DATA ANALYSIS OF CALGARY CITY TRAFFIC

[For year 2018]

ENSF 592 – Programming Fundamentals for Data

University of Calgary

Analysis by (Name/UCID):
Taruneesh Sachdeva [30114398]
Neha Singh [30109044]
Sanyam [30118797]

Contents

Table of Figures

1. Reading the Calgary boundary

Boundary coordinates of the Calgary city are read from the file 'City_Boundary_layer.csv', based on which a rectangle is drawn around the Calgary city.

2. Display Calgary boundary with a rectangle

Screenshot of the map with plotted boundary

Output obtained is as follows:

Figure 1: Calgary map with boundary

3. Divide Calgary into grids

Based on the boundary coordinates obtained, Calgary city is divided into 10*10 grids which are displayed below. For drawing the grids, the maximum and minimum latitudes and longitudes are divided into ten groups each, based on which the horizontal and vertical lines are plotted.

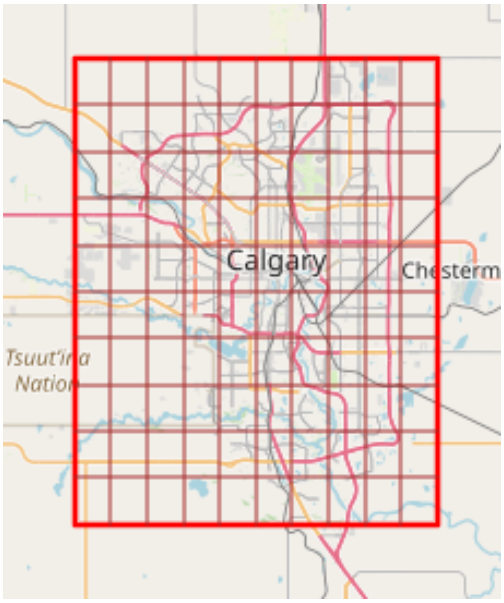

Figure 2: Calgary map divided into grids

4. Analyses of data with respect to various features

a) For Each grid area finding the various features for the year 2018.

The various feature calculated are along with their column names in data:

Feature / Column name

- Average speed limit – avg_speed
- Average Traffic volume – avg_volume
- Average number of traffic cameras – num_cameras
- Number of Traffic Signals – num_signal
- Number of Traffic Signs – num_signs

Sample data set of data frame created is as follows:

	num_cameras	avg_speed	avg_volume	num_signal	num_signs	num_accidents	cell_bounds
0	0	0.000000	0.000000	0	0	0	(50.842822, 50.8797823, -114.315796, -114.2702...
1	0	0.000000	0.000000	0	0	0	(50.842822, 50.8797823, -114.2702069, -114.224...
2	0	0.000000	0.000000	0	0	0	(50.842822, 50.8797823, -114.2246178, -114.179...
3	0	0.000000	0.000000	0	0	0	(50.842822, 50.8797823, -114.1790287, -114.133...
4	0	0.000000	2000.000000	0	54	0	(50.842822, 50.8797823, -114.1334396, -114.087...
5	1	55.063291	2509.433962	0	303	7	(50.842822, 50.8797823, -114.0878505, -114.042...
6	1	70.557621	11914.826498	0	1486	17	(50.842822, 50.8797823, -114.0422614, -113.996...
7	0	80.445545	17805.128205	0	753	13	(50.842822, 50.8797823, -113.9966723, -113.951...
8	0	0.000000	0.000000	0	27	0	(50.842822, 50.8797823, -113.9510832, -113.905...
9	0	0.000000	0.000000	0	5	0	(50.842822, 50.8797823, -113.9054941, -113.859...
10	0	0.000000	0.000000	0	0	0	(50.8797823, 50.916742600000006, -114.315796, ...

Figure 3: Data Frame with various features calculated

The daily visibility and temp is calculated by sampling the hourly data and getting the average of hourly visibility and temp values. The data so obtained is then merged with the accident count info, based on the date column present in both the data sets.

Sample set of data frame for daily weather conditions:

Feature / Column name

Count – Accident count

	Date	Count	Temp (C)	Visibility (km)
0	2018-01-01	12	-16.683333	42.570833
1	2018-01-02	30	-3.787500	40.891667
2	2018-01-03	26	-2.391667	39.212500
3	2018-01-04	22	-5.016667	40.891667
4	2018-01-05	20	-0.345833	40.891667
5	2018-01-06	10	3.666667	39.550000
6	2018-01-07	6	-0.166667	40.020833
7	2018-01-08	10	1.475000	39.212500
8	2018-01-09	14	-9.854167	16.370833
9	2018-01-10	30	-21.625000	10.587500

Figure 4: Sample set of daily weather

b) Analysis of data

Analyse the data and interpret what is the relation between the number of accidents and the above feature in 2018.

