Capstone Project

Raj M Sundaram
30 Oct, 2020

Accident Analysis

1. Introduction

Background

The traffic accidents observed in City of Seattle since 2013 till 2019 by the Traffic Management Division collects variety of attributes directly associated with the accidents such as roady type, number of vehicles, pedestrian involved, speeding, cyclist, intersection type etc. It also captures data of geographic, natural and other conditions applicable around the site of accident which might have played a role in the cause of accident, such as location, weather, road condition, lighting condition etc. So, it would be useful if all direct and indirect causes of accidents can be analyzed in conjunction so as appropriate predictions could be made based on prevailing weather conditions along certain route in order to prevent drivers plying onto routes with higher probability of collisions.

Problem

The problem would enable developing Location based predictive model to assess the probability of delays based on aforementioned attributes and thus help traveller to avoid roads with high probability of delays due to untoward traffic incidents.

Data

The collision data for Seattle has been available through Coursera website for the period 2013-2019 in csv. The data has contains 194,673 observations and 38 variables, including the location of collision, accident related data, weather and lighting conditions.

Data Cleaning

Though the data contains significant attributes about each accidents, however, its been observed there have been unfilled attributes rendering such records unfit for analysis.

Since we would like to identify the factors that cause the accident and the level of severity, we will use SEVERITYCODE as our dependent variable.

Also, there were some observations where location data was missing, as such those records must be removed from the dataset.

2. Exploratory Data Analysis

2.1 Accident vs Road type and Condition

The road condition attribute contains weather related information prevalent at the time of accident. So, it can be hypothesized that the number of accidents should be high in adverse weather conditions at the locations where concentrations of traffic is high. And this is verified through pivoting the data as show below, which does not correlate with the hypothesis. Instead the number of accidents observed are highest during Dry weather conditions.

	SEVERITYCODE	
ADDRTYPE	**Block**	**Intersection**
ROADCOND		
Dry	48596	20880
Ice	674	118
Other	56	12
Sand/Mud/Dirt	25	8
Snow/Slush	559	118
Standing Water	46	12

| Unknown | 9826 | 1013 |
| Wet | 17263 | 8579 |

Further assessment of data indicates the are number of accidents observed are highest on the Fridays (Day no 4) while it is lowest on Saturdays.
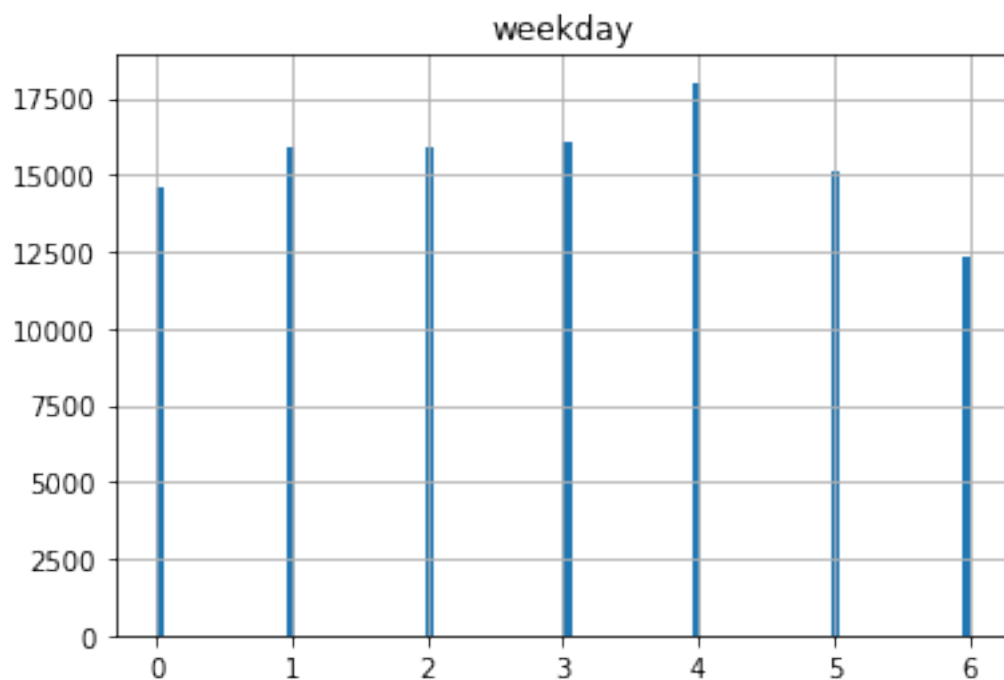

Consequently, it becomes imperative to dig deeper into the data to ascertain the factors resulting in high numbers of collisions on a fairly clement weather condition. So, a heat map has been generated for

each hour of the possible weather for each day of **day** on all condition the week.

