Seattle Car Accidents

PREDICTING SEVERITY INJURIES

PASCHALIDIS DIMITRIS

Business Understanding

Consequences of car accidents:

- there may be people with injuries or fatalities;
- displacement of police vehicles;
- displacement of rescue vehicles;
- congestion in the area;
- delays for car and bus users.

Business Understanding

Most interested of accident prevention / reduction:

- Seattle Department of Transportation

Data Understanding/Preparation

Data Wrangling:

- identify duplicate columns;
- eliminate unnecessary columns;
- verification and cleaning of NaN data;
- conversion of categorical features to numerical values;
- balance de values;
- normalize values;
- etc.

Results

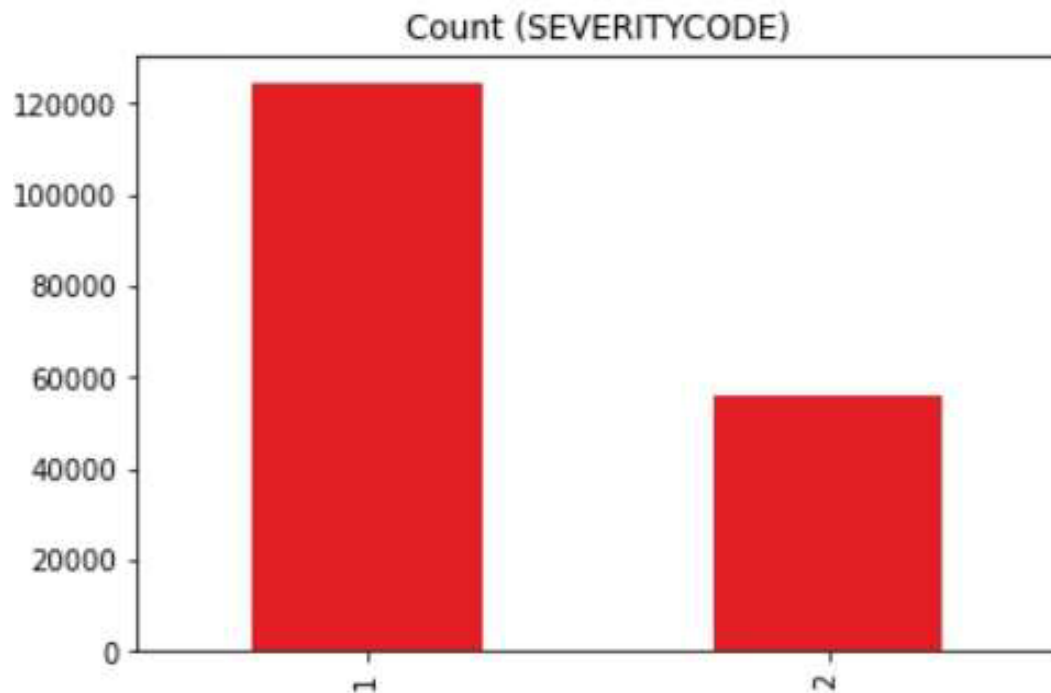Possibility of concentration of incidents on specific area. 2017
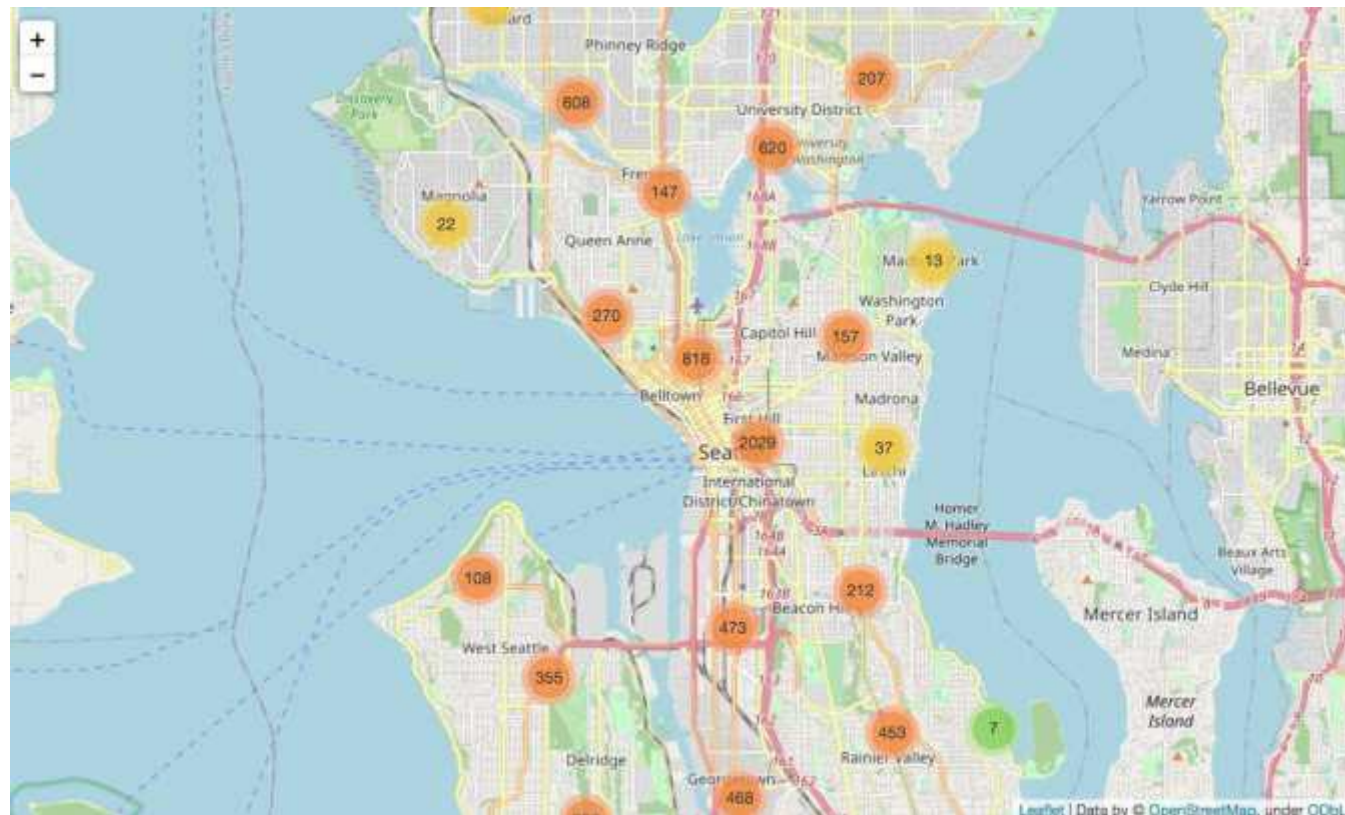
Results
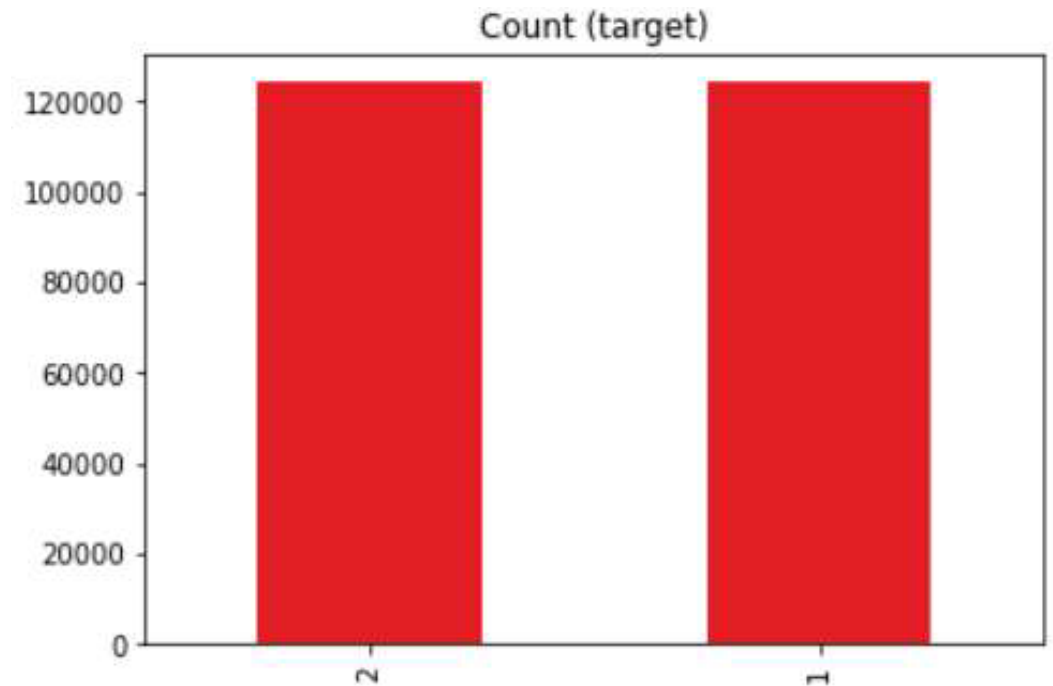
PossIbility of concentration of incidents on specific area. 2018
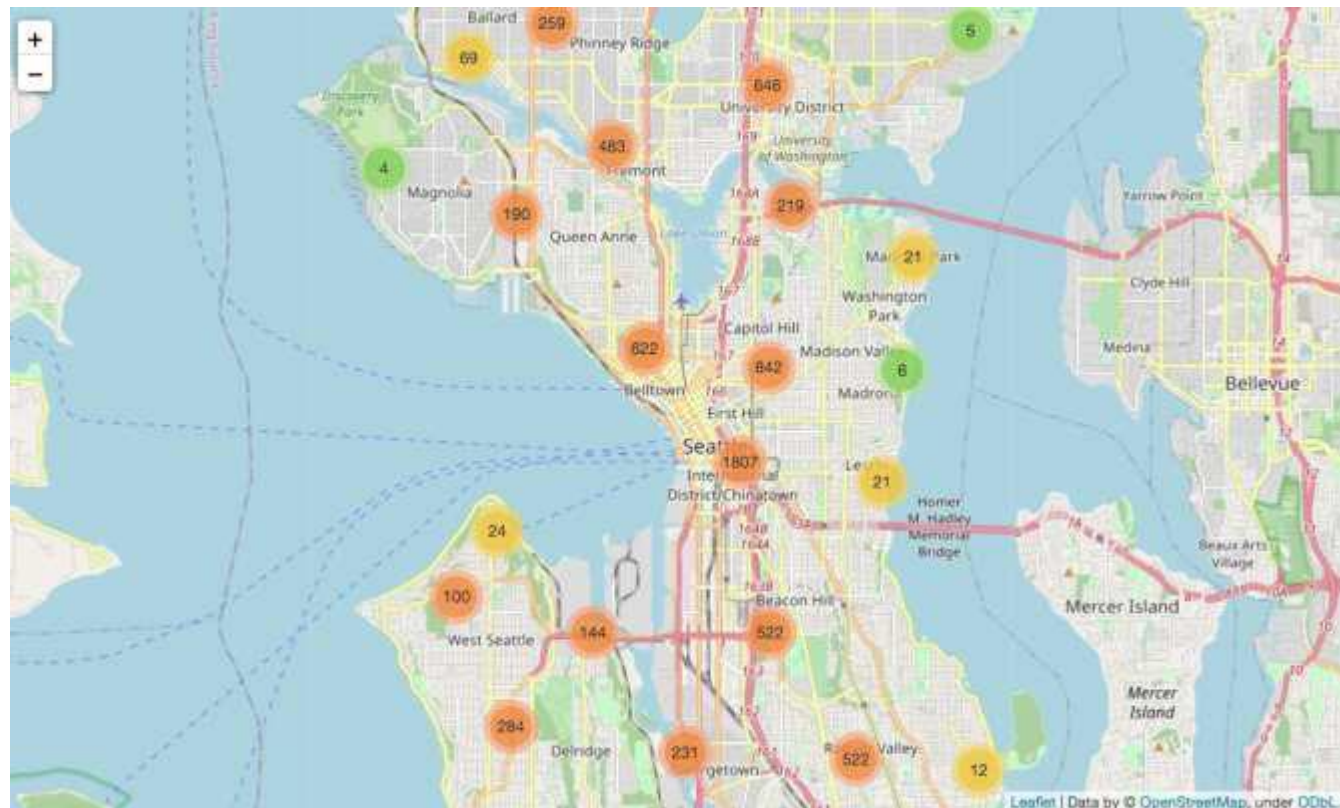
Methodology

