IBM Capstone Project

■ **Submitted by Sashank**

The business problem –

To analysis the Boston Crime data.

To find the different types of committed in the city, correlation between the different crimes.

This analysis is done to understand the nature of crime and the relationship between them.

Doing this could shed light on various reasons and possible ways to prevent some crimes. We are also analyzing crimes based on different neighborhoods of the city.

Data –

Source 1 –

https://data.boston.gov/dataset/crime-incident-reports-august-2015-to-date-source-new-system

Details Regarding the Source 1 -

TITLE	Crime Incident Reports (August 2015 - To Date) (Source: New System)
TYPE	• External • Tabular
DESCRIPTION	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. Records in the new system begin in June of 2015.

PUBLISHER	Department of Innovation and Technology
TEMPORAL NOTES	This dataset contains records of crime incident reports using the new system starting in June of 2015.
THEME	Public safety
LOCATION	Boston (all)
CONTACT POINT	Boston Police Department
CONTACT POINT EMAIL	mediarelations@pd.boston.gov
LICENSE	Open Data Commons Public Domain Dedication and License (PDDL)

The CSV contains – **17 Columns and 494021 Rows**

Columns Data Types -

Sl No.	Columns	Type
1	INCIDENT_NUMBER	Int
2	OFFENSE_CODE	Int
3	OFFENSE_CODE_GROUP	String
4	OFFENSE_DESCRIPTION	String
5	DISTRICT	String
6	REPORTING_AREA	String
7	SHOOTING	String
8	OCCURRED_ON_DATE	Date
9	YEAR	Int
10	MONTH	Int
11	DAY_OF_WEEK	String
12	HOUR	Int
13	UCR_PART	String
14	STREET	String
15	Lat	float

16	Long	float
17	Location	String

Source 2 - https://en.wikipedia.org/wiki/Boston_Police_Department

Contains the details regarding the different district codes.

Created a table from the above code to get the district names.

With the present data we have extensive records of crime which can be analyzed with the help of K Mean Method and clustered.