Analysis of City of Seattle Collision Data

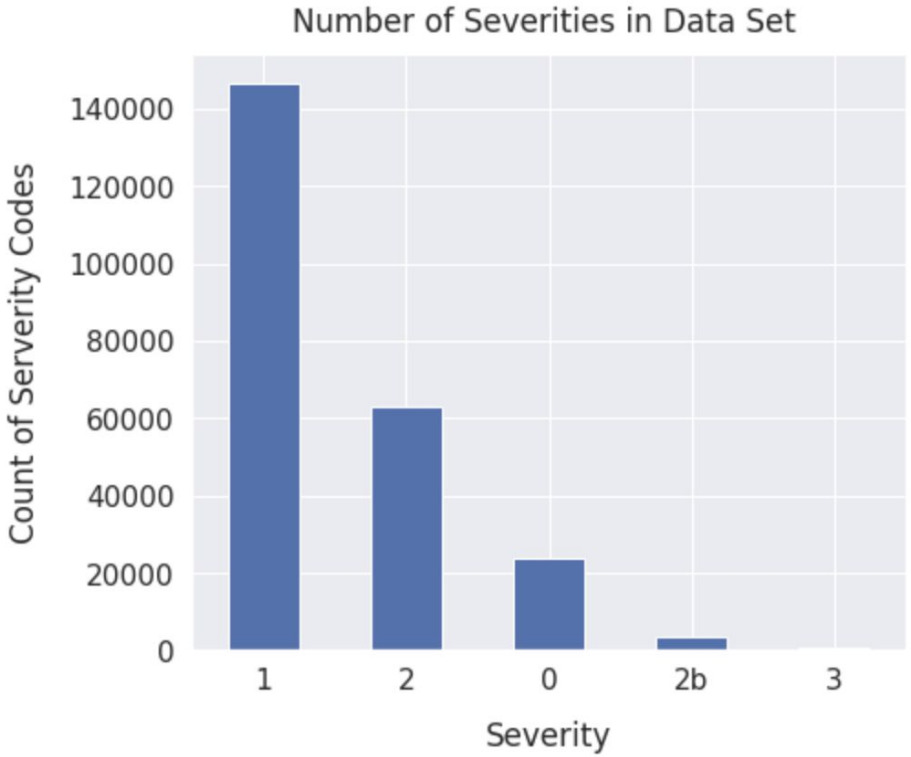
Data 5100
Group 5: Ava Delanty, Lizz Hunt, Akanksha Sharma, Anna Tsai

