**Capstone project overall story in your own words. Min 1000 words.**

**Introduction**

This project is about the crimes in India, in which we analyse the different states/UT, district and its crime rates. As we all know India is highly diversified country, it has different culture, people, languages also it is popular for the high population.

One of the biggest challenges of India is the increasing crime in different places. As the population increase and resources are limited is the one of cause of increasing crime. There are number of variables to cause crime in India. The main contributing factor is poverty; Unemployment, illiteracy, and a lack of basic necessities are all effects of poverty. This causes people to become frustrated and desperate, which ultimately end in criminal activity. If we analyse then get an idea about somehow those factors are associate with increasing population.

This project is divided into 4 phases, first phase is Data Collection, second phase is State/UT wise analysis, Third phase is analysis by SQL operations and last phase is using the Clustering Algorithm.

**Phase 1**

This is belongs to Data Collection, in which data of population, area and literacy rate of different states and UTs are collected and store in a excel file. Data of crimes already present in csv file.

The data of population is taken from Wikipedia, It consist only 1981, 1991,2001 and 2011 data and we need the data from 2001 to 2012, so from the previous data, the data for these years are predicted by using the best fit line of data and those predicted data has been used as for analyses. The area of different states taken from statics times (https://statisticstimes.com/geography/india/indian-states-area.php). As in mostly data the Andhra Pradesh and Telengana is a one state, so for the rest of the data we consider it is as one state. The Area of Dadra and Nagar haveli and Daman and Diu is taken as separately.

Literacy of different states and UTs id taken from Wikipedia. I got 2001, 2011 and 2021 data, so i have divided that data into all remaining years with equal increasing rate.

**PHASE 2:**

Phase 2 is about analysing the different crimes according to their year, population, area of states and UTs and literacy rate. The analysis is done by state/UT wise not district wise.

First, I have analysed Total crime v/s literacy. To analyse this, I have chosen the IPC crime because it covers almost all crime and then remove the rows having total in its district column. A scatter plot was plot between them and no such correlation we find through that. But if we see overall then find a negative correlation between them. It is analyse that Kerala, Gujarat and Bihar approx. same crime rate but having huge difference in its literacy rate. Kerala having the high literacy rate and Bihar having low literacy rate whereas Gujarat falls in moderate literacy rate. So, it is can be concluded that literacy rate is not a major factor which impact the crimes.

Now, I have analysed the types of crime with state and literacy rate. There are total 29 types of crimes are mentioned named as murder, attempt to murder ,culpable homicide, attempt to murder, rape , custodial rape , other rape , kidnapping & abduction , kidnapping and abduction of women and girls , kidnapping and abduction of others , dacoity , Preparation and assembly for dacoity , robbery , burglary , theft , auto theft , other theft , riots , criminal breach of trust , cheating , counterfeiting , arson , hurt/grievous hurt , dowry deaths, assault on women with intent to outrage her modesty , insult to modesty of women , cruelty by husband or his relatives , importation of girls from foreign countries , causing death by negligence and other ipc crimes.

I have analysed it with different states and it is concluded that Uttar Pradesh shows maximum murder, attempt to murder, culpable homicide not amounting to murder, kidnapping and abducation, kidnaping and abducation of women and girls, criminal breach of trust, Dowry deaths cases. Madhya Pradesh shows maximum number of Assault on women with intent to outrage her modesty, rape, other rape case and other IPC crime cases. Death by Negligence is maximum in Tamil Nadu and followed by Andhra Pradesh; both have 10% difference in Literacy rate. Tamil nadu have more literacy rate then A.P. Mizoram shows lowest crime of Cruelty by Husband or Relatives and West Bengal show maximum number of this crime followed by Andhra Pradesh, Rajasthan, U.P. and M.P, In UTs, Delhi shows maximum of this crime. Maharashtra shows maximum theft, robbery and burglary cases. Maximum Dacoity cases reported in Bihar followed by Maharashtra. It is concluded that women related crimes majorly happened in U.P., A.P., Rajasthan, west Bengal and Bihar. If we talk about UTs, Majorly different crimes are happened in Delhi except Riots, Pondicherry show the high riots as compared to Delhi.

Followed by, I have analysed for state v/s literacy and concluded that Kerala shows the maximum literacy rate i.e. 92.5% and Bihar shows the lowest literacy rate i.e. 55%. In UT, Lakshadweep shows the highest literacy rate i.e. 89.6% followed by Chandigarh, Delhi and Pondicherry approx. 84%.

