

---

---

# PREDICTING THE SEVERITY OF COLLISIONS

— Jessica Teng —

---

---

# INTRODUCTION

Background: This project utilizes the shared dataset provided by the Seattle Police Department and Traffic Records of all the collisions happened from 2004 - 2020.

Goal/Problem: With the number of car accidents and its consequences increasing, it is imperative to improve road safety and reduce the number of occurrences by developing a prediction model to predict car collisions severity .

Audience: Mainly targets Seattle residents and every driver that is responsible for their own and others' safety.

# DATA DESCRIPTION

- Target variable = “SEVERITYCODE”
  - 1 = property damage
  - 2 = injury
- 16 relevant columns/attributes:
  - SEVERITYCODE
  - ADDRTYPE
  - SEVERITYDESC
  - COLLISIONTYPE
  - PERSONCOUNT
  - PEDCOUNT
  - PEDCYCLOUNT
  - VEHCOUNT
  - INCDATE
  - INCDTTM
  - INATTENTIONIND
  - UNDERINFL
  - WEATHER
  - ROADCOND
  - LIGHTCOND
  - SPEEDING

# METHODOLOGY & ANALYSIS

- Information on the total number of each type of collisions and the different causes/conditions involved in the collisions

Number of collisions with property damage: 136485

Number of collisions with injury: 58188

Number of collisions with fatality: 0

Number of collisions where the driver was speeding: 9333

Number of collisions during the day: 116137

Number of collisions during the night with street lights: 48507

Number of collisions during the night with no street lights: 1537

Number of collisions during the night: 50044

Number of collisions during dusk: 5902

Number of collisions during dawn: 2502

Number of collisions where the road was dry: 124510

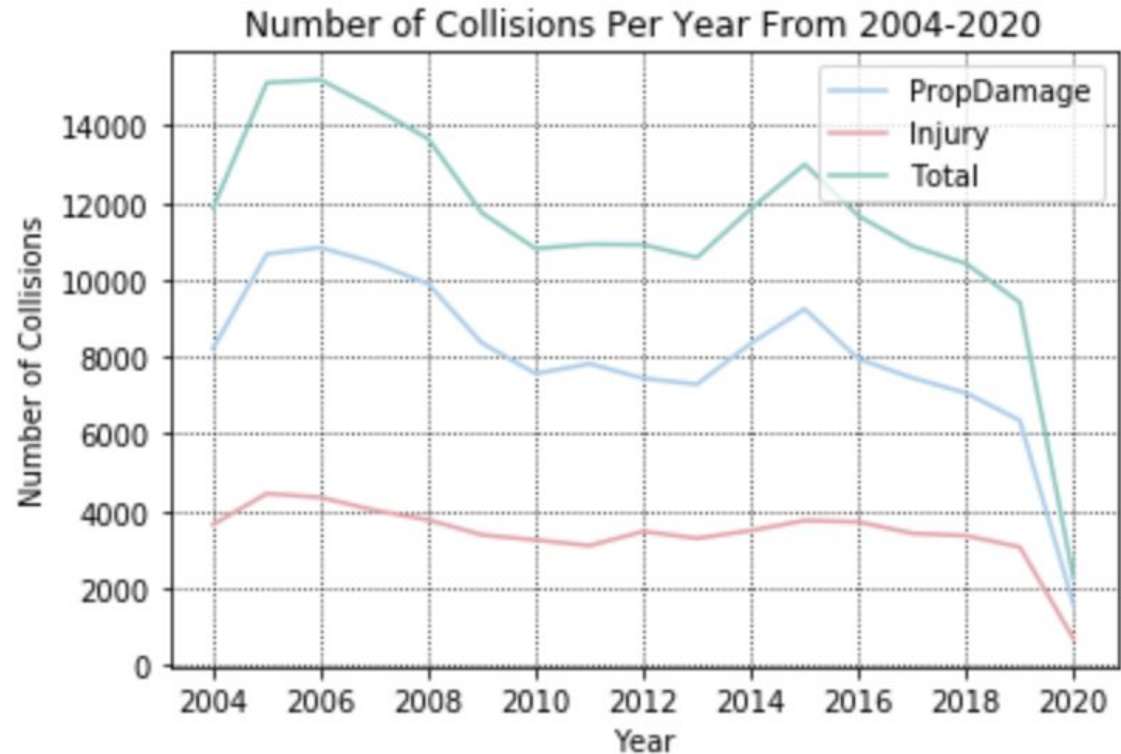
Number of collisions where the road was wet: 47474

