Hyperiondev

TASK

Exploratory Data Analysis on the Baltimore 911 Police Calls Data Set

Visit our website

Introduction

The dataset contains information on 911 calls for police assistance. The data is available in CSV format from data.baltimorecity.gov. The data for this project was a combination of the CSV files for 2017, 2018, 2019 and 2020.

The raw data once merged contained 4 923 669 entries and 19 features. Once cleaned the data contained. 4 853 171 entries and 19 features.

The data used contained information on where and when calls occurred, The reason for the 911 calls as well as the responsible police district.

Throughout the EDA, a number of questions were posed and answered through data summary and data visualisation.

DATA CLEANING

The data was read in using the pandas package in python by making use of the read_csv method. The data was then joined using the pandas concat function and merged on the 0 axis, which essentially allowed the individual data frames to be stacked by year from 2017 to 2020.

The data contained a 'callDateTime' feature which gave the date and time stamp of the call. When viewing the data frame using the .info method the feature returned as an object – this likely meant that Python and Pandas were viewing it as a string.

This feature was converted to a datetime object by using the pd.to_datetime method from pandas. Once the feature was converted lambda functions were used to extract parts of the date into their own columns. This would allow for easier group and analysis further on.

From the datetime object, four new features were created by applying lambda functions. These were:

- Hour
- Month
- Day
- Year

Month values ran from 1 to 12, ie January to December. However, day values ran from 0 to 6, ie Monday to Sunday. Interpreting the days feature was more difficult. To help aid easy interpretation and for clear axis labelling when graphing a dictionary was created. This dictionary called dmap had keys from 1 to 6 and values from Monday to Sunday. This dmap dictionary was then used to map new values onto the day feature. This changed the values from 0 to 6, to Monday to Sunday.

There were three columns which appeared to be unnecessary for analysis purposes. These were 'OBJECTID', 'callKey', and 'callNumber'. These would provide little information or grouping benefits. They are likely there as a reference for a particular call, 911 operator etc and are of use to the collators of the data, but of little to no use for analysis.

MISSING DATA

By using the msno package a data visualisation of all missing values were created. Where the visualisation contained missing values the graph displayed a white break from the solid grey.

From the visualisation, run from the sixth cell, it was determined that the 'VRIZones' feature was mostly missing data. As it was an object and text-based data, it would not have been possible to fill.

The decision was made to drop this feature from the data frame.

The remaining missing values rows were dropped. It was decided that this was the best approach given the type of data and also that these few records would have little impact if removed as the data contained over four million records.

DATA STORIES AND VISUALIZATIONS

During the Exploratory Data Analysis, a number of questions were posed. Each will be discussed below.

What is the most common day for calls?

To answer this question a seaborn countplot was used. The order of the data was not correct, ie running from Monday to Sunday. It ran in alphabetical order. To correct this an order command with a list from Monday to Sunday was passed into the plot function.

The plot showed that the total count of calls was similar throughout the week, increase from its work week low on Monday. The peak was reached on Fridays and then declined day on day through the weekend.

What are the call Priorities per day?

To answer this question a seaborn countplot was again used. This time with call priority passed in as a hue effect. This broke up the total calls per day into their priority categories. Non-Emergency calls were the highest every day of the week. While Fridays had the highest total number of calls per day, they did not have the highest number of non-emergency calls. Fridays' tend to have more High and Medium priority calls than other days.

What are the busiest hours for 911 Police calls?

The busiest times for Police 911 calls is between 16:00 to 18:00. With calls decreasing steadily until 22:00 and then spiking around 23:00. After 23:00 calls decline steadily again to their lowest levels between 04:00 and 06:00 with calls increasing as people start their days.

During 16:00 to 18:00, the top ten reasons for calls were:

- Business Checks
- Calls with no voice on the other side
- Disorderly
- Traffic Stops
- Accidents
- Community Engagements
- Common Assault
- Hot Spot Check

- Foot Patrol
- Narcotics

During the quietest times for calls between 04:00 and 06:00 the top ten reasons for calls were:

- Business Checks
- Calls with no voice on the other side
- Disorderly Conduct
- Silent Alarm
- Community Engagements
- Repossessions
- Hot Spot Check
- Repairs / Service
- Common Assault
- Accidents

A second plot was used to try and understand the trend of calls broken down into High, Medium, and Low priority. To see during which times more serious calls take place. A line plot was used to help determine this.

