Analysis of Seattle Police Department 911 Call Logs

James Glenister

Western Governors University

1. **Introduction**

According to the Seattle Times, Seattle’s metro area is includes 3.87 million people and is the 15th largest metro area in the nation (Balk, 2018). With a large and growing jobs market, it is critical that the Seattle Police Department secures the government funding to modernize and better protect and serve the citizens of Seattle.

1. **Data Analysis**

We were given a log detailing numerous 911 emergency phone calls with pertinent information regarding location, number of officers, and the kind of incident. Below the steps necessary for analysis are documented.

* 1. **Data Preparation**

The initial raw data was a sheet with 1,047 observations with 20 features documented. It is important to remember there are only 4 key features that need to be analyzed and extracted:

1. Number of incidents per date
2. Number of incidents per event type
3. Number of incidents per sector
4. Number of officers per incident

To determine date information, at scene time was deleted while event clearance date was preserved. At scene time had 72% of missing observations, so it would not be a useful feature to analyze.

To determine location data, only the sector was preserved. Census tract, latitude, longitude, incident location, and 100 block were all discarded. They provided too fine-grained data for what was necessary, and could present a security or ethics risk. This will be discussed further in the ethics section.

To determine incident data, only the Event Clearance Group was kept. All reference to the Initial Type was discarded. While there was not a robust mathematical reason for this choice, there is a strong correlation between the initial type and group and the clearance type and group. Since the clearance date was kept, and there was no missing values, the clearance group was chosen to be the main source of incident type data.

A single unique identifying feature was kept to easily identify observations. This feature was the CAD Event Number. There were no duplicates in this feature, the CAD CDW ID, or the General Offense Number so any would have made a good identifying key.

Lastly, the officer on scene data was analyzed for outliers. Even with a tight outlier test of ±2 standard deviations from the mean, no data was outside this range. The only observation discarded was Event Number 16000105361 since it lacked sector data.

* 1. **Data Sheets**

After the data was cleaned, it was fairly easy to create tables and charts based on the three metrics.

* + 1. **Date and Number of Event**

Figure 2: Table detailing information found in the above bar graph

|  |  |  |
| --- | --- | --- |
| Date | Number of Events | Frequency |
| 3/26/2016 | 243 | 23% |
| 3/27/2016 | 583 | 56% |
| 3/28/2016 | 219 | 21% |

A frequency analysis is also included. Based on the data, it is clear that the most incidents occurred on Saturday 3/27/2016. There are limited days to do any time series analysis on which will be important in section 3.

* + 1. **Number of Incident Occurrences by Event Type**

Figure 4: Chart detailing the information found in the bar graph above. A frequency analysis is included.

|  |  |  |
| --- | --- | --- |
| Event Clearance Group | Number of Events | Frequency |
| HARBOR CALLS | 1 | 0.1% |
| WEAPONS CALLS | 1 | 0.1% |
| MISCELLANEOUS MISDEMEANORS | 2 | 0.2% |
| BIKE | 2 | 0.2% |
| LEWD CONDUCT | 2 | 0.2% |
| PERSONS - LOST, FOUND, MISSING | 3 | 0.3% |
| ARREST | 3 | 0.3% |
| PROWLER | 3 | 0.3% |
| ANIMAL COMPLAINTS | 4 | 0.4% |
| PERSON DOWN/INJURY | 5 | 0.5% |
| FRAUD CALLS | 6 | 0.6% |
| THREATS, HARASSMENT | 6 | 0.6% |
| ROBBERY | 8 | 0.8% |
| PROPERTY - MISSING, FOUND | 12 | 1.1% |
| HAZARDS | 13 | 1.2% |
| NARCOTICS COMPLAINTS | 14 | 1.3% |
| OTHER PROPERTY | 16 | 1.5% |
| PROPERTY DAMAGE | 17 | 1.6% |
| AUTO THEFTS | 18 | 1.7% |
| SHOPLIFTING | 21 | 2.0% |
| ASSAULTS | 24 | 2.3% |
| BURGLARY | 25 | 2.4% |
| BEHAVIORAL HEALTH | 26 | 2.5% |
| NUISANCE, MISCHIEF | 39 | 3.7% |
| TRESPASS | 54 | 5.2% |
| LIQUOR VIOLATIONS | 55 | 5.3% |
| CAR PROWL | 58 | 5.6% |
| MOTOR VEHICLE COLLISION INVESTIGATION | 62 | 5.9% |
| FALSE ALACAD | 64 | 6.1% |
| SUSPICIOUS CIRCUMSTANCES | 150 | 14.4% |
| TRAFFIC RELATED CALLS | 164 | 15.7% |
| DISTURBANCES | 167 | 16.0% |

It is clear from the data that the vast majority of events that occurred were either general disturbances, traffic related calls, or other suspicious circumstance. In fact, those three events accounted for approximately 50% of all incidents that occurred in Seattle between the dates of March 26th and March 28th.

