COURSERA FINAL REPORT

COLLISION SEVERITY SUMMARY OF DATA

INTRODUCTION

According to online sources, the USA experiences an average of approximately 102 vehicle crashes per day. From approximately 6 million crashes a year, over 35,000 crashes are fatal. To improve the country's road safety, it is important to understand the causes of accidents as well as the conditions leading to the incident.

The data will be analyzed mathematically and graphically through data science methodologies such as data visualization tools and machine learning models in order to provide insight on traffic collisions in Seattle.

DATA ANALYSIS

ROAD CONDITIONS, **LIGHTING CONDITIONS**, AND **WEATHER CONDITIONS** will play the main factors for determinig the relationship with the **SEVERITY** of collision.

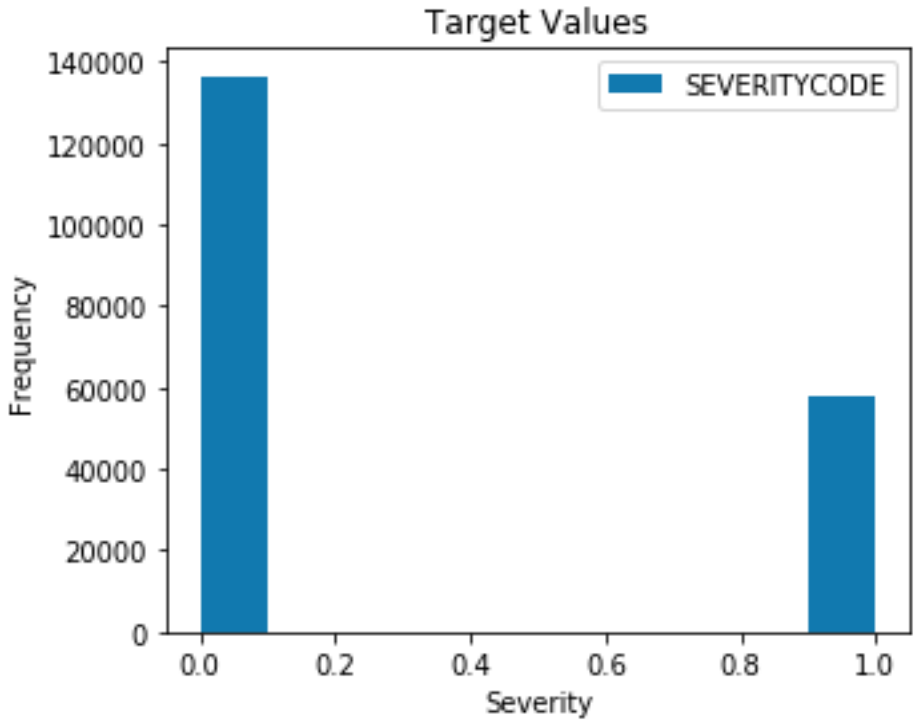
	SEVERITYCODE	WEATHER	ROADCOND	LIGHTCOND
0	2	Overcast	Wet	Daylight
1	1	Raining	Wet	Dark - Street Lights On
2	1	Overcast	Dry	Daylight
3	1	Clear	Dry	Daylight
4	2	Raining	Wet	Daylight
5	1	Clear	Dry	Daylight
6	1	Raining	Wet	Daylight
7	2	Clear	Dry	Daylight
8	1	Clear	Dry	Daylight
9	2	Clear	Dry	Daylight

