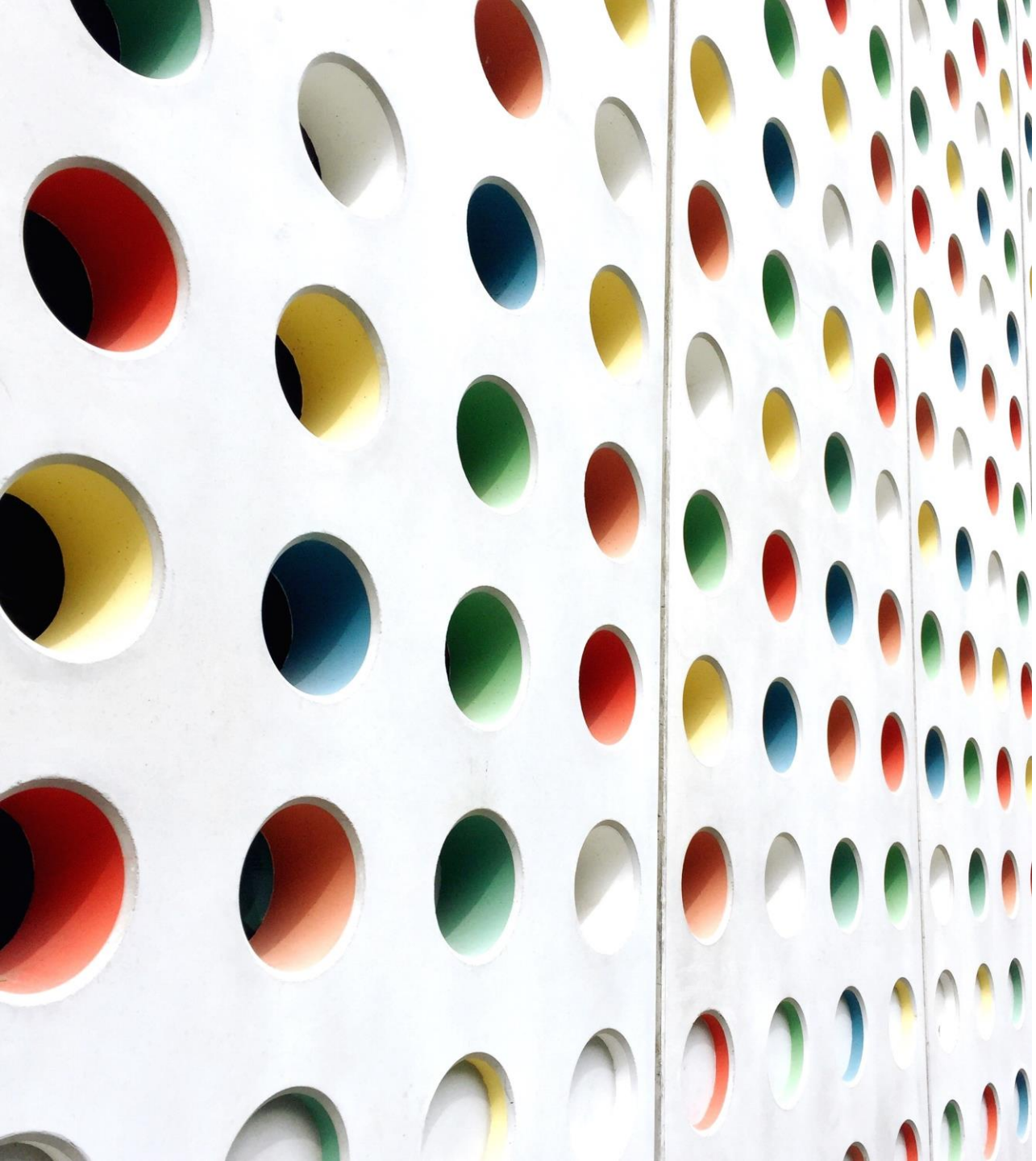


# Predicting Car Accident

REGINA RHODELIA B. MONASTERIAL

# Introduction

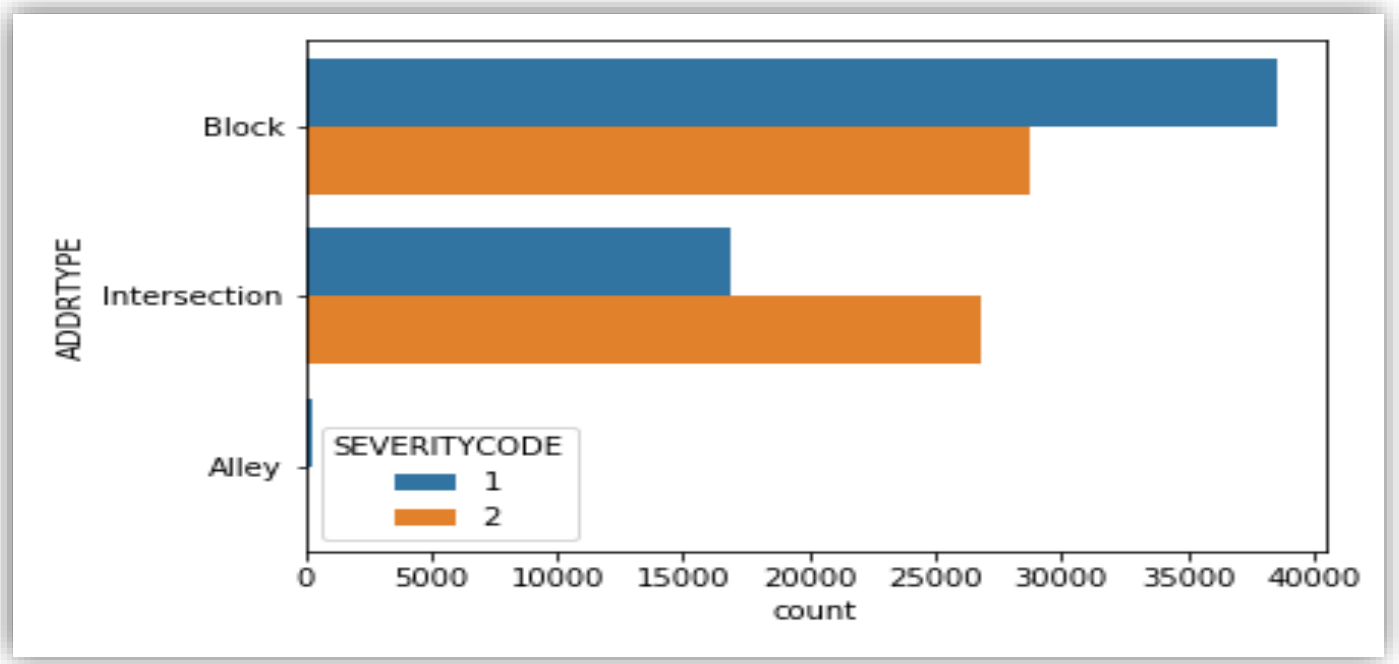
- Road accidents can occur to anyone anytime anywhere.
- Fortunately, data on road accidents are continuously collected and are made available to everyone.
- Maximize on the value of available information on location and other external conditions.
- The possibility and severity of an accident can be predicted and prevented thereby assisting local government units in publishing this information to travelers along their local areas.

# Data

- Dataset on collisions collected since 2004 to present, as provided by the Seattle Police District and recorded by Traffic Records
- 194,673 observations with 37 attributes
- Dependent Variable – SEVERITYCODE
- Independent Variables
  - Location - collision address type - ADDRTYPE
  - External conditions
    - The weather during the time of collision - WEATHER
    - Light conditions during the collision - LIGHTCOND
    - The condition of the road during collision - ROADCOND
    - The week of day as extracted from the date of collision - INCDATE