Number of collisions where the driver was under the influence: 9121

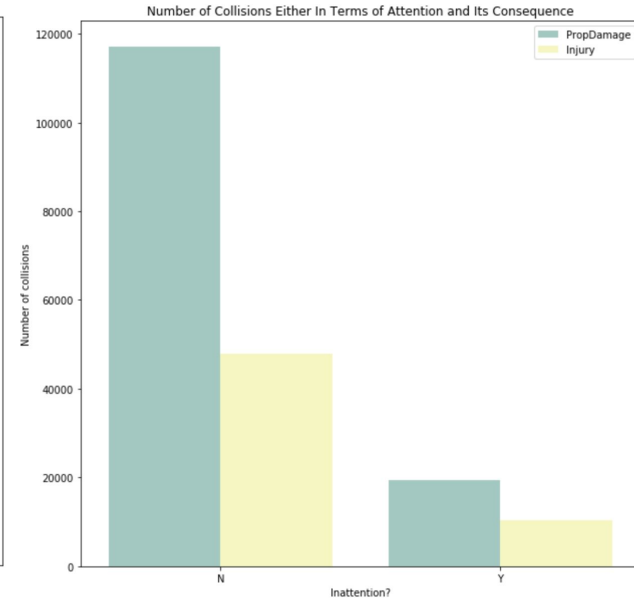
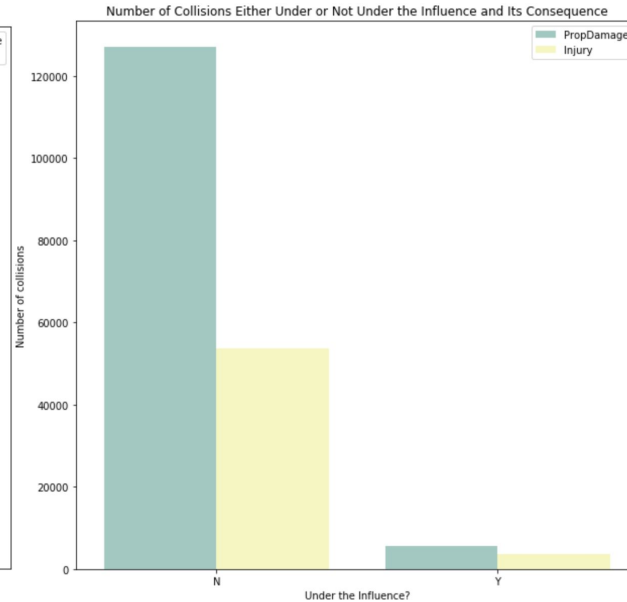
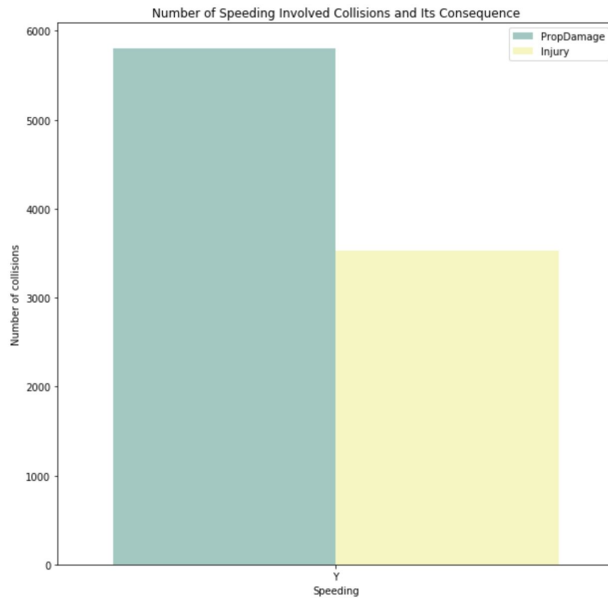
Number of collisions where the driver was not paying attention: 29805