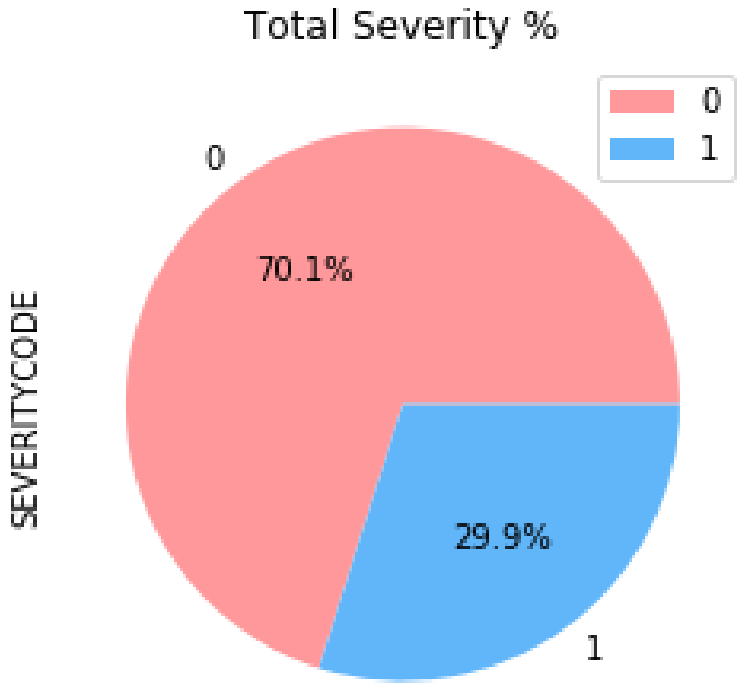
```
#Check Data
df_data_1.isna().sum()
```

```
SEVERITYCODE    0
WEATHER         0
ROADCOND        0
LIGHTCOND       0
dtype: int64
```

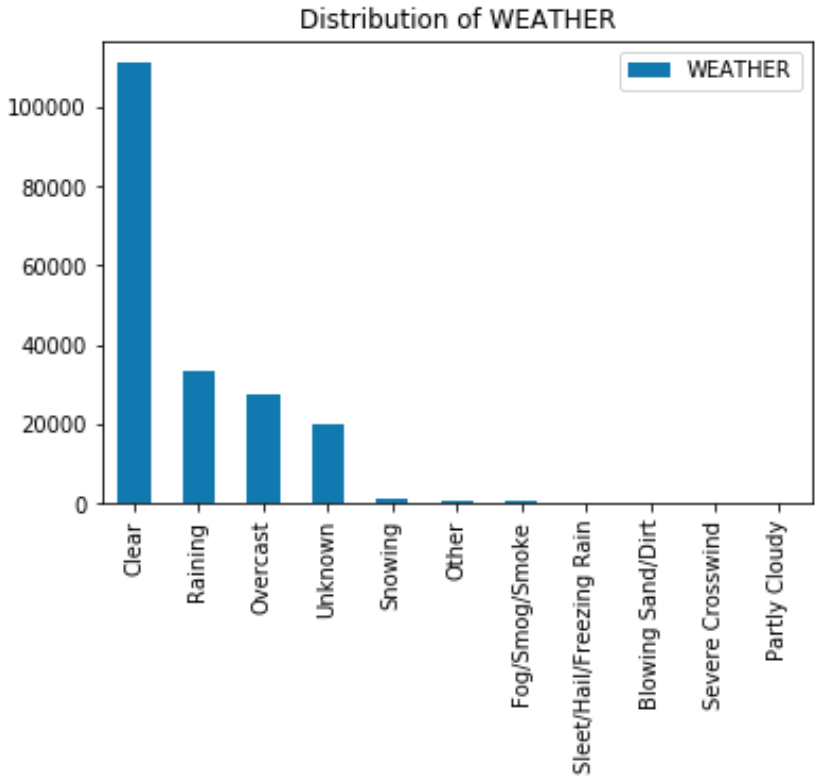
Cont.

