IBM Data Science Professional Certificate Capstone Project

Car Accident Severity

Raj Mukta Sundaram
Oct, 2020

Introduction

- The Capstone Project is the culmination of IBM Data Science Specialization course.

- This project is based on the analysis of Accident data provided through Coursera for Seattle.

- The traffic accidents observed in City of Seattle since 2013 till 2019.

- The Data contains 38 attributes pertaining to the accidents, some of major attributes include:

  - intersection type,

  - number of vehicles,

  - pedestrian involved,

  - speeding, cyclist,

- The data also captures data of geographic location, natural and other conditions applicable around the site such as - weather, road condition, lighting condition etc.

Business Problem

The problem entails developing Location based predictive model:

• to assess the probability of delays based on direct and indirect factors.
• to avoid roads with high probability of delays due to untoward traffic incidents.

The Data

- The collision data for Seattle has been available through Coursera.
- The data contains accidents for the period 2013-2019 in csv format.
- The data has contains 194,673 observations
- The data contain 38
- The direct factors associated with accidents include:
  - Location (x,y)
  - Date, time of incident
  - ADDRTYPE, INTKEY (Intersection type)
  - Person Count, Pedestrian Count, Vehicle Count
  - Speeding
  - Crosswalk
- The indirect factors associated with accidents include:
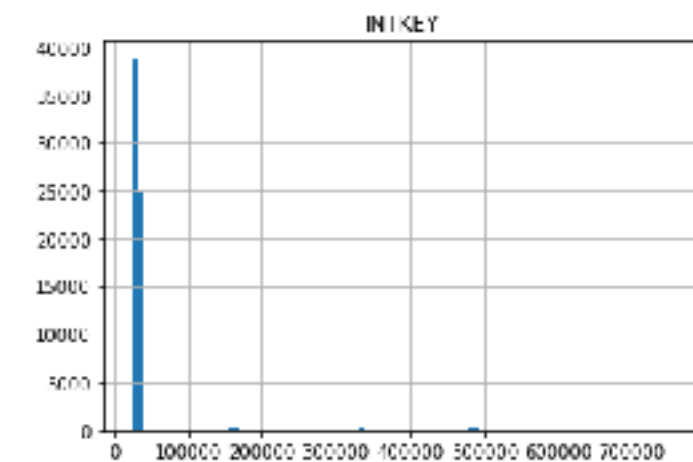 - Road Condition
  - Light Condition

Preprocessing

- From the Analysis of the data it has been there are several records unfit for analysis, due to the following reasons:

  - Date, Time information inaccurate / missing
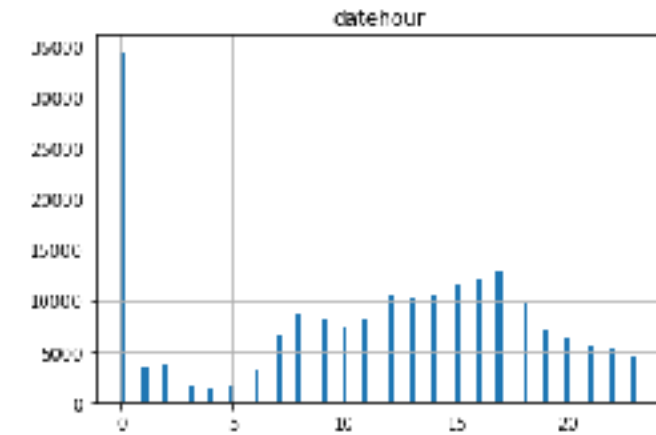
  - Intersection Key (location data missing)

- Data with null values were removed.

