report_interpretation

August 12, 2020

1 Interpretation

1.1 Grid data

The average speed limit, average traffic volume, traffic cameras, traffic signals, traffic signs and traffic incidents were each calculated for each section of a 10x10 grid accordingly. Results were combined into 1 dataframe for visualization analysis. Incident/volume ratio was plotted against average speed limit, traffic cameras, traffic signals and traffic signs respectively to evaluate possible correlations. (Incidents were normalized against volume as naturally higher volumes would lead to higher incident counts).

Incident/volume ratio is indicated by the red line in figures 2.21, 2.31, 2.41 and 2.51. This incident/volume ratio peaks near grids 35, 45, 55 and 65. These grids correspond to the downtown area, with grid 55 (the highest incident/volume ratio) being at the downtown centre of Calgary.

Bar graphs of traffic signals, traffic signs and traffic cameras, shown in cyan color (figures 2.2.1, 2.3.1, 2.4.1 and 2.5.1), followed a similar pattern as the red line representing incident/volume ratio. Grids with greater number of signals, signs and cameras resulted in higher incident/volume ratio. An overall strong positive correlation was shown between incident frequency and the number of traffic signals, signs and cameras. On the other hand, a moderate negative correlation was spotted between incident/volume ratio and average speed limit (figure 2.2.1). Grids with higher average speed limit, such as grid 60, resulted in lower incident/volume ratio. Whereas grids with significantly lower average speed limit, such as grids near downtown area (35, 45, 55, 65), resulted in much higher incident/volume ratio.

Within figure 2.3.2, a bar plot comparing the number of cameras in Calgary that were within ~ a 10m radius of an incident. This plot shows that about 77% of cameras were close to an incident, showing that cameras are at least being placed in good locations to catch incidents. However, when comparing to figure 2.3.3, only 1.4% of incidents occur near cameras. This shows that there are a lot more incidents than cameras in Calgary. On the other hand, 11% of incidents occurred within 10m of a signal. Table 2.4.2 and figure 2.4.5 indicates that ½ signals (intersections where there is only a traffic signal on 2/4 of the streets) have a 23.8 incident to signal count percent. This tells us that ½ signals are more likely to have an incident occur nearby them. However, because there are less ½ signals (21) than other signals (1010 and 257 for 'T-intersection' and 'Overhead Flasher' signals, respectively) this conclusion is not as concrete. It is also interesting that there is a 41.3% incident to signal count in the South Calgary quadrant (table 2.4.3 and figure 2.4.6), and in general, there are higher incident counts in southern Calgary. Part of this may be because downtown is included in South Calgary, and that is where the highest volume of cars is located, and everyone drives to/from downtown every day for work. A bubble plot of incident to signal count can be seen in figure 2.4.7, where each bubble is placed within its respective quadrant in Calgary. Figure 2.4.8 is a visual line plot that show south Calgary has having the highest incident count as well.

As for pedestrian button signals, table 2.4.4 shows an interesting statistic that there is a 20% higher chance of having an incident at a signal with no pedestrian button. The city usually places pedestrian buttons where pedestrians are common, so perhaps cars at these intersections are not often expecting to have a pedestrian crossing the street. Another explanation might be that most intersections without a pedestrian button are “uncontrolled” intersections, where cars need to regulate themselves and organize themselves when they can turn or drive through an intersection. Whereas an intersection with lights at each road will automatically regulate the traffic for drivers. This relates to possibly why ½ signals are the most common to have incidents, as they require drivers themselves to decide when to drive through the traffic or ‘turn into’ the traffic.

1.2 Weather data

Extra Weather data tables and figures can be found in section 2.6. Overall, these figures indicate that incidents are more common in January and February (figure 2.6.3), and that these two months are on average the coldest months in 2018 at -10oC on average. Figure 2.6.4 agrees with this conclusion, showing that on average 33 incidents occur per day during days of -20oC and -30oC, and any temperature greater than this has little effect on daily incident rates. Figure 2.6.7 agrees with the conclusions so far, showing that the highest daily incidents occur in January and February, but it also shows how visibility (>~3km) had a minor effect on incident rates, and is perhaps overshadowed by temperature. Visibility does however play a role in the daily incident count as seen I figure 2.6.8, where any visibility less than 3km results in an increase in daily incidents.