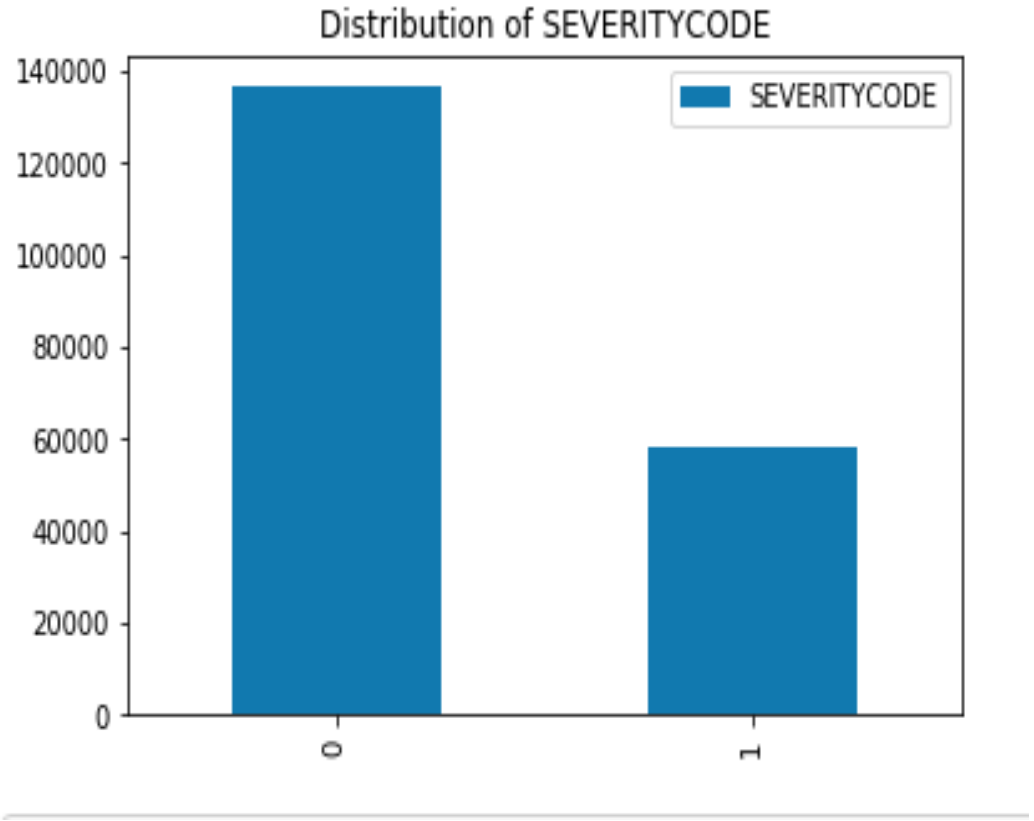
```
Accidents in each severity level:
```



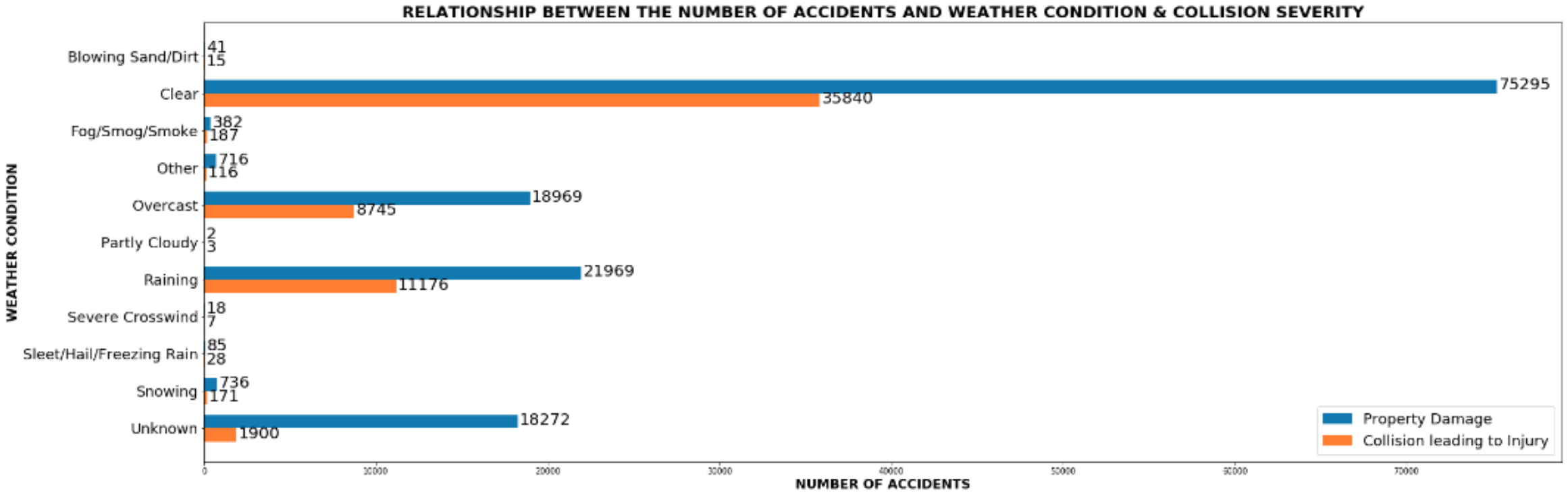


Cont.

