

Car accident severity data analysis of Seattle

Martynas Brazauskas

A description of the problem and a discussion of the background

- Serious situation with COVID-19 in the world should change travel habits. More people will start to avoid public transport, because of possibility to get sick. So some of them will start to travel to work on feet or by bike.
- Another reason, which will eventually change travel habits is the fact that the world is getting greener. People understand that they has to do something to reduce co2 emmission. Of course some of them is starting to use electric car, however, some people choose to travel on foot or by bike.
- I will use Collisions data from Seattle city to generate a few most worrying neighborhoods based on my criteria.

A description of the data and how it will be used to solve the problem.

Based on definition of my problem, factors that will influence our decision are:

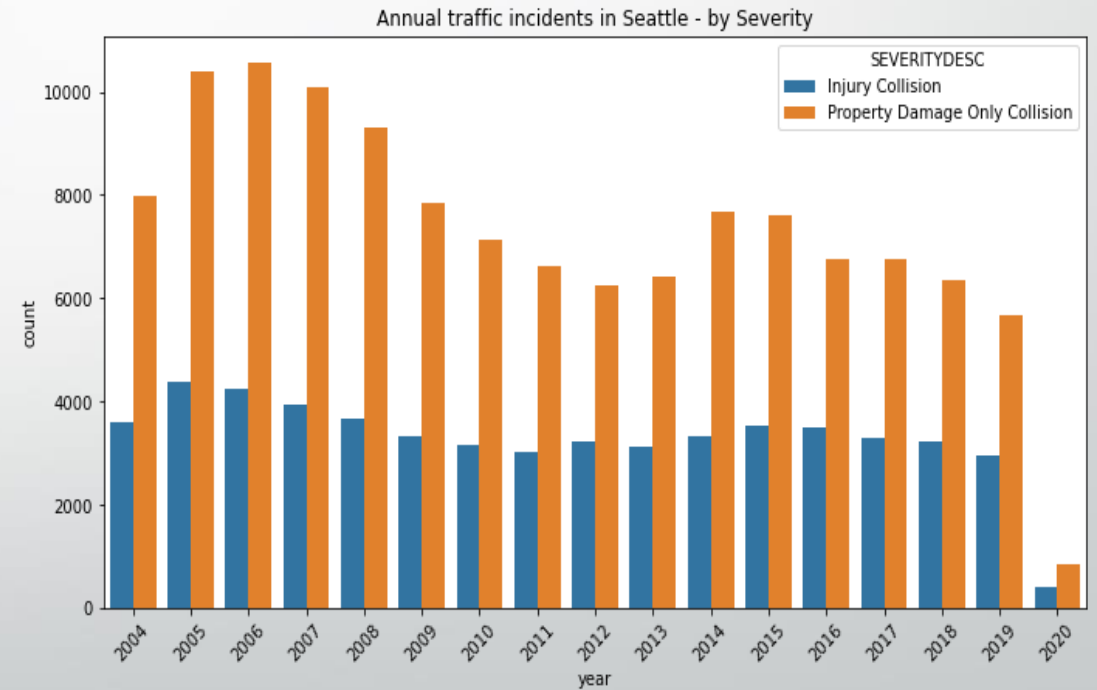
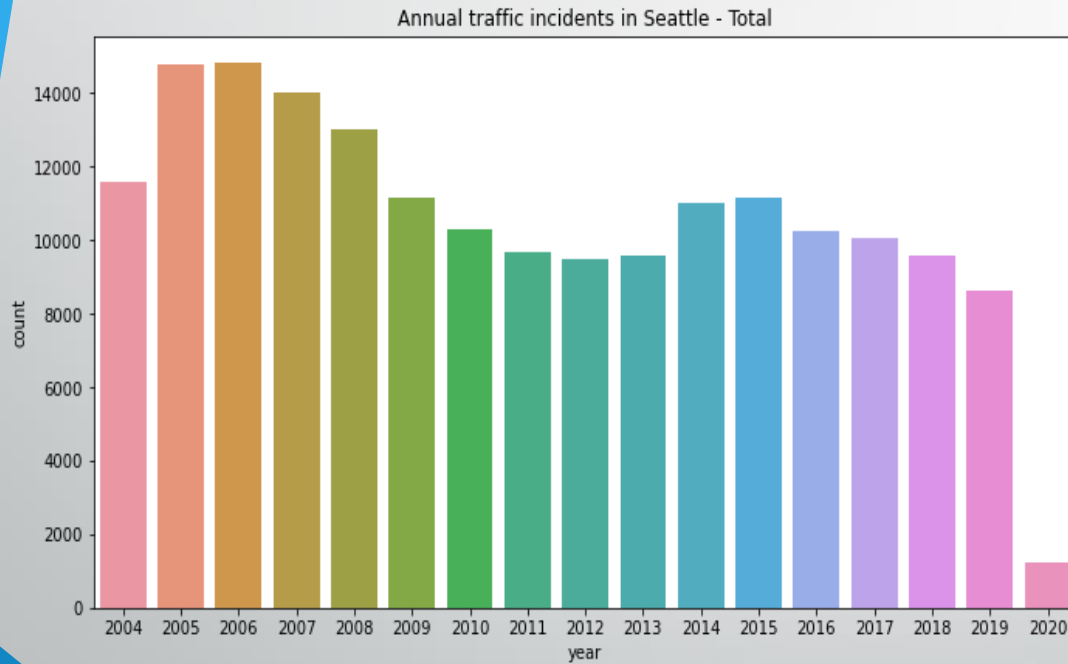
- number of collisions, which includes pedestrians and bicycles.
- additional factors may have contributed to the collisions (weather, road condition, light condition and etc.)
- seriousness of collisions.