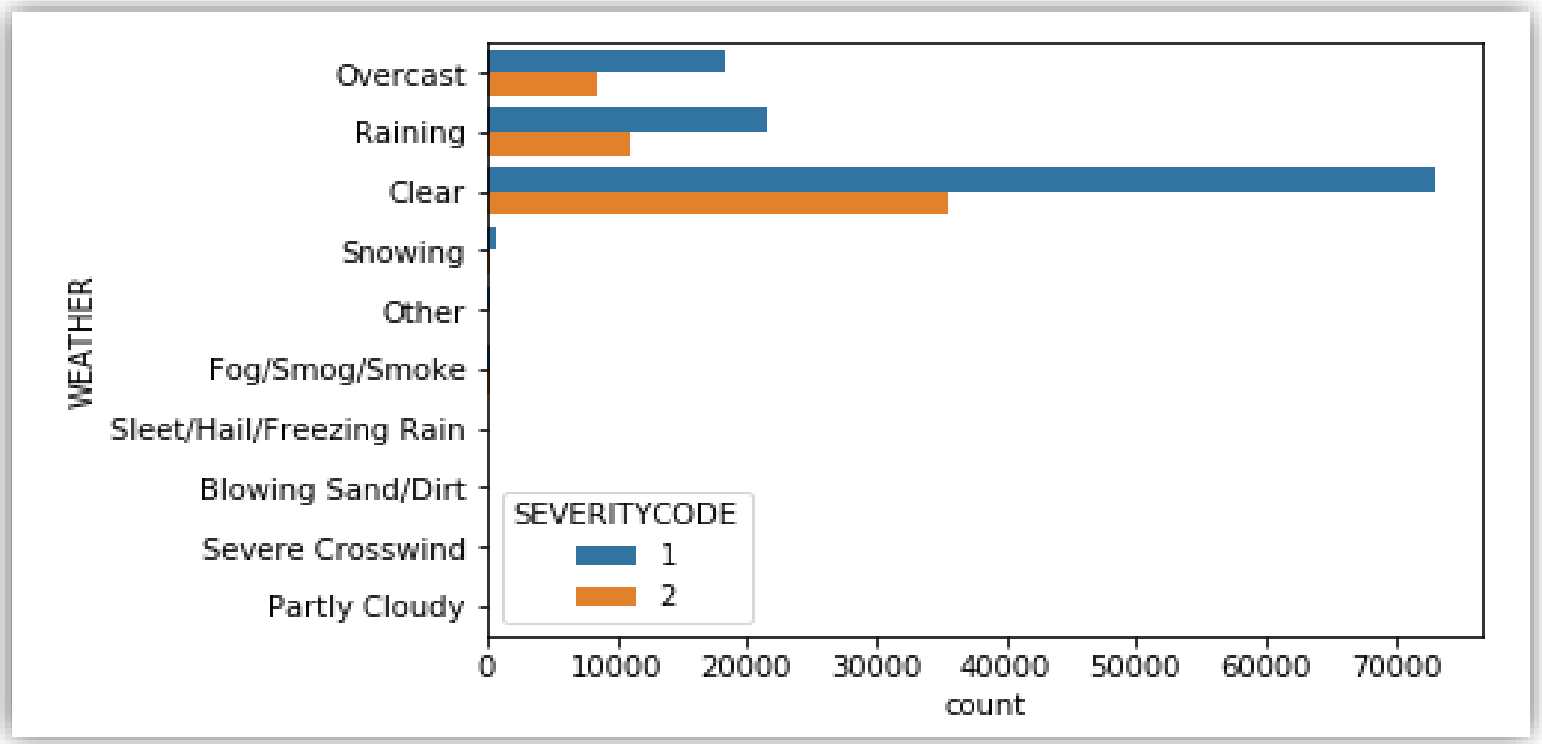
# Observations on Independent Variables

- There are more Severity Code 1 collisions along Blocks than on Intersections. Collisions of Severity Code 2 registered almost the same count on both Blocks and Intersections.



