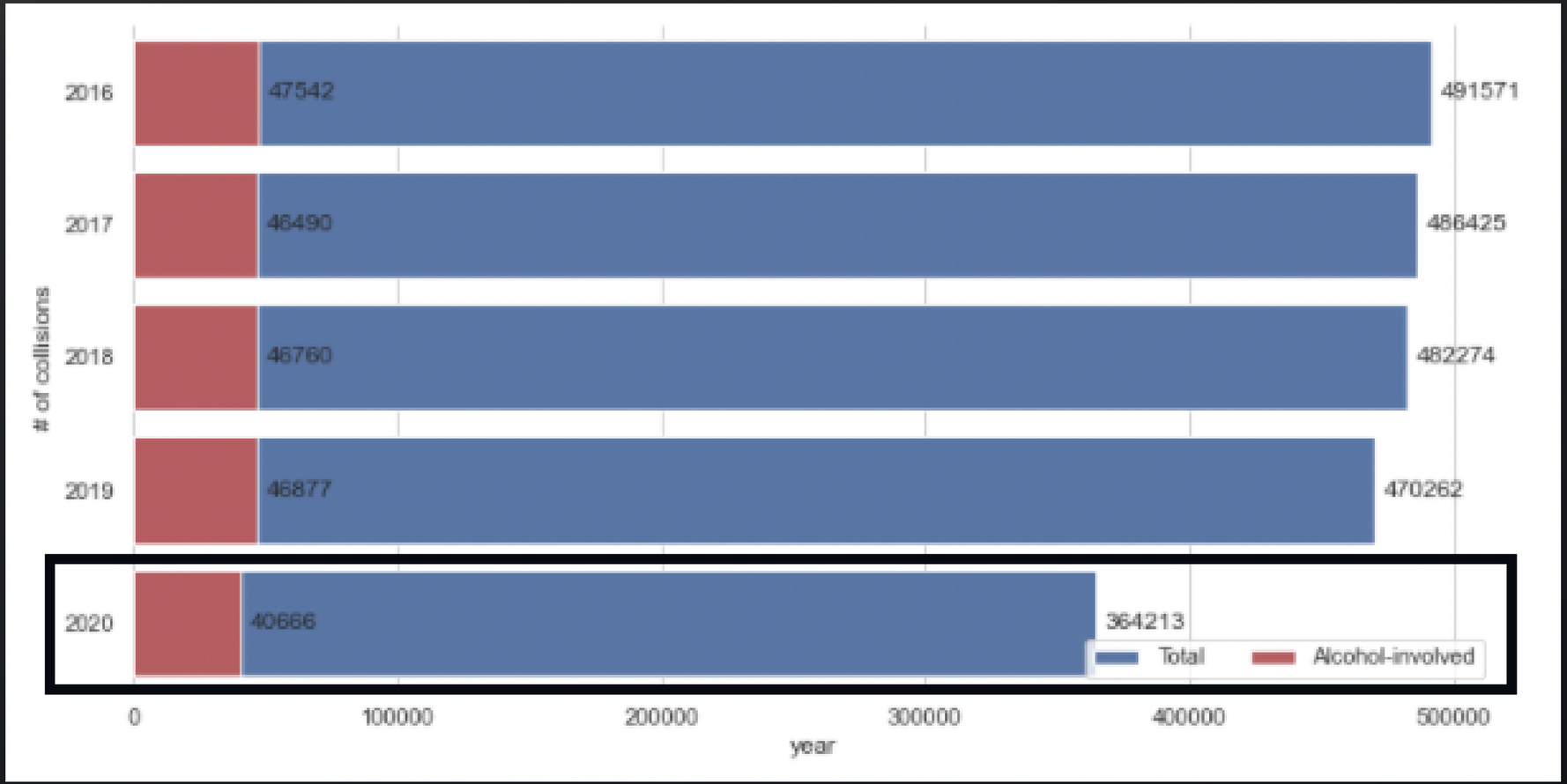
VEHICLE COLLISIONS OVER TIME

VIOLATION CATEGORY BASED ON LIGHTING



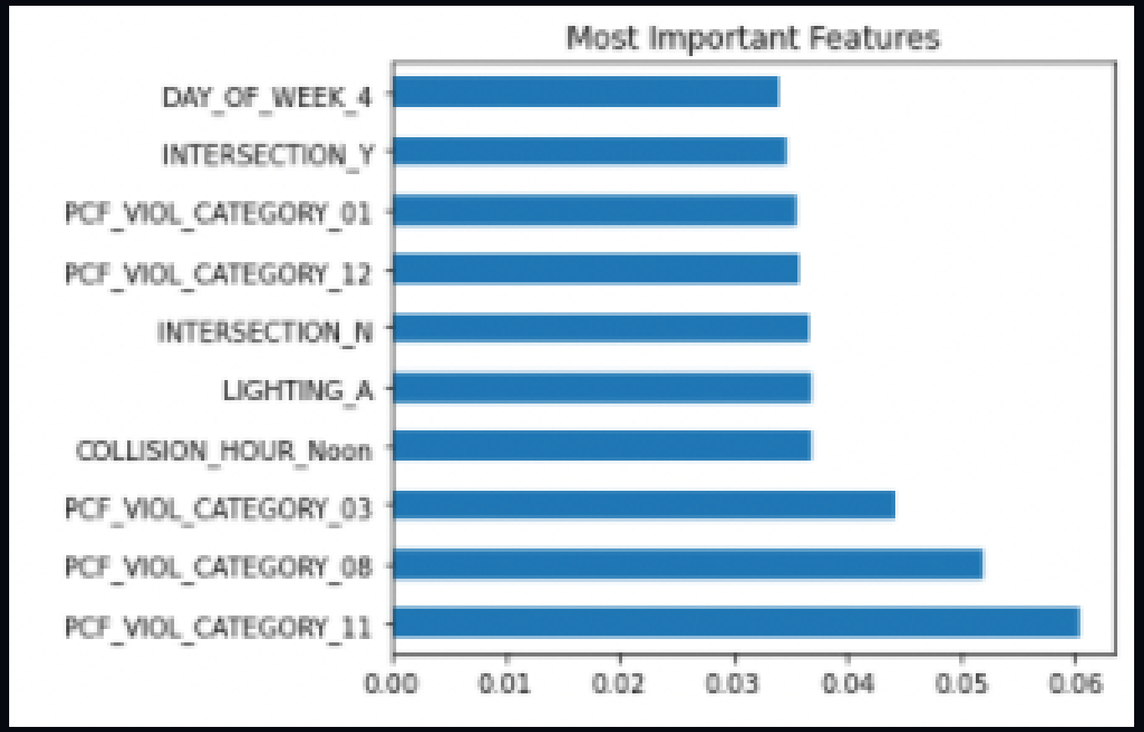


PARTY BY GENDER AND AGE AND RACE



ALCOHOL-INVOLVED COLLISIONS OVER TIME

```
Grid Search found the following optimal parameters:
model__bootstrap: True
model__criterion: 'entropy'
model__max_depth: 50
model__max_features: 'auto'
model__min_samples_leaf: 3
model__min_samples_split: 10
model__n_estimators: 100
pipeline test accuracy: 60.80%
```



Most Important Features

PCF_VIOL_CATEGORY_11: Pedestrian Violation
PCF_VIOL_CATEGORY_08: Improper Turning
PCF_VIOL_CATEGORY_03: Unsafe Speed
LIGHTING_A: Daylight
INTERSECTION_N: Not An Intersection

MODEL PREDICTIONS

Business Recommendations

- Explore different avenues and opportunities to reduce the speeds

- Better education and/or road signage for drivers to combat improper turning

FUTURE WORK

PREDICTIVE MODELING WITH MORE DATA/ MULTIPLE LABELS

OTHER CONSIDERATIONS

GATHER MORE DATA/MORE SCRUBBING

THANK

YOU