I decided to use various Machine Learning methods to analyze data and find the impact to collisions associated with pedestrians and bicycles.

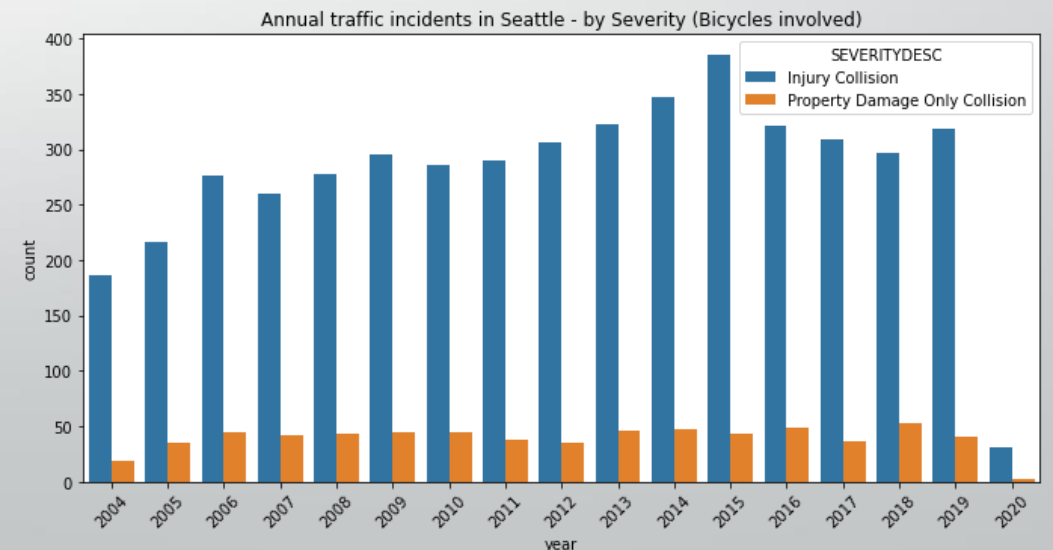
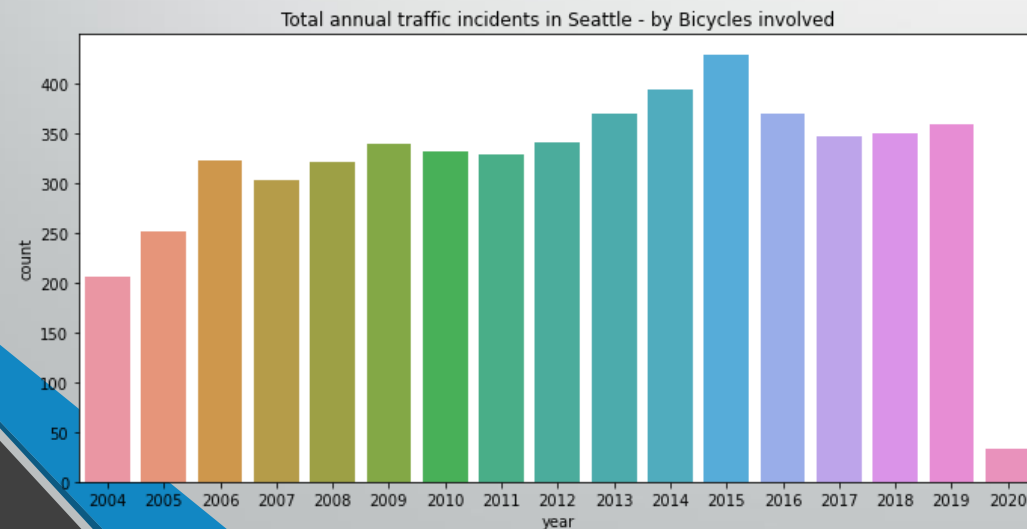