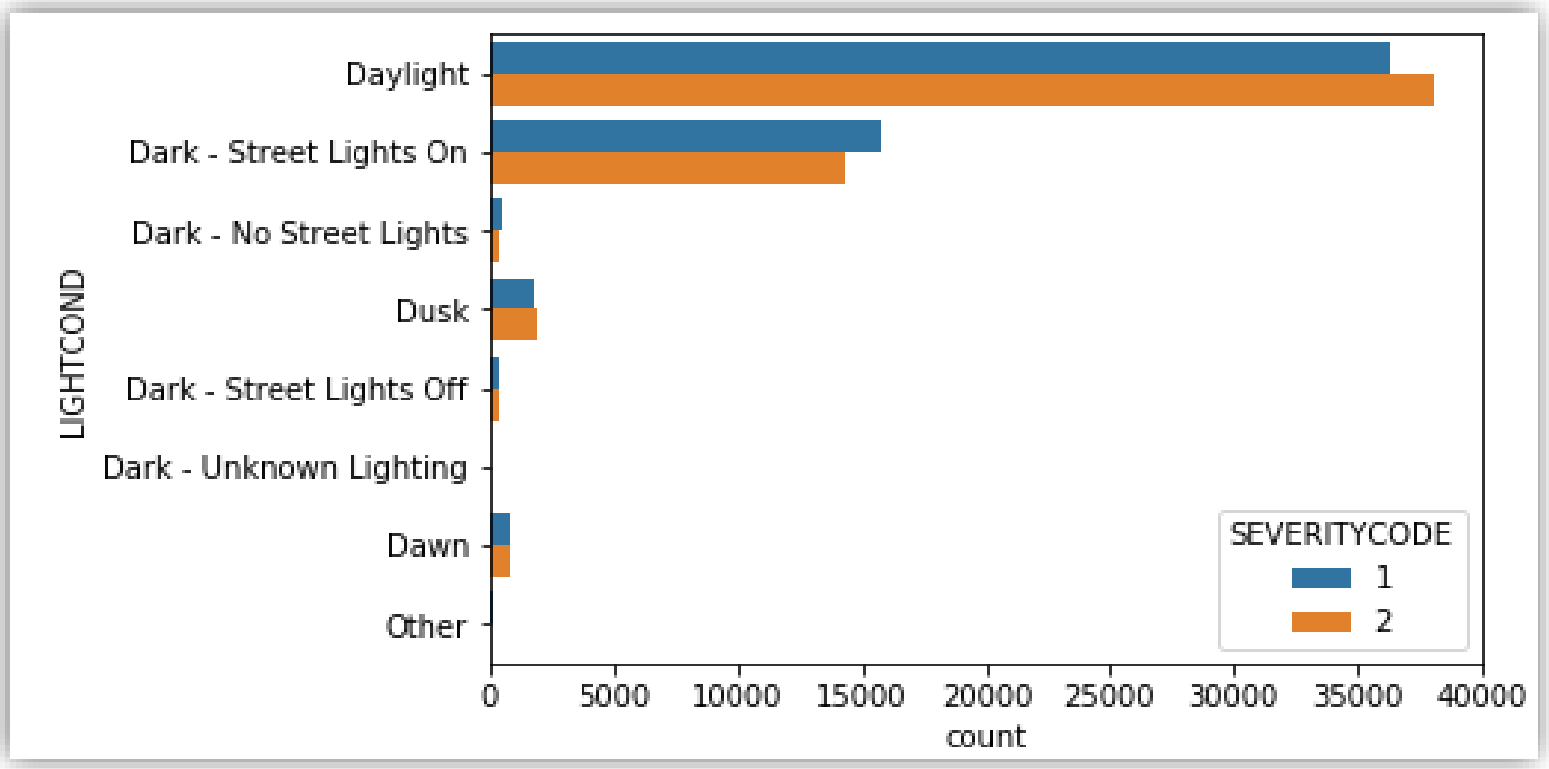
# Observations on Independent Variables

- Most number of collisions happen on Clear Weather.