I. Analysis of accident with cameras

Box plot representing relationship between number of cameras and number of accidents for each grid.

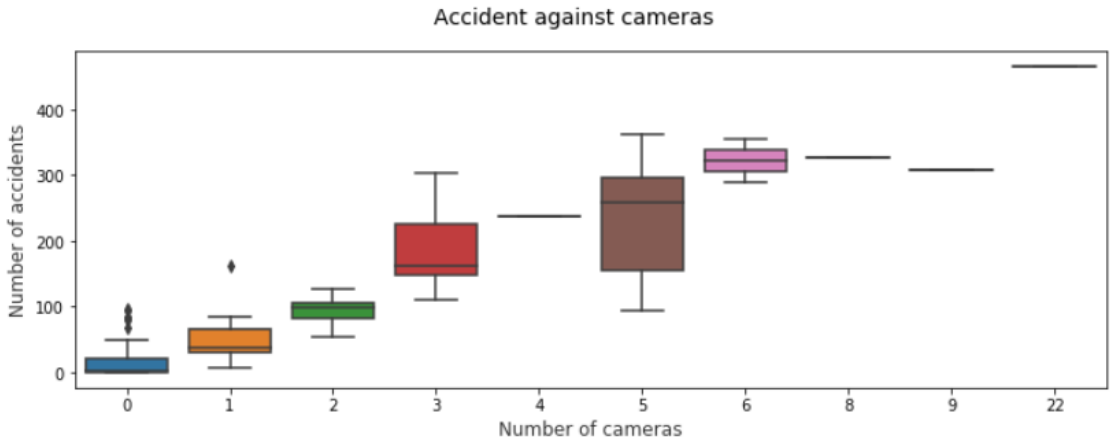

Figure 5: Boxplot of accidents and cameras for each grid

Spearman Coefficient for cameras and number of accidents – (+0.8097809780978098)

Conclusion:

It can be clearly interpreted from the box plot that the count of the cameras and the number of the accidents are directly proportional to each other, i.e. as the accident count increases, the camera count also increases. It signifies that the large number of cameras are installed in the areas where accident count is more. Further, a positive coefficient also supports the same.

II. Analysis of accident with volume

Box plot and scatter plot displayed below represents the relationship between number of accidents and average volume for each grid.

Spearman Coefficient for average volume and number of accidents – (+0.8167416741674167)

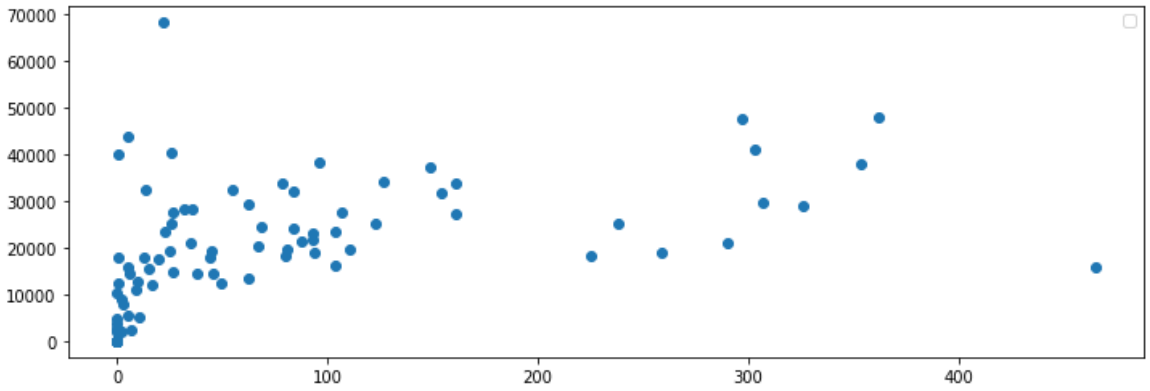

Figure 6: Scatter plot of average volume with accidents

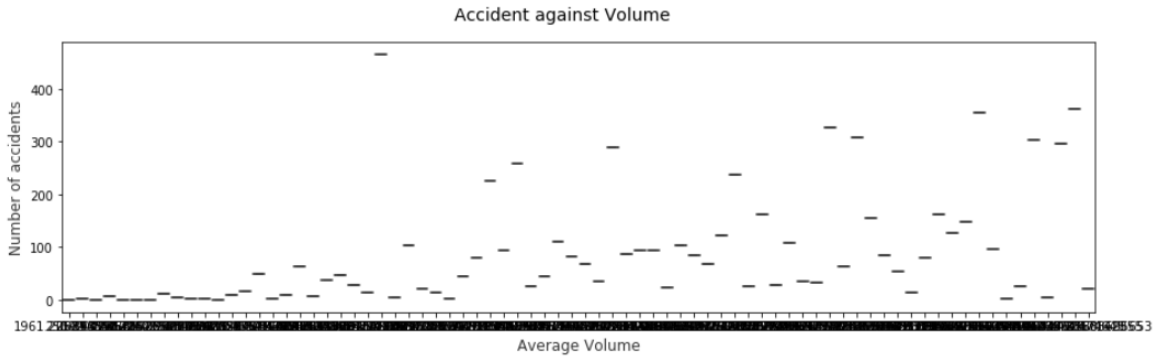

Figure 7: Boxplot of accidents and average volume

Conclusion:

It can be clearly interpreted from the box plot and the scatter plot that the average volume and the number of the accidents are directly proportional to each other, i.e. as the average volume increases, accident count also increases. It signifies that the volume of vehicles was higher where higher number of accidents took place. Further, a positive spearman coefficient supports the above analysis.