Introduction

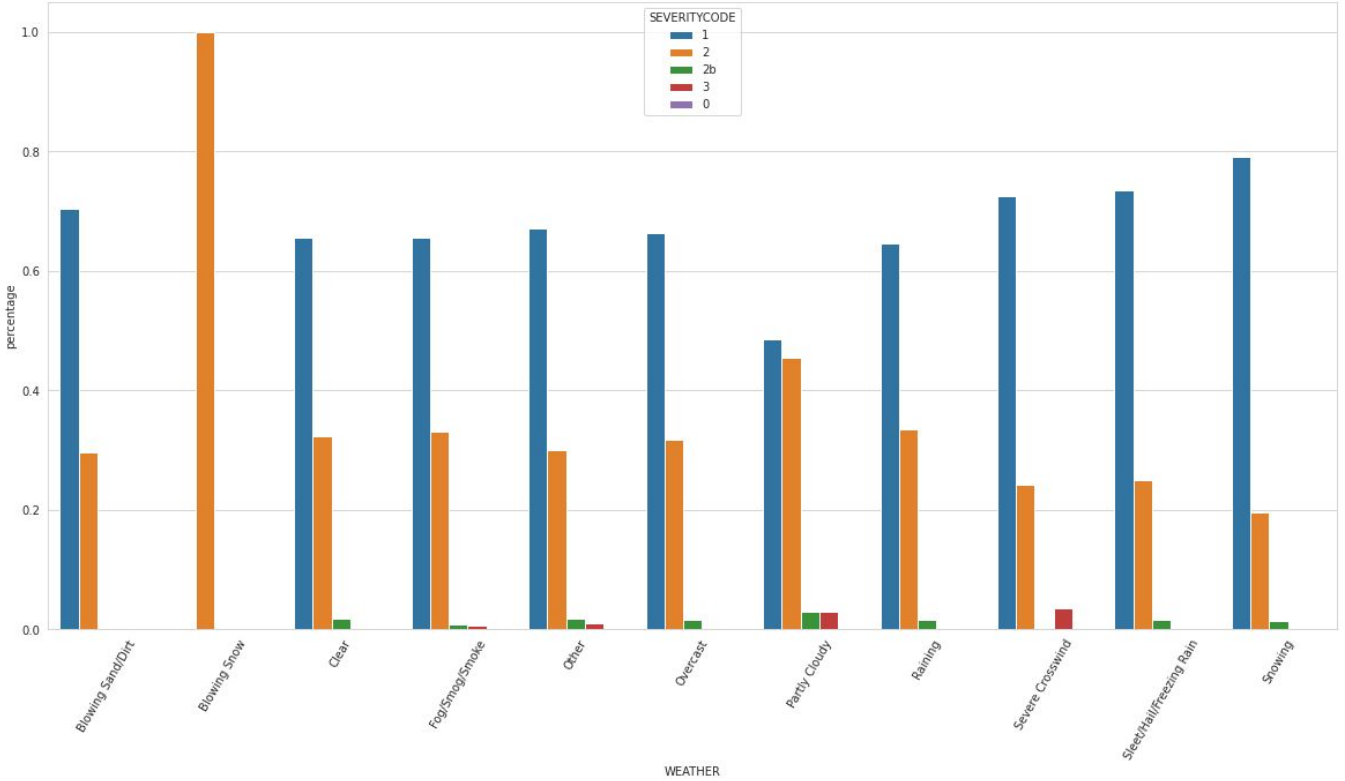
- In Seattle, traffic deaths or injuries have continued to rise.
- Speed limits and speeding are one of the leading factors in the severity of collisions (City of Seattle).
- Understanding the bigger picture.
- This project will use the City of Seattle’s collision dataset to analyze these different factors.

Data Description

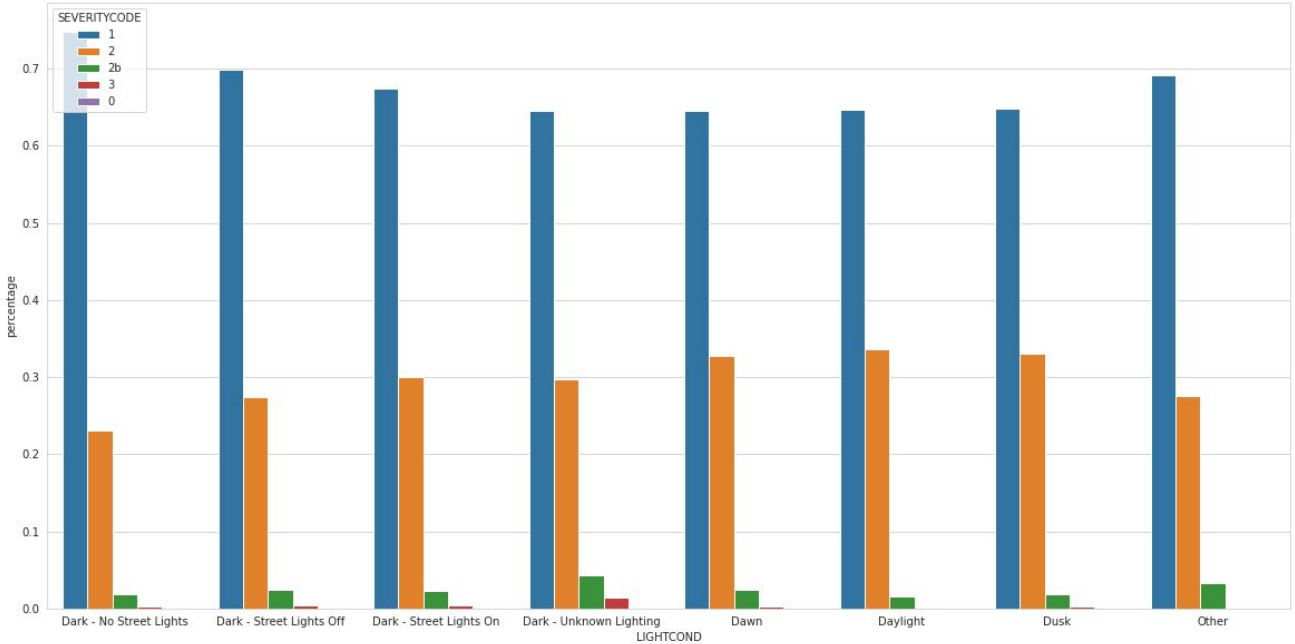
- The dataset includes all types of collisions from 2004 to Present day.
- The size of our dataset is 9,487,040 elements (237,176 rows).
- The dataset is from the City of Seattle Gov. site and is collected by the SDOT Traffic Management Division.
- Quality of the Data: NaNs.
- Data Processed: Extracting the csv file from site.
- Unsure on whether informed consent was given for data collection.
- No privacy was violated.



