

REFLECTION REPORT
CRIME ANALYSIS IN BOSTON CITY

Team 4

GIRIDHAR SAMINENI

DEVISRI POTHAKAMURI

ABHINAV PALLEPATI

BRAHMESWARI PAPASANI

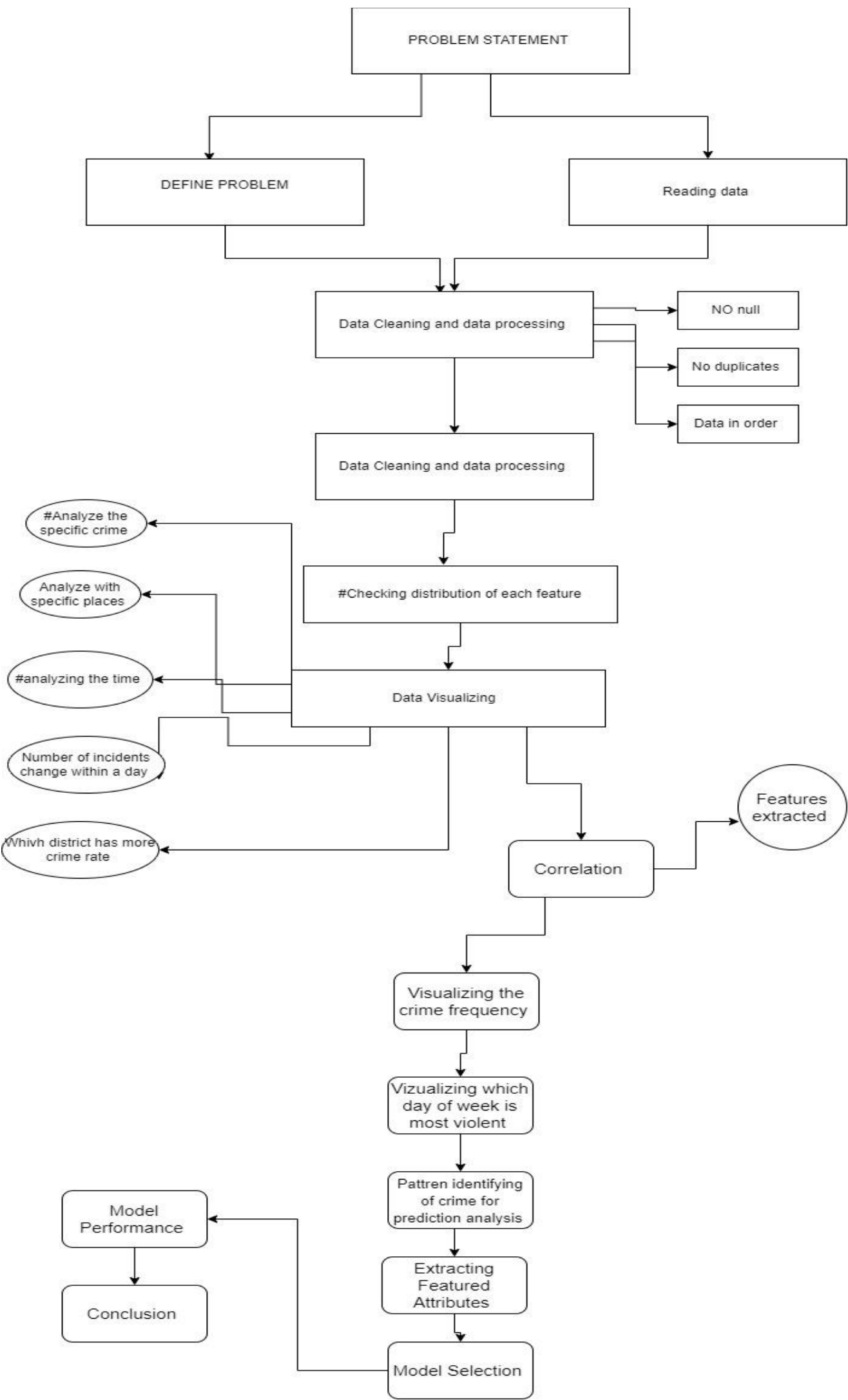
INTRODUCTION

It contains data from the Boston Police Department's (BPD) new crime incident report system from 2018 to 2022; it has fewer fields and is more focused on capturing the type of incident as well as the time and place it happened.

