Analysis of emergency calls

Team Trident

Ketan Joshi

MS Computer Science (Second year)

Pranit Meher

MS Computer Science (Second year)

Vishal Kole

MS Computer Science (Second year)

Motivation

Credits - https://thecomicninja.wordpress.com/tag/motivation/

Over 240 million calls are made to 911

HUGE dataset

Analysis and prediction can be a boon to the mankind

Could help first responders by analyzing the trend in the calls

Hindsight analysis of an event

Cut in the response time with preparedness in advance

System Architecture & Design

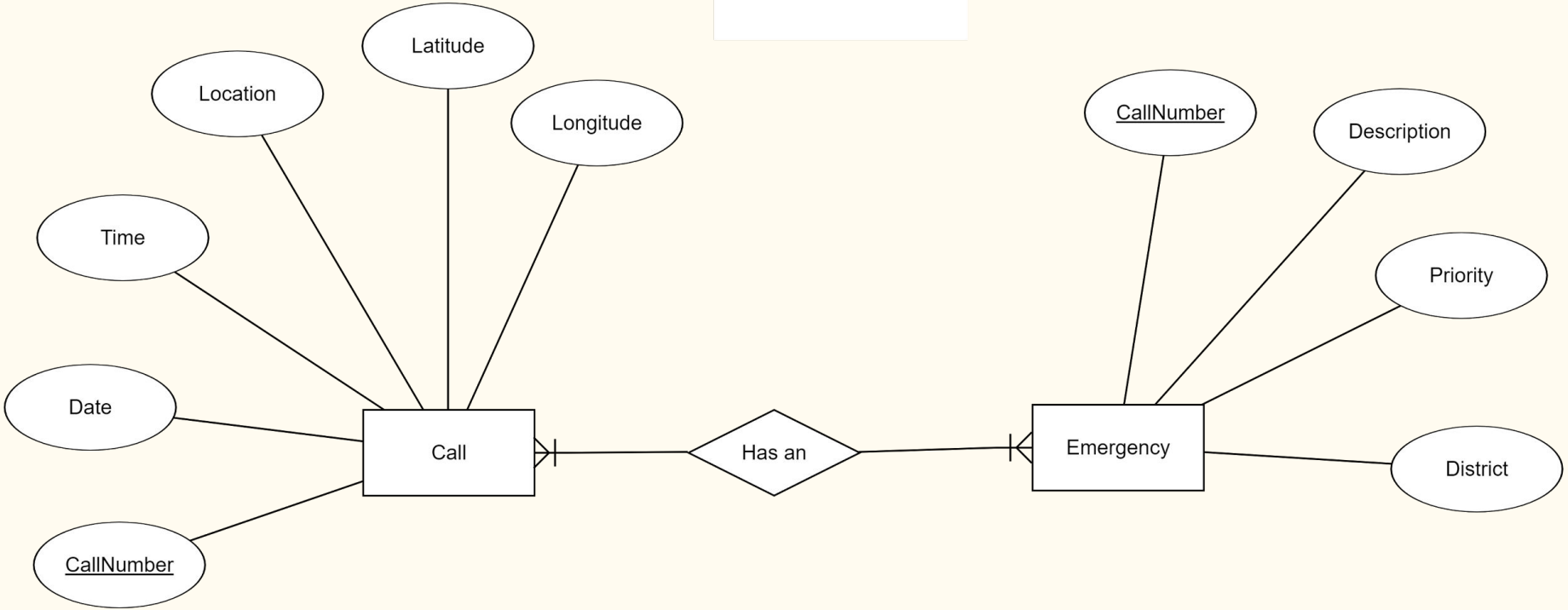
Entity Relationship Diagram

Gives insights about the entities in the dataset and their relationship

The dataset consists of 911 Emergency calls, hence it logical to create entities

Entities created: Call and Emergency

E-R diagram for the database: Baltimore 911 emergency calls.



