

Ohood Soliman

### Abstract

The main goal of this project is to find out the basic concepts of data science, know when and how to apply Exploratory Data Analysis on a data set. Therefore, it was important to collect a dataset from any open data resources and explore it by using common EDA techniques and tools.

This this project aims to describe the most common types of crimes, show the frequencies of crimes change over the day? week? Year?, clarify the day with the highest number of crimes and the most dangerous places in Boston. In fact, we analyzed this data that have been collected from Kaggle.com and visualizing the result in addition to gain insight into a future criminal works

### Design

The dataset "Crimes in Boston" were used in the study, 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". This project built to show the workflow of exploratory data science and classify crimes accurately, via analyzing and visualizing the data that would be enable the Boston Police Department to achieving safety and security. Data contains information about the crime such as date, location, crime group, crime code.

### Data

It contains about 316K criminal records and 16 columns determine:

'INCIDENT_NUMBER'	'OFFENSE_CODE_GROUP'	'LOCATION'		
'OFFENSE_DESCRIPTION'	'DISTRICT'	'REPORTING_AREA'		
'OCCURRED_ON_DATE'	'YEAR'	'MONTH'	'DAY_OF_WEEK'	'HOUR'
'UCR_PART'	'STREET'	'LAT'	'LONG'	

## Algorithms

### Feature Engineering

1. Mapping latitude and longitude coordinates so show crimes places in reality.
2. Cleaning data many times such as missing values and duplicate or unnecessary data.
3. Managing columns of data. For example, rename data columns and dropping columns.
4. Cleaning string data, like analyzing text data and checking for substrings.
5. Exploring new data. Such as add a new columns, display the data type of each column and figure out how these data help answer the

Project questions?

6. Diagnosing errors. For example, Create a custom Python functions to create new columns.
7. Comparing against group statistics.

### Tools

- Python and Jupyter Notebook
- Numpy and Pandas for data manipulation
- Matplotlib and Seaborn for plotting visualization
- Wordcloud for data visualization
- Folium for for visualizing geospatial data

## Communication:

### Number of crimes each year