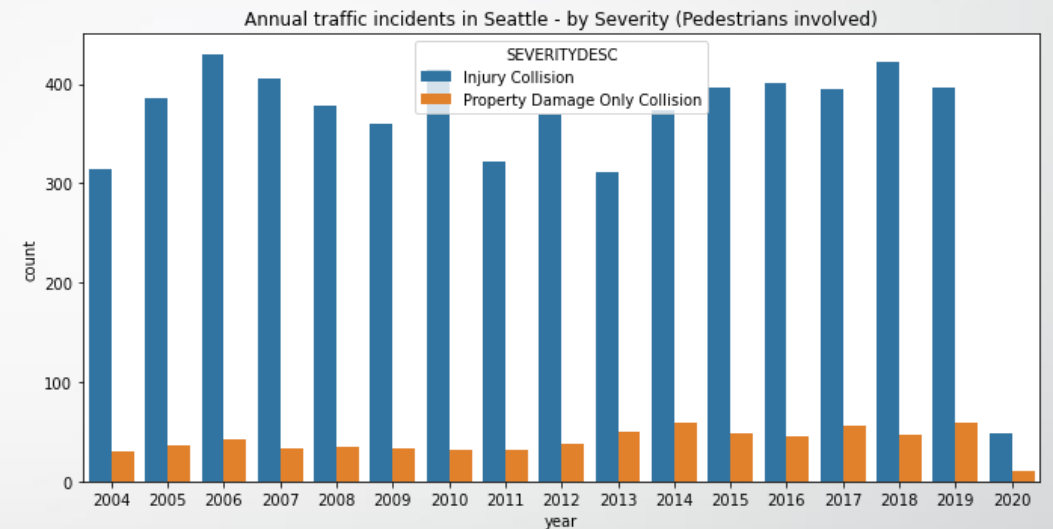
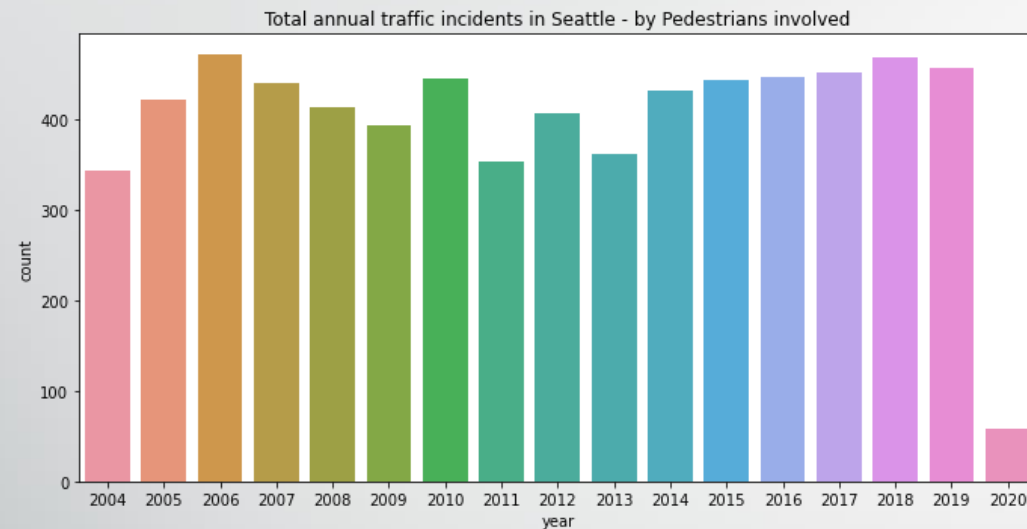
Following data sources will be needed to extract/generate the required information:

- ArcGIS Metadata Form of all year collisions data set from Seattle city.