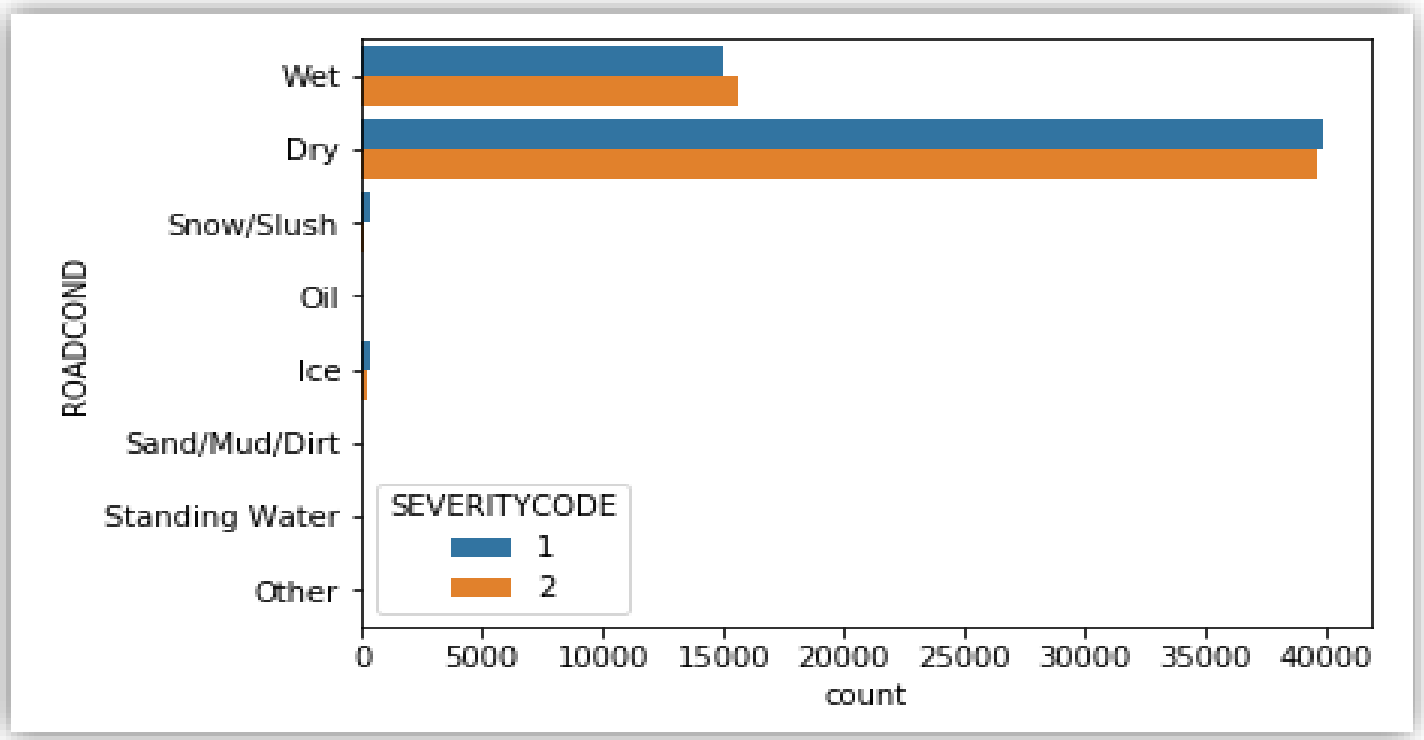
# Observations on Independent Variables

- Most number of collisions happen on daylight.



