Data Science Work

5. Recommendation System for crime prediction:

In this session, we discuss how the recommendation model helps to predict a crime or how the recommendation model comes into the picture to predict a crime. We also see that how this technique can be modelled as user and item.

5.1 User and Item:

In any recommendation system, there are two main factors are User and Item, in our crime prediction the two factors are time and area/location. Similar in any recommendation system to the rating a user gives to an item that reflects the interaction between the user and the item, the number of crimes in an area/location at a given time also reflects the criminal interweaving between the location and time factors. Thus, we consider time (year in our case) as the user and area/location as an Item. We have some reasons to select these two factor. Firstly, users are much more than items, items are in some limited quantity but time can be increase as time passes just like users but the location/area is limited it can be countable. Secondly, when we talking about the crime are mostly happen area/location comes in the picture it is more appropriate to represent the human factor as the area/location, rather than as a time. We will need a technique to solve our problem so, we used the collaborative filtering technique and content filtering technique in the recommendation system.

5.2 Collaborative filtering technique:

Collaborative filtering (CF) is a technique used in the recommendation system and the motive for collaborative filtering comes from the idea that people often get the best recommendations from someone with tastes similar to themselves. It is widely used in product recommendation systems. There are User-Item collaborative filtering and User-User collaborative filtering. With User-item CF, items with similar ratings are ranked and recommended to the user. With User-User CF, user similarity is calculated based on the rating they give to items, and items from similar users are ranked and recommended to the user. We take User-item CF because it is easier to recommend the item with the same taste as other users but while using User-User CF it is harder to recommend because Users opinions change with time.

The first step in the CF is to find the similarity matrix. To find the similarity using a cosine similarity:

$$SIM(i, j) = Cos(\theta) = \frac{\hat{\imath} \cdot \hat{\jmath}}{||\hat{\imath}|| * ||\hat{\jmath}||}$$

where row i is the embedding for user i, where row j is the embedding for item j. Higher the angle between the i and j, lower the cos value, lower they are similar to each other

and Lower the angle between the i and j, higher the cos value, higher they are similar to each other.

In our Collaborative filtering system the user can search the crime number our system will show him the District or location where the exact crime number has been searched that is the half part that the user search and the result is shown what they want and In the half part our system will recommend them extra 30 District or location which are highly similar to the user search. Fig.1. show some part of the code how we achieve the Collaborative filtering system.

```
In [170]: rate = {}
          rows_indexes ={}
          for i, row in pivot_m.iterrows():
              rows =[x for x in range(0,len(pivot_m.columns))]
              combine = list(zip(row.index,row.values,rows))
              rated = [(x,z) for x,y,z in combine if str(y) != 'nan']
              index = [i[1] for i in rated]
              row_names = [i[0] for i in rated]
              rows_indexes[i] = index
              rate[i] = row_names

In [172]: items_dic ={}
          for i in range(len(pivot_table.T.index)):
              item_idx = item_indices[i]
              col_names = pivot_table.T.index[item_idx].tolist()
              items_dic[pivot_table.T.index[i]] = col_names

In [173]: topRecs ={}
          for k,v in rows_indexes.items():
              item_idex = [j for i in item_indices[v] for j in i]
              item_dist =[j for i in item_distances[v] for j in i]
              combine = list(zip(item_dist,item_idex))
              diction = {i:d for d,i in combine if i not in v}
              zipped = list(zip(diction.keys(),diction.values()))
              sort = sorted(zipped,key = lambda x:x[1])
              recommendations = [(pivot_table.columns[i],d) for i,d in sort]
              topRecs[k] = recommendations
```

Fig.1. Code for achieve Collaborative Filtering System

Fig.1.1 shows that the user enter the crime count and our system will recommend him the district with exactly same number of crime count.

```
In [176]: def getrecommendations(user,number_of_recs = 30):
              if user > len(pivot_table.index):
                  print('out of range, there are only {} users,try again'.format(len(pivot_table.index)))
              else:
                  print("there are all the DISTRICT  you have viewed view:\n\n{}".format('\n'.join(rate[user])))
                  print()
                  print("we rcommend to view these too:")
              for k,v in topRecs.items():
                  if user == k:
                      for i in v[:number_of_recs]:
                          print('{} with similarity: {:.4f}'.format(i[0],1-i[1]))

In [177]: getrecommendations(200)
          there are all the DISTRICT  you have viewed view:

          ADILABAD
          AHMEDABAD COMMR.
          AIZAWL
          ALAPUZHA
          ALIGARH
          ANAND
          AURANGABAD RURAL
          BAREILLY
          BARMER
          BELLARY
          BETUL
          BHABHUA
          BHARUCH
          BHAVNAGAR
          BHOPAL
          CHITTOOR
          DARRANG
```

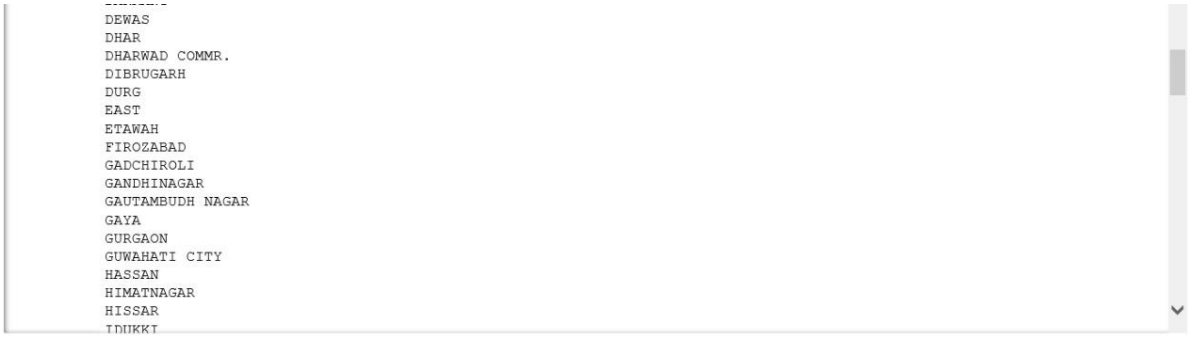


Fig.1.1 District that user search through Crime Count

Fig.1.2 show the 30 District which is recommended by us with their similarity which user search earlier.

Fig.1.2 District that we recommend the user

Here, you can see that the “HATHRAS” has the higher similarity with the user search District

It is 0.7181 and list is going on in decreasing order. The 30 District that we recommend to the user is totally mobile the administrator of the recommendation system can change according to their suitability.

5.3 Content filtering technique:

Content filtering is one of the technique in the recommendation system. The content or attributes of the things you like are referred to as "content." The motive behind content-based filtering is to classify products with specific keywords, learn what the user likes, look up those terms in the database, and then recommend similar things. This type of recommender system is hugely dependent on the inputs provided by users.

In our content-based recommendation system users search with a specific District and district is connect with the year, state, crime and crime count so if user search any district we recommend them to district which are related with the year or crime count or state or crime so the output maybe larger.

Fig.2. the recommendation system show the District which are related to users search which based on the crime, and there count, state and year also. In content based recommendation system there is no need show the detailed part the administrator can show only the related only. You can see in Fig.2.1 only related search part only.

```
In [68]: dataset["combine_feature"] = dataset.apply(combine_feature,axis=1)
         print(dataset["combine_feature"].head())
```

```
Searching_District: DISTRICT                ADILABAD
STATE_UT           ANDHRA PRADESH
YEAR                         2001
Crime                      MURDER
CrimeCount                    101
combine_feature              None
Name: 0, dtype: object
Searching_District: DISTRICT               ANANTAPUR
STATE_UT           ANDHRA PRADESH
YEAR                         2001
Crime                      MURDER
CrimeCount                    151
combine_feature              None
Name: 1, dtype: object
Searching_District: DISTRICT                CHITTOOR
STATE_UT           ANDHRA PRADESH
YEAR                         2001
Crime                      MURDER
CrimeCount                    101
```

Fig.2 District recommended by the system with details

```
In [20]: item_name = [("ANANTAPUR"),("ANANTNAG"),("ALLAHABAD")]
         similar_item = pd.DataFrame()

In [21]: for DISTRICT in item_name:
             similar_item = similar_item.append(get_similar(DISTRICT),ignore_index=True)
         similar_item.head()
         similar_item.sum().sort_values(ascending=False)
```

```
Out[21]: ZUNHEBOTO
         GOLAGHAT
         GUMLA
         GULBARGA
         GRP(RLY)

         NAUGACHIA
         NASIK RURAL
         NASIK COMMR.
         NARSINGHPUR
         24 PARGANAS NORTH
```

Fig.2.1 District recommended by the system

In Fig.2.1 the system recommended the District which are related to there crime count and it show the only district not their details.

5.4 Hybrid Recommendation System:

Hybrid recommender systems combine two or more recommendation strategies in different ways to benefit from their complementary advantages. Collaborative technique

Show the same item have same taste of other person too and Content technique is show what the user search in there past and there are more recommendation system present in our surrounding. The motive to use the hybrid recommendation system to recommend the more accurate and specific item to user so, that user feel friendly with the environment.

There are different hybridisation Designs:

- Parallel use of several system
- Monolithic exploiting different features
- Pipelined invocation of different systems

Just by putting different recommendation system in different position it will recreate a hybrid recommendation system. The motive of all hybrid recommendation system is to recommend more accurate item.

6. Time series analysis and forecasting:

Time Series Analysis and Forecasting states that any information periodically recorded with time can be used for forecasting a future event related to the information. In India where criminal activities take, place more frequently. By applying modern technology forecasting techniques to these cities crime data, future crime rates can be forecasts. This project analyses crime data and gives various visualizations for easy understanding of the results. It also uses past 11 years crime data from official website of NCRB to forecast future crime rate. For Time Series Forecasting in this paper forecasting method 'ARIMA' (Auto Regressive Integrated Moving Average) has been used.

To visualize how crime evolved over time in the India we plot the number of Crime per year from 2001 to 2012. The graph (Fig. 1) shows that Crime in the India has been increased year after year with continuous Incline.

'ARIMA' is also consider as a non-recommendation system. It consists of an autoregression model (AR) that captures lagged patterns, and a moving average model (MA) that captures long terms trends.