All three follow a similar pattern until around 16:00. At 1600 Low and Medium priority calls start to plateau while High priority calls increase sharply from 16:00 to 17:00. After 17:00 all priority level calls decrease with high priority calls decreasing the least and Low and High priority calls dropping off significantly. Between 22:00 and 23:00, High priority calls experience a spike while Low and Medium priority continue to decline.

What are the proportions of calls by priority?

Of total calls Non-Emergency make up the largest contribution with 42%. These are calls such as business check ins, community engagements, parking and loud music complaints.

The second biggest category are medium priority calls. These contribute 29.7% to total calls. These are calls for reasons such as disorderly conduct, suspicious persons, burglary, and family disturbances.

The third largest category are Low priority call, these contribute 19.2% of total calls. Calls under Low priority are for reasons such as transport requests, general investigations, juvenile disturbances and intoxicated persons.

High priority calls make up the one of the smallest contributions at 10.9%. Calls under High priority are for reasons such as silent alarms, auto accidents, common and aggravated assaults, armed persons, fire and shootings.

Emergency and Out of Service had negligible contributions with less than 0.1% combined.

What are the most common reasons for a 911 call?

The ten most common reasons for 911 calls, visualised with pandas bar plot, are:

- Business Checks
- No voice
- Disorderly Conduct
- Traffic Stops
- Community Engagement
- Common Assault
- Auto Accident
- Hot Spot Checks
- Silent Alarm
- Foot Patrols

What is the Police District Responsible for the most calls?

The police district responsible for responding to the most calls is the Northeastern Police District. When breaking down the total calls by priority this police district receives the highest number of High and Medium and Low priority calls, while it has relatively few non-emergency calls compared to other districts. The call composition of this police district is therefore higher calls of more serious incidents.

What are the top 10 Neighbourhood where Incidents take place?

The Neigbourhoods with the highest number of calls were visualised with a horizontal bar chart. The Neighbourhood with the highest number of calls is Downtown Baltimore. The majority of calls received in this Neighborhood are Medium priority which are events like assault, family disturbances, narcotics, silent alarms and burglary. The ranking of the top ten districts can be seen in the plot under the same heading in the Jupyter Notebook task_18_EDA.ipynb.

How do calls trend during a year?

By plotting a line chart of total calls by month, it was evident that Baltimore experience increased calls during their winter months when they experience cold and snow. What appears to have caused the sharp increase was Business check

ins. This was determined by splitting out the top 10 reasons for calls in months April and May and analysing the trends.

How have calls trended throughout the Years?

From 2017 to 2020 the number of 911 Police calls has increased sharply,
The biggest increase taking place from 2019 to 2020. This was due to a large increase in Business Check-ins from 2018 and a new call category – Community Engagement from 2019. Community Engagement was the second highest call reason in 2019 despite being a new call reason.

The increase of these two call reasons was due to the Baltimore Police Departments focus on community engagement and community policing after the death of Freddie Grey while in Police Custody.

Read about it here:
https://www.policechiefmagazine.org/from-crisis-to-community-policing/
https://www.bbc.com/news/world-us-canada-32400497

Conclusion:
The analysis has given a break down of how calls can be broken down by day, month, year, priority and district. Presenting the composition and trend of each.

However, the ultimate story in the data, was the tragic story of Freddie Grey.
The incident was totally unknown to me when beginning my data exploration. When identifying unique patterns in the data, these were followed up with online searches and investigations. The first was getting an understanding of the seasons in Baltimore to understand why calls increased sharply during certain months.

The second investigation was why community engagement as a call description only appeared from 2019 and why business check-ins increased so drastically. The data showed how sharply calls increased from 2019 onwards. After further investigation into why Community Engagement and Business Check-ins had increased so much it was found to be due to Baltimore Police focusing on Community Engagement and Community Policing after the death of Freddie Grey in 2015. The Police, through these initiatives hoped to restore the Police department and community's relationship.

THIS REPORT WAS WRITTEN BY: Steven O'Ehley