III. Analysis of accident with signs

Strip plot displayed below represents the relationship between number of accidents and the count of signs for each grid.

Spearman Coefficient for the number of signs and number of accidents – (+0.865523576240049)

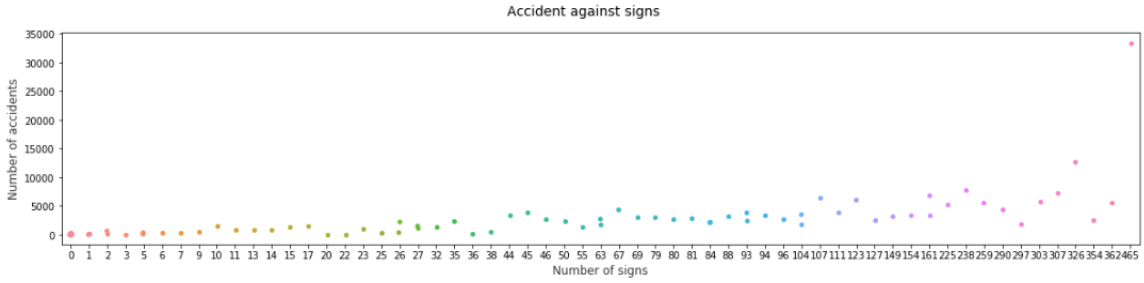

Figure 8: Strip plot of accidents and traffic signs

Conclusion:

As seen, the count of signs is slightly increasing proportionally for accident count in the plot, it signifies that the number of signs installed at different locations is more at locations where a greater number of accidents took place. Further, a positive spearman coefficient supports the above analysis.

IV. Analysis of accident with signals

Strip plot representing relationship between number of accidents and count of signals for each grid.

Spearman Coefficient for signals and number of accidents – (+0.5438703870387038)