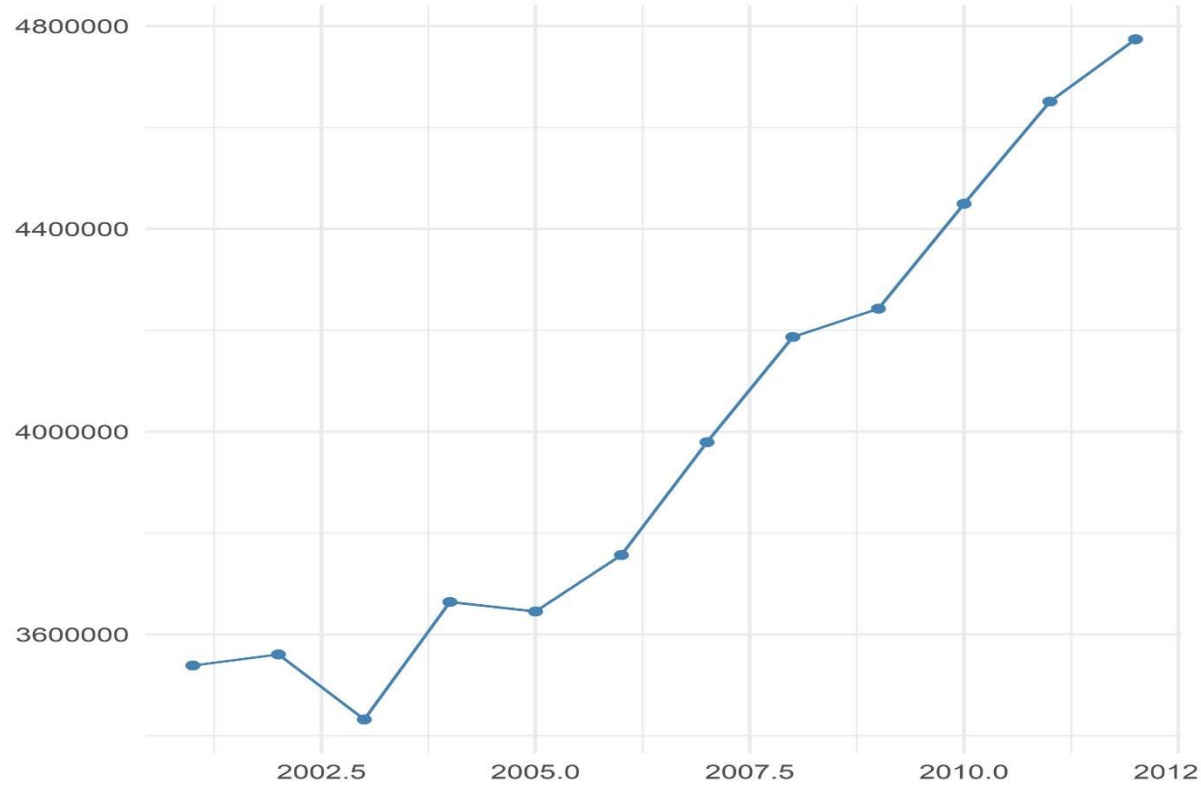


Fig. 1 Number of total IPC crimes versus year

As we can see that the graph is continuously increasing 2001 to 2012. We will continue observing this thing deeply in current dataset from 2013 to 2020.

Fig.2 depicts the annual frequency of crimes per type and their trend. The most common types of crimes are theft and hurt_grevious_hurt they are continuously increasing from 2016 to 2017.

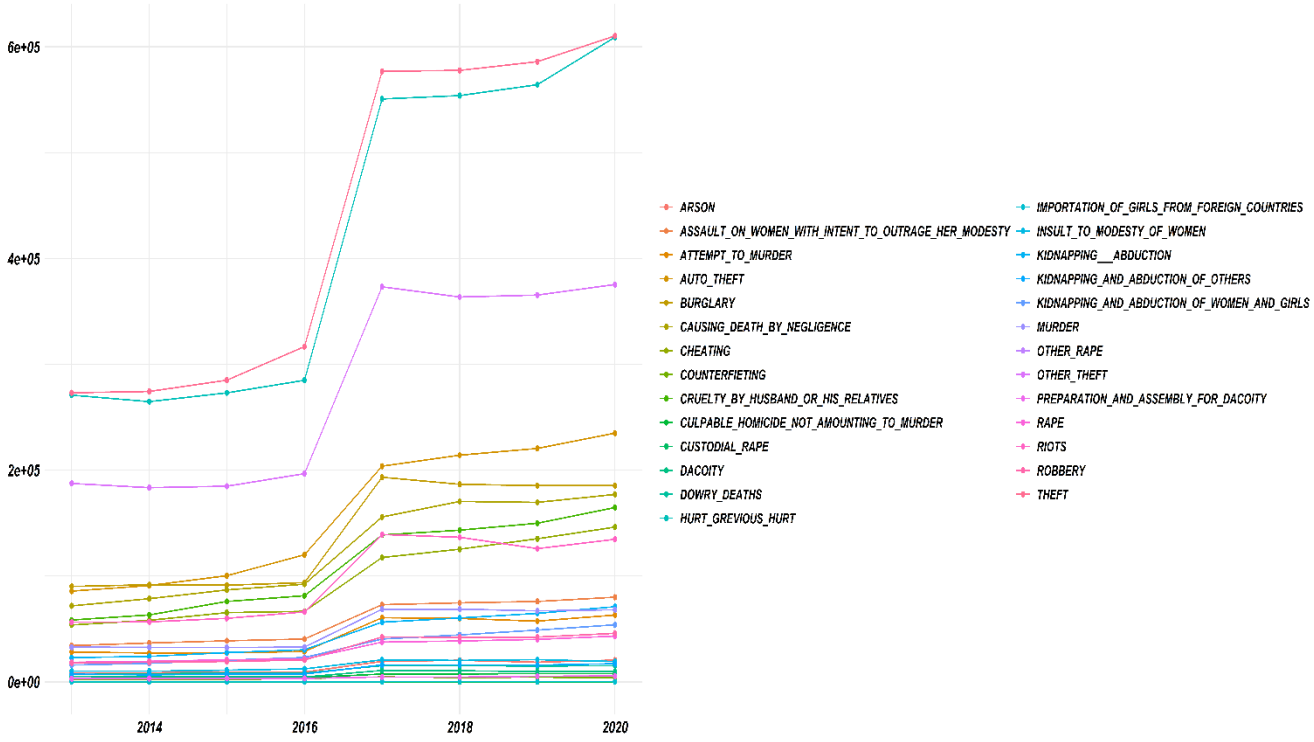


Fig.2 Crimes evolution per type of crime 2013-2020

As we can see in Fig.2 there are lot of crime and they created a mess because most of them are lesser in number to visualize it more clearly, we would see only those crime whose crime count are greater than 10,000 or 10k and to see there trend individually we have Fig.3. In that figure we can some of the crimes are decreasing, some of them are increasing and some of the crimes are start increasing rapidly after some years like Kidnaping and kidnapping of women and girls.

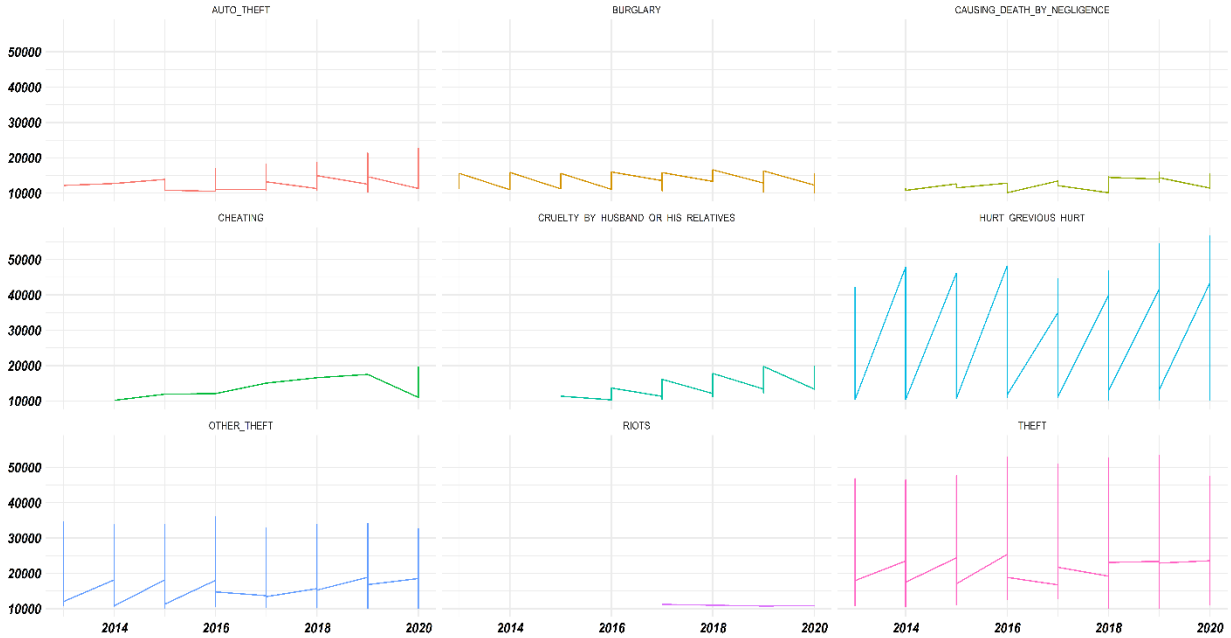


Fig.3 Crime evolution per type individually and who's crime count >= 10k

Now we want to know that the flow of crime count(>10k) throw the states in India from 2013 to 2020 to know this we use Chord diagram (The primary use of chord diagrams is to show the flows or connections between several entities (called nodes). Each entity is represented by a fragment (often coloured or pattered) along the circumference of the circle. Arcs are drawn between entities to show flows). See Fig.4. there are 14 states which has crime count greater than 10k.

