CapstoneProject_CarAccidentSeverity
Final Submission

By

Sri Soumya Gudla

Business and Integration Arch. Analyst

Accenture.

Contents

CapstoneProject_CarAccidentSeverity Final Submission 1

1. BUSINESS PROBLEM 1
2. DATASET 1
3. DATA VISUALISATION 2
4. METHODOLOGY 4
5. RESULTS 4
6. CONCLUSION 5
7. REFERENCE 5

1. BUSINESS PROBLEM

Considering the traffic conditions in the Seattle city let's consider a business model like installing accident prone areas signboard at the areas where more severe accidents occur. To predict the top 10 vulnerable places it is required to collect the accident data. The dataset is obtained from SDOT traffic management system available from 2004 to present. In order to predict the most vulnerable places to accident the following factors are considered i.e. time of occurrence, weather, fatalities, traffic delay, property damage etc. So, considering these factors the data has to be segregated and cleaned for predicting the accident severity at those 10 vulnerable places.

2. DATASET

This dataset is about past accident data. The Data-Collisions.csv data set includes details of 194673 accidents from 2004 to 2020. It has a total of 38 fields. The following fields are as below

- OBJECTID:- ESRI unique identifier
- SHAPE:- ESRI geometry field
- INCKEY:- unique key for the incident
- COLDETKEY:-Secondary key for the incident
- ADDRTYPE:- Collision address type: Alley\Block\Intersection
- INTKEY :- Key that corresponds to the intersection associated with a collision
- LOCATION:-Description of the general location of the collision
- EXCEPTRSNCODE
- EXCEPTRSNDESC
- SEVERITYCODE:- A code that corresponds to the severity of the collision(3—fatality,2b—serious injury,2—injury,1—prop damage,0—unknown)
- SEVERITYDESC:-A detailed description of the severity of the collision
- COLLISIONTYPE:- Collision type
- PERSONCOUNT :- The total number of people involved in the collision
- PEDCOUNT: - The number of pedestrians involved in the collision. This is entered by the state.
- PEDCYLCOUNT: - The number of bicycles involved in the collision. This is entered by the state.
- VEHCOUNT: - The number of vehicles involved in the collision. This is entered by the state.
- INJURIES: - The number of total injuries in the collision. This is entered by the state.
- SERIOUSINJURIES: - The number of serious injuries in the collision. This is entered by the state.
- FATALITIES: - The number of fatalities in the collision. This is entered by the state.
- INCDATE: - The date of the incident.
- INCDTTM: - The date and time of the incident.
- JUNCTIONTYPE:- Category of junction at which collision took place
- SDOT_COLCODE:- A code given to the collision by SDOT.

- SDOT_COLDESC:- A description of the collision corresponding to the collision code.
- INATTENTIONIND:-Whether or not collision was due to inattention. (Y/N)
- UNDERINFL: - Whether or not a driver involved was under the influence of drugs or alcohol.
- WEATHER: - A description of the weather conditions during the time of the collision.
- ROADCOND: - The condition of the road during the collision.
- LIGHTCOND: - The light conditions during the collision.
- PEDROWNOTGRNT: - Whether or not the pedestrian right of way was not granted. (Y/N)
- SDOTCOLNUM: - A number given to the collision by SDOT.
- SPEEDING: - Whether or not speeding was a factor in the collision. (Y/N)
- ST_COLCODE:- A code provided by the state that describes the collision. For more information about these codes, please see the State Collision Code Dictionary.
- ST_COLDESC:- A description that corresponds to the state's coding designation.
- SEGLANEKEY: - A key for the lane segment in which the collision occurred.
- CROSSWALKKEY: - A key for the crosswalk at which the collision occurred.
- HITPARKEDCAR: - Whether or not the collision involved hitting a parked car. (Y/N)

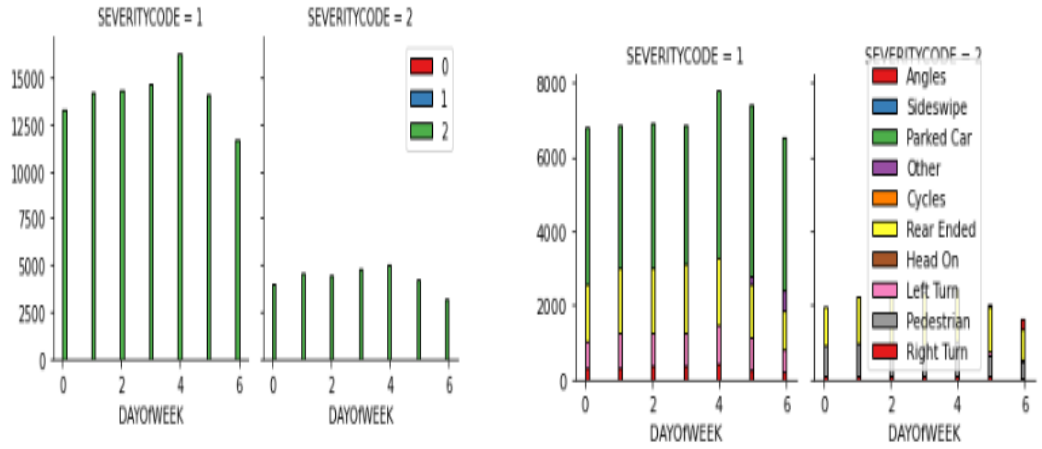
So the data has to be prepared before building a model. Data Wrangling is the process of converting data from the initial format to a format that may be better for analysis. The steps followed are as shown below.
* Identifying and handling missing values
* Data Evaluation
* Data Standardization

Once data is cleaned it can be used to visualize and built the model for the business problem stated above.