Exploratory Data Analysis: Weather



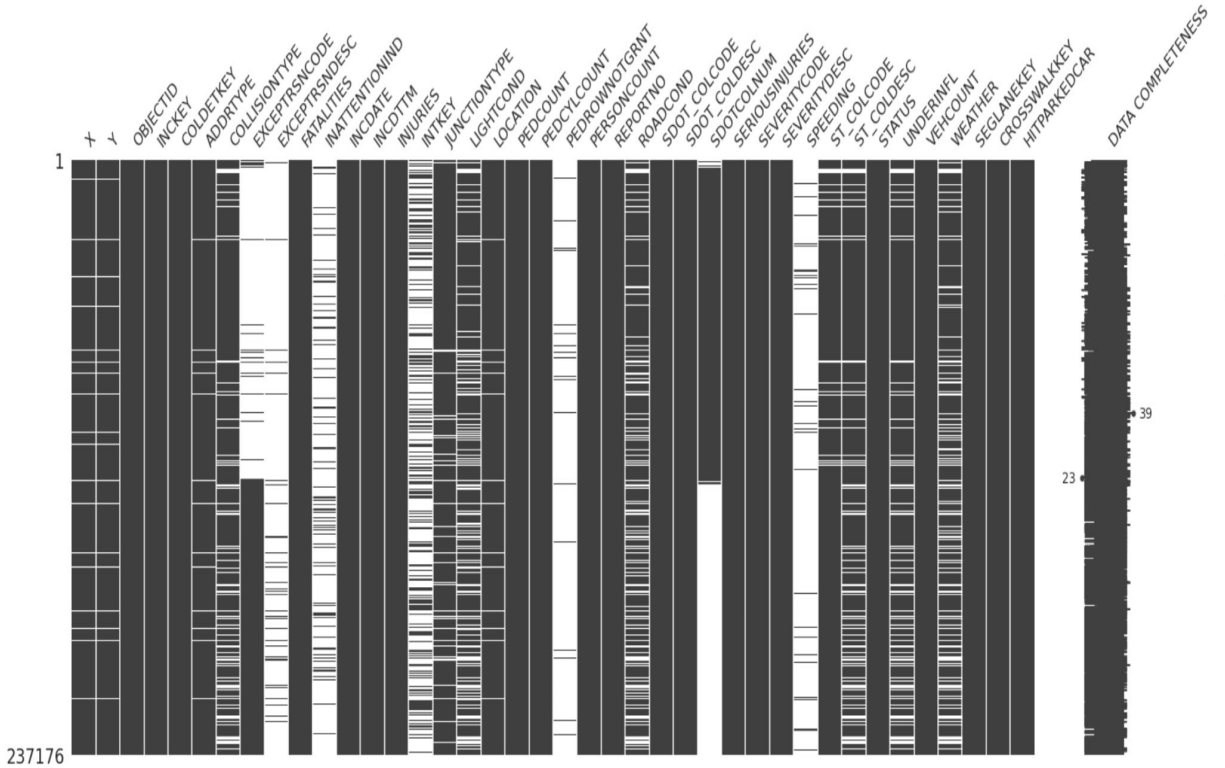
Exploratory Data Analysis: Light condition