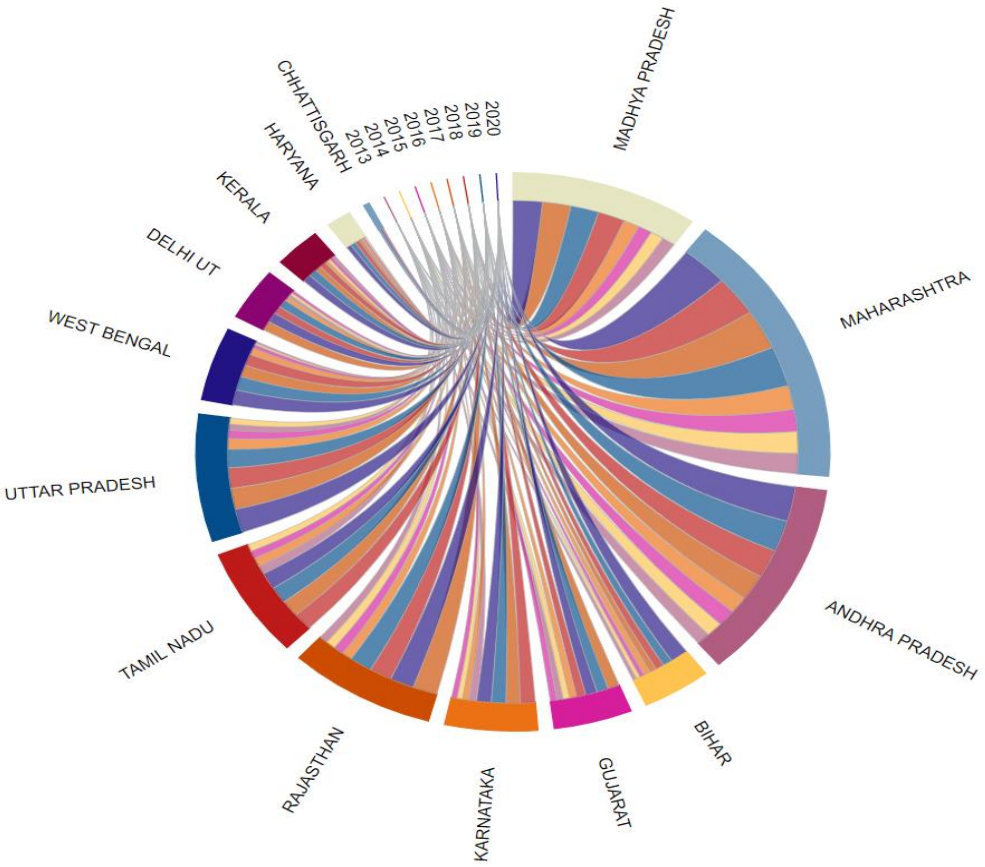
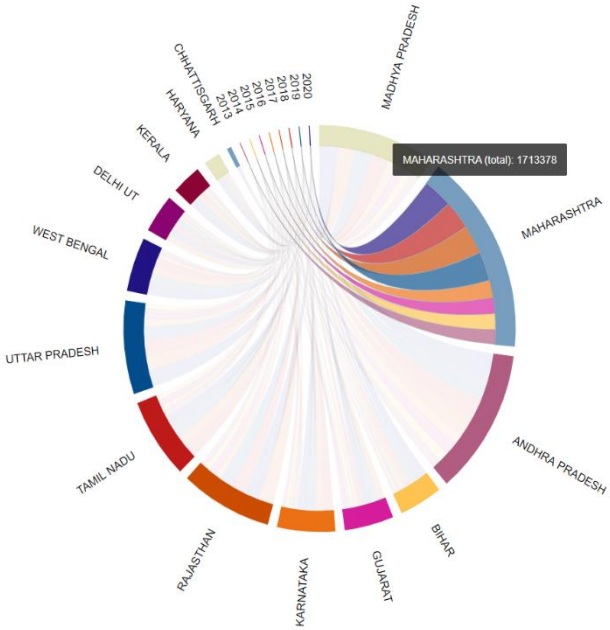


Fig.4 Flow of crime count in India from 2013 – 2020

If you want the total number of crime count of any specific state just point on that state the graph will show the crime count like in Fig.4.1. you can see that the total number of crime count of Maharashtra is 1713378 in 2013-2020, if we point on the other state it will the total crime count of that state. The line connect the arc of state and the year are in sequence the first line of the arc show that which year has the maximum number of crime count like Fig.4.2. we can see that the first line of the state Maharashtra show that in 2020 the crime count is maximum (284988) and the sequence is going on decreasing order last line of the arc has minimum (139334) number of crime count Fig.4.3.



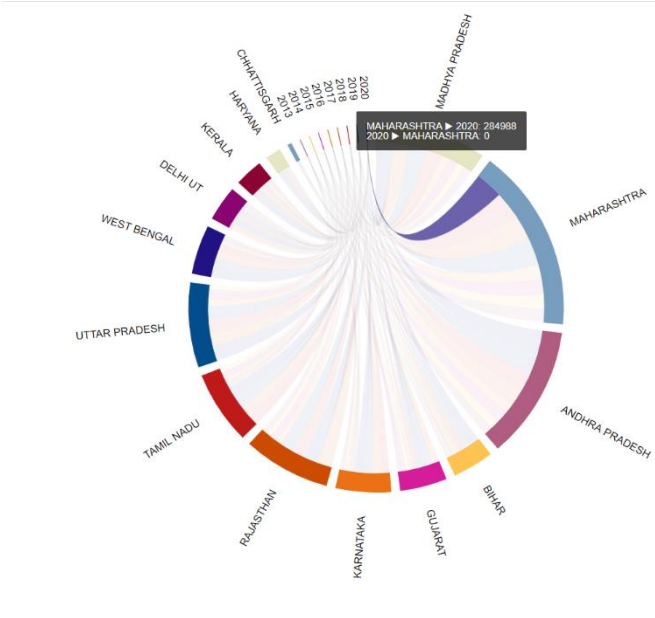


Fig.4.1 Total Crime Count of Maharashtra

Fig.4.2 In 2020 Maharashtra has the Maximum crime count

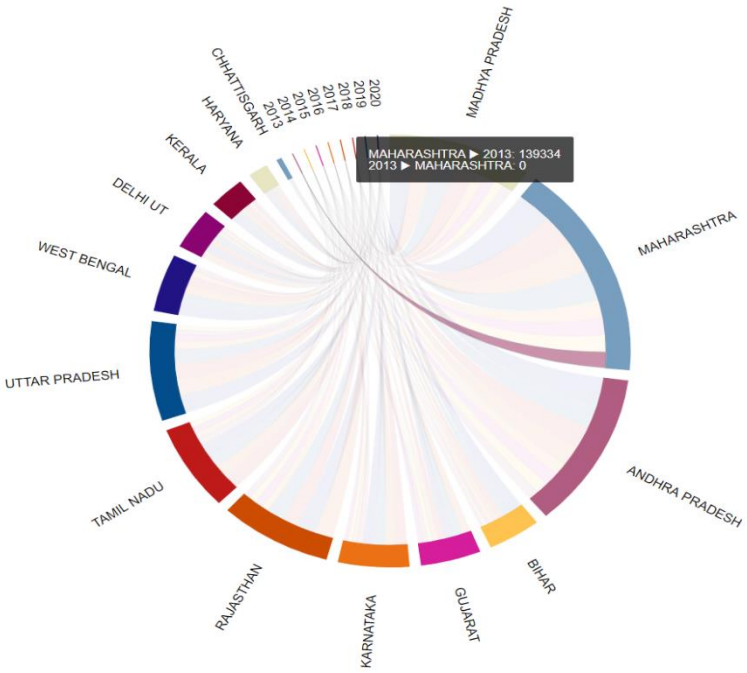


Fig.4.3 In 2003 Maharashtra has the Minimum crime Count

If we visualise the distribution of crimes per location and type in Fig.5, we can see that some types occur in specific States. To see that we use the heat map it shows that more the colour is darker more the specific crime has occur in specific state.

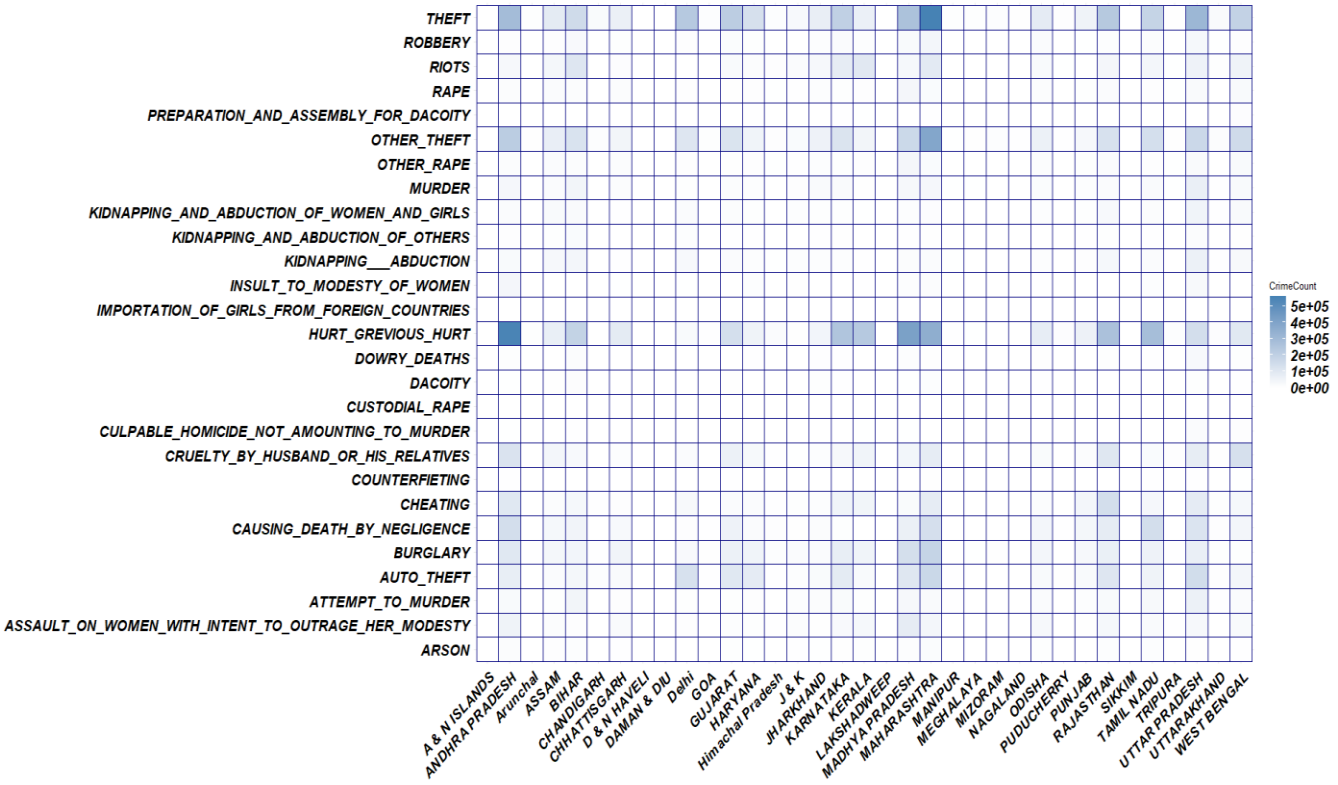


Fig.5 Types of crime Vs States

Now to see the which type of crime is mostly happen in India Fig.6 is show the trend that of the maximum crime. Theft and Burglary are the maximum crime which follow the increasing tend and the other crime are also the maximum crime but they are in some of the states in some years.