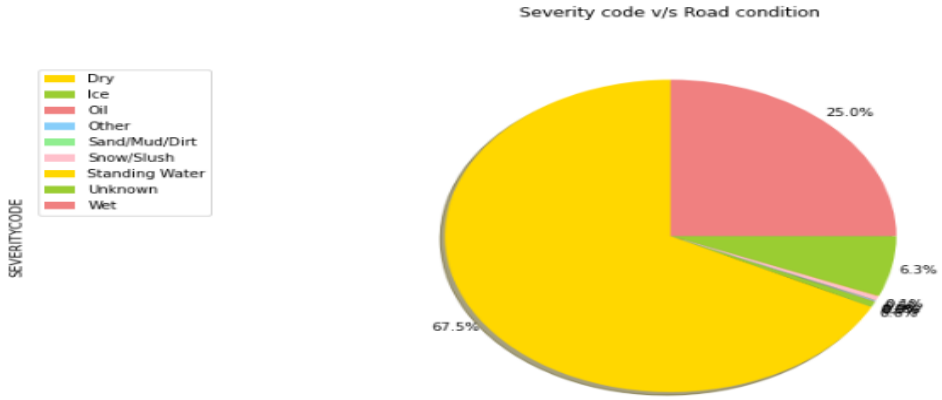
3. DATA VISUALISATION

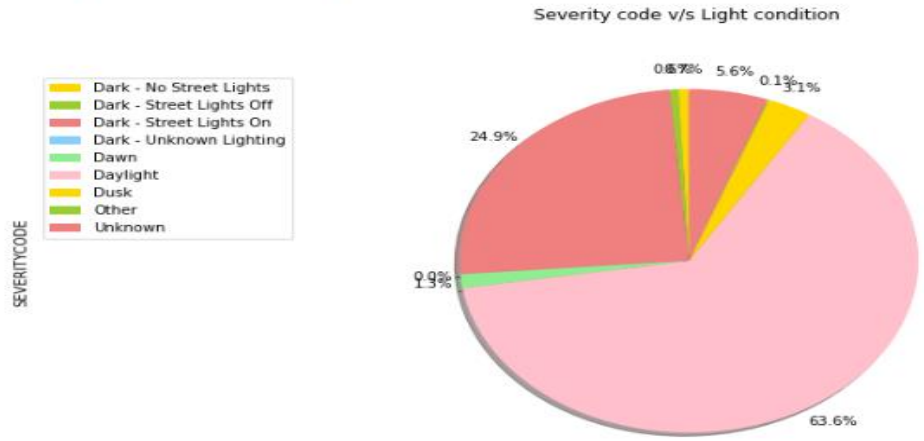
To identify top 10 accident prone locations i am considering location field as a dependent variable for rest of the fields. So in order to observe the pattern of these occurrences am considering Histogram plot to observe the frequency of accidents occurred for each day of the week, most frequent type of collisions and ADDRType category based on severity code. While pie chart is used for the visualizing the weather, road condition and light conditions at the time of accident based on severity code.

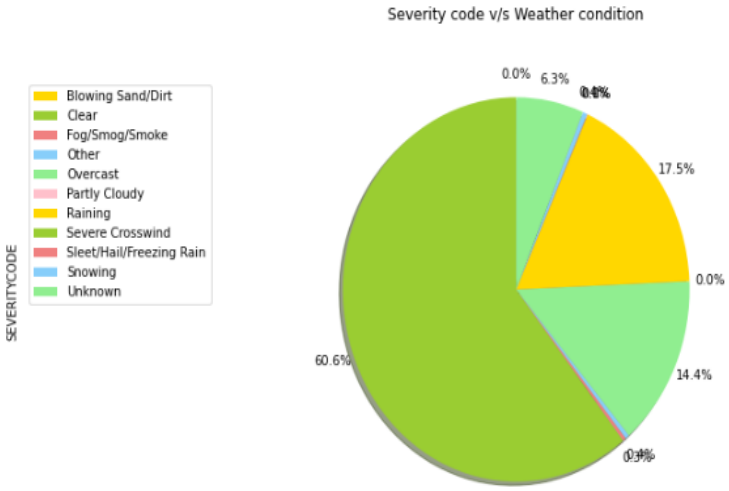
Plot to identify the max.number of accidents occurred at which day of the week for ADDRTPE and Collision Type



Pie Charts to observe the weather, road and light conditions at the time of accidents based on severity code







4. METHODOLOGY

So to determine the top 10 accident prone areas, the major factors influcing the accidents are Day of the week, weather, light condition, road condition and address type. Based on these conditions the top most accident prone locations can be identified. Hence I used KNN and Decision tree methods to test and train data and evaluate the model. Since the accuracy of the data model is high when we use these methods

In order to prepare the data for the above mentioned data we need to convert the categorical data into numerical data

Then divide the data to train and test data, then execute the methods

We will then evaluate the model by calculating f1 score, jacquard index

5. RESULTS

After running the KNN and DT methods in python the following results are obtained as data evaluation results.

Methods	F1 score	Jacquard Index
KNN	0.0060	0.0091
DT	0.00072	0.0157

6. CONCLUSION

Hence we will be able to identify the locations using the correct data fit model.

7. REFERENCE

- https://github.com/srisoumya02/Coursera_Capstone/blob/master/Coursera_Capstone/CapstoneProject_CARAccidentSeverity.pptx
- https://github.com/srisoumya02/Coursera_Capstone/blob/master/Coursera_Capstone/CapstoneProject_CarAccidentAssignment%20(1).ipynb