An Analysis of Emergency Calls

Group Name: Trident

Ketan Joshi
Rochester Institute of Technology
Rochester , New York
ksj4205@rit.edu

Vishal Kole
Rochester Institute of Technology
Rochester , New York
vvk3025@rit.edu

Pranit Meher
Rochester Institute of Technology
Rochester , New York
pxm3417@rit.edu

ABSTRACT

Analysis of emergency calls is a crucial issue which needs special attention considering the rate of development. Using the time, location, priority, and reason of emergency calls, the type of emergencies prominent in an area could be analyzed and necessary measures could be taken. We analyze and extract information[1] from the Baltimore city. emergency calls database based on time of emergencies, nature of emergencies and location of emergencies by performing clustering and plotting over heat-maps and graphs.

The information obtained can be used to make necessary arrangements by the emergency preparedness organizations like the police themselves.It could also help in strategic city planning. Our research helps such organizations be well prepared in advance to tackle emergencies efficiently and thereby contributing for the welfare of the society.

KEYWORDS

Big Data, Clustering, 911, Emergency calls, Baltimore, Analysis

1 PROJECT MOTIVATION

Over 240 million calls are made to 9-1-1 in the United States of America. The rate of these calls are increasing drastically every year. These calls comprise of various types of emergencies like medical emergencies (cardiac arrests, unconsciousness etc), criminal emergencies (Murders, Thefts etc), Accidents (Vehicle accidents, gas leaks etc) and even minor crises. It is very important to record these emergencies and classify them under categories with other information like location, time, contact details and the action performed etc.

This information can be used to perform data analysis and extract the information which is not obvious and straightforward. Links between the emergencies and their location etc can be inferred. We focus our research on this domain because of its crucial importance in the welfare of human society. This research will help organizations handling emergencies like the police, fire departments, hospitals etc be better prepared ahead of time and handle emergencies efficiently. The can also help them incorporate laws or establish mechanisms to curb such emergencies.

[1] Copyrights and courtesy to the City of Baltimore.

2 DATABASE DESIGN

2.1 Data-set Description

The database [1] is the information collected from actual 911 emergency call in the city of Baltimore.
Database instances : 2798001
Number of attributes: 9.

Table 1: Attribute description.

Attribute Name	Attribute Description	Attribute Datatype
CallDateTime	Date and time of the call	DateTime
Priority	Priority of the call	String
District	Districy acronym	String
Description	Desciprion of the call	String
CallNumber	Call number (Unique key)	String
IncidentLocation	Address of the call	String
Location	Longitude and latitude	String

2.2 Data cleaning

- After iterating over the instances the findings suggested that, many instances had missing attributes which were removed.
- The attributes "date" and "time" were separated.
- The attribute priority was mapped to the numerical values of the corresponding priorities which were : low, medium, high, non-emergency.
- The findings insinuated that there were 18 districts in the data-set, which were mapped to the numerical value from 0 to 17 each value for a district.
- The attribute "location" consisted of both longitude and latitude which were separated.
- Also one of the prominent thing that we found out is that the description of the calls has spelling errors which makes it difficult to for us to interpret the description even though the description implies same things.
- There were lot of consecutive quotation marks which interfered with the quote character of the database. We had to filter those and remove all the unnecessary characters from the description. The database was cleaned to remove

all missing values, columns segregation was performed using python scripts. Redundancy constraints were examined to remove.

- One of the key problems we faced was that, the address field in data consisted of the block number,street,zip separated by comma hence even though entire address belonged to the same field the program considered as separate values. Using python script the data was divided into tables into CSV files as explained in the relational schema diagram.

2.3 Database Architecture

The database provides the information about the call and the emergency and not the person who has called 911. Hence conceptually the important entities in the scenario are the call location, time, and the emergency priority. We have split the database into two entities viz. Call and entity with 'has an' emergency.