PROBLEM STATEMENT

The actual problem started in the city of Boston. Being one of the largest cities in the world it is difficult to manage the crime activities patrolling team in Boston. Since people there, are suffering from many of the issues and are subject to crimes., it is essential to take some safety measures. The city itself is known for many criminal activities, gang wars, and a large homeless population. So, it is extremely dangerous. This project aims to improve public safety in the city. By using machine learning techniques the following model will predict future crime reports by analyzing the previous data. The data is trained in the model which is been taken from the sources of BPD that are written by hand and there are some inaccuracies with the data. The location, scheduling, and routines of police officers can be made more effective with the help of predictive analysis, however, would improve patrolling in the Boston area. The following are some of the references that explain the seriousness of crime control in Boston.

FLOWCHART



SUMMARY

By using Pandas, we analyzed and visualized the open data of Boston Crime Incident Reports.

Turns out Pandas is indeed a very powerful Python package in terms of extracting, grouping, sorting, analyzing, and plotting the data. From bring the raw unstructured data, merging them together and brought into one dataset.

Later brought with research question where the motivation of the project is build. In building the prediction model it is imprtent to visualize the data.

The following are the outcomes from this milestone crime is more number of times happening that is investigate person by this we can say that the city of boston has more number of investigate person crimes from the years 2018 to 2022 and street washington has more number of crimes the investigative crime is more that to in the street of washington st.The 9th month has the highest amount when compared the crimes with all the 12 months with around 2500 number of crimes approximately. 'Investigating person' has been most reported in Boston in the last years.

Whereas, 'wepon fire arm sale','Burgylary','Kidnaping' are some of the least reported crimes which shows the rate of occurence of these crimes may be very low in the past years.

CONCLUSION

See visualization (Total number crime for each year)

It is hard for predict. It is not possible with this data. We need to more date. For example, promenade zone, street type and so on. The given dataset has only 13 number of coulms although we have done dataslicing (i.e, occured_date trasnformed to date and year separately), It will be very hard to predict the future anaysis with this present dataset. To find any features, prediction for a model we need to have more number of attributes with more featured values so that the pattern for prediction can be identified easily with more accuracy. However, the model had secured a good score, but in the reality it is failed to recognize the pattern for the future analysis.

Not more evaluted (District need to be evaluted more)

While performing location based data analysis it is got to know that there is more redundant data and prediction is very hard since the location attribute can be more classified.

Moreover, we merged all the datasets, perfomed data cleaning opertions, organized all the attributes in an order, and understood them thoroughly. Then we analyzed the data deeeply by performing data analysis techniques. We found few intresting things that cant be visualized unless and until you know the data completely.

It is exciting while transforming the raw data to data visualization, before performing data visualization the data we are unable to read but after every thing had changed and now we can visualize the data effectively.

Using correlation techniques, our team have successfully identified the relation between the each featured column and able to interpret the actual correlation between each of the attributes. Futher, performing more data visualization, we identified which attributes are good for model prediction.

Now the main task has started, since our dataset is full of non numerical values, for any model it is not possible to predict without the non numerical values. So, our team had put in their efforts in converting non numerical values to numerical values and presented the featured attributes.

Finally, this was an intresting journey all through the semester, We derived many things and implemented them in our real world. Also, we have seen the beauty of datascience and got to know how it is transforming and developing the world. During this journey, we had faced many errors especially while building the model, the only motivation we have was to get an outcome ultimately.

Thanks for providing this fantastic course with good curriculum

--Regards Team 4