**Fundamental Analytics Task: Police Department Analysis**

Heidi L. Petersen

College of Information Technology, Western Governors University

C740: Fundamentals of Data Analytics

Dr. Emelda Ntinglet

July 29, 2020

Fundamental Analytics Task: Police Department Analysis

The given data set covers the different type of 911 call incidents over the period of about 48 hours. It has 20 features listed and 1,026 observations. Many features were irrelevant to the analysis at hand and had to be removed for a more concise overview. Two outliers skewed the regression line and also were removed for a more concise and accurate presentation of the data. There was an average of 1.89 officers present at each of the incidents, which does not reach the minimum 2.5 officers per incident required for additional funding.

Data and Model

In reviewing the data, 14 of the 20 features were irrelevant or repetitive. The “At Scene Time” feature was unnecessary due to 646 of the 1046 cells being blank. There was not enough to meaningfully contribute information. “Cad Event Number”, “General Offense Number”, and “Census Track” were not relevant to the analysis. The description of each of the events was covered adequately with the “Event Clearance Group” feature so following were needless: “Event Clearance Code”, “Event Clearance Description”, “Event Clearance Subgroup”, “Initial Type Description”, and “Initial Type Subgroup”. Similarly, “District Sector” covered the location of the events so “Hundred Block Location”, “Zone/beat”, “Longitude”, “Latitude”, and “Incident Location” were redundant. The most vital data was captured in the remaining 6 features.

One of the observations (Incident 1702543) was missing significant data so it was removed completely from the analysis.

The data is skewed from its best fit. The point (1,1) and (125, 165) are outliers. To know whether to include the outliers or not, the “key is to examine carefully what causes a data point to be an outlier.” (Libretexts) The point (1,1) was the only incident of its type with only one officer responding. With so many incidents of the other events occurring, it seems that the event might be an error. (125, 165) all occur in section H. It would be of use to examine what the barrier to having more officers responding in that area.

The standard error including these two points is 15.45. Whereas without the outliers the standard error is much better at 7.25. P-values show that there is significance with either regression line, but the regression line without the outliers captures the information better. The slope of the second regression line is steeper and corresponds better to most points. R2 goes from 0.8795 with outliers to 0.9591 without the outliers.

Results

Most of the incidents occur on March 27. (See Table 1 and Figure 1) This may be due to incomplete data for the other two days submitted. On March 26, only the latter half of the day was included, and March 28 only included up until midday. One cannot conclude if one of the days was more incident filled than another.

Three of the events were most prominent. (See Table 2 and Figure 2) The events that occurred most often were traffic related calls, disturbances, and suspicious circumstances making up more than 46% of the reported events.

A few sectors contained more incidents than others. (See Figure 3) About 12% of all events took place in Sector H. The next closest sectors, Sector M had about 9% of the incidents and Sectors B and E each had about 8% of the occurrences, respectively. Conclusions about sector safety are not helpful due to not knowing the size and population of each sector.

The line of regression shows a positive correlation between the number of events in the sector and the number of officers responding the scene.

The precinct had an average of 1.89 officers at each incident. The governor offered additional funding for departments that had at least 2.5 officers at each incident. The department needs to increase their presence at each call. Limitations exist in the recommendation one should consider. It might be fruitful to compare more than 48 hours of data for officers on the scene for calls. The data also did not indicate whether the officers responding to each call were a unique group or not.

Precautions about Sensitive Information

The given data set contains sensitive information. The exact locations must be encrypted or removed. It is possible that if the wrong parties were to view the 911 calls, they could use the data to a criminal advantage. I.E. Knowing which areas have the most reported crime, knowing which crimes occur most often in each area of the city, and so forth. If the report is to be distributed, it may be of use to randomize the sectors and remove the identifying information.

References

Libretexts. (2020, July 27). 12.7: Outliers. Retrieved July 29, 2020, from https://stats.libretexts.org/Courses/El\_Camino\_College/Test1/12:\_Linear\_Regression\_and\_Correlation/12.07:\_Outliers

Fundamentals/Statistics for Data Analytics V5. (2018). Retrieved July 25, 2020, from https://learn.zybooks.com/zybook/WGUC740V52018

Table 1

Events per Date

|  |  |
| --- | --- |
| Date | Number of Events |
| 3/26/2016 | 243 |
| 3/27/2020 | 583 |
| 3/28/3030 | 219 |

Table 2

Incidents by Type

|  |  |
| --- | --- |
| Event Type | Incidents |
| Animal Complaints | 4 |
| Arrests | 3 |
| Assaults | 24 |
| Auto Thefts | 18 |
| Behavioral Health | 26 |
| Bike | 2 |
| Burglary | 25 |
| Car Prowl | 58 |
| Disturbances | 167 |
| False Alacad | 64 |
| Fraud Calls | 6 |
| Harbor Calls | 1 |
| Hazards | 13 |
| Lewd Conduct | 2 |
| Liquor Violations | 55 |
| Miscellaneous Misdemeanors | 2 |
| Motor Vehicle Collision Investigation | 62 |
| Narcotics Complaints | 14 |
| Nuisance, Mischief | 39 |
| Other Property | 16 |
| Person Down/Injury | 5 |
| Persons - Lost, Found, Missing | 3 |
| Property - Missing, Found | 12 |
| Property Damage | 17 |
| Prowler | 3 |
| Robbery | 8 |
| Shoplifting | 21 |
| Suspicious Circumstances | 150 |
| Threats, Harassment | 6 |
| Traffic Related Calls | 164 |
| Trespass | 54 |
| Weapons Calls | 1 |

Table 3

Number of Event Per Sector

|  |  |
| --- | --- |
| District Sector | Number of Events |
| B | 83 |
| C | 44 |
| D | 60 |
| E | 86 |
| F | 35 |
| G | 39 |
| H | 125 |
| J | 41 |
| K | 64 |
| L | 38 |
| M | 91 |
| N | 53 |
| O | 31 |
| Q | 62 |
| R | 60 |
| S | 44 |
| U | 52 |
| W | 37 |

Figure 1

Figure 2

Figure 3