IDENTIFYING WAYS OF REDUCING SEVERE ACCIDENTS IN SEATTLE

Coursera Capstone Project

Diderico van Eyl

1. **Introduction**

The problem being studied for this project is to explore ways that may be able to reduce the number of severe accidents in the Seattle area. By observing historical data relating to accidents, features of accidents can be observed, future accidents can be predicted and by observing predicted future accidents, policies can be implemented to reduce the number of severe accidents. This problem would be of interest to insurance companies and government agencies authorized to impact laws of drivers in Seattle.

As further described in the description of the data available for this project, there is meaningful historical data of accidents that have occurred in the Seattle area and that provides a meaningful amount of experience that can be used to perform tasks that identify correlations between accidents and accidents’ respective features. Using machine learning techniques such as linear regression or other approaches, a model can be developed and implemented with a performance measure such as the prediction of accidents. More particularly, the severity of an accident and features such as the collision type, weather conditions, location of accident, road conditions can be analyzed to determine how such features may affect future accidents. Although the data is limited to Seattle, findings from this study may be used to provide general guidelines for other cities as well.

1. **Data**

The dataset being used includes the dataset that is provided by Coursera and was downloaded from this link: <https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv>. Metadata for the dataset can be found here: <https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Metadata.pdf>.

The data has 37 attributes (columns) of numerous accidents and 194, 673 rows. A preliminary review of the data revelas that the severity of the accidents have been categorized in two categories: Severity of “1” and Severity of “2” where the level 2 accidents are the most severe. Accidents involving Severity of 1 include those that are “Property Damage Only Collisions” and accidents involving a Severity of 2 include those that are “Injury Collision.” A preliminary review of the data also reveals that “Level 1, Property Damage Only Collisions” reveals that there are the following subcategories with the indicated number of accidents: blank: 3863, Angles: 21050, Cycles: 671, Head On: 1152, Left Turn: 8292, Other: 17591, Parked Car: 45325, Pedestrian: 672, Rear Ended: 19419, Right Turn: 2347, Sideswipe: 16103. Level 2 (Injury Collision) accidents include the following features with the indicated number of accidents: (blank) 2082, Angles: 27248, Cycles: 9488, Head On: 1744, Left Turn: 10822, Other: 12224, Parked Car: 5324, Pedestrian: 11872, Rear Ended: 29342, Right Turn: 1218, and Sideswipe: 5012

The dataset can be imported into a data frame into a Jupyter notebook, pandas and numpy libraries can be imported into the notebook and the features of the dataset can be identified using Python’s “head” and “shape” commands. Below are screenshots that show features of the data:

* A screenshot of a social media post

  Description automatically generated

Above-mentioned features of Severity Level 1 and 2 accidents can be visualized as follows:

Severity Level 1: Property Damage Only Collisions:

* Severity Level 2 Injury Collisions

1. Methodology

The exploratory data analysis that was performed, inferential statistical testing that was performed, and what machine learnings were used are discussed.

Exploratory Data Analysis Performed:

The above-mentioned dataset was analyzed by various filters to better understand how different variables corresponded to accidents using different severity codes. The dataset was uploaded to Excel’s PowerQuery, PowerPivot and using pivot table analysis, the relationship between various attributes was observed. Also, the data was uploaded to a Jupyter notebook using Python code and various aspects of the dataset were evaluated.

1. Results

When filtering the data by collision types, it was observed that there were 58,188 Level 2 accidents and 136,485 accidents, as summarized below:

|  |  |  |  |
| --- | --- | --- | --- |
| **Collision**  **Type** | **Level 2 Severity Accidents** | **Level 1 Severity Accidents** | **Total** |
| Rear Ended | 14,671 | 19,419 | 34,090 |
| Angles | 13,624 | 21,050 | 34,674 |
| Other | 6,112 | 17,591 | 23,703 |
| Pedestrian | 5,936 | 672 | 6,608 |
| Left Turn | 5,411 | 8,292 | 13,703 |
| Cycles | 4,744 | 671 | 5,415 |
| Parked Car | 2,662 | 45,325 | 47,987 |
| Sideswipe | 2,506 | 16,103 | 18,609 |
| (blank) | 1,041 | 3,863 | 4,904 |
| Head On | 872 | 1,152 | 2,024 |
| Right Turn | 609 | 2,347 | 2,956 |
| Total | 58,188 | 136,485 | 194,673 |

Accidents by State Collision Codes

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| State Collision Code | State Collision Code Name | Level 1 Severity Accidents | Level 2 Severity  Accidents | % of Total Level 2 Accidents |
| 10 | Entering at an Angle | 21030 | 13624 | 23.41% |
| 14 | From Same Direction - Both Going Straight - One  Stopped - Rear End | 14333 | 11362 | 19.53% |
| 28 | From Opposite Direction - One Left Turn - One  Straight | 5682 | 4631 | 7.96% |
| 45 | Bicycle | 572 | 4117 | 7.08% |
| 50 | Struck Fixed Object | 10239 | 3290 | 5.65% |
| 13 | From Same Direction - Both Going Straight - Both  Moving - Rear End | 4556 | 3052 | 5.25% |
| 0 | Vehicle Going Straight Hits Pedestrian | 311 | 2565 | 4.41% |
| 32 | One Parked - One Moving | 41349 | 2143 | 3.68% |
| 2 | Vehicle Turning Left Hits Pedestrian | 180 | 2001 | 3.44% |
| Total |  | 98252 | 46785 | 80.40% |

Weather & Accidents

To observe how road conditions related to accidents, the data was filtered:

|  |  |  |
| --- | --- | --- |
| **Road Conditions** | **Level 1 Severity Accidents** | **Level 2 Severity Accidents** |
| Dry | 84328 | 40064 |
| Wet | 31689 | 15754 |
| (blank) | 4257 | 1061 |
| Unknown | 14061 | 749 |
| Ice | 936 | 273 |
| Snow/Slush | 834 | 167 |
| Other | 88 | 43 |
| Standing Water | 85 | 30 |
| Oil | 40 | 24 |
| Sand/Mud/Dirt | 52 | 23 |
| Unknow | 15 |  |
| Unkn | 13 |  |
| D | 12 |  |
| Sno | 1 |  |
| Unkno | 20 |  |
| U | 13 |  |
| Dr | 7 |  |
| W | 3 |  |
| We | 2 |  |
| Snow/ | 1 |  |
| Unk | 10 |  |
| Un | 18 |  |

Regression analysis was conducted to see if the severity of accidents could be predicted from the ST\_COLCODE. The following

A screenshot of a cell phone

Description automatically generated

The result suggests that the data is not suitable for linear regression analysis and for predicting the severity of accidents.

1. Discussion

Observations suggest that, based on the sample space, it was more likely to have less severe, as there were more Level 1 accidents than Level 2 accidents. It was also observed that there was a correlation between the collision type and the number of accidents.

Importantly, collisions arising from rear ended and angles, for instance, accounted for 28,295 Level 2 accidents, **or about 49% of total severe Level 2 accidents** Surprisingly, it was observed that most accidents occurred in dry conditions, as indicated below

1. Conclusion

* Rear-Ended & Angle Accident situations are the most prevalent Level 2 severity accidents.
* From a policy perspective, measures to reduce the number of rear-end and angle crashes should be explored
* Automakers should be informed of the results and **challenged to develop safety features that reduce the likelihood of rear-end and angle collisions.**