Every year annual number of accident is decreasing.

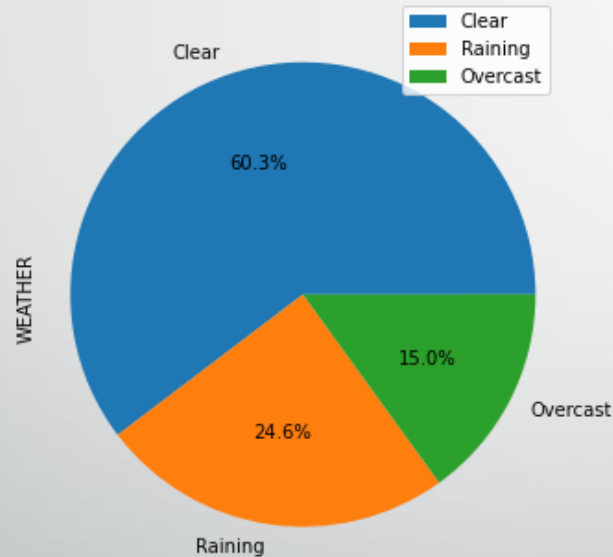


Pedestrians involving collision isn't decreasing and bicycles involving collision is increasing.

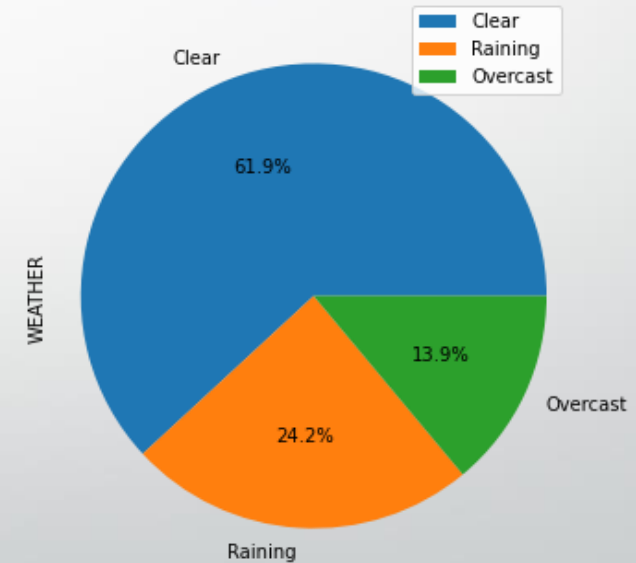


It can be said, that weather wasn't main reason of seriousness of collisions, which involves pedestrians and cyclist.