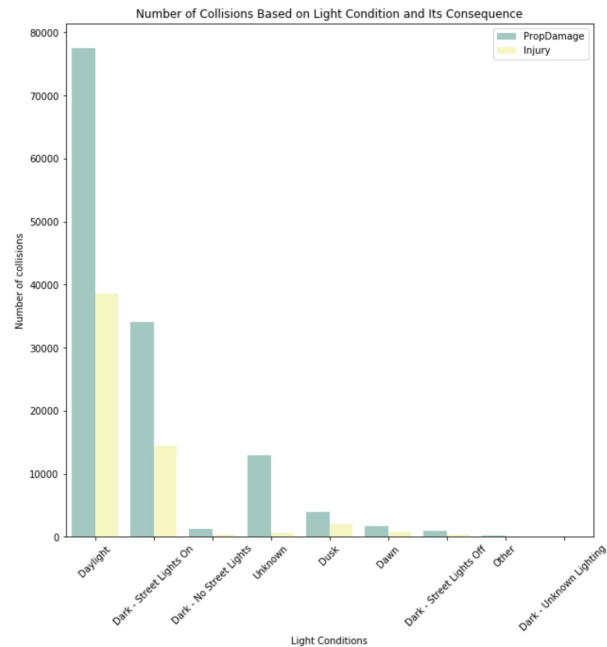
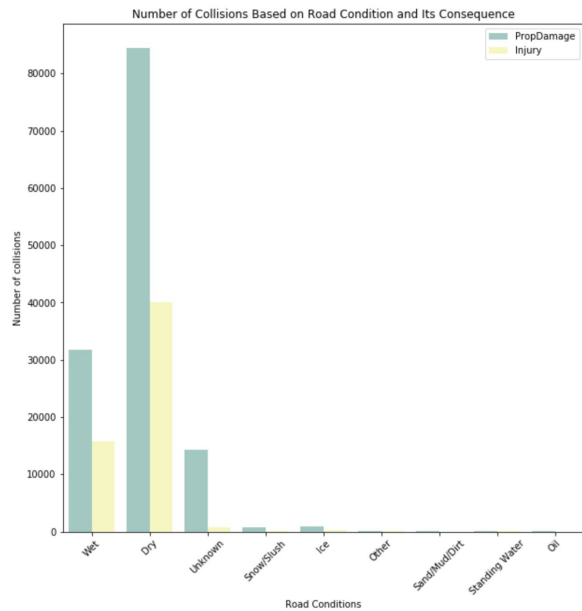
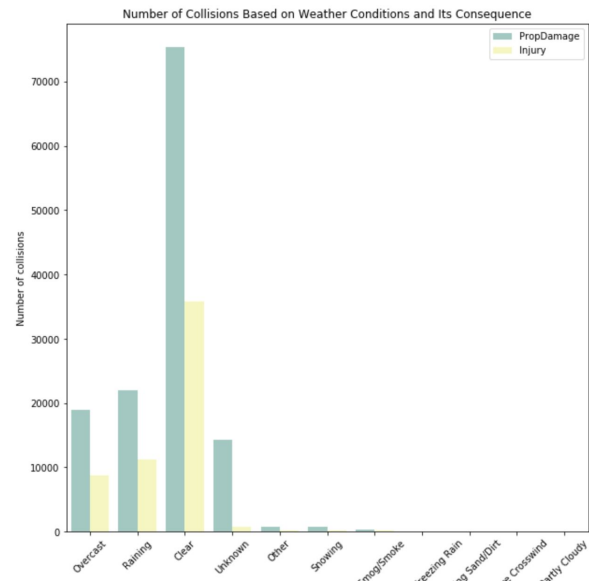
- Collisions with respect to the SEVERITYCODES throughout the years from 2004-2020

SEVERITYCODE	PropDamage	Injury	Total
Year			
2004	8218	3647	11865
2005	10665	4450	15115
2006	10838	4350	15188
2007	10439	4017	14456
2008	9893	3767	13660
2009	8356	3378	11734
2010	7563	3245	10808
2011	7820	3099	10919
2012	7440	3467	10907
2013	7287	3290	10577
2014	8351	3490	11841
2015	9243	3752	12995
2016	7945	3714	11659
2017	7454	3419	10873
2018	7061	3358	10419
2019	6350	3062	9412
2020	1562	683	2245

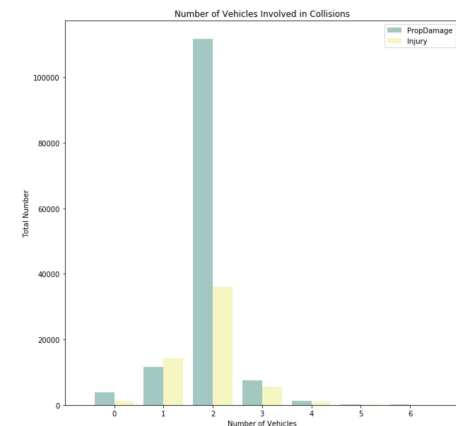
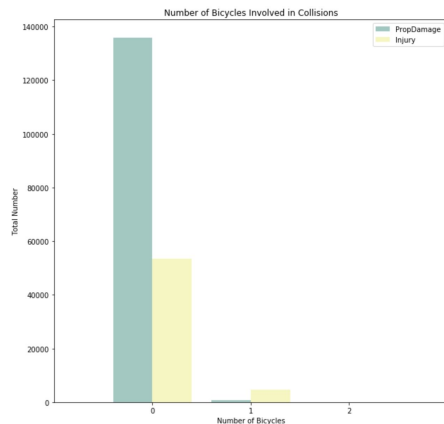
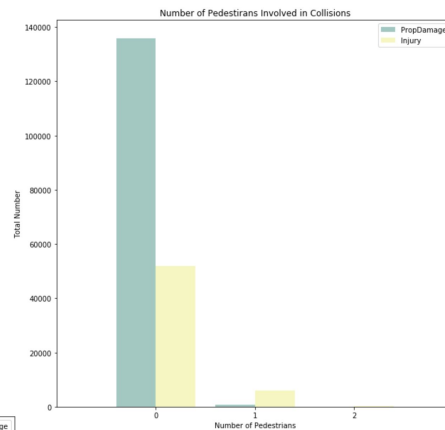
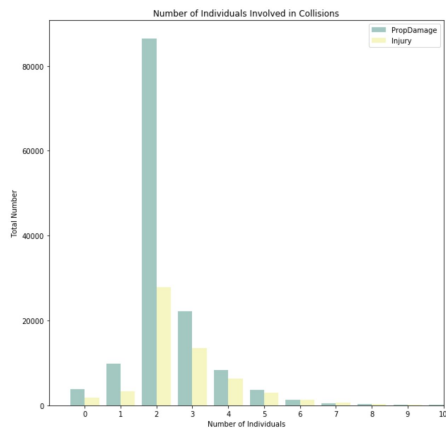