Exploratory Analysis

- Accident vs Roadtype and Road 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.
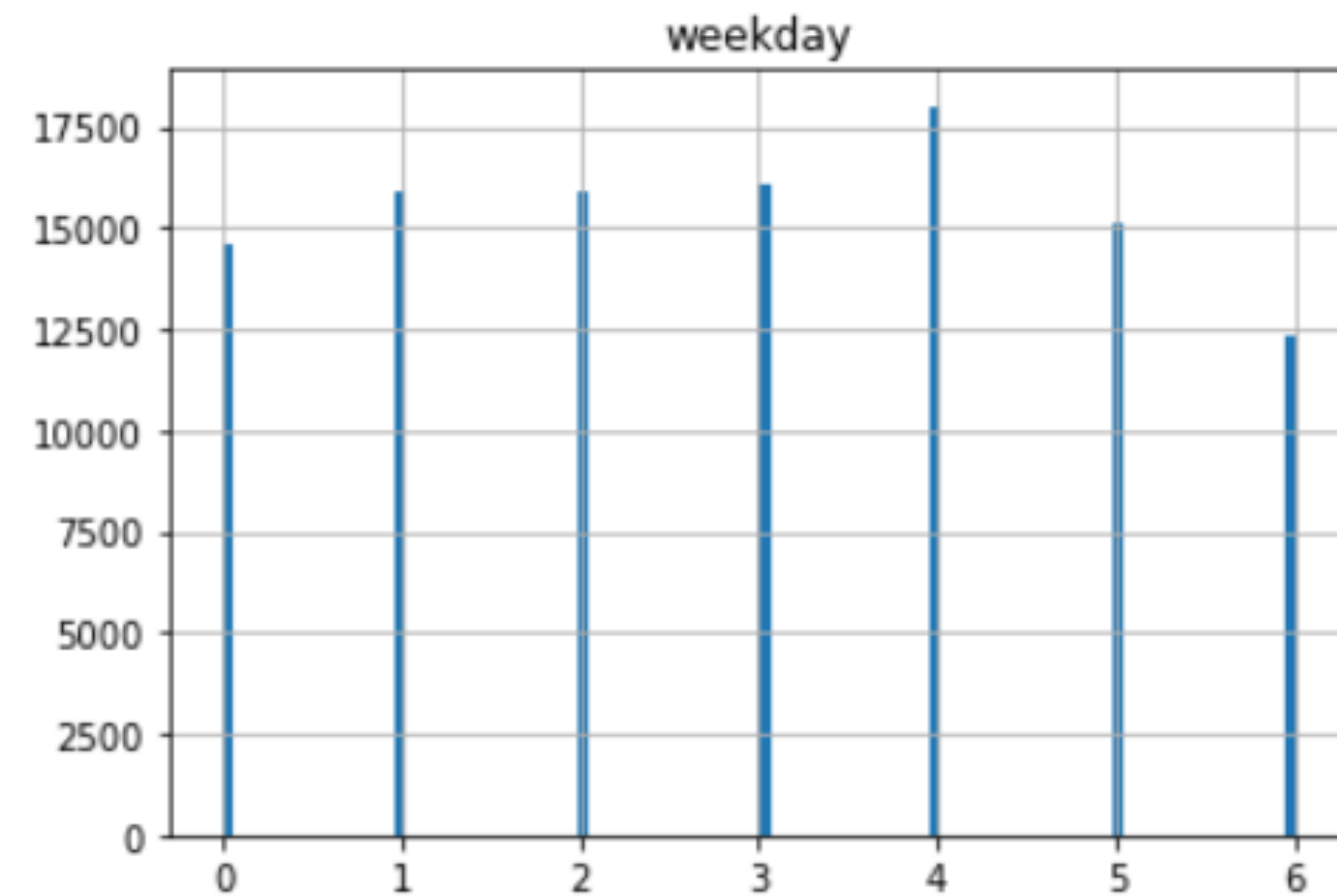
|	SEVERITYCODE	
ADDRTYPE	**Block**	**Intersection**
ROADCOND		
Dry	48596	20880
Ice	674	118
Other	56	12
Sand/Mud/Dirt	25	8
Snow/Slush	559	118
Standing Water	46	12
Unknown	9826	1013
Wet	17263	8579