Of all this data so far, the most important seems to be the time of day. Figure 2.6.9 show the total incidents vs. the time of day (on a 24 hour clock). It is obvious that most incidents occur at 8:00am and 5:00pm, which is exactly when rush hour occurs downtown. The signal data, shows that the highest incident to signal ratio is in downtown, the traffic volume and incident heat maps also show that most of the incidents and traffic volume is located downtown. Final speed data shows that speed is not a major factor in the frequency of incidents, this also correlates to downtown where cars are on average driving at 50-60km/hr, but they have to stop and go frequently at traffic signals and pay attention to when they can turn right onto another street at a red light. The traffic is very dense in downtown at 5:00pm and 8:00am, people are either sleepy or excited to get home, yet they need to be alert and stop/go at frequent intersections, this is a recipe for having incidents, especially in January and February when the weather is cold and snowy.

1.3 Conclusions

Based on the interpretations above as well as taking into the spearman coefficients here are the conclusions for each independent variable:

Cameras (spearman = 0.73): While cameras have a high positive correlation with incidents, this is mainly due to the fact that many incidents occur within downtown Calgary where the cameras are located. Taking a closer look into the location, only 1.4% of incidents occur near a camera (10m radius). (However, the placement of the cameras are good as 77% of them were able to capture an incident).

Signals (spearman = 0.92): Signals has the highest correlations with incidents. This makes sense as most signals are placed at intersections and “stop and go” locations where drivers must pay attention to their surroundings as well as potentially self-regulate.

Signs (spearman = 0.88): Signs share the same reasoning with signals.

Speed (spearman = -0.51): Interestingly, speed has a negative spearman coefficient which seems counter intuitive. Faster speeds generally mean less time to react, however, if we look at our average speed limit data to incident/volume ratio, we see that most accidents occur in the downtown area where the speed limit is much lower compared to highway speeds. However, one important thing to note is that this does not measure the severity of the incident (will be mentioned in the following section).

Visibility (spearman = -0.13): Overall, the visibility seems to have little (negative) to no correlation based on the spearman coefficient. However, if we bin the date, we can see that visibility has an impact on incidents with less than ~3km of visibility. After this amount, there seems to be very little correlation which makes sense as >5km of visibility is relatively good (which causes the low spearman value as most points are above this visibility).

Temperature (spearman = -0.02): Similarly with visibility, temperature seems to have little to no correlation based on the spearman coefficient. However, plotting average temperature versus month shows us that the coldest months (January and February) have the most incidents. Binned data also agrees with this (-20C to -30C having the most incidents). The cause of the low spearman coefficient is that most the data is lies above these low temperatures.

1.4 Limitations of our analysis/future improvement:

During our analysis, we made several assumptions which could limit the usefulness of our data. We will address these here for potential future improvement:

1. We treat all incidents as the same (not enough granularity):

A multi-car collision incident is treated the same as single car incident. Instead maybe some sort of index of severity would provide more insight. (More data would also be needed, as current data only tells us that there was an incident and not the type/severity). Along this note, a pedestrian incident is the same as a car on car incident.

Reducing accidents could be difficult (it’s impossible to reduce all accidents to 0) but instead maybe a good compromise is to reduce severity instead (that’s why we need more data for a severity index). A good outcome would be if most accidents were minor.

2. The 10x10 grid:

Our 10x10 grid, while useful so we are able to analyze data, is an arbitrary (and maybe even biased) grouping structure. Instead maybe an unsupervised learning/grouping algorithm to group attributes (such as location of incident, or even weather groups) could possibly provide more insight. Some popular choices are k-means (but we would have to choose amount of groups in this algorithm) or DBSCAN (which would find groups and number of groups for us) could be used to group data together.

3. Driver competency:

We don't know the competency of the drivers (i.e. they are all treated the same). Instead of traffic and weather data, another large factor in reducing incidents can be in educating the driving population. This could also be the more cost efficient solution.

4. More data:

Finally (and the most obvious), using more data points for the analysis, comparing year to year changes could provide us with valuable insight, however was out of the scope for this project.