* + 1. **Sectors and Total Number of Events**

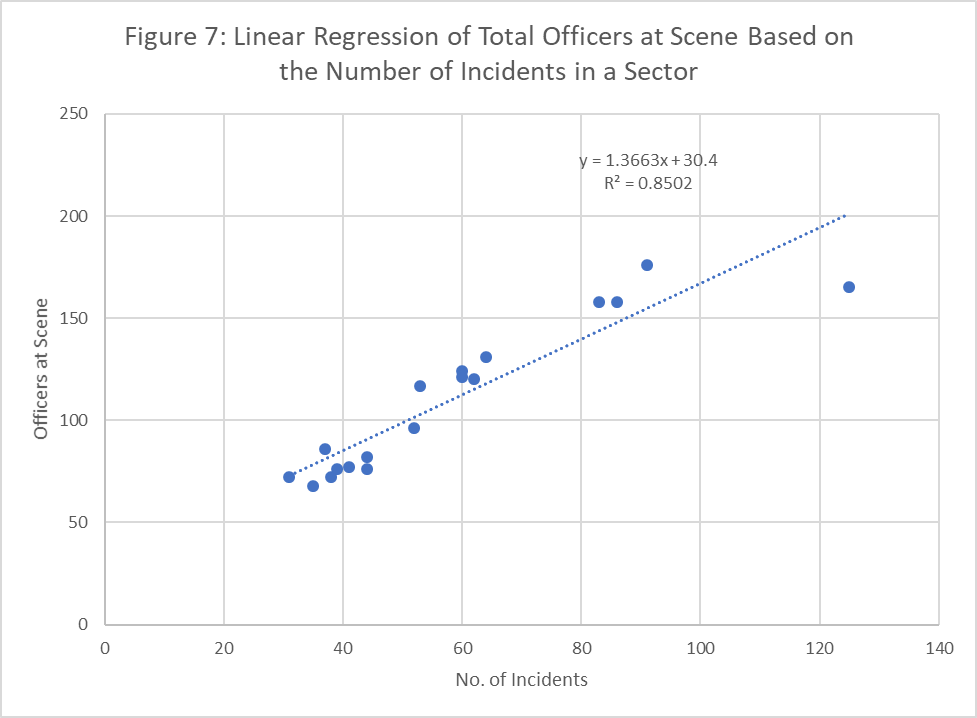
Figure 6: Sector analysis data for figure 5

|  |  |  |
| --- | --- | --- |
| Sectors | Number of Events | Frequency |
| O | 31 | 3% |
| F | 35 | 3% |
| W | 37 | 4% |
| L | 38 | 4% |
| G | 39 | 4% |
| J | 41 | 4% |
| C | 44 | 4% |
| S | 44 | 4% |
| U | 52 | 5% |
| N | 53 | 5% |
| D | 60 | 6% |
| R | 60 | 6% |
| Q | 62 | 6% |
| K | 64 | 6% |
| B | 83 | 8% |
| E | 86 | 8% |
| M | 91 | 9% |
| H | 125 | 12% |

This sector data will later be used to help make our linear regression model.