Balance of the Data:

Before normalization

After normalization

Count (SEVERITYCODE)

Count (target)

Prediction results

Decision Tree

```
In [61]:  from sklearn.metrics import jaccard_score
          from sklearn.metrics import f1_score
          from sklearn.metrics import log_loss
```

```
In [62]:  from sklearn.tree import DecisionTreeClassifier
          DT_model = DecisionTreeClassifier(criterion="entropy", max_depth = 4)
          DT_model.fit(X_train,y_train)
          DT_model
```

```
Out[62]:  DecisionTreeClassifier(criterion='entropy', max_depth=4)
```

```
In [63]:  yhat = DT_model.predict(X_test)
          yhat
```

```
Out[63]:  array([2, 2, 1, ..., 1, 1, 2])
```

```
In [64]:  #DT_yhat = DT_model.predict(X_test)
          print("DT Jaccard index: %.2f" % jaccard_score(y_test, yhat))
          print("DT F1-score: %.2f" % f1_score(y_test, yhat, average='weighted') )

          DT Jaccard index: 1.00
          DT F1-score: 1.00
```

Conclusion

▶ prediction with an accuracy index close to 100%;

▶ possibility to extract more insights from the dataset;

▶ Data Scientist:

    ▶ vast field to be explored;

    ▶ more studies for professional improvement.