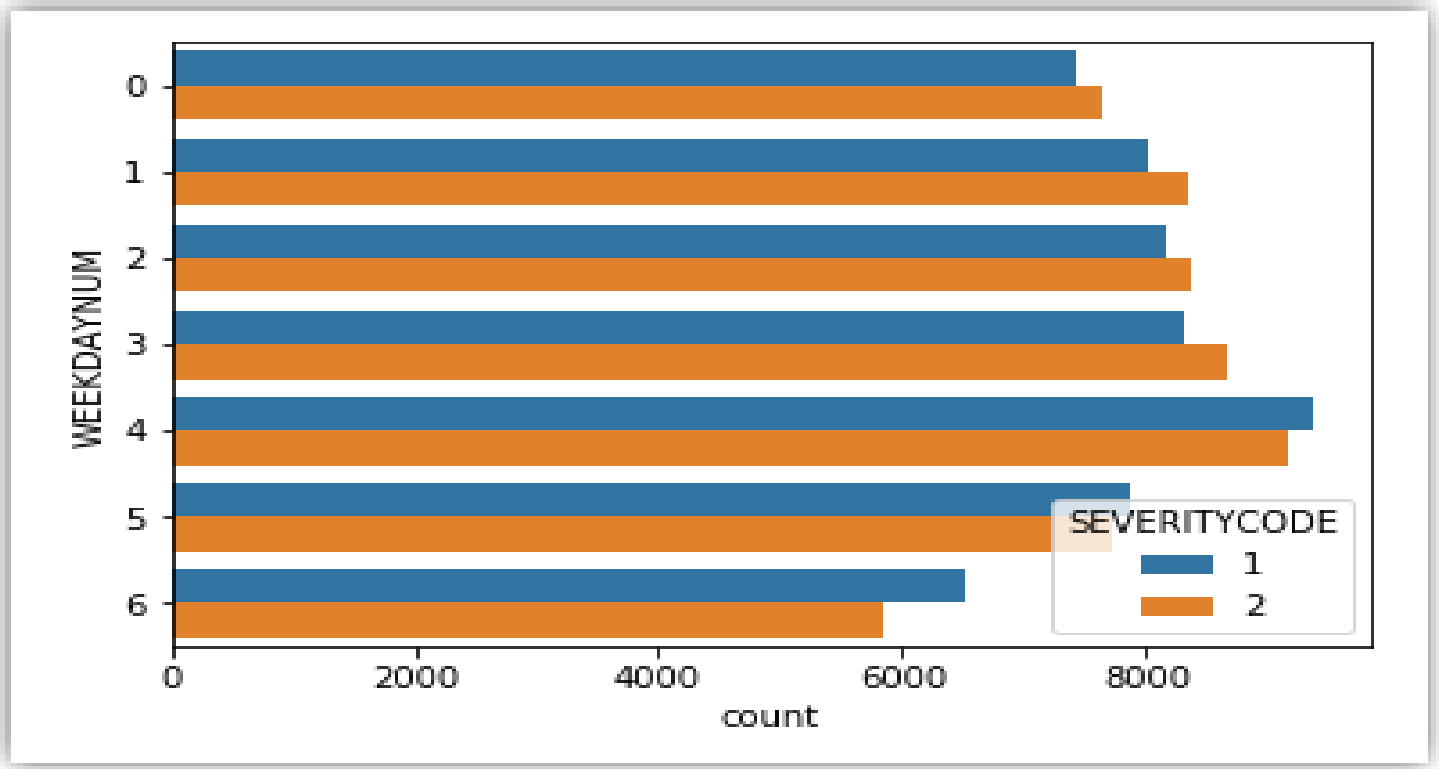
# Observations on Independent Variables

- Most number of collisions happen on dry roads.