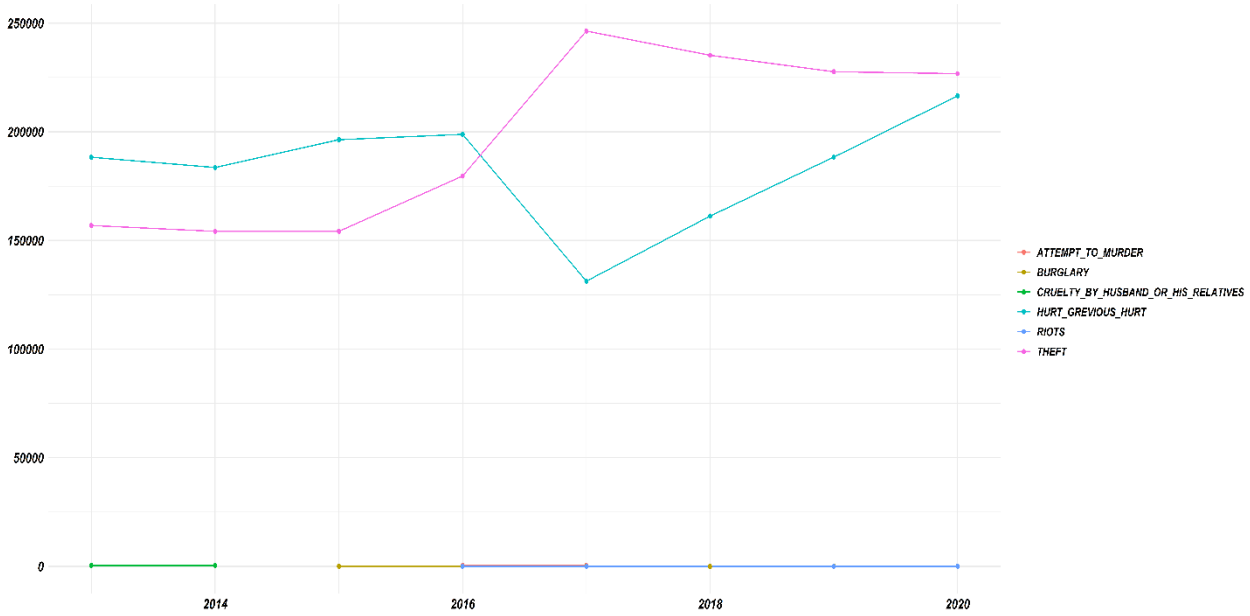


Fig.6 Trend of Maximum Crime in States in 2013 - 2020

Fig.7 show the flow of the maximum crime in states with the years throw chord diagram. More details about the graph are given in Fig.7.1, 7.2, 7.3.

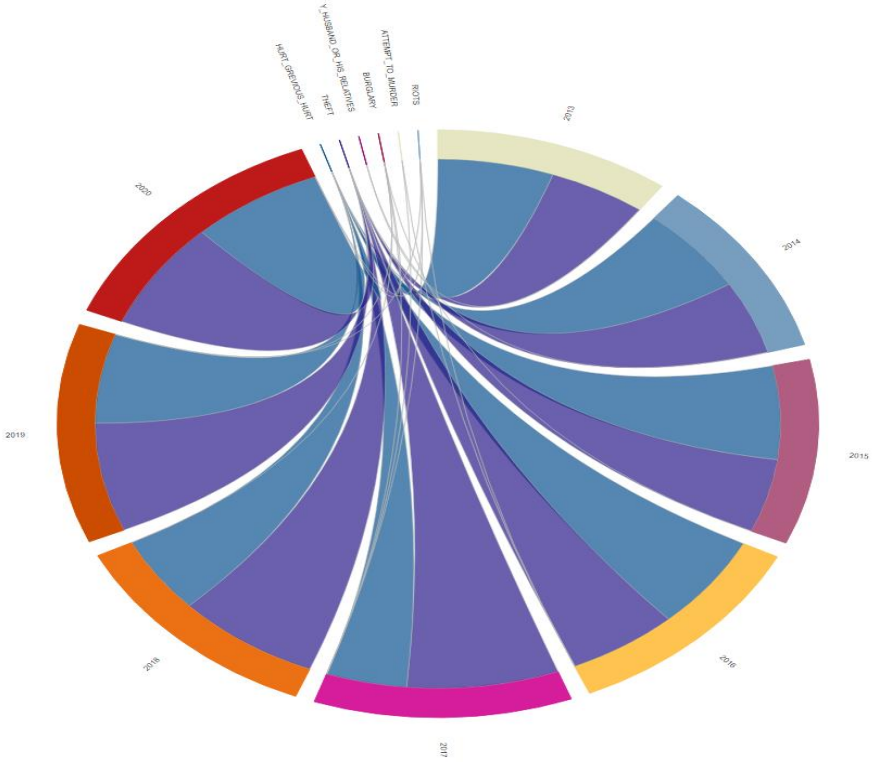


Fig.7 Maximum crime type in states from 2013 - 2020

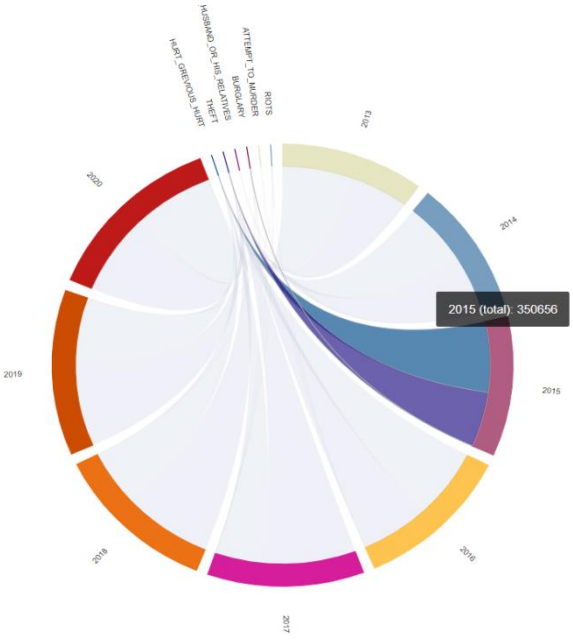


Fig.7.1 Total maximum crime in 2015

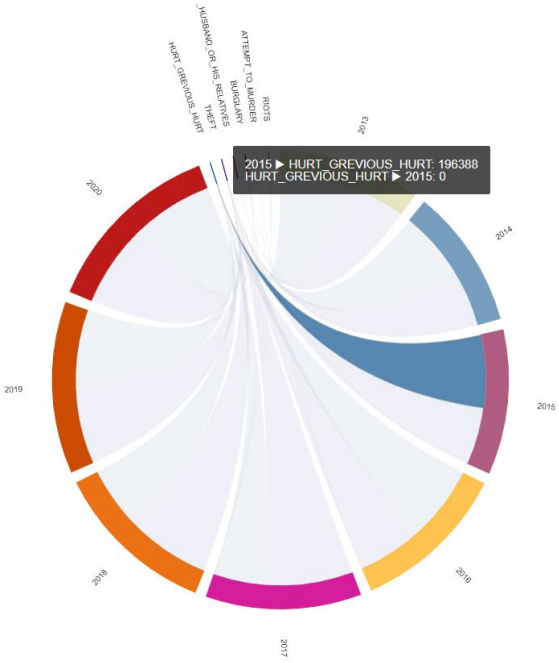


Fig.7.2 Hurt_Grevious_Hurt is the Maximum Crime happen in 2015

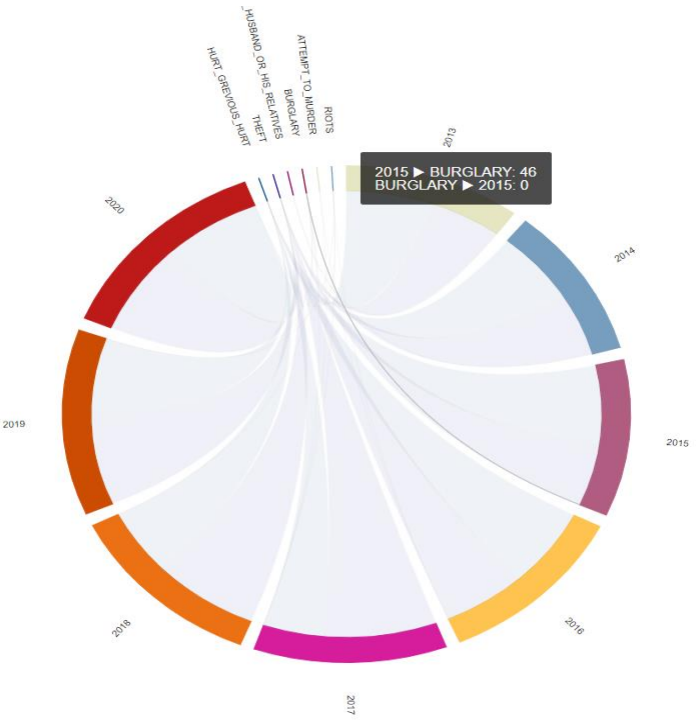


Fig.7.3 Burglary is the minimum in 2015

we visualise the distribution of max crimes per location and type in Fig.8, we can see that some types occur in specific States. To see that we use the heat map it shows that more the colour is darker more the specific max crime has occur in specific state.

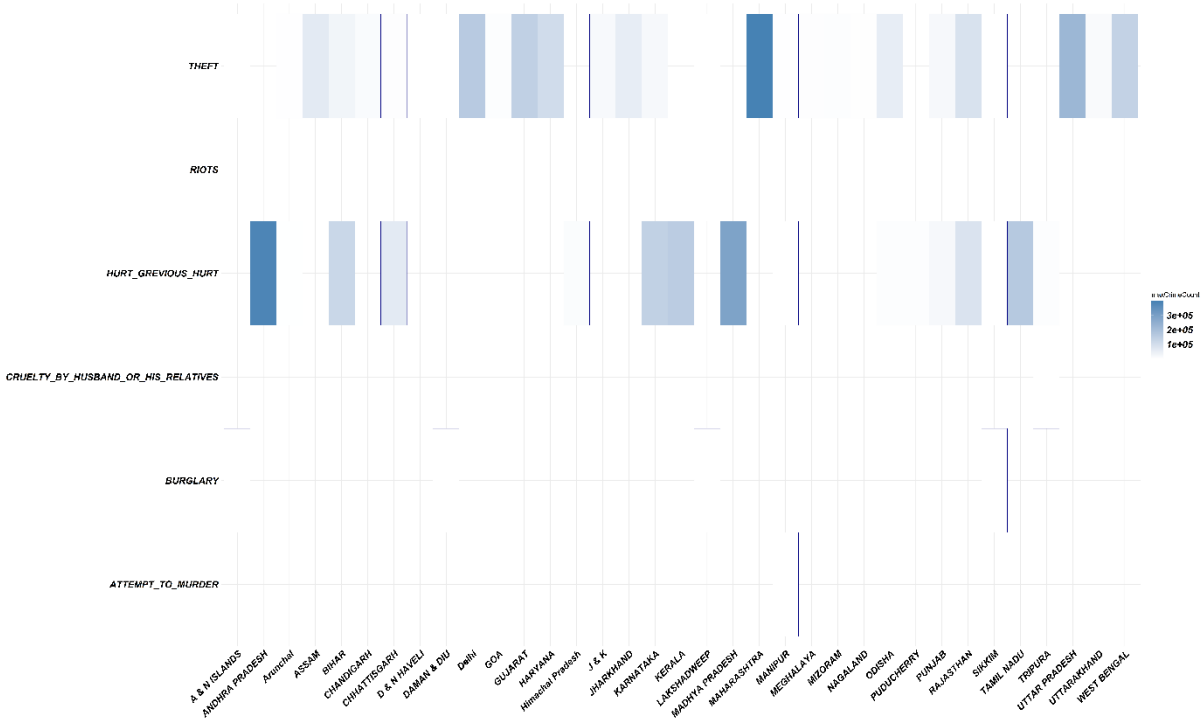


Fig.8 Max Crime Vs States

Now to see the which type of crime is mostly happen in India Fig.9 is District wise Max crime it is differ from state wise max crime because state is collection of many districts.