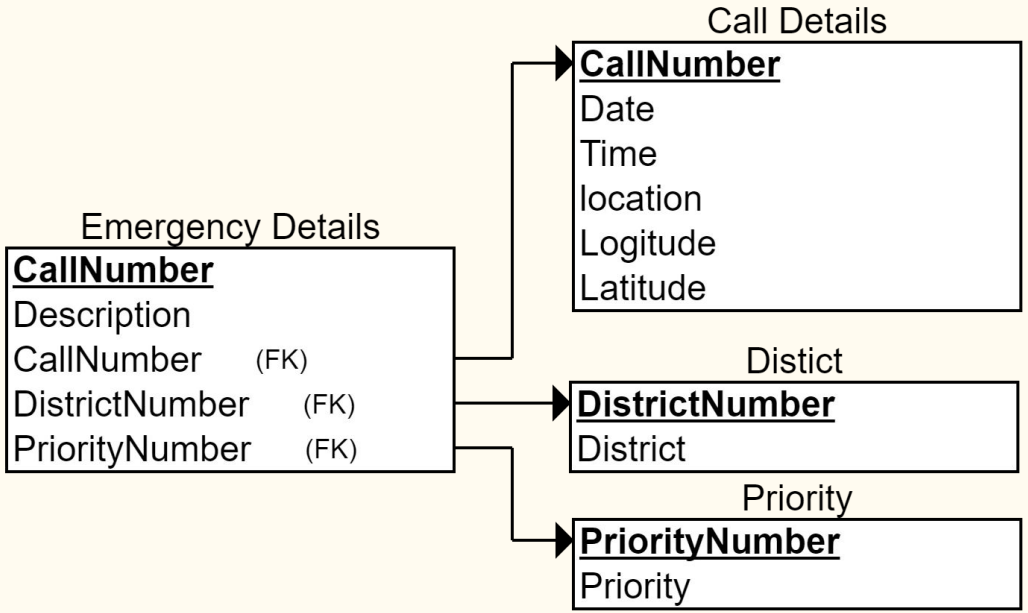
Relational Schema

Relational schema gives insights about how the data will be stored in the tables

Split the database into tables to minimize the redundancies

Handled referential integrity. E.g. The tables emergency details contain foreign keys Call number, priority number , district number

Relational Schema For the Database : Baltimore 911 emergency Calls.



Dataset selection

Credits- Data Science Central
https://twitter.com/DataScienceCtrl

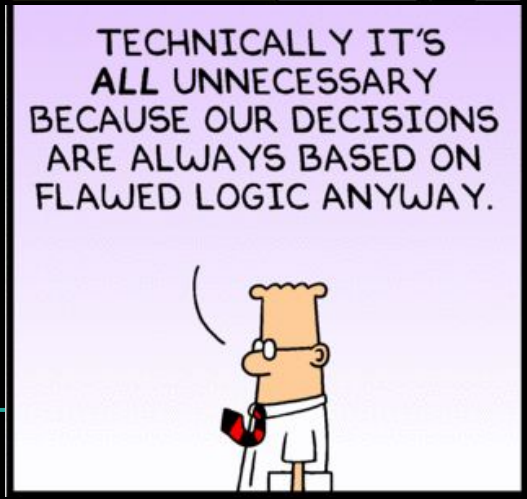
We wanted to work with large datasets and not limit ourselves.