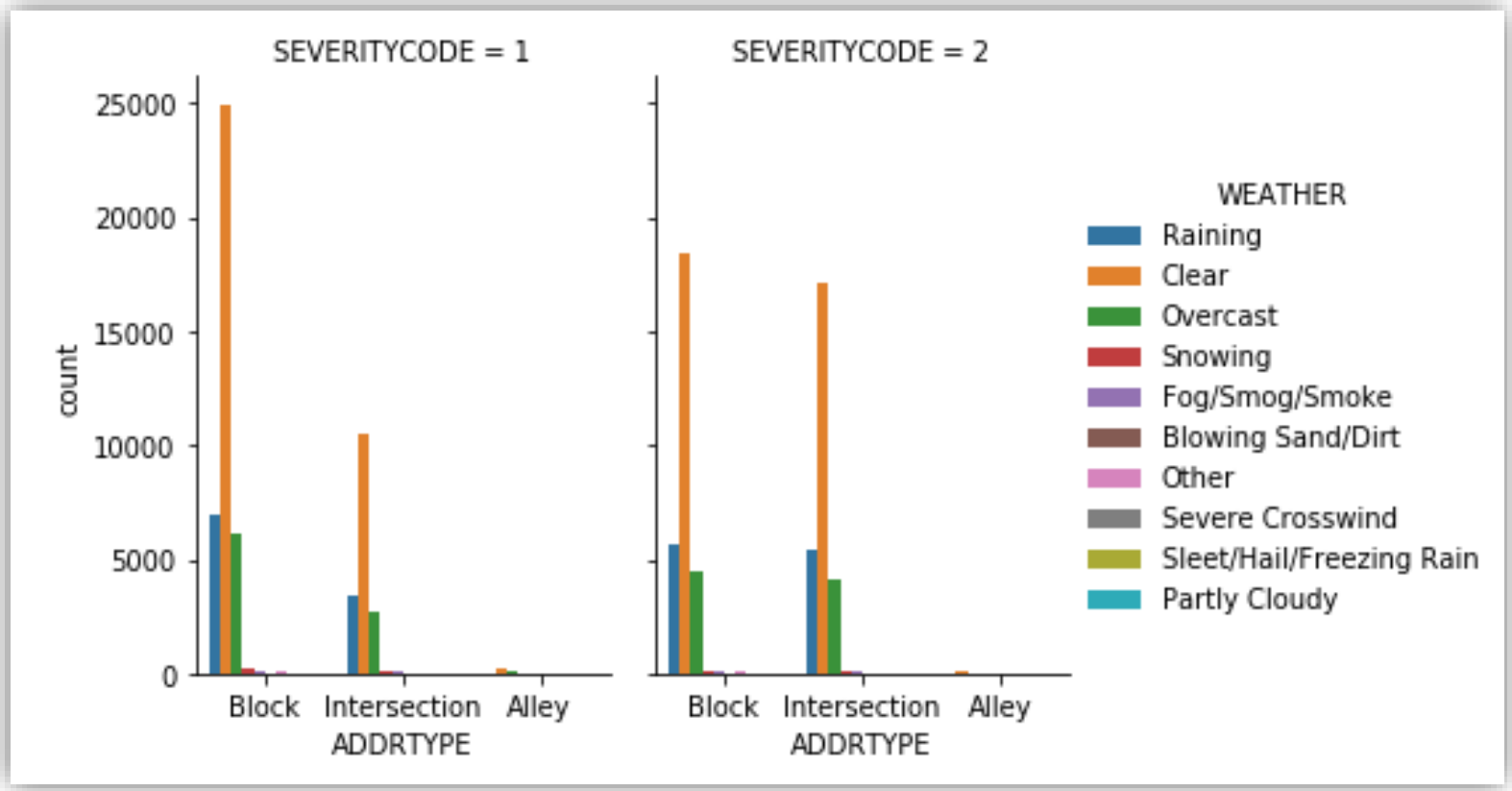
# Observations on Independent Variables

- There is less collisions during weekend and surprisingly, there is an upward trend on collisions on weekdays.



