



US Automobile Accidents - A Presentation and Analysis

By: Elan Wilkinson, Parker Christenson, and Melissa Short

AAI-500-01

Group 8



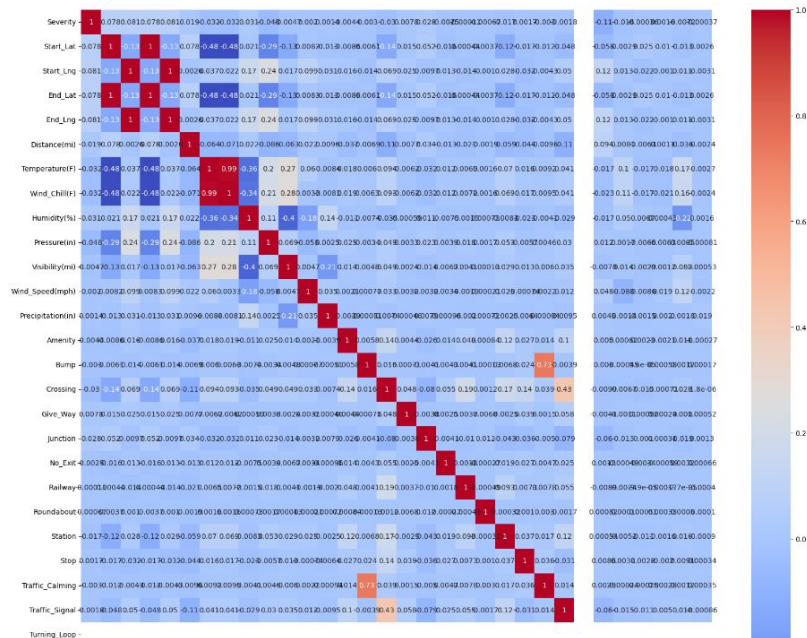
Introduction

- US Accidents (2016-2023)
- 7.7 million accident records.
- Analyze different factors and contributions to car accidents
- Prediction helps understand patterns in car accidents
- Predict number of automobile accidents for the future.



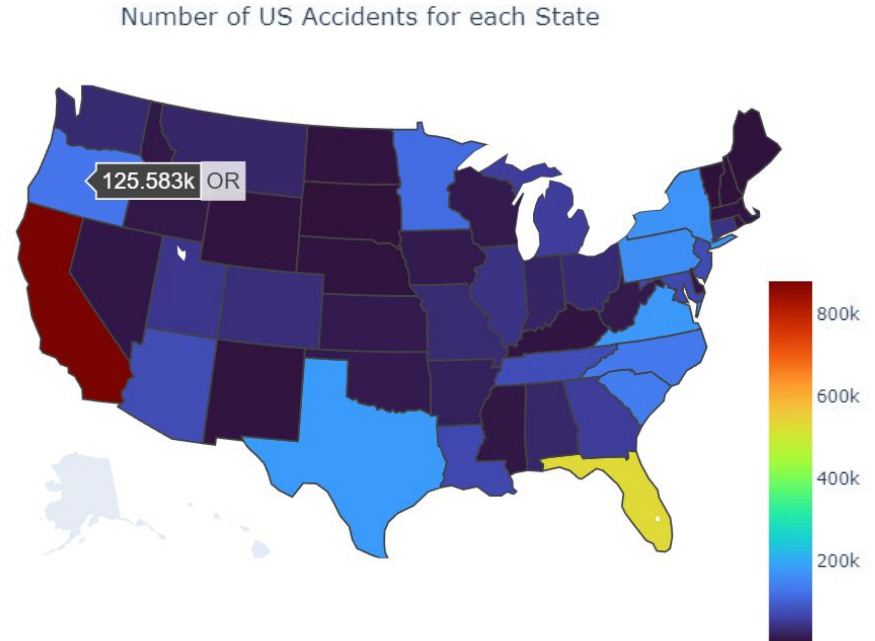
Data Cleaning/Preparation

- Original data set is 6,599, 480 rows and 46 columns of data
- Drop 2019-2023; data file size too large
- Drop empty data columns (5 columns)
- Referencing data columns (severity, location, weather, traffic, etc.)
- Change values True False to 1s and 0s
- Drop missing values in rows or data
- Add minute, hour, day, month, weekday to Data Frame