District has different max crime.Fig.9 is show the trend that of the maximum crime. Theft and Burglary are the maximum crime which follow the increasing tend and the other crime are also the maximum crime but they are in some of the states in some years.

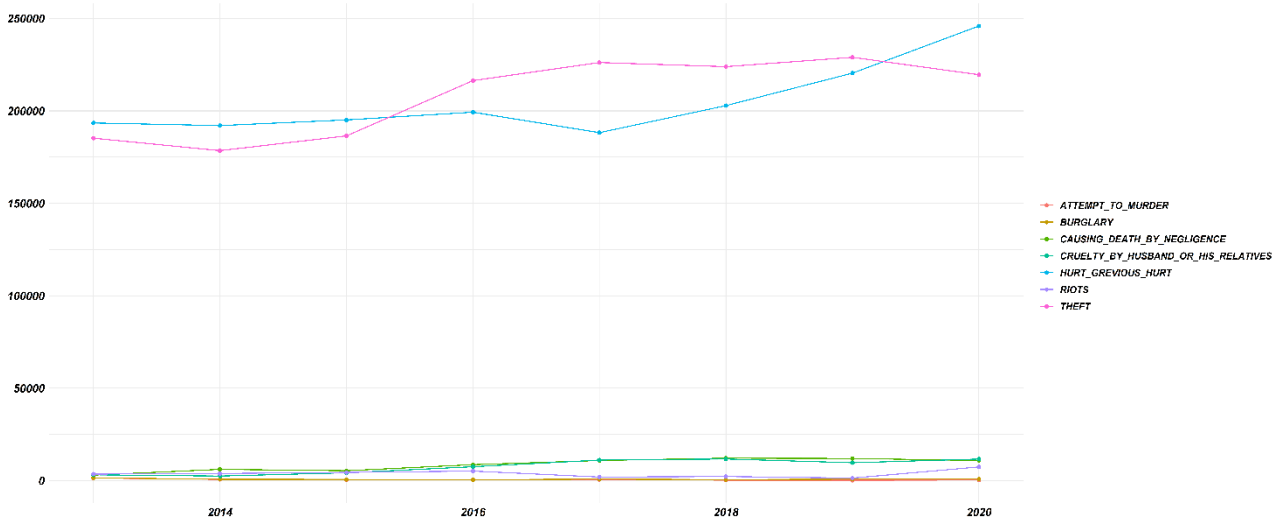


Fig.9 Max Crime Evolution per type in 2013 – 2020 District wise

As we can see in Fig.9 there are lot of Max crime and they created a mess because most of them are lesser in number to visualize it more clearly, we would see only those crime whose max crime count are greater than 1000 or 1k and to see there trend individually we have Fig.10. In that figure we can some of the crimes are decreasing, some of them are increasing and some of the crimes are start increasing rapidly after some years.

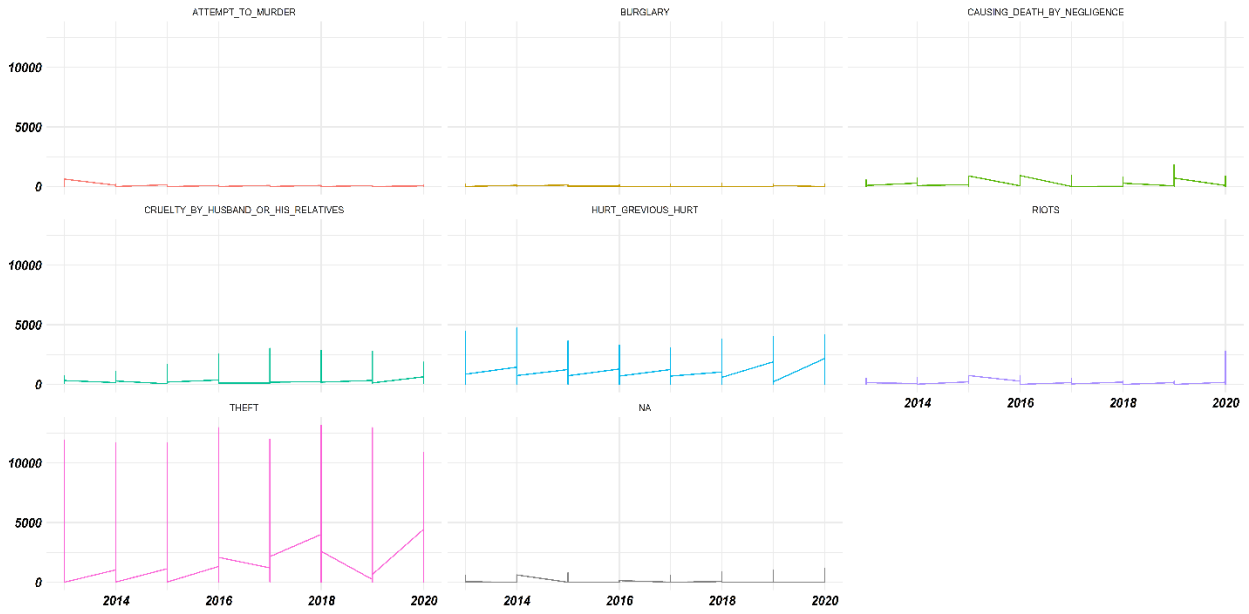


Fig.10 Max Crime evolution per type individually District Wise and who's crime count >= 1k

Fig.11 show the flow of the maximum crime in District with the years throw chord diagram. Hurt_Grevious_Hurt are the max crime because there arc cover a large part circle and theft are too. We can see that in Fig.11.2 hurt_grevious_hurt are maximum in 2020 with max crime count 245884 and in Fig.11.3 show that it is minimum in 2017 with max crime count 188210, like wise this we can see the all max crime type district wise.

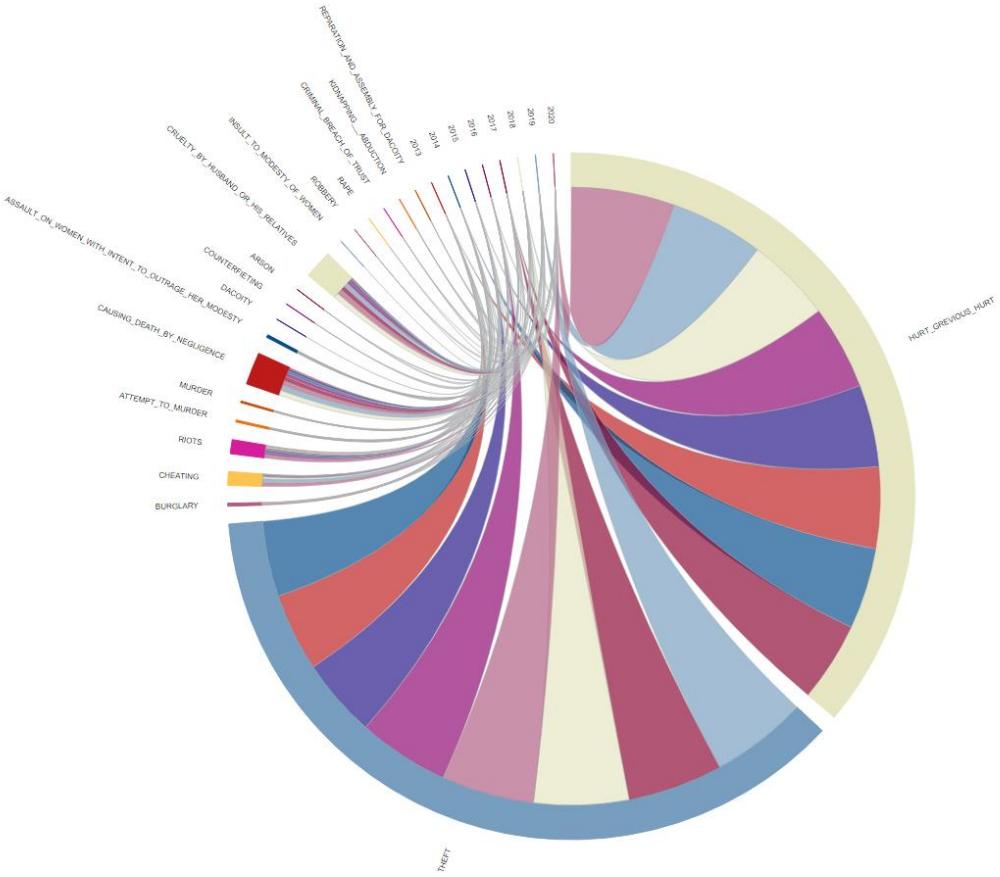
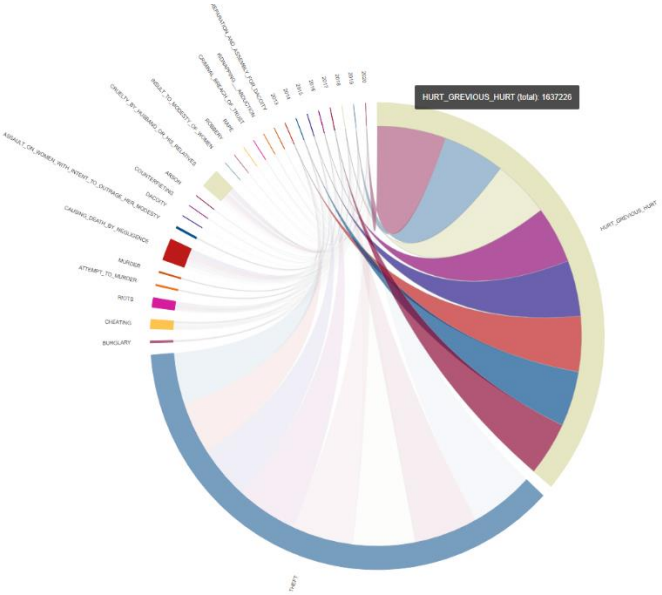


