**Data**

This data set contains statistics, in arrests per 100,000 residents for assault, murder, and rape in each of the 50 US states in 1973. Also given is the percent of the population living in urban areas.

**Business Problem**

**Objective:** Target States which need high priority to control crime rates.

**Constraints:** High Urban Population

|  |
| --- |
| **Data Dictionary** |

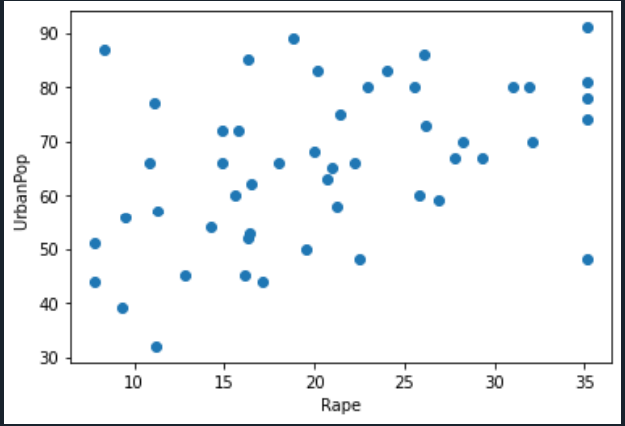
|  |  |  |  |
| --- | --- | --- | --- |
| **Name of Feature** | **Description** | **Type** | **Relevance** |
| X | 50 U.S States where crime occured | Quality, Nominal | Irrelevant |
| Assault | |  | | --- | | Numeric assault arrests |   (per 100000) | Quantity, Interval | Relevant |
| Murder | |  | | --- | | Numeric murder arrests |   (per 100000) | Quantity, Interval | Relevant |
| UrbanPop | Numeric Percent Urban Population | Quantity, Ratio | Relevant |
| Rape | |  | | --- | | Numeric rape arrests |   (per 100000) | Quantity, Interval | Relevant |

**Data Pre-processing**

* Imported crime\_data .csv dataset.
* Na values --- no
* Duplicate values --- no.
* Outliers --- present in Rape.
* Winsorization technique is used to remove outliers.
* Standardise the all columns except X column.

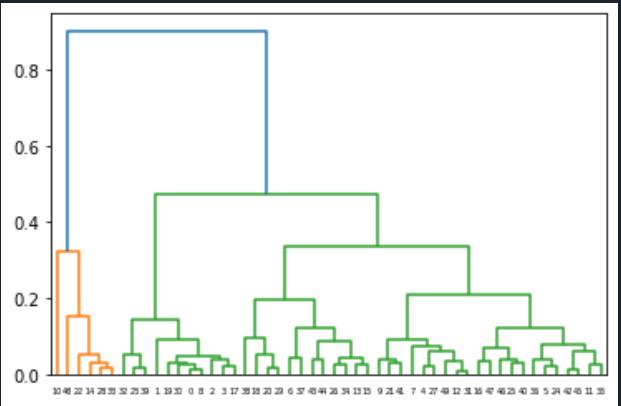
**Exploratory Data Analysis**

* Symmetrical --- all except Rape is Positively skewed
* Platykurtic kurtosis --- all columns
* Scatterplot --- Rape and UrbanPop



**Model Building**

* Found distance between 2 records using Euclidean Distance.
* Found disimilarities using farthest neighbour using distances.
* Plotted Dendrogram.



* Cut the Dendrogram into 3 clusters.
* Forming matrix with 2 clusters.
* Converted matrix into dataframe #moving data into “crime\_data" dataset
* Combining these data sources into a summary for data analysis using aggregation function.

**Solution**

* Cluster 2 states having high urban population with high assault and rape values. So Cluster 2 countries are targeted to reduce assaults and rapes.
* Cluster 1 states are having high murder rate, so cluster 1 is to target to reduce murders.
* Cluster 3 states are low assault, rape and murder rates.

Priority order of target to control crime: Cluster 2>Cluster 1>Cluster 1