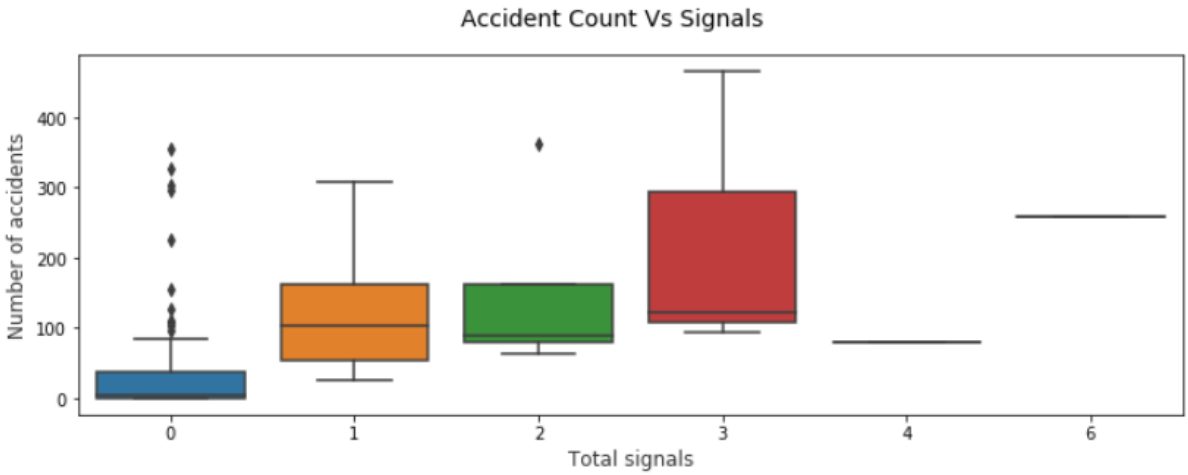

Figure 9: Boxplot of number of accidents and signals

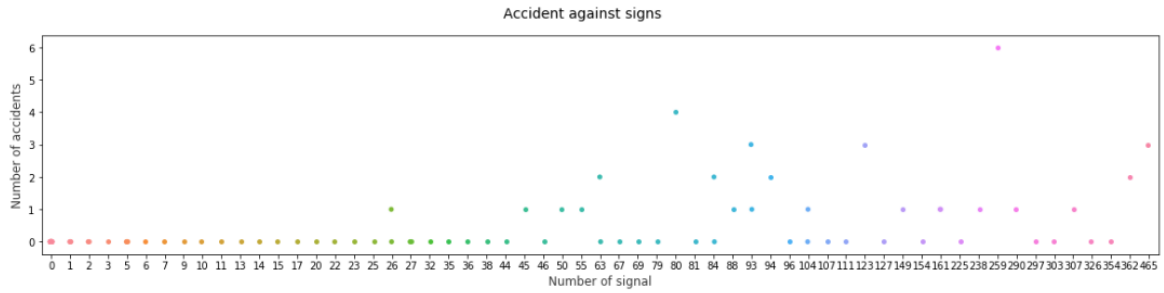

Figure 10: Strip plot of accidents and traffic signals

Conclusion:

As seen, the count of signals is increasing proportionally for accident count in the plot, it signifies that the number of signals installed at different locations is more at locations where a greater number of accidents took place. However, there are few exceptions in the middle, where count is significantly larger. Such exception is also displayed in the spearman coefficient as the value for coefficient is positive, but it is in the middle of 0 and 1, which also shows not such strong correlation.

V. Analysis of accident with speed limit

Point plot analysis for speed limit and number of accidents for days in year 2018.

Spearman Coefficient for signals and number of accidents – (-0.30293040293040296)

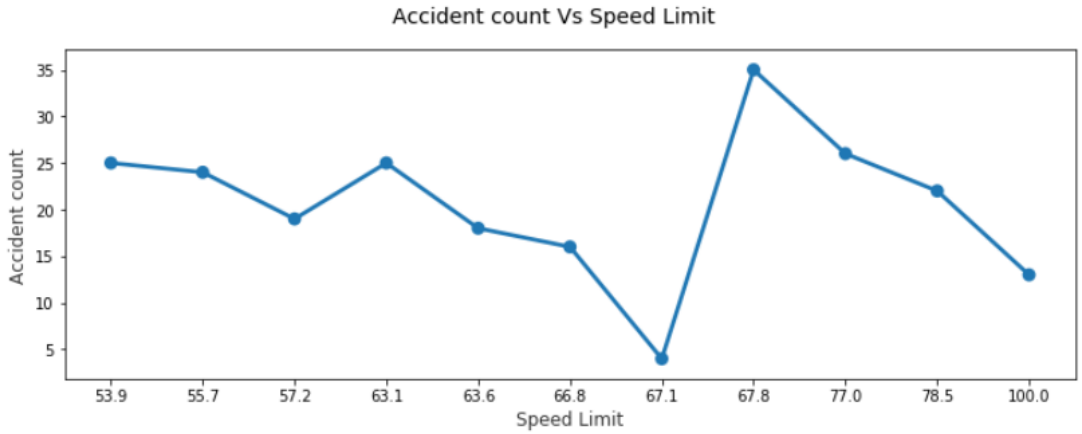

Figure 11: Point plot of speed limits and number of accidents

Conclusion:

The analysis of the speed against the accidents does not provide a constructive conclusion. The point plot does not represent a specific behaviour. Similarly, the spearman correlation value

between speed and the accidents is -0.30244, which gives the output opposite to what was expected.

VI. Analysis of accident with weather

Scatter plot representing relationship between number of accidents and visibility for each day of year 2018.

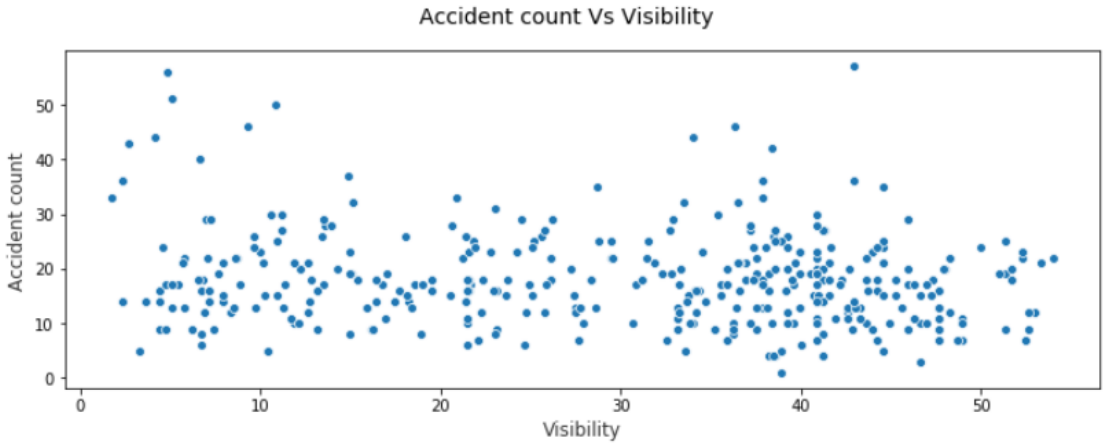

Figure 12: Scatter plot of accident count and visibility

Conclusion:

The relationship between the visibility and the accidents is not very strong. Although as observed, the scatter plot dots have risen to higher count of accidents for lower values of visibility. However, overall, the accidents count seems to remain similar. Similarly, spearman correlation value: -0.12389 shows the weak relationship.