Our dataset contains nearly 3 million records from the city of baltimore

Live dataset which is updated daily

https://data.baltimorecity.gov/Public-Safety/911-Calls-for-Service/xviu-ezkt

Data Cleaning



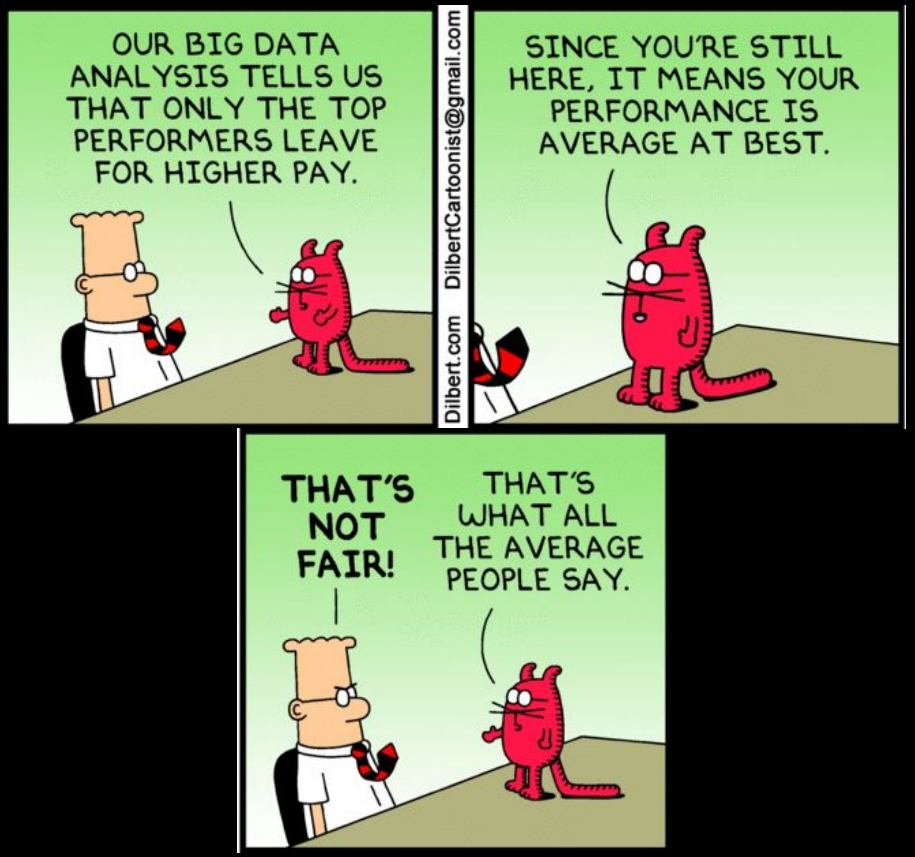
Credits - http://dilbert.com/

Many instances with missing attributes were removed

Two attributes spliced together e.g. date and time , longitude and latitude, which were segregated into two different attributes.

Database Normalization viz. Priority and District mapped to numerical value so that it will be useful for data analysis part.