After this a year wise analysis has been done and concluded mostly states/UT shows drop in total crime in 2003 and 2005. Maximum crime shown in 2012, after 2005 the crime rate is increase yearly.

Lastly in this step, analysed type of crime vs. literacy. And conclude that Types of crime do not directly depend on literacy rate. The different states are involved in different crime and having different literacy rate. Few stated like U.P, Rajasthan, A.P and M.P having almost same literacy rate that comes in the category moderate and showing the nearby number of crime rates whereas Tamil Nadu and Kerala have high literacy rate but show similar number of crimes.

In followed step I have analyse crime according to its area and population. The Total crime and area shows the strong positive co-relation that means on increasing the area there is high chance of increasing of crime but is not valid for all states but most of the states having this relation. Chhattisgarh and Tamil nadu having approximately same area but have huge variation in crime rate. Lakshadweep has the smallest area and lowest total number of crimes. Madhya Pradesh has highest number of crimes and has moderate area. Rajasthan has the maximum area, but shows the less crime as compare to few states having less area than Rajasthan. Overall population shows the positive correlation with total crimes. States/UT having low population shows the less total crime. Lakshadweep has the lowest population and crimes. UP shows the high population but the total crime is less than few states which have moderate population. If we analyse for UTs only then Delhi shows the high population and high crime rates

**Phase 3**

Phase 3 is about doing analysis by using SQL operations.

From phase 3 we conclude that:

1. Mushidabad of west Bengal shows the maximum number of rapes and kidnapping in 2009 year.

2. District Dahoh shows the maximum dacoity and maximum robbery i.e. 29 and 32 respectively

3. Lowest number of murder happened in district Kondagaon and Mungeli

4. In Bihar, Patna sees the most murders each year. And if see other cases also, it is conclude that most of district which shows maximum murders are belong to the capital of city or having economic value. For example Mumbai commr. Of Maharashtra, Bangalore commr. Of Karnataka and etc. One of the reasons can be overcrowded population.

5. Bihar shows the high number of murder, Mostly murder cases were happen in 2004 in Bihar but over the year it shows the decrease in number of murders as compared year 2004 cases.

**Phase 4**

Phase 4 is based on the unsupervised learning, in which Clustering ML is used. The 3 clusters are form on the basis of total crime and total area of different districts. The cluster defines the area according to its crime rate.

From this phase it is analysed that there are few states and UT comes under fully peaceful area. That means, that states/UT have the low crime rate. Few examples are Sikkim, Goa, Dadra nagar haveli, Daman and Diu, Lakshadweep etc, the reason can be not overcrowded population.9 states having the sensitive area, and Kerala shows the high percentage of sensitive area as compare to other states. Andhra Pradesh shows the high percentage of moderate area as compare to other states.

Mumbai is the most unsafe district and longleng is the safest district. Most of the cities are metropolitan, having the high population which make the city overcrowded whereas the area is not that large, that may be reason to increase crime rate in this area. Because most of the city is economic hub, so many people migrate to different city to these cities, which cause the higher crime rate. The fact that migration frequently causes a sharp rise in population density, which can strain available resources, housing stock, and social services, increasing poverty, unemployment, and social inequality. These circumstances can play a role in a number of crimes, including violent crime, theft, and property crime. This crime can be reducing by proper surveillance, proper following of rule regulations, spreading awareness among people about their rights especially village’s women.

**Conclusion:**

Population and area of state playing an important role in increasing crime rate, both are correlated to each other whereas if we see literacy rate, it not clear that it is having some direct relation with all states because It is analysed that Kerala, Gujarat and Bihar approx. same crime rate but having huge difference in its literacy rate. Kerala having the high literacy rate and Bihar having low literacy rate whereas Gujarat falls in moderate literacy rate. From phase 2 it is concluded that in UT Delhi have the high population and high crime rate where in literacy rate it is come at third position, Lakshadweep shows the high literacy rate and lowest crime and lowest area. A & N have the maximum area. In states, Kerala and Bihar show maximum and minimum literacy rate respectively whereas both show approx. same crime rate. Goa have the lowest area and Rajasthan have largest area the crime rate in goa is low. U.P show high population with moderate total crime rates and Sikkim show low population with low crime rate. Maximum crime shown in 2012, after 2005 the crime rate is increase yearly. Form phase 4, it is concluded most of the cities are metropolitan, having the high population which make the city overcrowded whereas the area is not that large, that may be reason to increase crime rate in this area. Because most of the city is economic hub, so many people migrate to different city to these cities, which cause the higher crime rate.