Fig.11 Max Crime type in District in 2013 – 2020



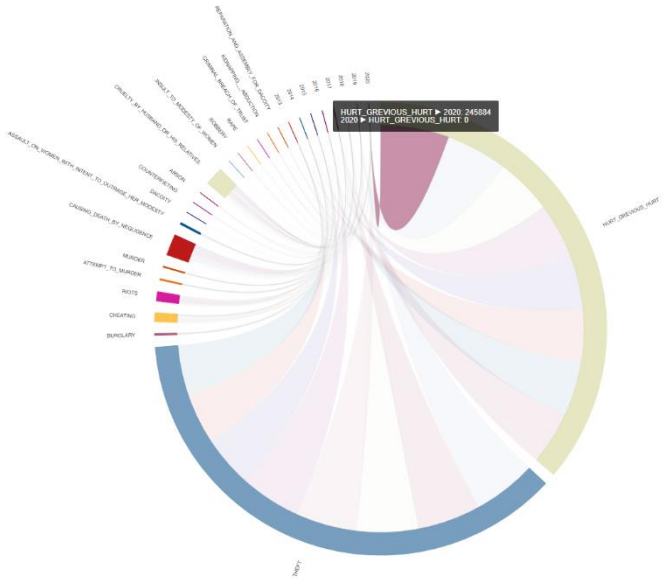


Fig.11.1 H_G_H are maximum with 1637227 Total crime count

Fig.11.2 H_G_H are maximum in 2020

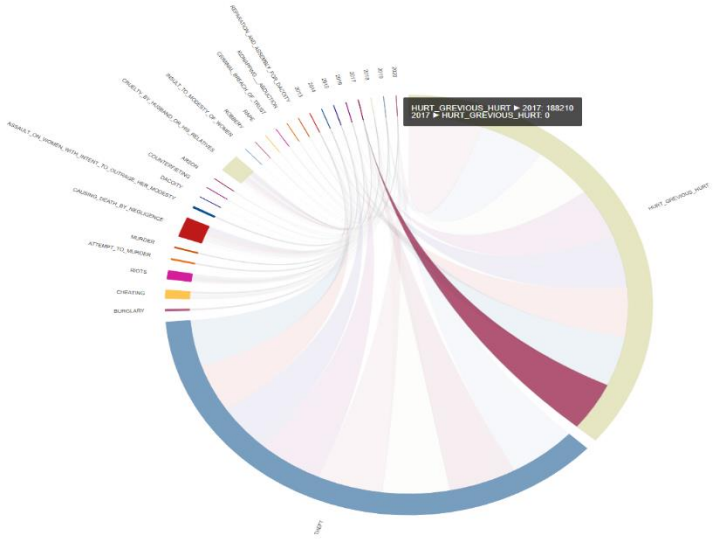


Fig.11.3 H_G_H are minimum in 2017

As we visualise the distribution of crimes per location earlier, state VS crime and max crime Vs state by using heat map but here we can't visualize the district Vs crimes with any graphs because In India there are 718 District these are very high in number to visualize with any graph. Every states has many district so take any one state and visualize there district Vs crime, while taking any random state we see the Top 10 most crime state and take number one (1) crime state in 2013 – 2020. We can see in Fig.12 that Madhya Pradesh has the high crime between 2013 – 2020 so we will visualize the district of Madhya Pradesh only.

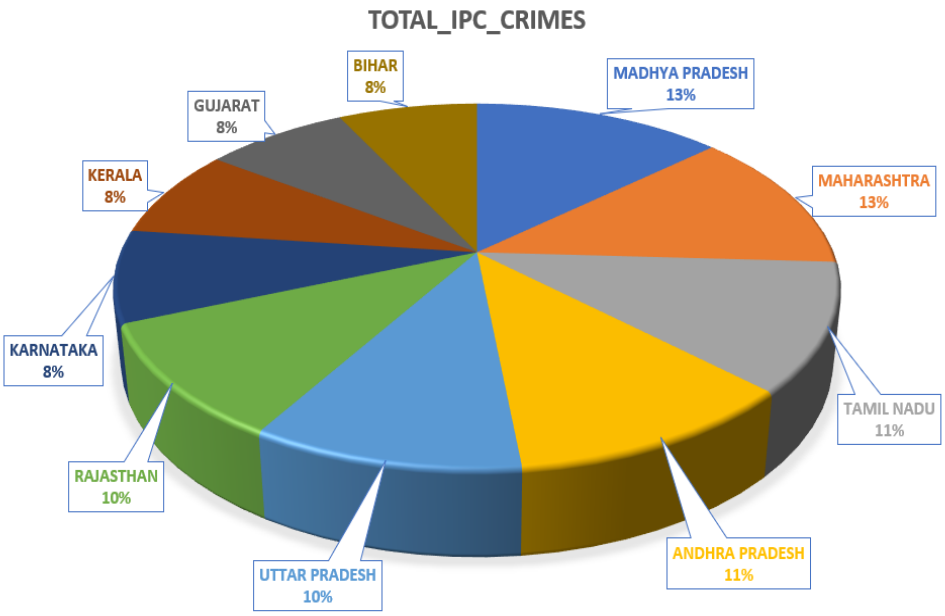


Fig.12 Top 10 Crime States in India between 2013 – 2020

Fig.13 depicts the annual frequency of crimes per type and their trend of Madhya Pradesh. The most common types of crimes are theft, hurt_grevious_hurt are in irregular in trend but most of the time is it's increasing. Hurt_Grevious_Hurt are the crime type which are highly occurs other than other type of crime type.

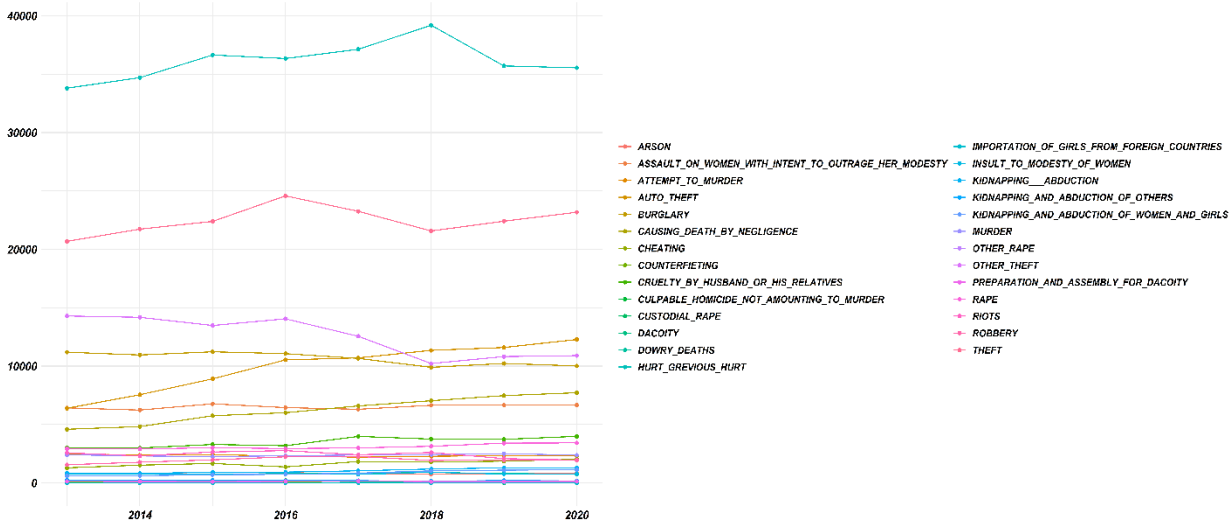


Fig.13 Crime evolution of Madhya Pradesh per type of crime between 2013-2020

In the Fig.13 most of the crime are very less so to visualize it more clearly, we neglect the crime type who's crime count is less than 1k or 1000. In Fig.14 we see the crime types individually and crime count is greater than 1k.

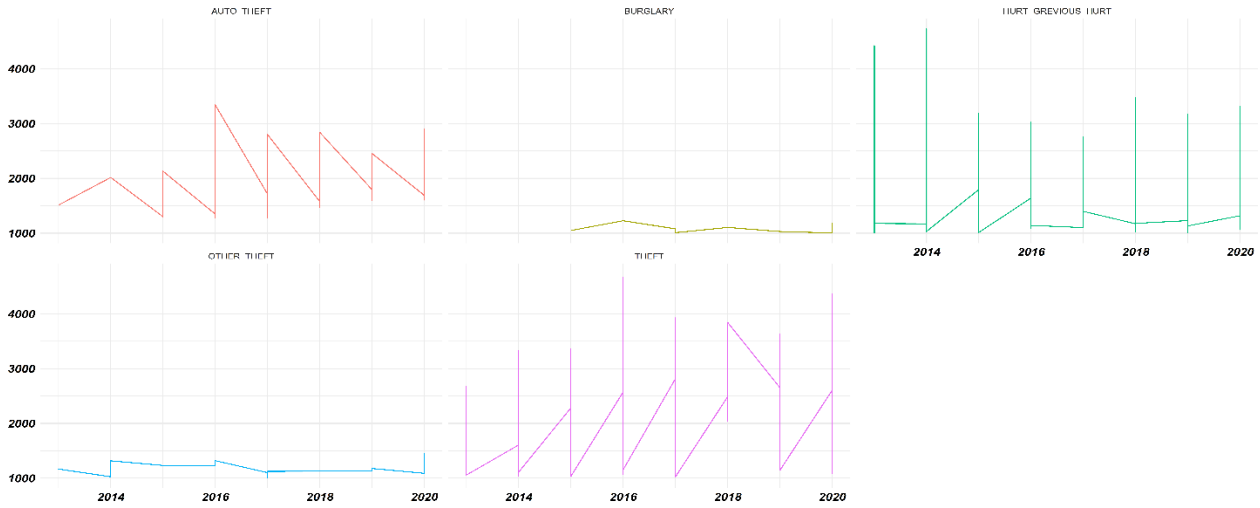


Fig.14 Crime evolution of MP per type individually and who's crime count >= 1k

Fig.15 show the flow of crime count of all District of Madhya Pradesh from 2013 – 2020 if we want to see the specific district with total crime count in 2013 – 2020 just point on that district or we can also see that in which year the crime count is maximum or minimum of that district like we see in above.