-

Exploratory Analysis

- Weekdays vs Number of Accidents

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

Final Analysis

- Heat map has been generated for **each hour of the day** on all possible weather condition for each day of the week.
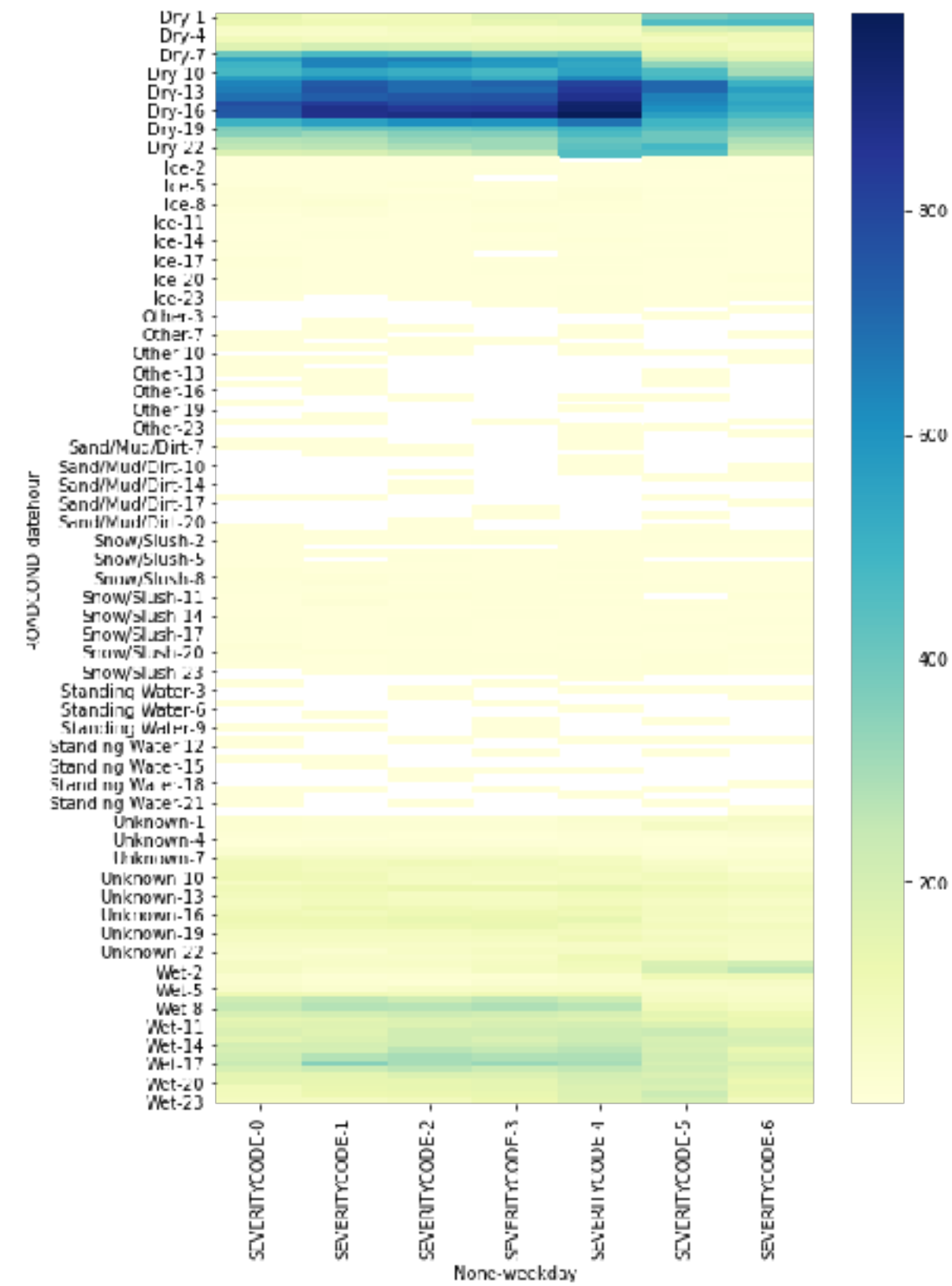
Final Analysis

- The heat map depicts, that morning and evening rush hours on the working days have unprecedented levels of collisions, .
- During the weekend (Fridays, day 5) rush hour incidents are highest.
- Wet weather (bottom part of heat map) has some impact on the collisions,
- But wet weather is significantly low compared to the dry conditions.

Discussion

- The analysis suggests that probably weather has no significant role in the collisions.
- The accidents are more attributable to human factor
- The main factor could be attributed to high volume of vehicles during rush hours.
- Vehicles rushing to workplace at high speeds could be causing more accidents at intersections by jumping traffic lights, colliding with vehicles stuck in traffic jams.
- Predictive models, would enable to advice drivers to be cautions on certain intersections where huge number of vehicles are converging.
- The model should advice drivers should to be extra cautious on weekends when the excitement to rush return home on the last working day is high.