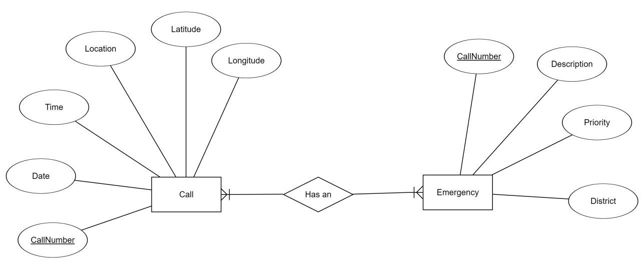

Figure 1: Entity - Relationship Diagram.

the ER-diagram in the figure-1 provides clear insights about entities and their role in the database, Relational schema gives insights about how the data will be stored in the tables. We have

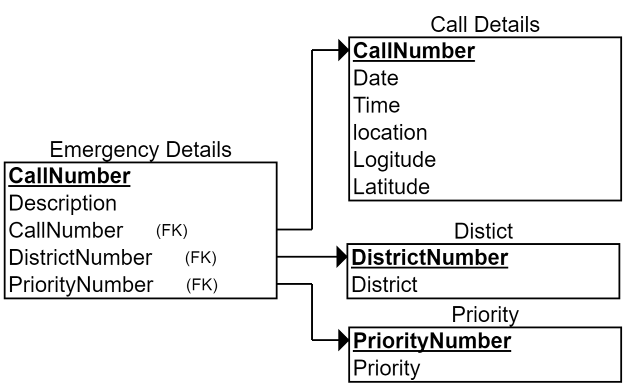

Figure 2: Relational Schema .

split the database into four tables as shown in the figure-2, in the relational scheme diagram.

2.4 Design details

- We have mapped the string data priority and district to numerical value so that the database can have numerical value which will be good for data analysis and achieves database reduction.
- We have split the data in such a way to minimize the redundancy.

3 DATA ANALYSIS

The data-set that we have generated has many numerical attributes viz. Time, priority number, district number, longitude and latitude, which are useful in performing data analysis. We have performed a variety of data analysis to extract the insight of the data and we have also performed visualization of data to have a comprehensive idea and gain as much information as we can about the 911 calls in the city of Baltimore. As said earlier, we have majority of numerical data hence we decided to move forward with the unsupervised data analysis technique which is 'clustering' . The data has naturally occurring groups, hence it will be interesting to see how clustering partitions the data naturally, hence we chose clustering.

K-means Clustering

K means clustering is an unsupervised clustering algorithm. The algorithm has simple working: The algorithm picks random k data points as centroids of the cluster. Each data point belongs to the cluster with the minimum distance from the corresponding centroid.Thus we get k clusters. A new centroid is picked up randomly from the existing cluster members. This process is performed till no new centroids This process is performed till there is no change in the cluster centroid.

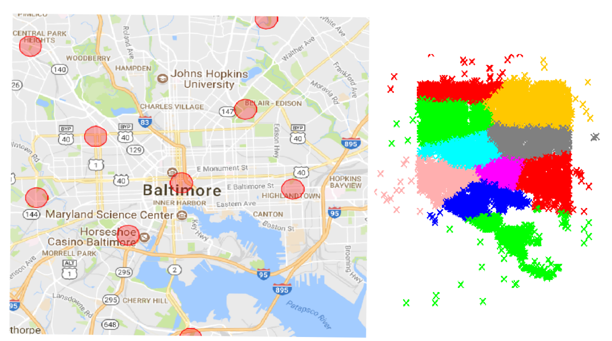

Figure 3: Cluster centers on Google Maps and clusters formed in Weka. K=10

3.1 Location Analysis

For the given task, in the data-set we had the information about the longitude and latitude i.e the exact spot from where the police received the call. We decided to use this information to get clusters of the spot i.e to find the spots in the city from which majority of the calls are received. For this we applied K - mean clustering algorithm with value k = 10 , k =50 with target attributes as longitude and