The heat map that morning evening rush the working have depicts, and hours on days

unprecedented collisions, and particular the (Fridays, day 5) numbers is Though the wet (bottom part of has some impact on the but its significantly low dry conditions. levels of in weekend when the highest. weather heat map) significant collisions, compared to the
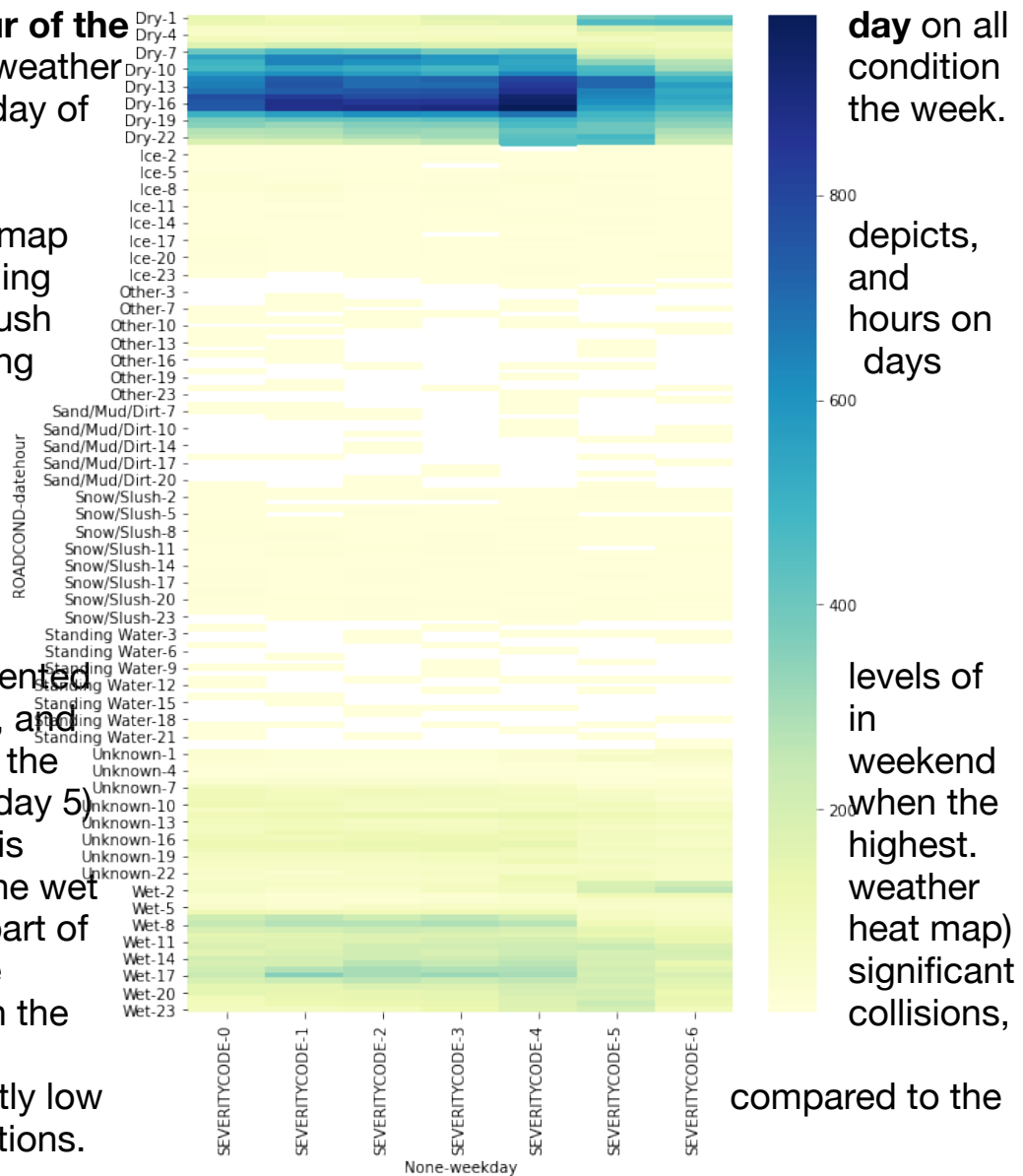
Conclusion

The analysis of accident data only suggests that probably weather has no significant role in the collisions. Instead, it appears to be human factor, due to high volume of vehicles wherein vehicles rushing to the workplace at high speeds could be causing more accidents at intersections by jumping traffic lights, colliding with vehicles stuck in traffic jams. As such, the predictive models, would enable to advice drivers to be cautions on certain intersections where huge number of vehicles are converging. In particular, drivers should be extra cautious on weekends when the rush to return home on the last working day is more exciting.