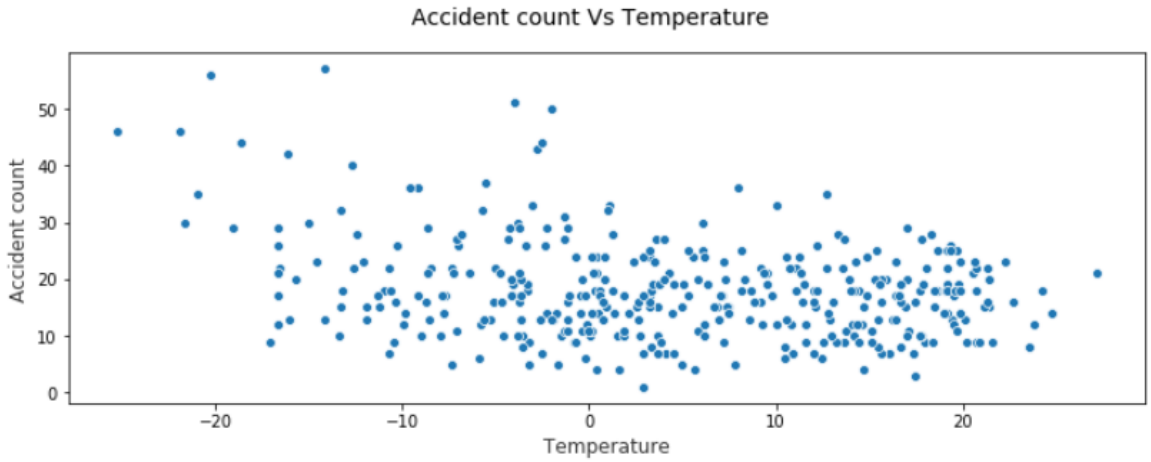

Figure 13: Scatter plot of accident count with temperature

Conclusion:

Although the relationship between the temp and accidents seems to be stronger than the one with the visibility, but it also does not give a much stronger insight. However, the scatter plot shows that count of accidents has increased for lower values of temperature. Similarly, spearman correlation value: -0.16212 indicates the same behaviour. Thus, it can be safely concluded more accidents took place on days when temperature was low.