- Same concepts and relationships but with different potential causes/conditions





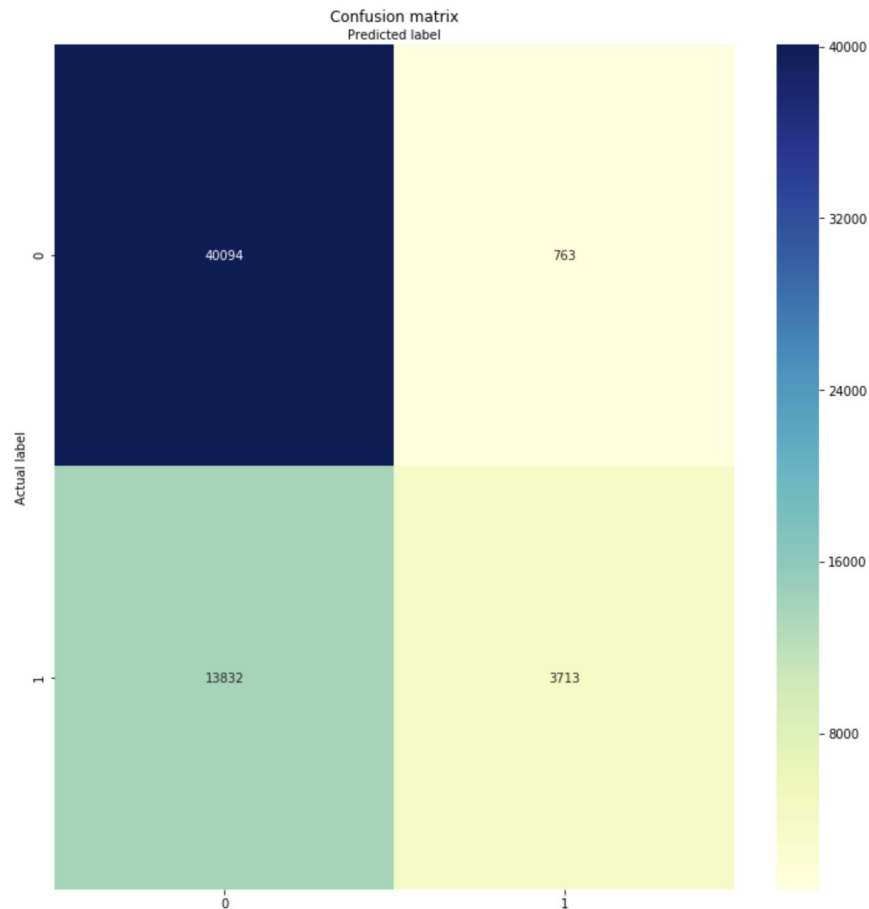
- Total number of vehicles, bicycles, pedestrians, and individuals involved in the collisions with respect to the SEVERITYCODES





# LOGISTIC REGRESSION

- The chosen classification algorithm
- Pre-processing & selection
- Training, testing, splitting dataset
- Model development
- Fitting the model with data
- Prediction
- Model evaluation via confusion matrix



# RESULTS/DISCUSSION

- Logistic Regression score of about 0.75

**Accuracy: 0.7500941748570255**

- Most of the collisions happened with:
  - Property damage
  - Drivers not under the influence, not speeding, and paying attention
  - Clear weather, raining, and overcasting
  - Dry road conditions
  - Daylight or night time but with street lights on
  - 2 individuals and 2 vehicles
  - 0 pedestrians and 0 bicycles involved

# CONCLUSION

- Useful prediction model
- With 0.75 accuracy score, there is definitely room for improvement
- Can improve with more relevant and precise information and if there are less “NaN” or missing values