Analysis



Credits - http://dilbert.com/

K-Means Clustering based on

Location

Priorities

Histogram of Time of events based on

Overall Emergencies

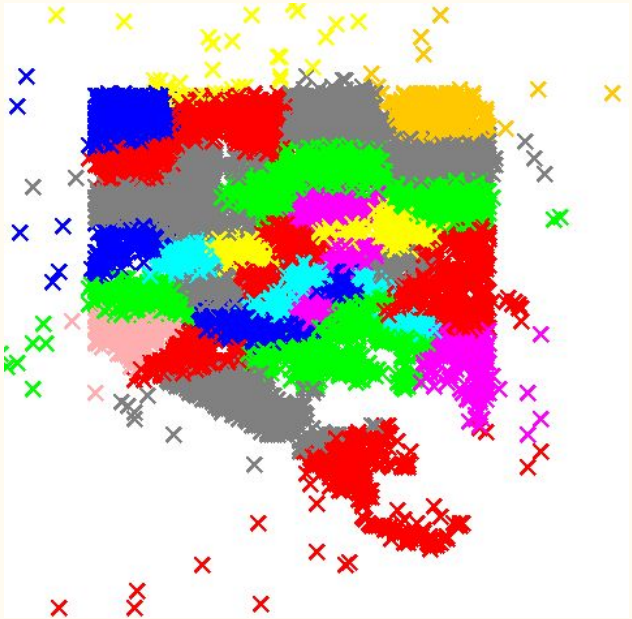
Priority wise

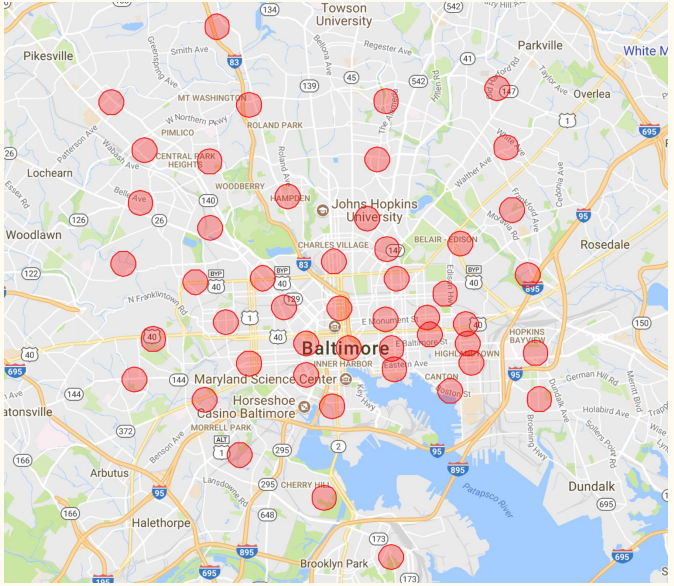
Heat Maps based on

Bucket size

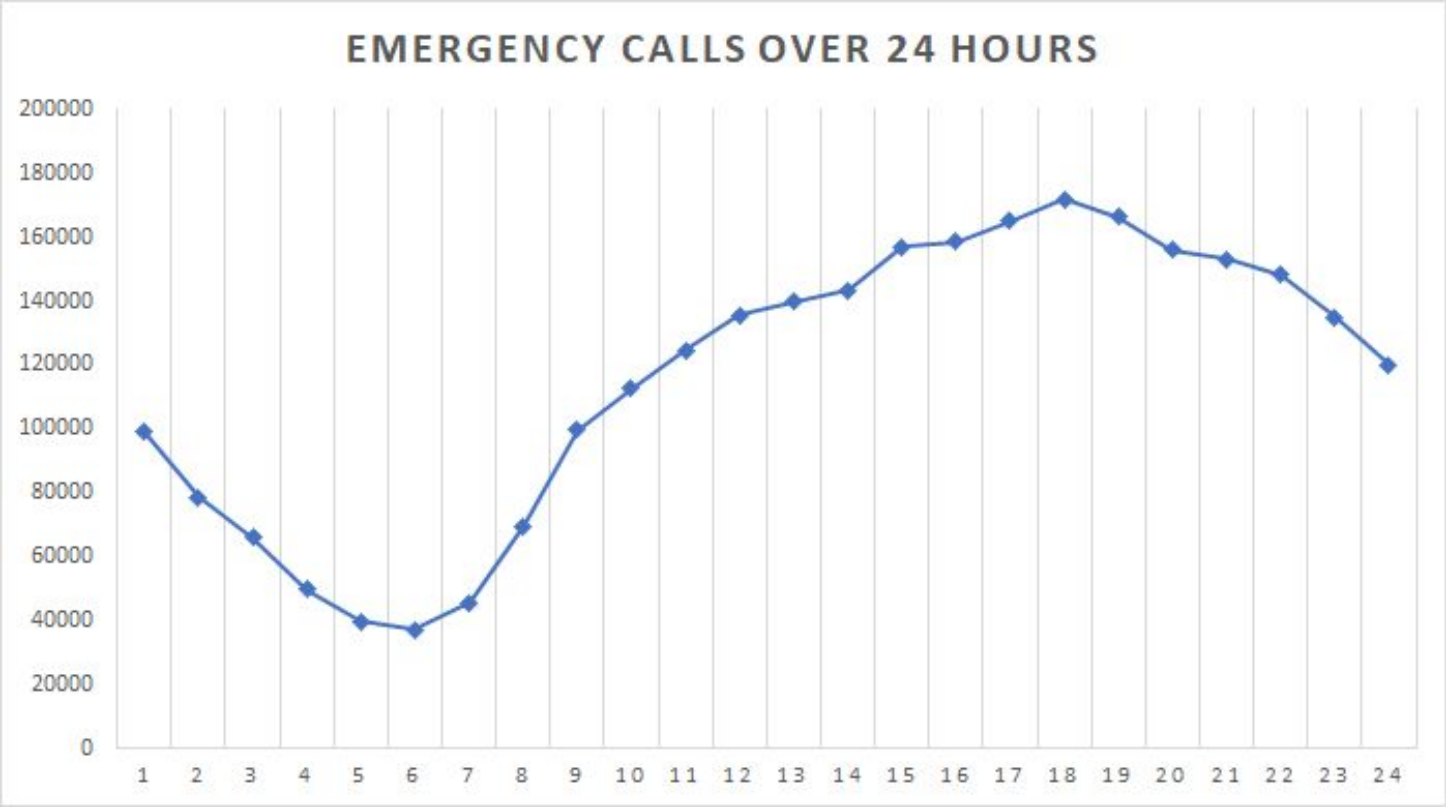
Priority Wise

K-means



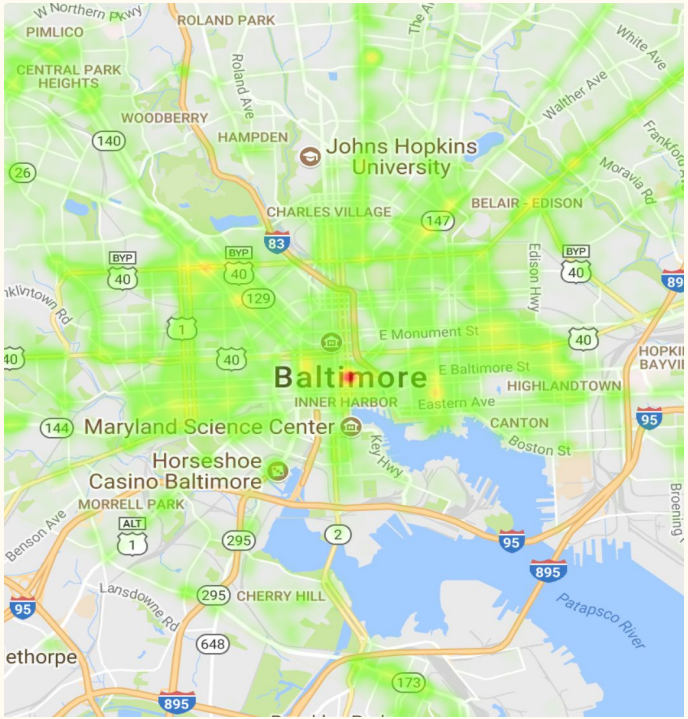

Towson University
Pikesville
Parkville
Overlea
Lochearn
Johns Hopkins University
Woodlawn
Rosedale
Baltimore
Maryland Science Center
Horseshoe Casino Baltimore
Arbutus
Dundalk
Halethorpe
Brooklyn Park

Line Graph


EMERGENCY CALLS OVER 24 HOURS
200000
180000
160000
140000
120000
100000
80000
60000
40000
20000