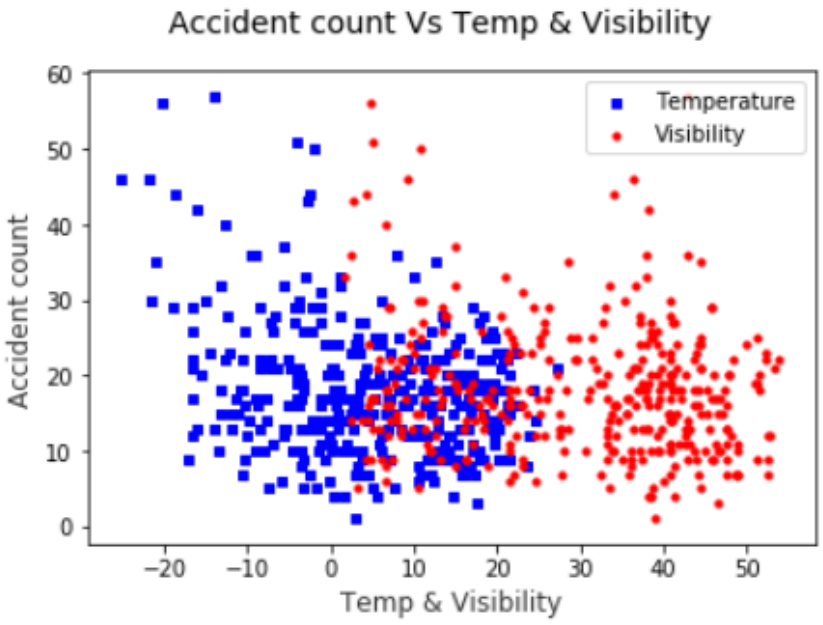

Figure 14:Comparison of temperature and visibility against accident count

VII. Heatmap for number of accidents in each grid

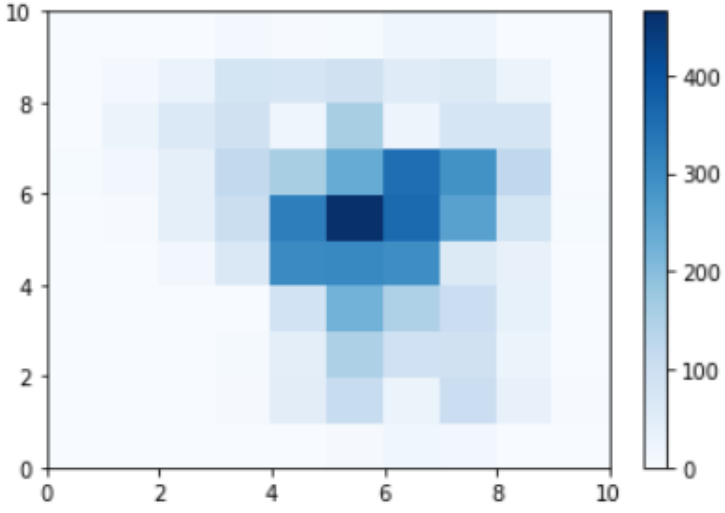

Figure 15: Heat map of number of accidents in each grid

VIII. Heatmap for average volume of traffic in each grid

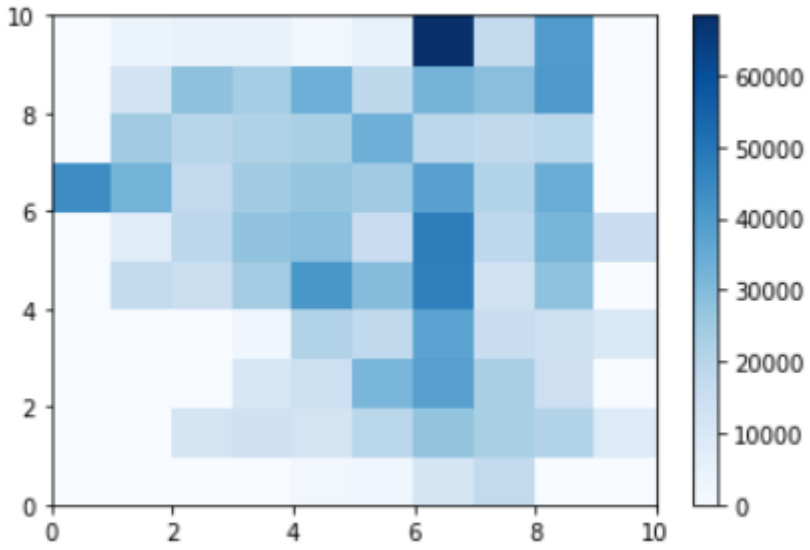

Figure 16: Heat map of average volume for each grid

IX. Heatmap for camera count in each grid

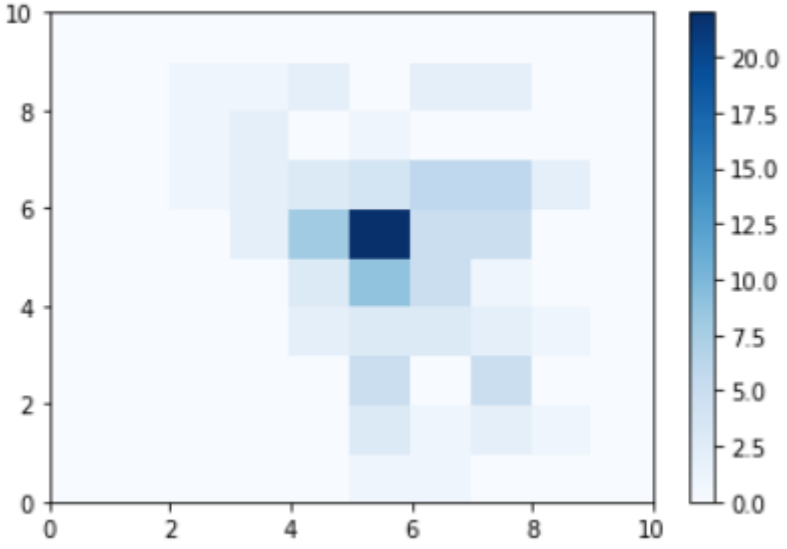

Figure 17: Heat map for camera count in each grid

X. Heatmap for average speed limit of traffic in each grid

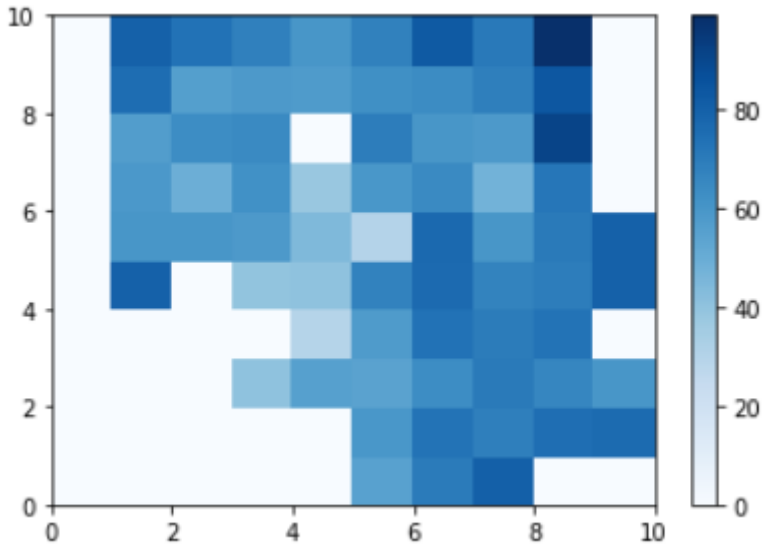

Figure 18: Heat map for speed limit for each grid