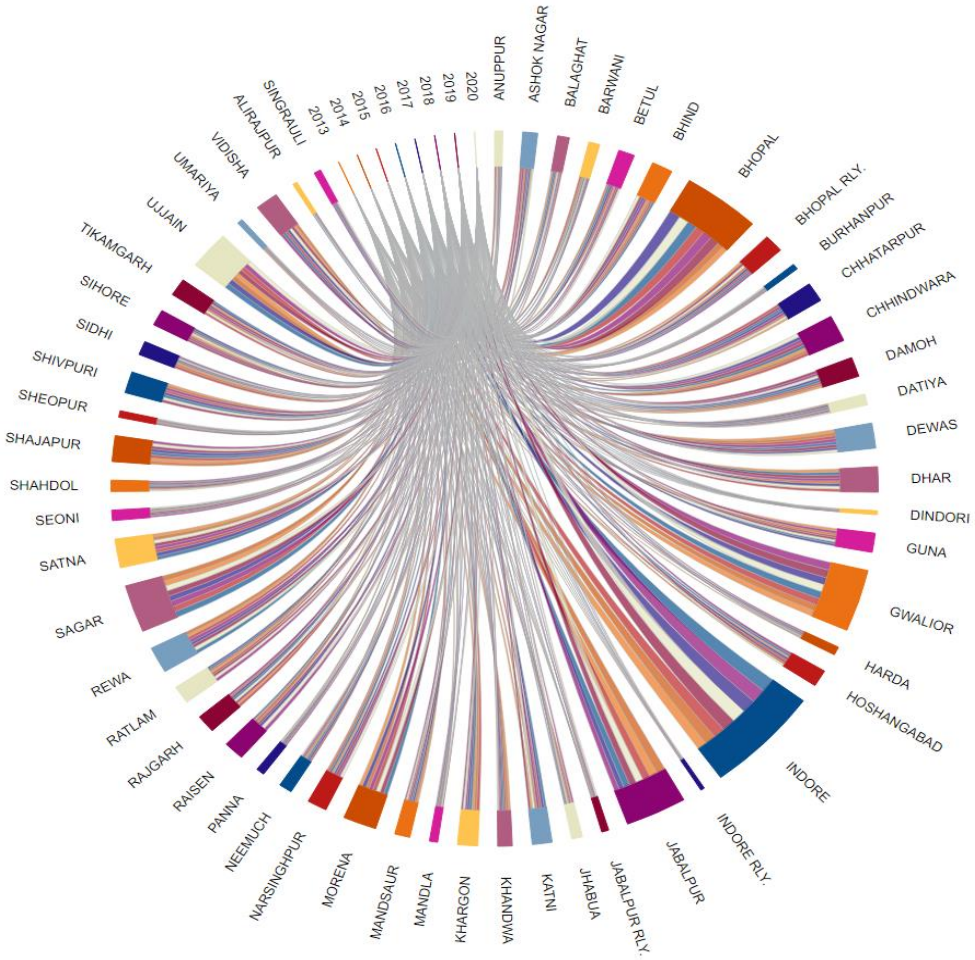


Fig.15 Flow of crime count in Madhya Pradesh from 2013 – 2020

If we visualise the distribution of crimes per location and type in Fig.16, we can see that some types occur in specific District. To see that we use the heat map it shows that more the colour is darker more the specific crime has occur in specific district. All of the crime type lies in the range between is 0 – 40000 in 2013 – 2020.

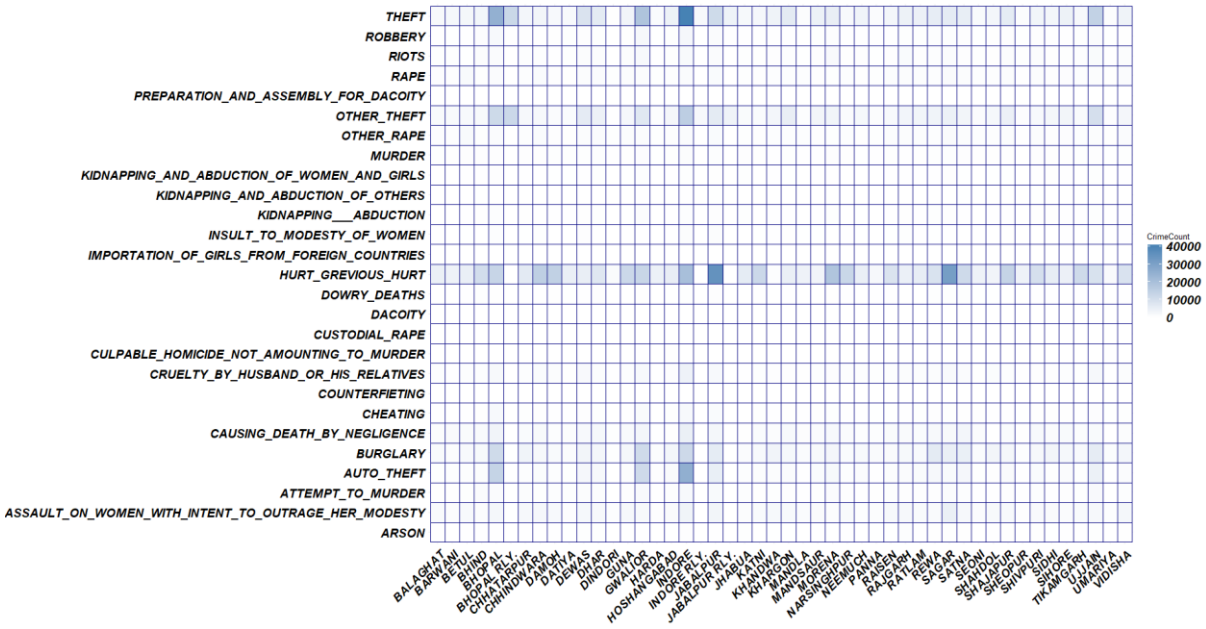


Fig.16 Crime Type in District of MP Vs District of MP

Now to see the which type of crime is mostly happen in India Fig.17 is show the trend that of the maximum crime. There are only four crime type which is mostly happen in all district. In which Theft and Hurt_Grevious_Hurt are the maximum crime which follow the increasing tend and the Burglary crime are firstly increase then decrease than goes constant and Assault on women are very less in count even it is not seen in the graph.

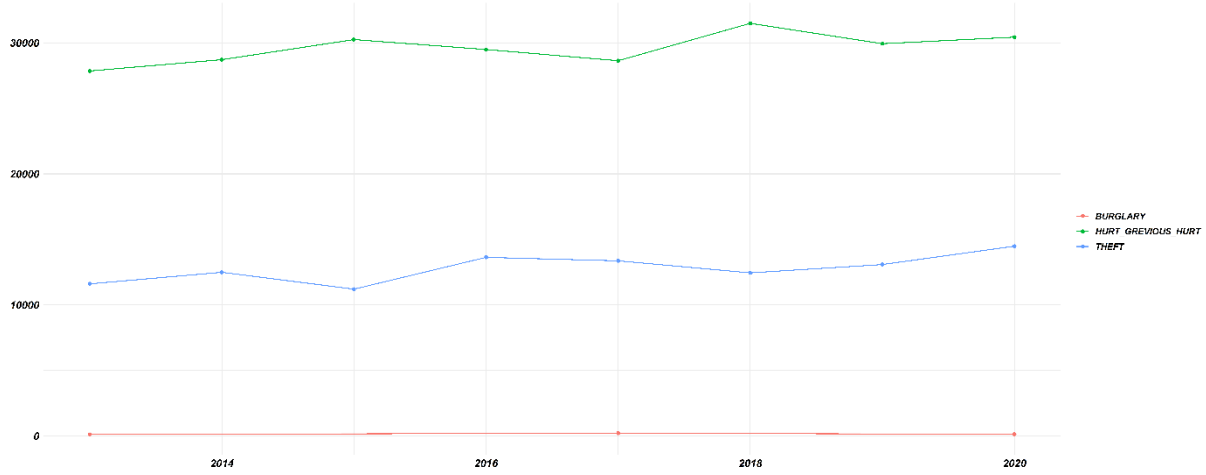


Fig.17 Max Crime Evolution per type in 2013 – 2020 in the District of MP

Fig.18 show the flow of Max crime type of all District from 2001 – 2012 if we want to see the specific crime type is maximum or minimum in which year just point on that crime type like we see in above.

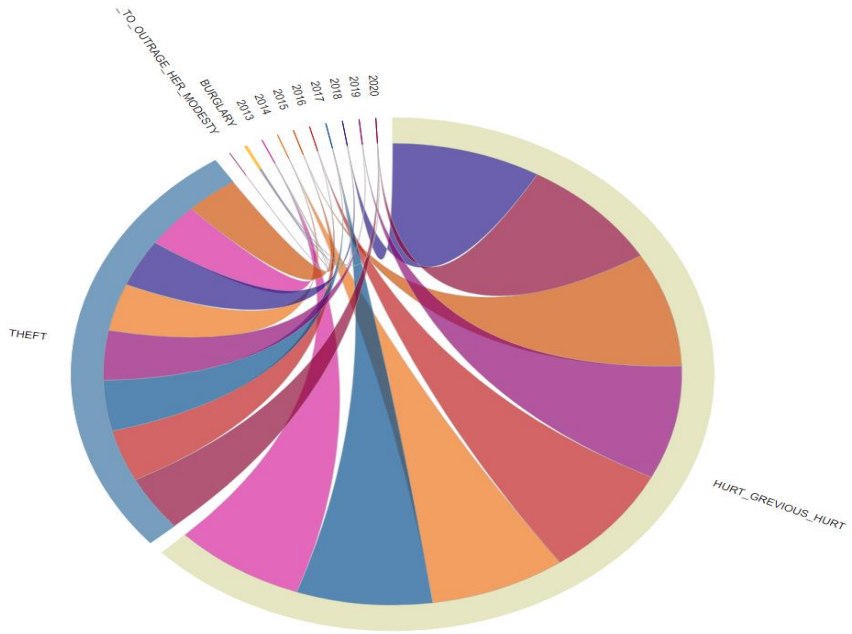


Fig.18 Maximum crime type in district of MP from 2013 – 2020

we visualise the distribution of max crimes per District and type in Fig.19, we can see that some types occur in specific District of MP. To see that we use the heat map it shows that more the colour is darker more the specific max crime type has occur in specific district.