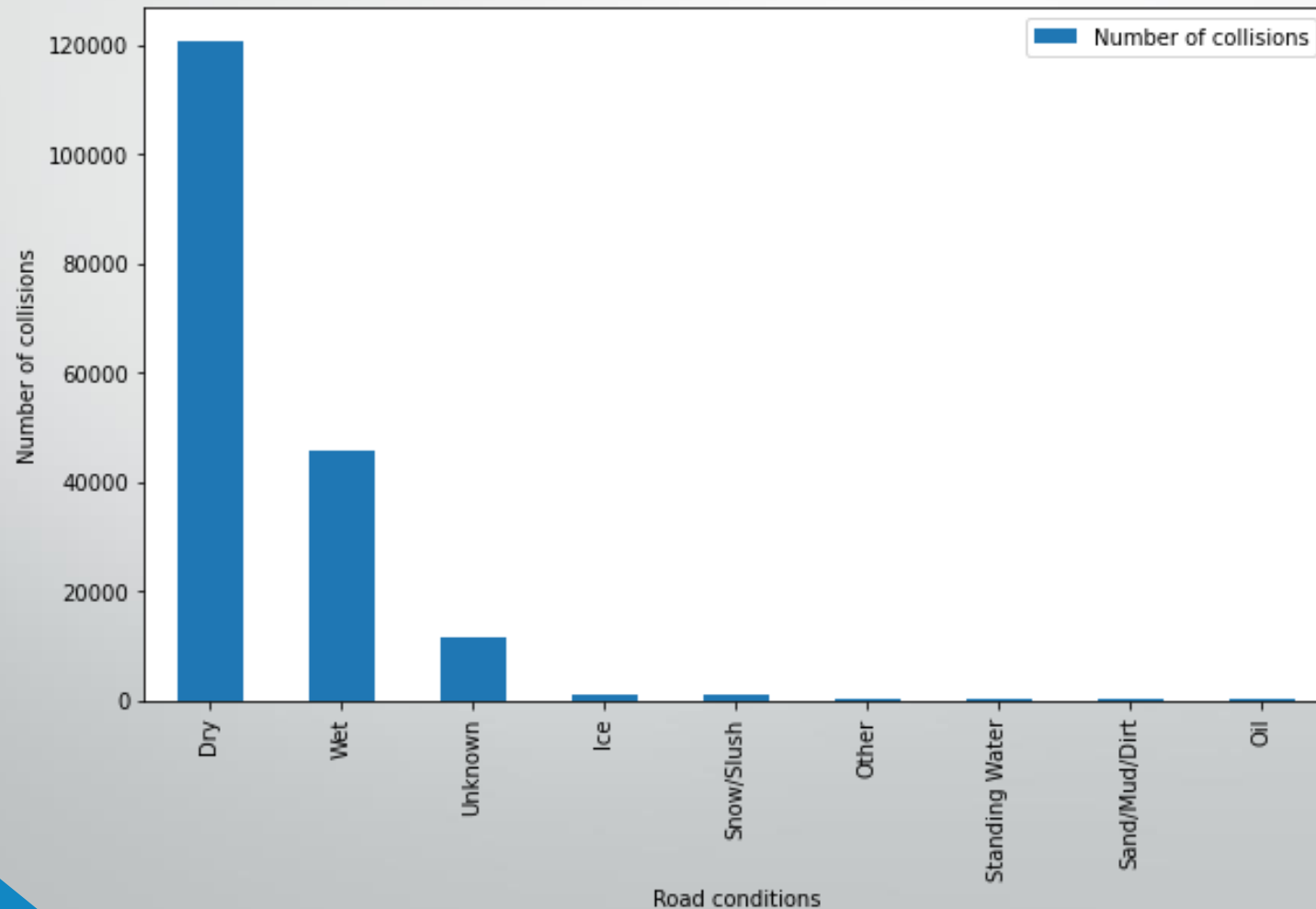
Collisions by weather, when at least 1 pedestrian is involved