XI. Heatmap for count of signals in each grid

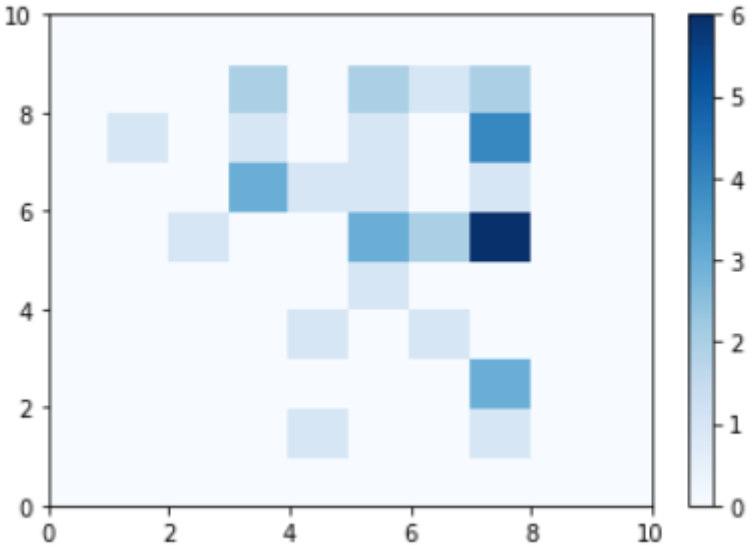

Figure 19:Heat map for count of signals for each grid

XII. Heatmap for count of signs in each grid

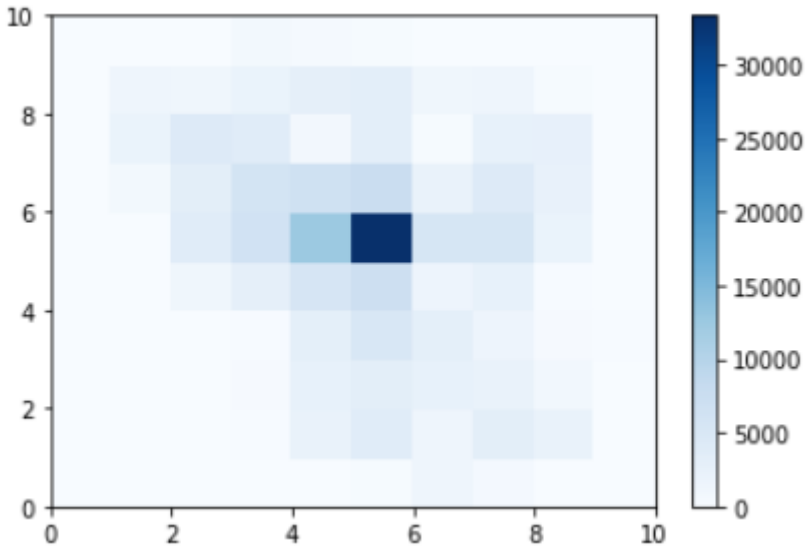

Figure 20:Heat map for count of signs for each grid

5) Visualize – Speed limit in relation to roads

Prepare the dataset of coordinates and speed limit from the csv provided for year 2018. To fetch list of individual coordinates from the column multiline, apply get_coordinates.

Define the 'color dict' to contain the set of predefined colours to be used for the plot for speed limit.

Used colormap to represent the data set in 3D Colo space. Add the colormap to the Calgary map.