* 1. **Linear Regression Model**

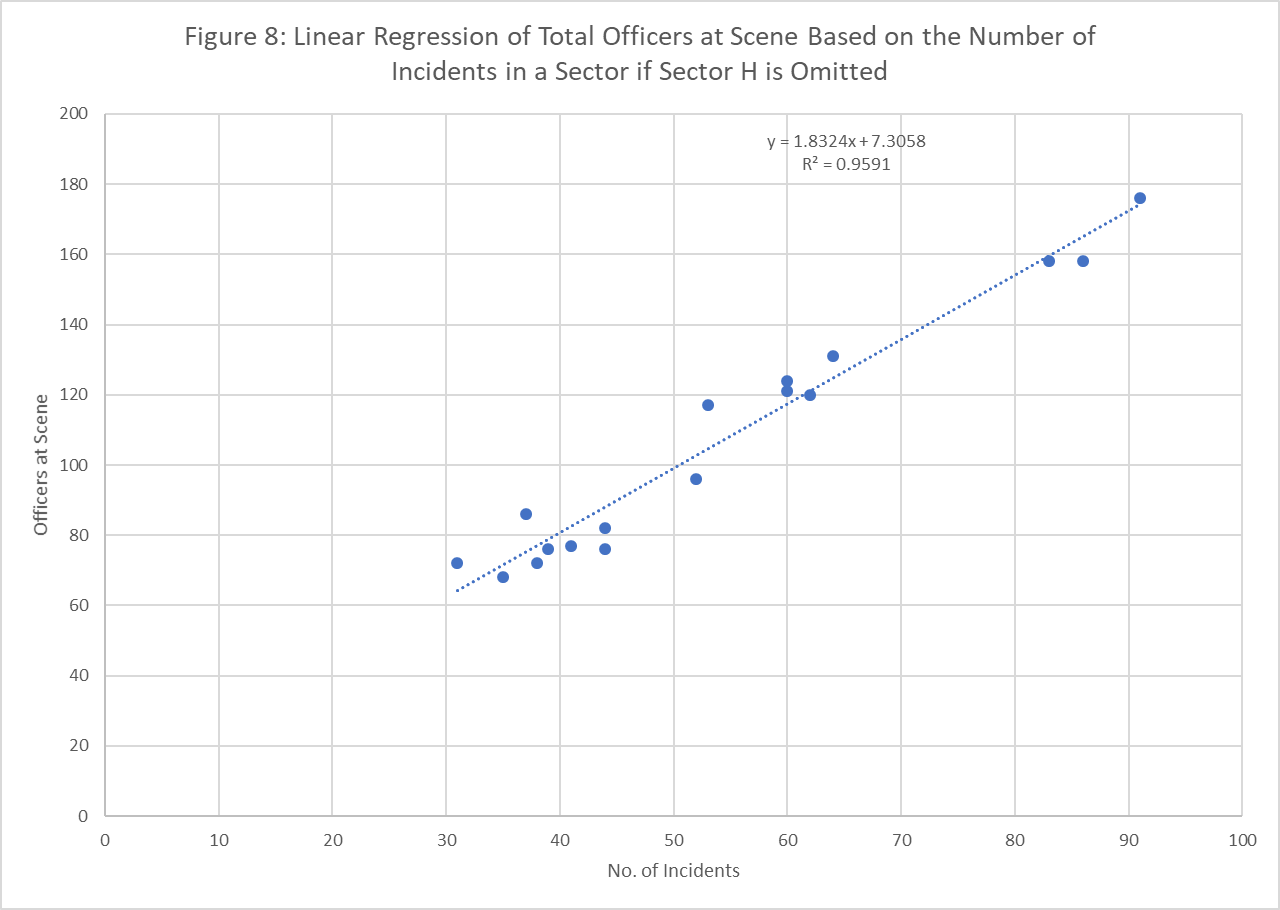
A regression model was made that predicted the total number of officers at the scene based on the number of incidents in a sector. The regression model is given below:



This linear regression model models the data decently well. 85% of the variation in the data is explained by the model. The p-value for the slope is which is significantly less than 0.05. This allows us to reject the hypothesis that the slope is 0. It is not worthwhile to analyze the intercept since it states if there were no incidents there would still be 30.4 officers at the scene.

* 1. **Outliers**

There was only one significant outlier in the data set. The point in sector H (125,165) is a significant outlier in the model. In the below graph we show the regression model if that point is omitted:



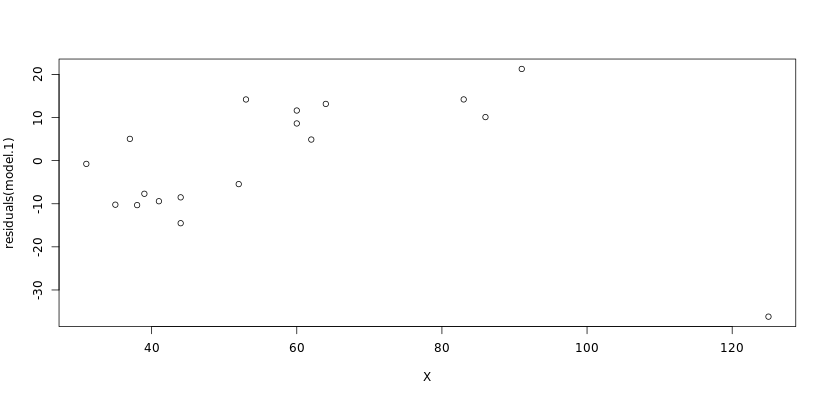
With the simple omission of Sector H’s data point, the coefficient of determination jumps up to 95.9%, indicating that the number of officers per incident is much closer to 1.83 officers per incident instead of 1.37 officers per landing. This is much closer to the true mean of 1.89 officers per incident used later in our analysis.

* 1. **Residual Plot**

According to Zybooks, there are 4 main tests to look for when analyzing residuals (Zybooks, n.d.). They are the “mean of zero” assumption, the “constant variance” assumption, the “normality” assumption, and the “independence” assumption.

