Name: Ganji Kiran; **Email ID**: ganji.kiran99@gmail.com; **College**: NIT Calicut; **Ph.No**:9495176026;
Team Name: Data_Cruncher; **No. of team members**: 1(Individual)

METHODOLOGY

- Google Fusion tables and MS Excel were used in the preparation and analysis of the data given.
- Topic chosen: Statistical Sherlock.
- The data was segmented into 3 main categories namely “Crime Data”, “Police Data “and “Crime Genre” wherein a part of the first sheet document fell into crime genre category.

Crime Data

- Year wise data of total number of crimes reported, segregated state-wise are created and histograms are plotted to observe the trends.
- A dataset consisting of total number of cases, state-wise, spanning over ten years was created and histograms are plotted to observe the overall trend.
- Histograms pertaining to year wise total crimes are plotted (does not include the state based segregation) to observe the crime growth across the country.
- Combining the entire data, a graph was plotted showing all crimes committed within different states year-wise.
- Calculations like growth percentages, crime cases reported, segregations based on the persons committing crimes are done and included in the info graphic.

Police Data

- Data sets consisting of year wise police data, segregated state wise, are created and respective histograms are plotted to observe the trends.
- Above given procedure is repeated for both police strength and sanctioned police to get a clear trend on difference between sanctioned and police strength.
- State wise sanctioned police and police strengths, segregated year wise, are created and histograms were plotted.
- Different strengths of police according to cadre are created and segregated on the basis of gender.
- Calculations like percentages of police strengths, sanctioned police and cadre of the police in different states are calculated.

Crime Genre

- The datasets pertaining to different educational and financial backgrounds of juveniles arrested are created from the first sheet.
- From the third sheet, different age groups of the population committing crimes are segregated and put into separate documents.

- Related histograms, pie-charts and graphs are constructed to observe the trends in the different sets of population performing the crimes.
- Crime type and population type data sets are mapped to create a dataset. Calculations related to the maximum and minimum reported cases in each category are constructed.
- The datasets that have been created in the beginning have been mapped with the state and year which helped in giving helpful insights.

ALLOCATION OF THE POLICE:

The main factors considered in allocating police for different states are:

- **Crime growth rate of individual states**: Dataset of reported crimes mapped year wise is created. The growth rate is calculated and a histogram is constructed for to observe the trends of the growth rate in different states. Particularly last 3 years were taken into the account for the final result deduction.
- **Current allocation of Police**: It is fair to assume that, Indian government allocates the police, year-wise, based on the reports of the previous year. So the current allocation remains a prominent factor and acts as a base which can be improved upon.
- **Crime genre growth rate**: Of the different available crime categories like dacoit, theft, murder etc. the data set has been prepared to observe the trends in the respective category in the consecutive years. Though the data had less impact on the final conclusion, it provided one of important trend regarding the Dacoit cases in Bihar.
- **Obvious ones**: Few states like Uttar Pradesh, Maharashtra, Madhya Pradesh etc. are known for their notoriety. Hence it is fair to assume that, since there is steady population growth, there is a steady crime growth rate and hence a higher percentage of police should be allocated.

Based on the above factors primary and secondary hot lists are created which required an increase in allotment. Primary hot list states require a higher increase in the police allotted than the secondary hot list states. Since the exact percentage of increase cannot be determined by the data available, a rough estimate of 2% increase for primary and 1% increase for secondary is suggested as probable solution.