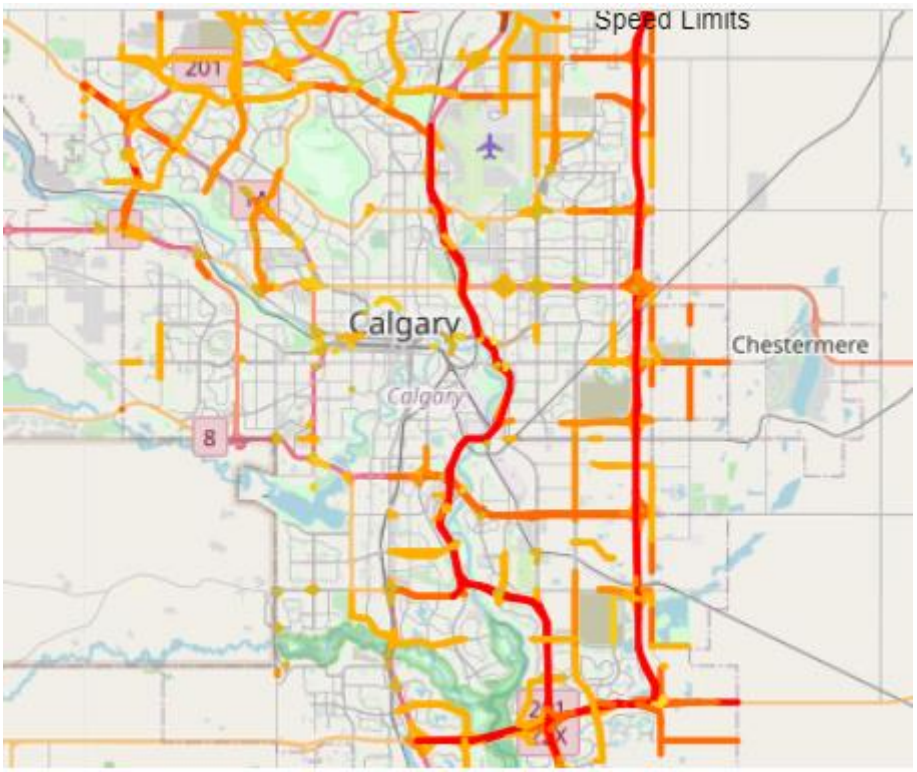

Figure 21: Speed limit representation year 2018

6) Visualization – Traffic heat map

Traffic heatmap of 2018

To prepare the dataset, refine the 'multilinestring' column to get set of individual coordinates for each location and volume.

Group the data on basis of latitudes and longitudes and take sum of volume for each set of coordinates.

Plot the heatmap for the data on the Calgary map where heatmap signifies the volume density at different locations on the map. As visible from the density of the plot around the Downtown area, the part of the city being the busiest, the traffic volume is high there.

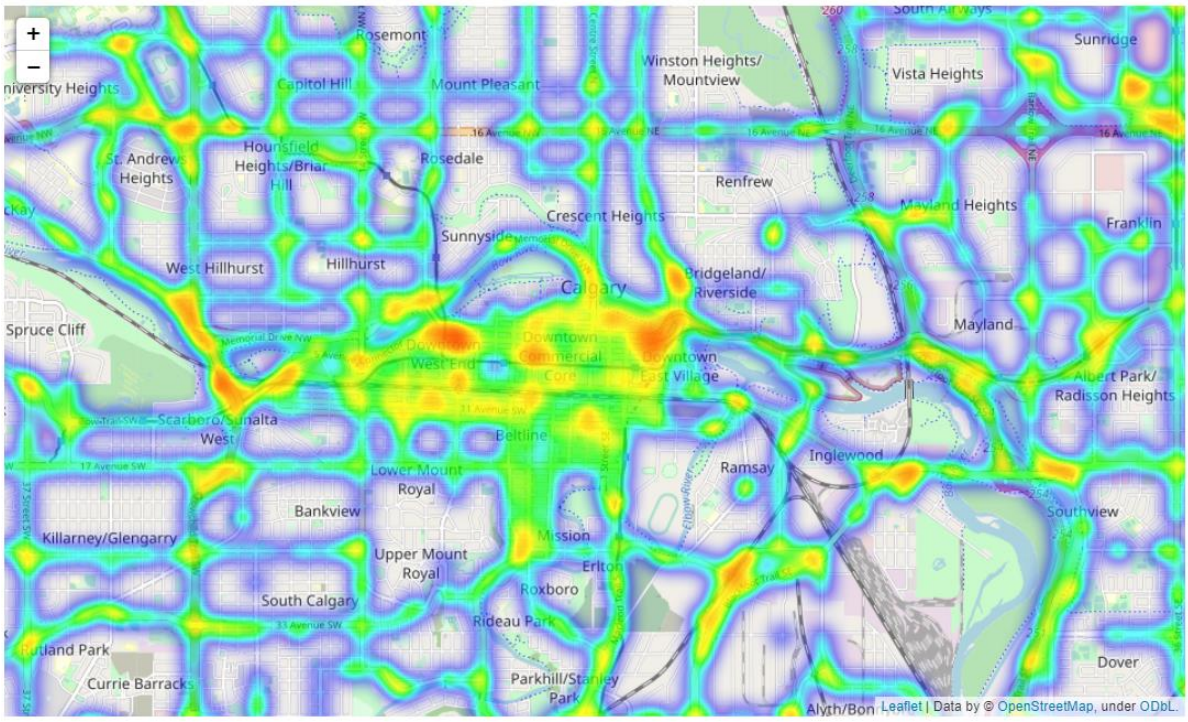

Figure 22: Traffic heatmap of 2018