The first two assumptions can be assessed using the below residual plot:

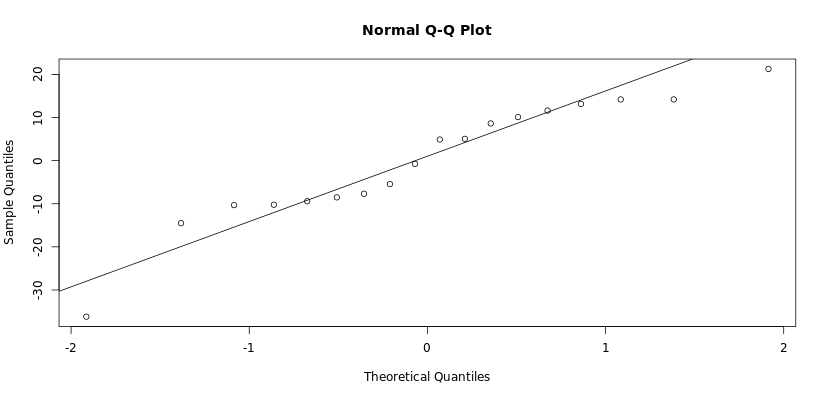
**Figure 9:** Residual plot of that maps the number of incidents to the residuals at that point



There are several things to note. First, the residual at point 125 is incredibly large with respect to the rest of the residuals. This highlights that the point described by Sector H is such an outlier and therefore ok to be omitted in the model. Second, from a quick visual scan, it is clear the mean of the residuals is not approximately not zero. Third, as the residuals seems to be increasing, the constant variance assumption does not seem to hold.

The normalized residual plot is given below:

**Figure 10:** Residual plot of that maps the number of incidents to the residuals at that normalized point



As the residuals lie close to the normal line, the normal assumption holds. The independence assumption likely does not hold. There is a strong linear correlation between each residual. This leads to a conclusion that there is likely another linear term that would help make for a better model.

1. **Simulation and Recommendation**

Part 2 of the analysis revolved around calculating potential officer at scene numbers to see if the Seattle Police department qualified for the governor’s funding initiative. What follows is an analysis of the current officers per incident levels and possible future qualifications.

* 1. **Current Qualification**

A Monte Carlo simulation was run to find the probability of 2.5 or more offers would be at the scene of each incident. 5000 incidents were simulated. The mean and standard deviation were calculated from the officers at the scene for each incident logged in the 911 call logs. The below data table shows the results of one round of simulations:  
**Table 11:** Table relating the results of the Monte Carlo simulation for officers at the scene.

|  |  |
| --- | --- |
| Mean | 1.890 |
| Standard Deviation | 1.189 |
| Average Officers on Scene | 1.92 |
| Calls with more than 2.5 Officers | 1591 |
| Probability to have more than 2.5 Officers | 0.32 |

Based on the number of incidents, and the low probability that an incident will have more 2.5 officers per incident, it is unlikely that the Seattle Police Department will qualify for the funding initiative.

* 1. **Future Qualifications**

Unfortunately, there are only 3 data points in our time series, so it is impossible to be able to forecast precisely. However, with the three points we have, it is possible to say there is not an increasing trend. Likely, the officers per incident rate will stay fairly close to 1.88-1.92 officers per incident. Therefore, the Seattle Police Department is unlikely to qualify for the funding initiative in the future.

* 1. **Ethics**

While the data is publicly available from the Seattle Police Department website (IT, 2018), there are still some ethical issues that need to be addressed. From the location data and incident type, it could be possible to mine identities of parties involved, breaking anonymity. For our type of analysis, it was not necessary to have a fine level of location information. To better anonymize the data, the location information, barring the sector feature, could be omitted. As we are being hired by the Seattle Police Department to help them secure funding, there could be a financial conflict of interests as well.

1. **Conclusion**

From this project, it is clear that Seattle, in this moment in time, do not have enough officers per incident to secure the Governor’s additional funding now or in the near future. Incidents are concentrated in a few sectors, and a majority of incidents are covered by a few event types.

Sources

Balk, G. (2018, March 26). *Seattle just one of 5 big metros last year that had more people move here than leave, census data show*. Retrieved from The Seattle Times: https://www.seattletimes.com/seattle-news/data/seattle-just-one-of-5-big-metros-last-year-that-had-more-people-move-here-than-leave-census-data-show/

IT, S. (2018, June 27). *Seattle Police Department 911 Incident Response*. Retrieved from City of Seattle Open Data Portal: https://data.seattle.gov/Public-Safety/Seattle-Police-Department-911-Incident-Response/3k2p-39jp

Zybooks. (n.d.). *6.6 Model Assessment*. Retrieved from Zybooks: Fundamentals of Data Analysis: https://learn.zybooks.com/zybook/WGUFundamentalsOfDataAnalytics/chapter/6/section/6