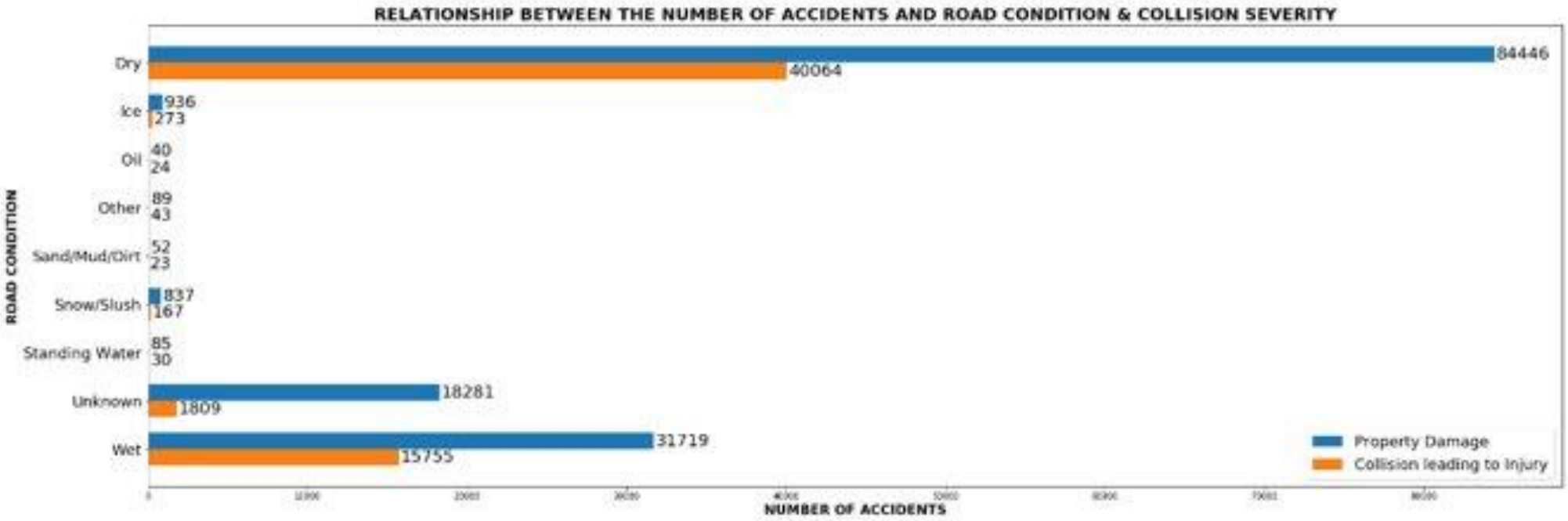

Distribution of WEATHER
WEATHER
100000
80000
60000
40000
20000
0
Clear
Raining
Overcast
Unknown
Snowing
Other
Fog/Smog/Smoke
Sleet/Hail/Freezing Rain
Blowing Sand/Dirt
Severe Crosswind
Partly Cloudy


Distribution of SEVERITYCODE
SEVERITYCODE
140000
120000
100000
80000
60000
40000
20000
0
0
1

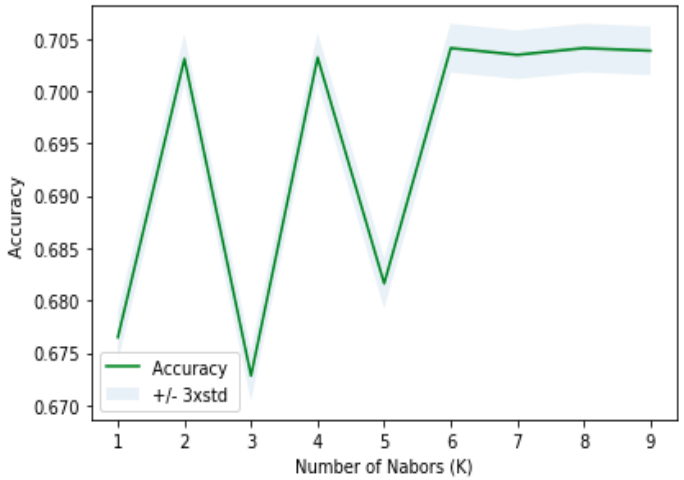
RELATIONSHIP BETWEEN THE NUMBER OF ACCIDENTS AND WEATHER CONDITION & COLLISION SEVERITY