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

Heat Map


10^4
10^3
10^2
10^1
10^0


PIMLICO
ROLAND PARK
CENTRAL PARK HEIGHTS
WOODBERRY
HAMPDEN
Johns Hopkins University
CHARLES VILLAGE
BELAIR - EDISON
Baltimore
INNER HARBOR
Maryland Science Center
HIGHLANDTOWN
CANTON
Horseshoe Casino Baltimore
MORRELL PARK
CHERRY HILL
Patapsco River

K-Means Clustering would help

Identify small and large clusters of emergencies based on location and priorities over Baltimore which could help us find areas which need special attention

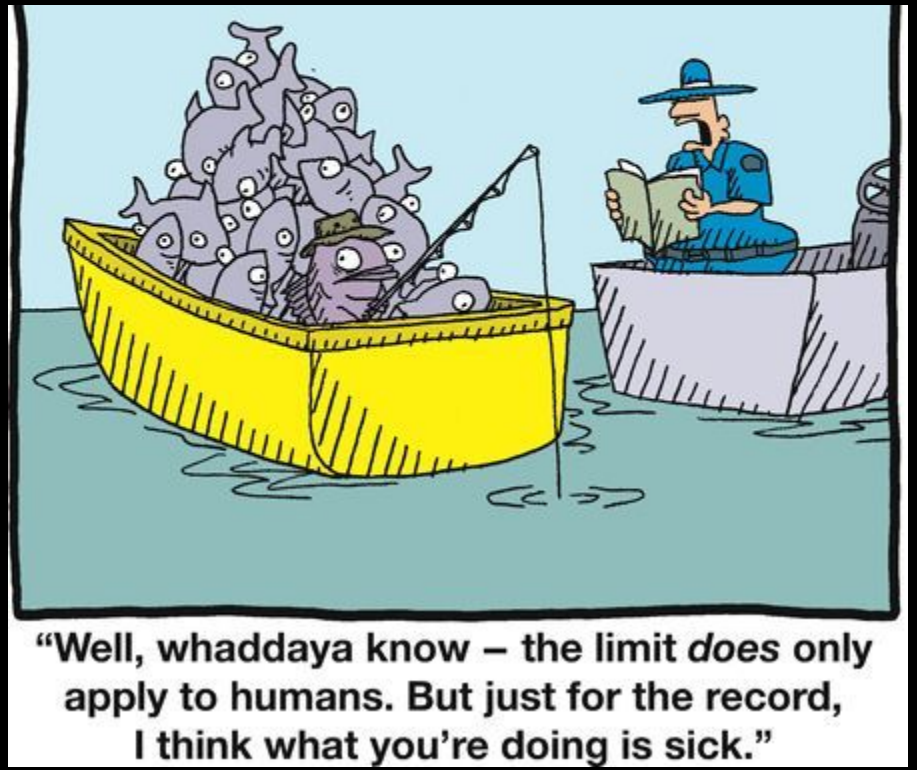
Histogram of Time of events would help

Understand the time of day during which the specific emergencies are prominent again based on locations and priorities