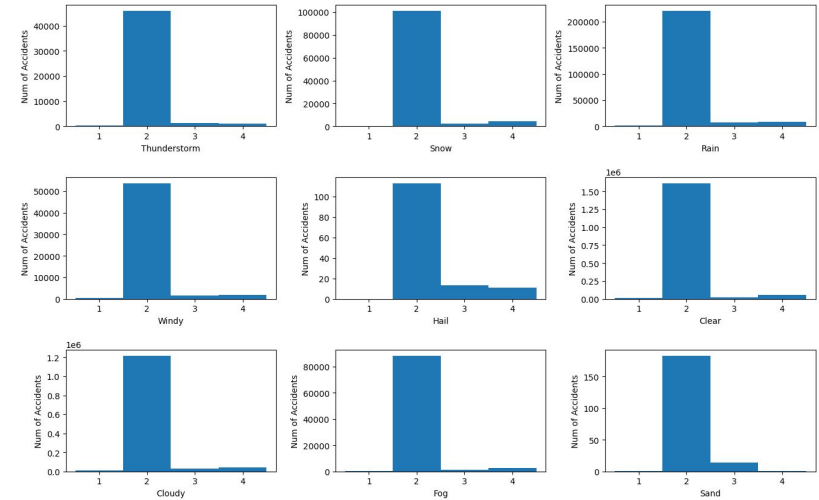
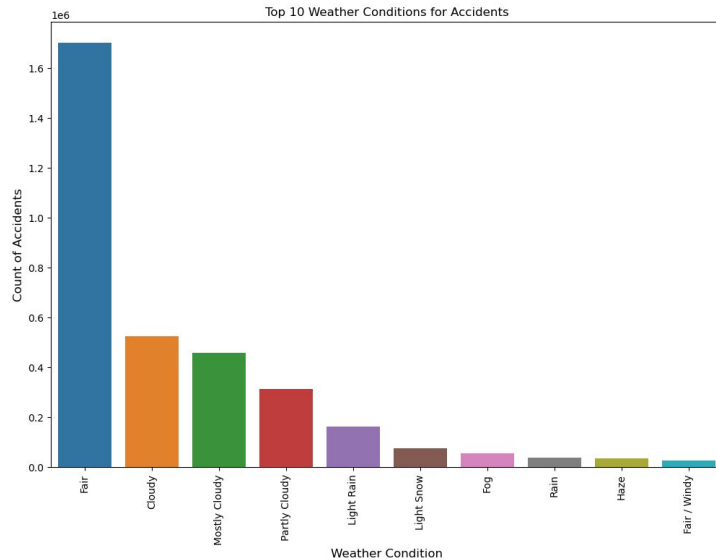


Exploratory Data Analysis

- Analysis of volume of accident data in relation to each state showed a significantly larger number of accidents reported in California and Florida (both of which are among the states with the largest populations)



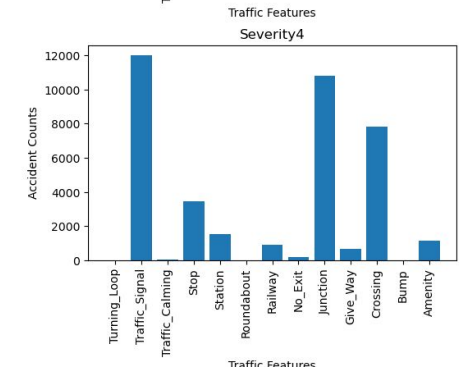
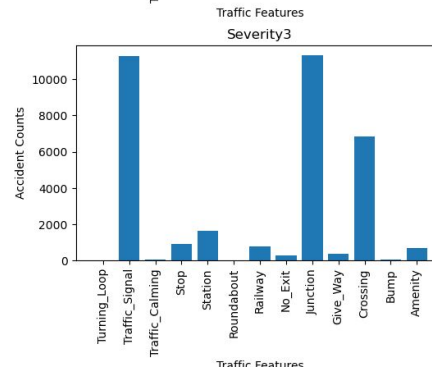
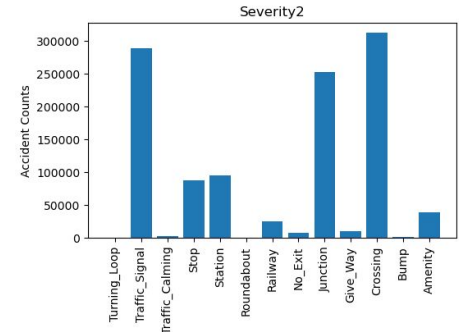
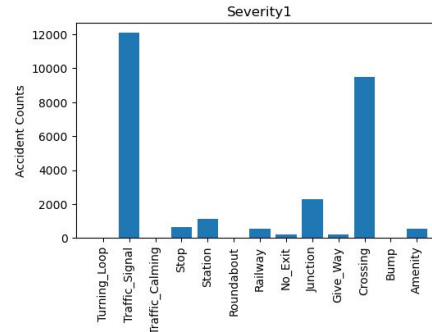
Exploratory Data Analysis



A view of histograms charting severity by weather type reflects the overrepresentation of severity 2 in the dataset, but does offer the insights of significantly more high severity accidents with hail, even more than thunderstorms or snow, although the number of accidents in dataset with hail was substantially lower.