Data Preparation and Initial Analysis

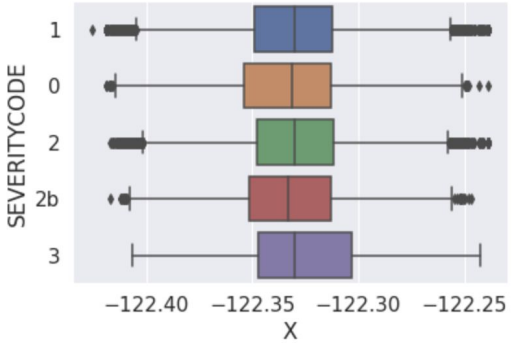
- Handling missing values

- Imputation:
  - Missing Values
  - Hit Parked
  - Speeding
  - Under Influence of drugs



Data Cleaning

- Columns we removed from the dataset:
  - Latitude and Longitude
  - Fatalities
  - Serious Injuries
  - Injuries

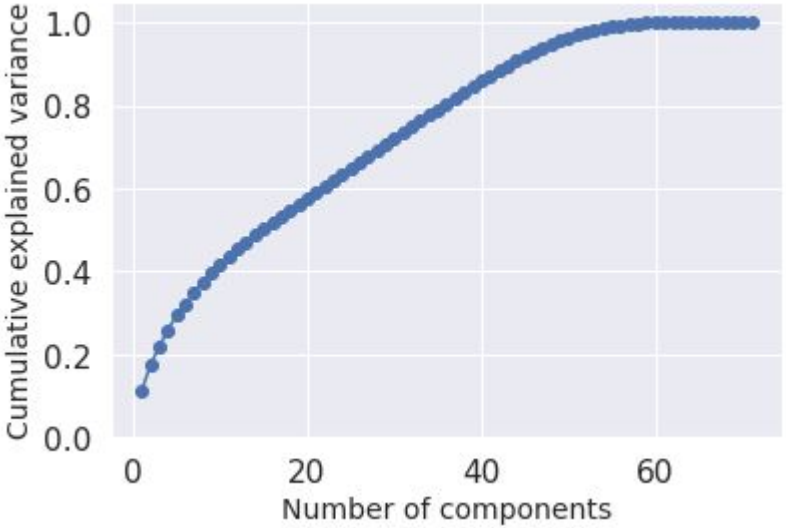


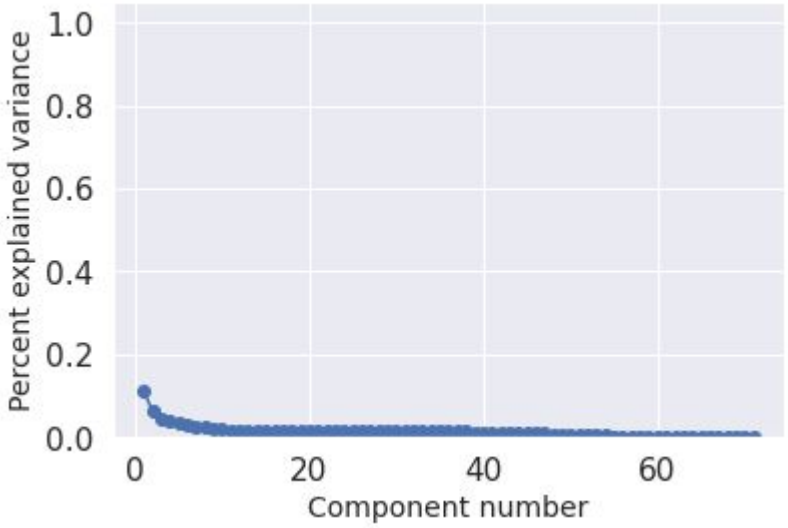
- One hot encoding for the categorical variables

Pair Plot

- Image shows a subset of the pair plot, hue is severity code
- Can see a lack of defined clusters, data looks almost categorical

Principal Component Analysis





Principal Component Analysis (cont.)

