Project Proposal-Crime in Boston

أكاديمية سدايا
SDAIA Academy

Introduction

- This is first time to do a mini-project in (Exploratory Data Analysis) during my scientific journey. In SDAIA Academy, I learned the most significant tools in Data analysis and we are here to implement the first steps of our projects, which is about 'Exploratory Data Analysis'.
- EDA is an approach of data science process as it concerns in the beginning of collecting data that you are interested to analyze it and get much more familiars with every characteristics consist in your datasets. Then, manipulating and processing these datasets to get more knowledge from EDA. In the final step, you have a vision to make your decision and visualizing your datasets

Data Description

- According to Kaqqle Website, I chose this dataset: "Crimes in Boston". URL: "https://www.kaggle.com/AnalyzeBoston/crimes-in-boston"
- It talk about "Crime incident reports are provided by Boston Police Department (BPD) to document the initial details surrounding an incident to which BPD officers respond. This is a dataset containing records from the new crime incident report system, which includes a reduced set of fields focused on capturing the type of incident as well as when and where it occurred".

Dataset Content

- It contains about 316K Rows and 16 columns. Records begin in June 14, 2015 and continue to September 3, 2018.

Question/need:

- What types of crime are most common?
- Does the frequencies of crime change over the day? Week? Year?
- Which time of the day the most of crime happened?
- Where are different types of crime most likely to occur?
- I would like to participate this project with the "Ministry of Interior's services" in the second Absherthon challenge". As this challenge focus on " Processing, analysing and linking big data with a heterogeneous structure related to crimes to form homogeneous groups that can be read and interpreted automatically and reduce dependence on the manual worker.

Features

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 319073 entries, 0 to 319072
Data columns (total 17 columns):
 #   Column               Non-Null Count   Dtype
---  ------               --------------   -----
 0   INCIDENT_NUMBER      319073 non-null  object
 1   OFFENSE_CODE         319073 non-null  int64
 2   OFFENSE_CODE_GROUP   319073 non-null  object
 3   OFFENSE_DESCRIPTION  319073 non-null  object
 4   DISTRICT             317308 non-null  object
 5   REPORTING_AREA       319073 non-null  object
 6   SHOOTING             1019 non-null    object
 7   OCCURRED_ON_DATE     319073 non-null  object
 8   YEAR                 319073 non-null  int64
 9   MONTH                319073 non-null  int64
 10  DAY_OF_WEEK          319073 non-null  object
 11  HOUR                 319073 non-null  int64
 12  UCR_PART             318983 non-null  object
 13  STREET               308202 non-null  object
 14  Lat                  299074 non-null  float64
 15  Long                 299074 non-null  float64
 16  Location             319073 non-null  object
dtypes: float64(2), int64(4), object(11)
memory usage: 41.4+ MB
```

Goals

- I hope to model this project professionally and add distinctive features, for example, upload the map inside the Jupyter and determine the crime locations to predict the future crimes and avoid it by tightening security in places where crimes are common.

Tools\Libraries:

- Python and Jupyter Notebook
- Numpy and Pandas for data manipulation
- Matplotlib and Seaborn for plotting visuialization
- PowerPoint for presention
- Maybe need additional tools.