Exploratory Data Analysis

- Investigation into the correspondence of the presence of certain environmental features and the severity of the accident was also performed, with all categories having a high number of traffic signals, junctions typically present in severities greater than 1, and greater representations of the presence of stop signs in severity 2 and severity 4 accidents.



Model Selection

- Data standardized using standard scaler
- Floats were converted to integers
- Data split into training set (80%) and testing sets (20%)
- Random Forest classification model chosen to predict accident severity given weather, location, and presence of variety of traffic features
- We chose the random forest model due to it being a robust classifier with a high level of accuracy



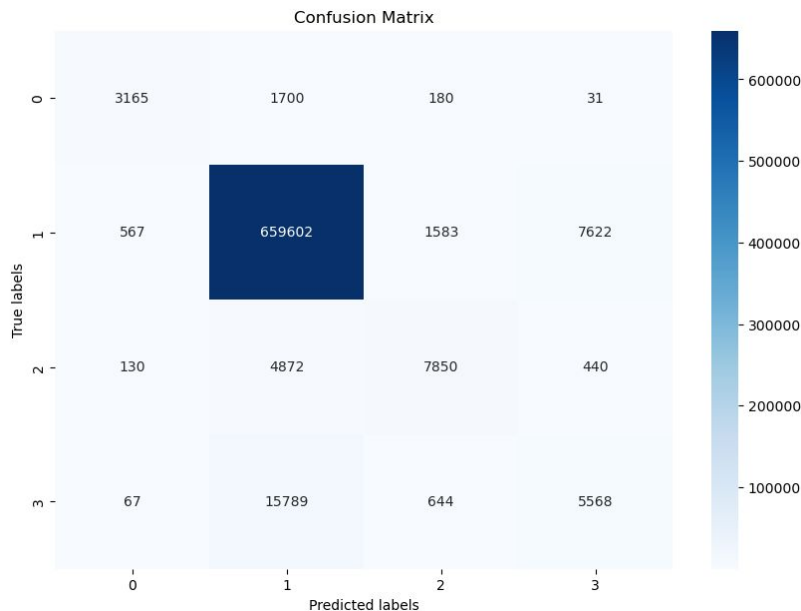
Final result

Model Analysis

Designed 3 SkLearn models in order to achieve a final graded accuracy of 95%

We attempted to make changes to all three of the models, by varying the training data formatting

The best choice out of all three of the models was in fact the first model created



Judging the models accuracy

SkLearn has a library that allows for the classification report to show some key metrics right after the development of the model

All of these metrics mean something, but the precision column shows us the accuracy of predicting each of the severity levels

Accuracy: 0.9526					
	precision	recall	f1-score	support	
1	0.81	0.62	0.70	5076	
2	0.97	0.99	0.98	669374	
3	0.77	0.59	0.67	13292	
4	0.41	0.25	0.31	22068	
accuracy			0.95	709810	
macro avg			0.66	709810	
weighted avg			0.95	709810	

Conclusions/Recommendations

We would like to see an overall changing on the National Standard of Car Accidents for a much more accurate model

When the model was built, the 2 rating was the most common 'Severity' level, and the model was trained to predict a 2 more than anything else, hence the models short coming for recognizing any other kind of accident

Lastly, we would like to see more features that are able to predict the severity of the Accident, such as amount and kind of vehicles involved, Speed, and impact position

References

Agresti, A., & Kateri, M. (2022). *Foundations of statistics for data scientists: With R and python*. CRC Press, Taylor & Francis Group.

Moosavi, S. (2023, May 28). *US accidents (2016 - 2023)*. Kaggle.
<https://www.kaggle.com/datasets/sobhanmoosavi/us-accidents>

Scikit-Learn. (n.d.). *GitHub - scikit-learn/scikit-learn: scikit-learn: machine learning in Python*. GitHub. <https://github.com/scikit-learn/scikit-learn>