Modeling Approach

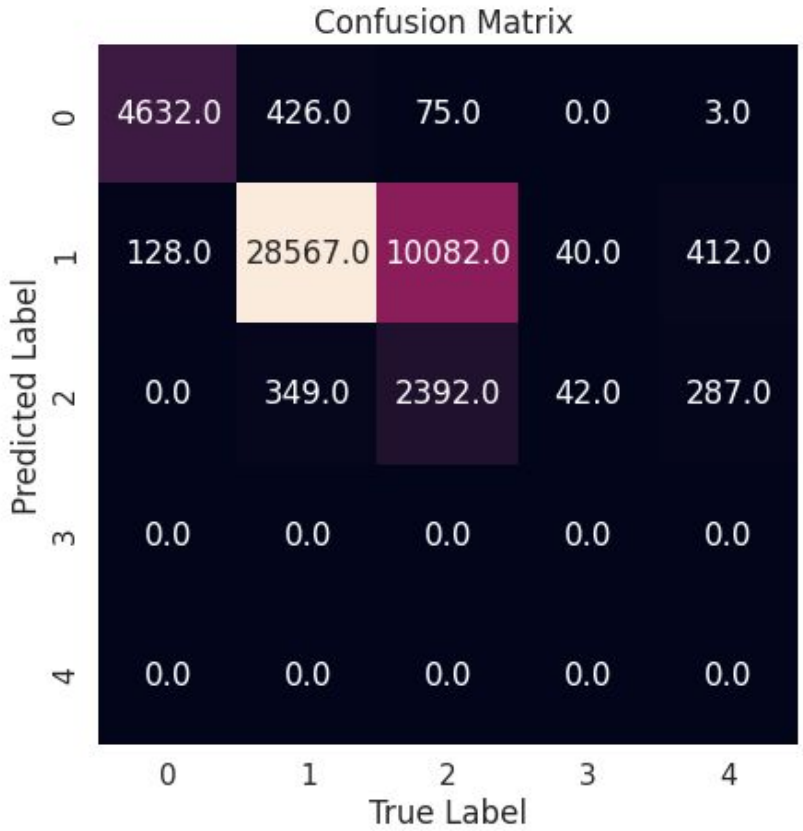
- **Predictors**
- **Response: Severity Code (5 categories)**
- **Type of problem**
  - Supervised Classification problem
  - Logistic Regression
    - L1 Regularization
  - Random Forest Tree
    - Hypertuning based on number of parameters

Modeling Results–Logistic Regression vs Random Forest

Model	Accuracy	Specificity	Sensitivity
Logistic Regression	75%	0.994	0.934
Random Forest	75%	0.973	0.974