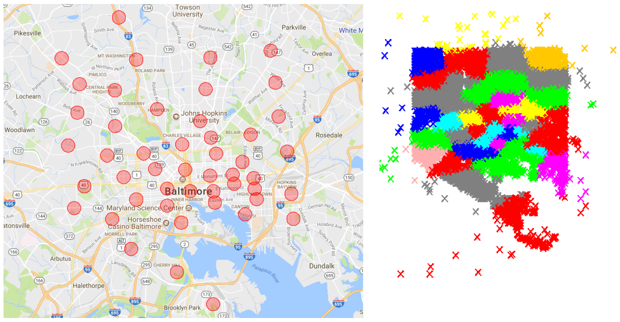

Figure 4: Cluster centers on Google Maps and clusters formed in Weka. K=50.

latitude. The K- means algorithm returned K centroids. The K means algorithm was implemented in 'Weka'. These centroid returned in from 'Weka' are essentially a pair of longitude and latitude on the world map. These pairs were plotted on the map of city of Baltimore. The heat-map of the location were generated using a python script on the actual google maps. The locations highlighted on the heat-maps are actual precise locations obtained from the 2.7 million instances of the data. From the heat-map highlighted locations we can find out the highly busy locations and locality with high number of calls

Figure 5: Large bucket analysis with all priorities.

3.2 Priority Analysis

The calls that the police department of Baltimore received is divided into the priority of the calls viz. Low priority, medium priority, high priority and non emergency with priority numbers 0,1,2,3 respectively. Clustering of the priority

We performed K-means clustering on the data-set with priority as target variable.

The priority analysis returns the cluster locations based on the priority. We can see the Heat-map generated,each priority has given

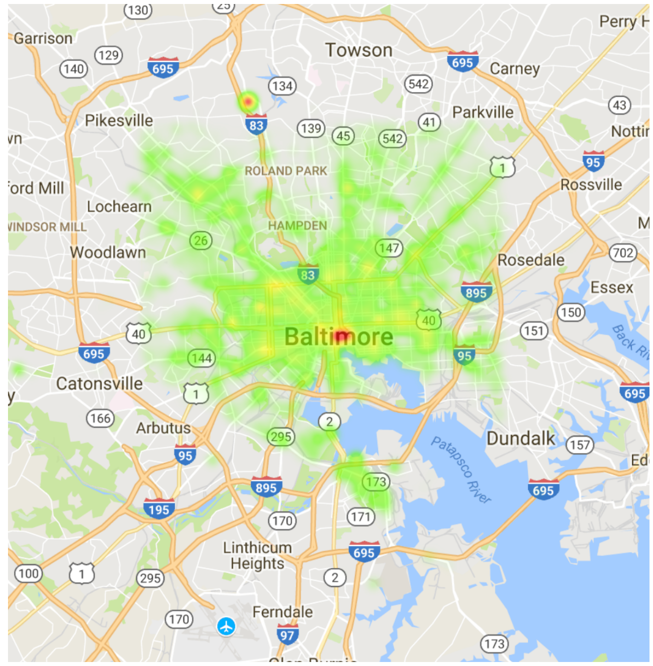

Figure 6: Heat-Map analysis of the complete database

a different color which helps us comprehend which locations have called wit high priority calls, or non emergency calls.

The visualization based on the priority of calls tells us the details of each locations, this information can help the police department to center the attention to certain locations where crime rates are high or here low priority calls are received etc.

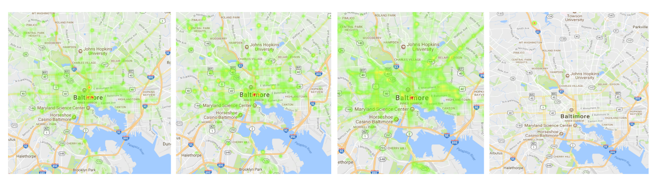

Figure 7: Heat-Map with priority (left to right) Low, Medium, High, Non-Emergency

3.3 Histogram

Histogram of hourly Calls: We are analyzing the behavior of the calls that are being received. A histogram is a representation of the numerical data. The histogram of the hourly calls is plotted on graph with number calls vs hours. The peaks on the graph implies those are the hours with high number of calls.