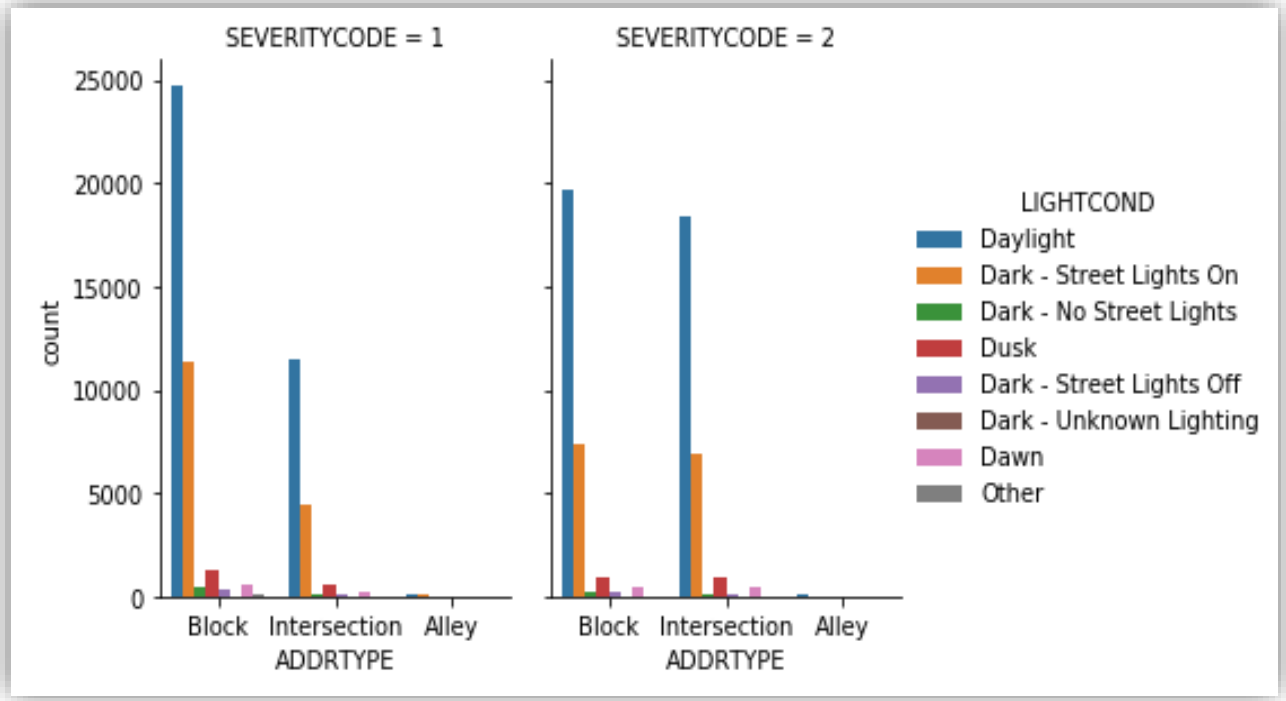
# Observations on Independent Variables

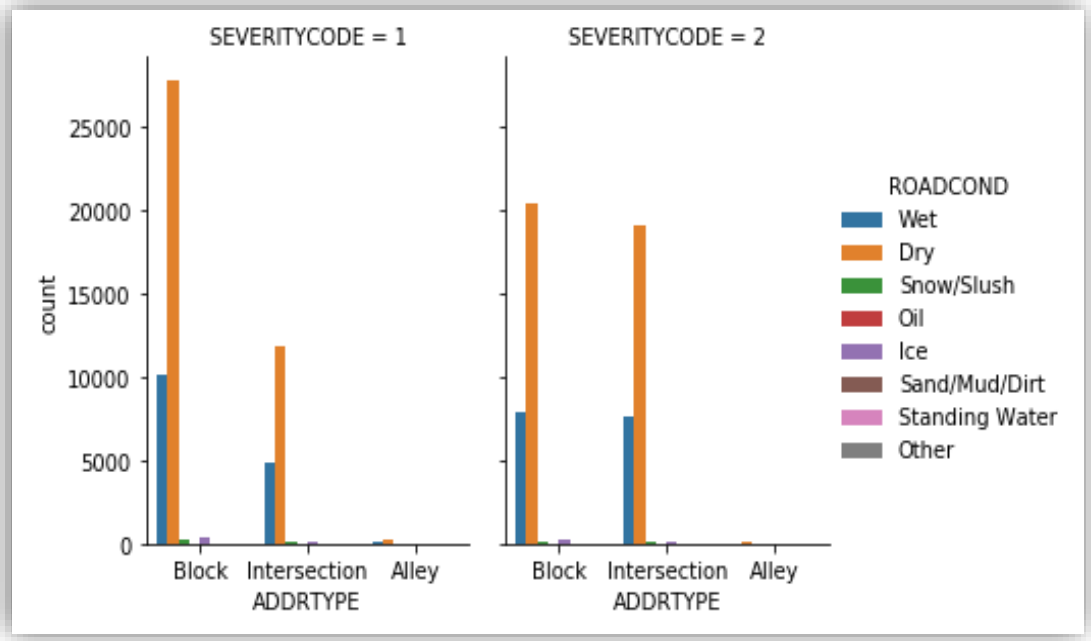
- Most collisions happen on blocks on a clear weather.



# Observations on Independent Variables

- Collisions happen on dry roads that are well lighted, mostly on blocks.



# Result

- Using the F1-score of our algorithms, SVM gave the highest result, though not that significant difference from the other F1-scores.
- The accuracy of the Logistic Regression is based on the Logistic Loss of 0.675.
- The result is not as good as we expected because the accuracy of the models is not very high.

| Algorithm | Jaccard | F1-score | LogLoss |
|---|---|---|---|
| KNN | 0.5615 | 0.5605 | NA |
| Decision Tree | 0.5896 | 0.5844 | NA |
| SVM | 0.5898 | 0.5847 | NA |
| LogisticRegression | 0.5896 | 0.5845 | 0.675 |

# Discussion

Valuable insight on how most of the collisions occur on the following conditions, though non-inclusive:

- Clear Weather
- Dry Road
- Daylight

Analysing the time of the incident would tell us that there is less collisions during weekend and surprisingly, there is an upward trend on collisions on weekdays.

Looking at the address/location of the collision, there are more Severity Code 1 collisions along Blocks than on Intersections. Collisions of Severity Code 2 registered almost the same count on both Blocks and Intersections.

With these observations, we cannot overemphasize the importance of safety and vigilance even at the most ideal driving situations, encouraging local governments in ensuring that protection and order on roads are in place.

# Conclusion

We were able to show how accident can be predicted by using collected and available data on collision.

Although our analysis has given us some good insights, the accuracy of our models is not that optimum.

Perhaps this can be improved by considering other features for analysis.