Heat Maps would help

Easily segment the regions with high crime rate or other type of emergencies

Limitations

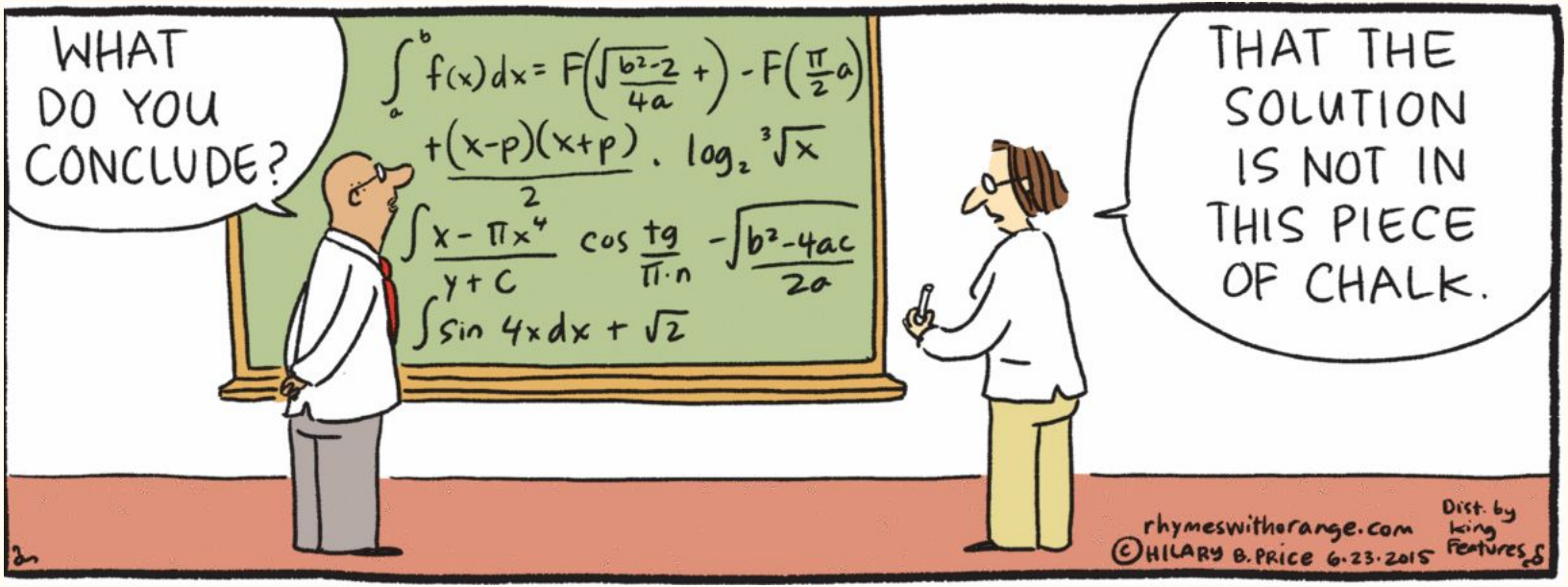


The dataset had vague description about the crime, in addition to that the description had spelling mistakes

hence we could not perform analysis based on crime or the emergency which occurred.

The size of the database quite large for the processing power.

In future we may use a parallel application to perform faster result.



Conclusions

Many conclusions can be inferred from the generated visualizations

The outskirts of the city face less number of emergencies as compared to the core part of the city

The rate of emergency calls increase post 3:00 pm

High priority emergency calls are very frequent from 9:00 am and plummet post 12:00 am

More number of clusters are found near the downtown of the city implying special attention needed in that area

Learning

No database is perfect. We learned how to efficiently clean a database

Learnt to efficiently use R, Python, Weka for Data Cleaning, Clustering, Heat Map Generation

Intrinsics of clustering (especially K-means clustering)

Thank You

Questions?