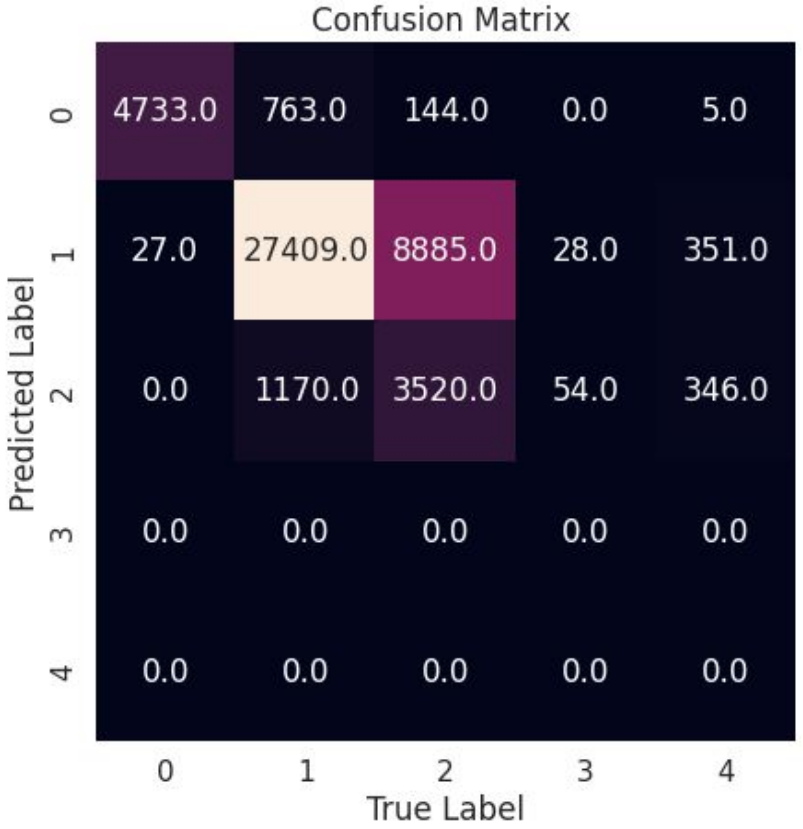
Random Forest Performance

	Precision	Recall	F1-Score	Support
0	0.9	0.97	0.94	4,760
1	0.73	0.97	0.83	29,342
2	0.78	0.19	0.31	12,549
2b	0	0	0	702
3	0	0	0	82



Logistic Regression Performance

	Precision	Recall	f1-score	Support
0	0.84	0.99	0.91	4,760
1	0.75	0.93	0.83	29,342
2	0.69	0.28	0.40	12,549
2b	0	0	0	702
3	0	0	0	82



Logistic Regression Model Sensitivity

Predicting Unknown Severity	
Parameter	**Sensitivity**
Person Count	-2.51
Vehicle Count	-2.23
Status Unmatched	+2.64
Junction Type Unknown	+1.72

Predicting Property Damage	
Parameter	**Sensitivity**
Pedestrian count	-2.95
Cyclist Ped Count	-2.248
Weather Blowing Snow	-1.48
Collision Type Parked Car	+2.51
Collision type Sideswipe	+1.621181

Predicting Injuries	
Parameter	**Sensitivity**
Collision Type Parked Car	-1.84
Collision Type Sideswipe	-1.13
Collision Type Cycles	+1.112
Collision Type Pedestrian	+1.102

Conclusion

- Random Forest and Logistic Regression had similar performance overall
  - Logistic Regression was better at discerning property damage from injuries
- Lack of data regarding crashes with fatalities lead to poor performance classifying what parameters lead to death.
  - Unable to relate speeding to fatalities as SDOT does.

Questions?

Works Cited

Citations

Anderson, Hans and Denkman, Libby. “Even with Vision Zero, traffic fatalities remain high in Seattle.” *KUOW*, https://www.kuow.org/stories/even-with-vision-zero-traffic-fatalities-remain-high-in-Seattle. Accessed 15 October 2022.

City of Seattle, Department of Transportation. *Collisions- All Years*. City of Seattle, n.d. https://www.seattle.gov/Documents/Departments/SDOT/GIS/Collisions_OD.pdf

City of Seattle, Department of Transportation. *2021 Traffic Report*. City of Seattle, February 2022.

“Demystifying the Safe Systems Approach.” *Vision Zero Network,* https://visionzeronetwork.org/resources/demystifying-the-safe-system-approach/. Accessed 15 October 2022.

“Speed Limits.” *City of Seattle,* https://www.seattle.gov/transportation/projects-and-programs/safety-first/vision-zero/speedlimits. Accessed 15 October 2022.

“Vision Zero.” *City of Seattle*, https://www.seattle.gov/transportation/projects-and-programs/safety-first/vision-zero. Accessed 15 October 2022.