From the information we can extract the information, that if for a particular hour the department receives large number of calls, then the department can allot more number of employees for those calls. Hence this information is useful.

Figure 8: Heat-Map analysis of the complete database with large bucket size (clockwise) Low, Medium, Non-Emergency, High

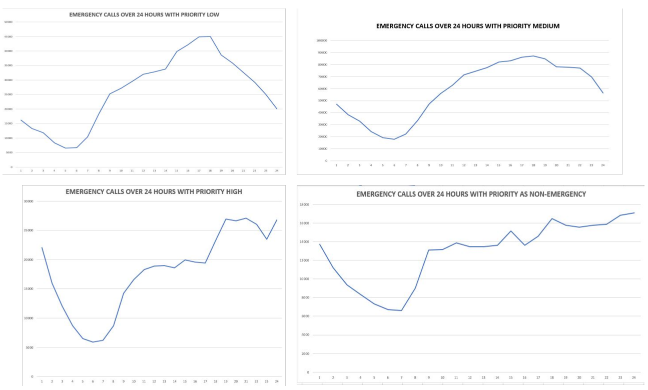

Figure 9: One dimensional Histogram of the database with time with priorities (clockwise) Low, Medium, Non-Emergency, High

3.4 Histogram of priority

The histogram of priority is another aspect to understand behavior of the calls that are received. The priority histogram tells us about the which priority calls we receive the most. The peaks from the histogram tells about which priority calls are received the most. This information will help the department to be prepared for certain priority calls. for e.g a particular hour gets maximum number of high priority calls, So the department team will be ready or will be well prepared in advance.

4 CONCLUSIONS FROM ANALYSIS

Few of the many conclusions that could be inferred from our analysis are as follows:

(1) Information of areas in Baltimore from where high priority calls are more frequent, the locations with maximum number of crimes, the time of the day when the number of emergency calls are high and many other types of information can be analyzed from the research.
(2) On the basis of the visualizations, it is evident that the downtown area of Baltimore is more concentrated with crimes as compared to its suburbs.
(3) Also, the information generated by the heat maps and clusters give precise details of the crimes in all the locations of the Baltimore City.
(4) Also, The time based histogram signifies that for all the crimes over a 24 hour period, the city experiences a high crime rate especially from 3:00 pm to 10:00 pm. This also stands true for crimes with priority as medium and low.
(5) On the contrary, the time based histograms also suggest that for calls for crimes with priorities as high and non-emergency are high throughout the day from 9:00 am and steeply decrease post 12:00 am.
(6) With such inferences, the police and other emergency preparedness organizations like fire stations and hospitals can be more alert for certain times of the day and special care can be taken to avoid such incidents.

5 LESSONS LEARNED

With successful completion of our project, we learned the following concepts and values:

(1) No database is perfect. We learned how to efficiently clean a database and remove redundant, corrupt and useless information from a database.
(2) While performing data cleaning, we learned the language R and make efficient use of Python for data cleaning purposes.
(3) Also, we learned the intrinsic of clustering and especially K-means clustering.
(4) To perform clustering, we learned how the tool Weka developed by the University of Waikato can be a very simple yet powerful tool to perform operations like Clustering and Classification etc.
(5) Moreover, we got a hands-on experience with handling big data and extracting valuable information from it.

6 FUTURE WORK

The database we selected had similar labels with many spelling errors and synonyms conveying the same meaning. One of the future works would be to segregate these attributes with proper labels such that the data can be clustered on the basis of data too. This research can also be extended by making a versatile application which can accept a database of emergency calls from any city and perform analysis on the basis of relevant information from the database.

From the Visualization: k-means - 10, 50 clusters (cluster centers plotted on map), Heat Map with priorities, location based heat-map, time based histogram

REFERENCES

[1] 911 calls, Baltimore City. *URL:https://data.baltimorecity.gov/Public-Safety/911-Calls-for-Service/xviu-ezkt.* ©2017 City of Baltimore.