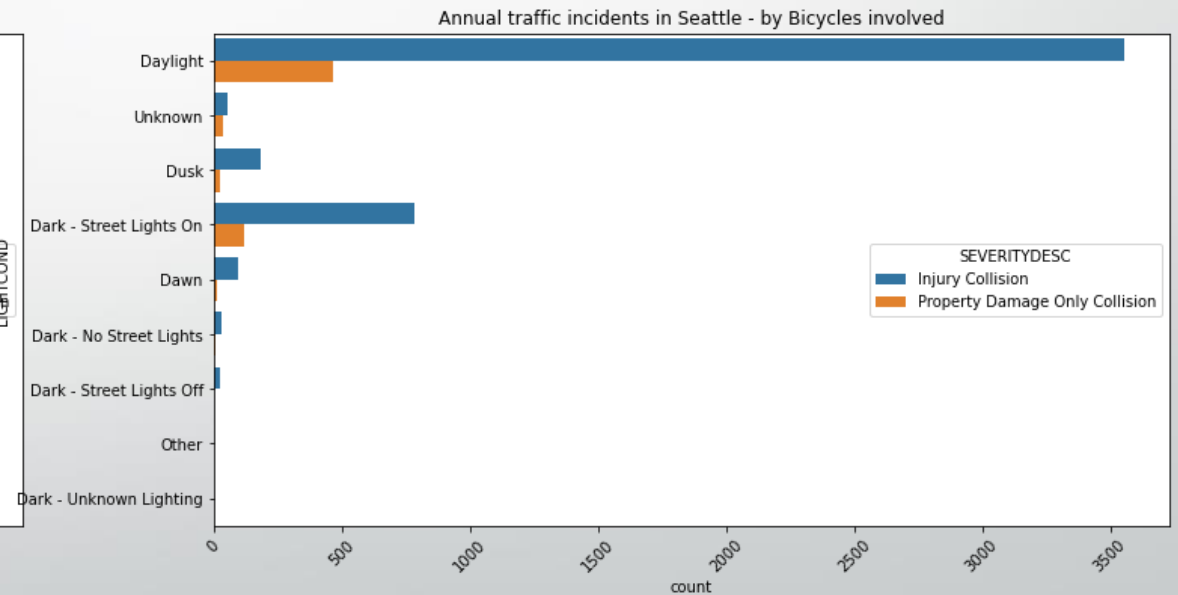
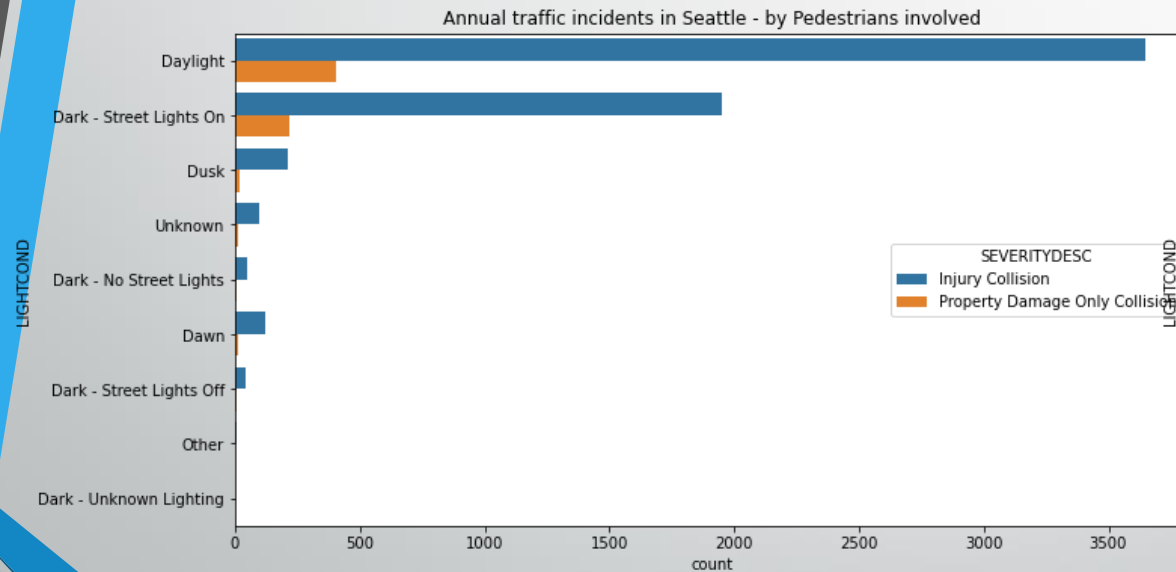
Collisions by weather, when more than 1 pedestrians is involved



Road conditions wasn't main reason of seriousness of collisions.



Most of accidents happens in daylight and when dark-street lights on.



Methodology

- WEATHER, ROADCOND, LIGHTCOND, INATTENTIONIND, SPEEDING and location (X,Y) was used as attributes to classify SEVERITYCODE.
- KNN, SVM, Logistic Regression and Decision Tree machine learning algorithms was used for data analysis.

Results

Algorithm	Jaccard	F-1 Score
KNN	0.652913	0.616497
DT	0.634698	0.619784
SVM	0.688232	0.561136
LR	0.688232	0.561136

- Two models (Logistic regression, Support Vector Machine) performed very similar.
- K-Nearest Neighbours stood out with a slightly higher F-1 score, but lower Jaccard index.
- SVM and Logistic regression were able to meet 68,8% accuracy (measured by Jaccard index). But these two algorithms are characterized by lowest accuracy in estimating it by F-1 score.
- According to average of accuracy measures, best performance was reached with KNN algorithm. This algorithm will be used to test pedestrians collision data and cyclist collision data.

Results and discussion

KNN	Jaccard	F-1 Score
Pedestrians	0.895007	0.845420
Cyclist	0.871058	0.816186

- According to this research, created model was even better for classification of collisions severity, when pedestrians and cyclist is included. I have reached almost 90 % accuracy for pedestrians involved collisions and 87 % accuracy for bicycles involved collisions. These results are important for the safety of pedestrians and cyclists in big cities.

Conclusion

- As a result, more often people are using public transport, bicycles and travel on foot in big cities. For this reason, we need better understanding of collisions, which includes pedestrians and cyclist.
- Historically, data shows that total situation of collisions is getting better. But situation of collisions when pedestrians or bicycles is involved doesn't change over the year.
- Research results shows that weather, road condition, light condition, inattention, speeding and location factors is very important to classify collisions severity.
- Understandably, more detailed analysis and more detailed data are needed to achieve better results.