RELATIONSHIP BETWEEN THE NUMBER OF ACCIDENTS AND WEATHER CONDITION & COLLISION SEVERITY



KNN MODEL

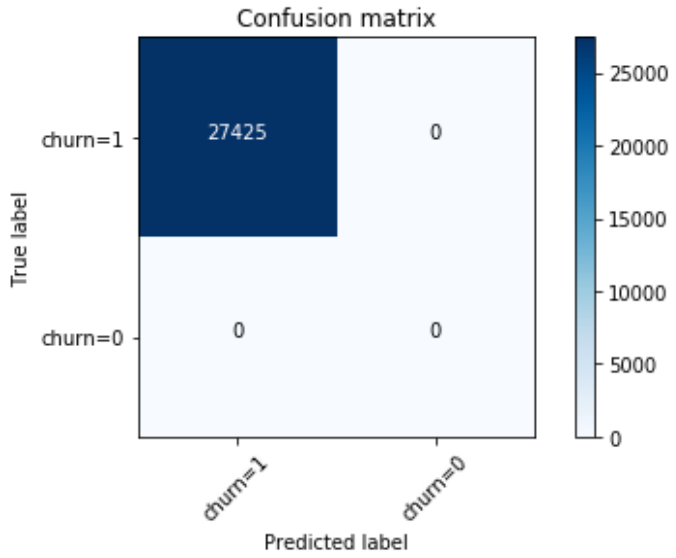


In [69]:
```
from sklearn.metrics import jaccard_similarity_score
from sklearn.metrics import f1_score

KnnJaccard = jaccard_similarity_score(y_test, yhat)
KnnF1 = f1_score(y_test, yhat, average='weighted')
print("K-Nearest Neighbors Average F1-score: %.7f" % KnnF1 )
print("K-Nearest Neighbors Jaccard Score: %.7f" % KnnJaccard )
```

```
K-Nearest Neighbors Average F1-score: 0.5822889
K-Nearest Neighbors Jaccard Score: 0.7043534
```

LOGISTICAL REGRESSION MODEL- CONFUSION MATRIX AND LOG-LOSS




```
n [54]: from sklearn.metrics import log_loss
        log_loss(y_test, yhat_prob)

ut[54]: 0.6068093485491171
```

DECISION TREE ACCURACY

In [70]:
```
###Decision Tree
```

In [72]:
```
from sklearn.tree import DecisionTreeClassifier
dtc = DecisionTreeClassifier(criterion="entropy", max_depth = 9)
dtc.fit(X_train, y_train)
```

Out[72]:
```
DecisionTreeClassifier(class_weight=None, criterion='entropy', max_depth=9,
            max_features=None, max_leaf_nodes=None,
            min_impurity_decrease=0.0, min_impurity_split=None,
            min_samples_leaf=1, min_samples_split=2,
            min_weight_fraction_leaf=0.0, presort=False, random_state=None,
            splitter='best')
```

In [73]:
```
yhat_dtc = dtc.predict(X_test)
```

In [75]:
```
# Accuracy:
len(yhat_dtc[yhat_dtc==y_test])/len(y_test)
```

Out[75]:
```
0.7043534095287017
```

RESULTS & CONCLUSIONS

1- Based on all the results generated, the Decision tree model displayed an accuracy of 0.70435, while the KNN gave off an Average F-1 score of 0.58 and a Jaccard score of 0.70435 and logistical regression also gave the same Jaccard score, similar to the Decision tree's accuracy and that of the KNN.

2- The graphs showed an important relationship between the number of accidents and road conditions and severity of collision. On 'DRY' road conditions, there was an estimate of 84,446 collisions that lead to property damage and 40,064 that lead to injuries; the 'DRY' road served the highest number of accidents. Taking a look at the weather conditions, 'CLEAR' weather resulted in the highest number of accidents with 75,295 leading to property damage and 35,840 leading to injuries.

3- Based on all the collisions that took place, 70.1% of them lead to property damage while 29.9% lead to injuries.

4- To conclude, the results give beneficial insights on the severity of vehicle collisions with focus on the different factors such as weather, road and light conditions. In terms of modeled results and values, they can be improved with more accurate and precise input data