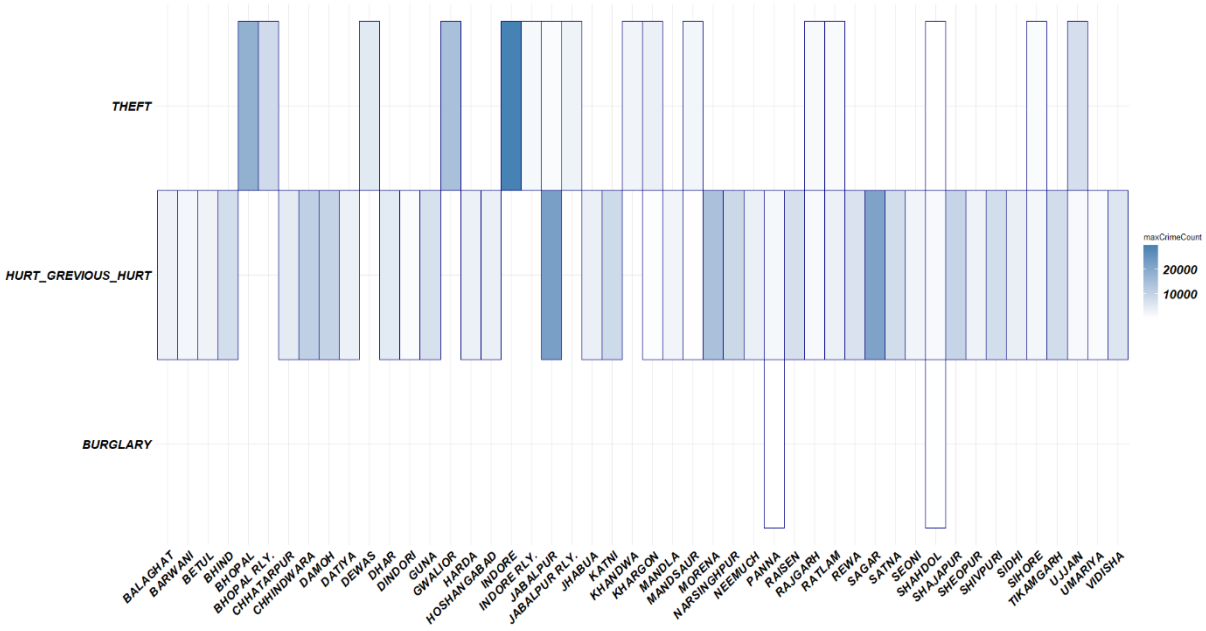


Fig.19 Max Crime Vs District of Madhya Pradesh

Spatial Temporal:

It describes a phenomenon in a certain location and time — for example, in our case we can see that the crime counts increases over the years through geographical area of top 40 districts in India (see below example images).

In the diagrams below, we can see that the larger the red area the more crime counts. We can easily see the difference between the two figures that the density of the criminal counts in 2009 is less than in 2020.

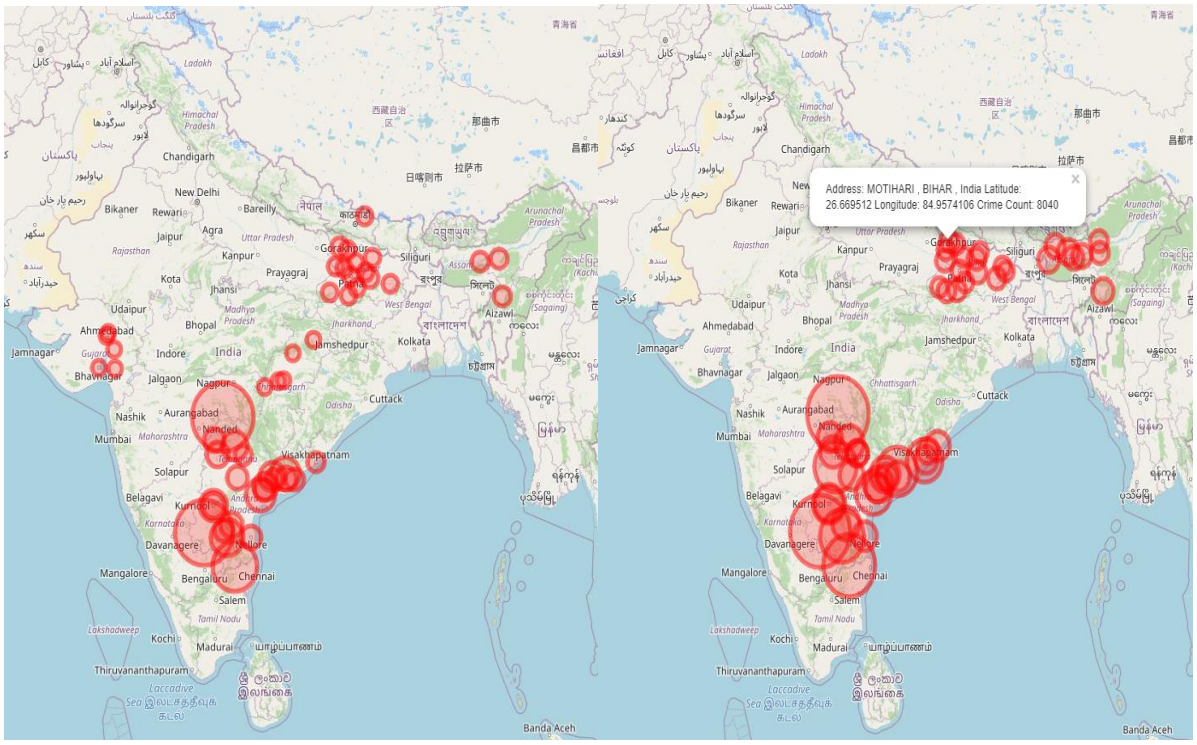


Fig.1 Top 40 districts of India in 2009 **Fig.2 Top 40 districts in India in 2020**

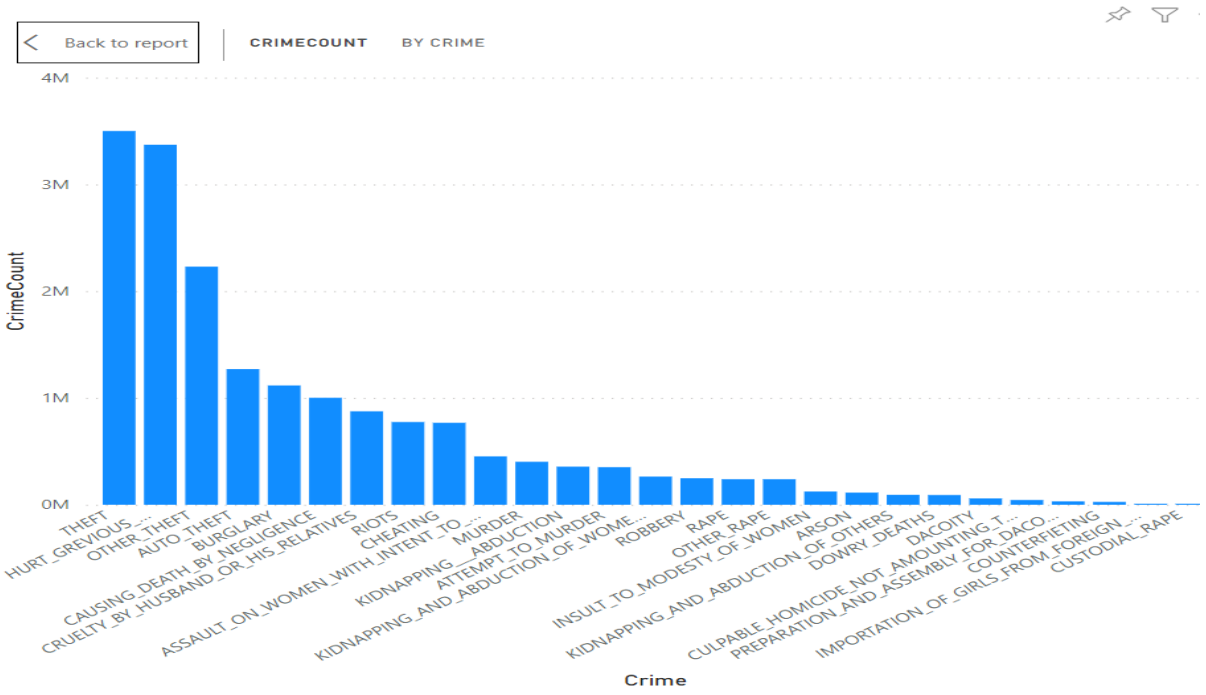


Fig.3 Crime Vs Crime Count in decreasing order

In figure 3, we can see that the Crimes that has been occurred in a greater number of times.

To visualize it more deeply we analyze the states and years from which the crime count effect it more.

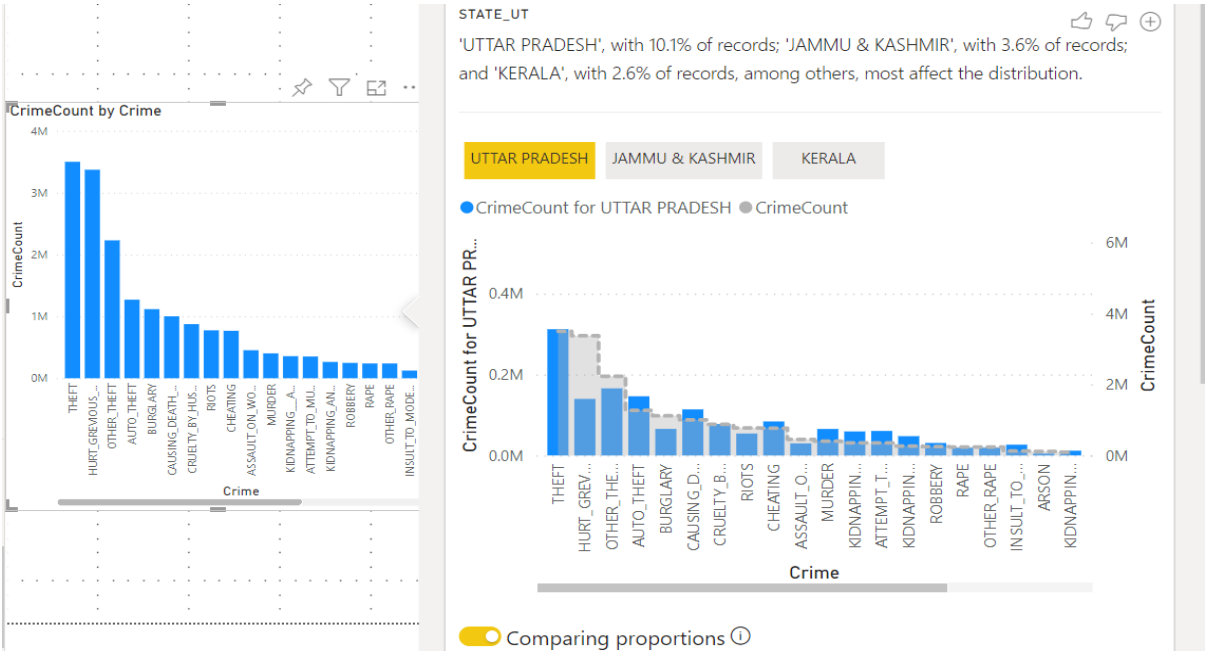


Fig. 4 Which States affect the distributions most

As we see in the Figure 4 that the 'Uttar Prades' with a 10.1% of records, 'Jammu and Kashmir' with a 3.6% of record 'Kerala' with a 2.6% of a record among others most affected distributions and Figure 5, we can see that the in 2009 with 7.9% of record 2010 with 8% of record and 2020 with 9% of record among others most effective distribution.

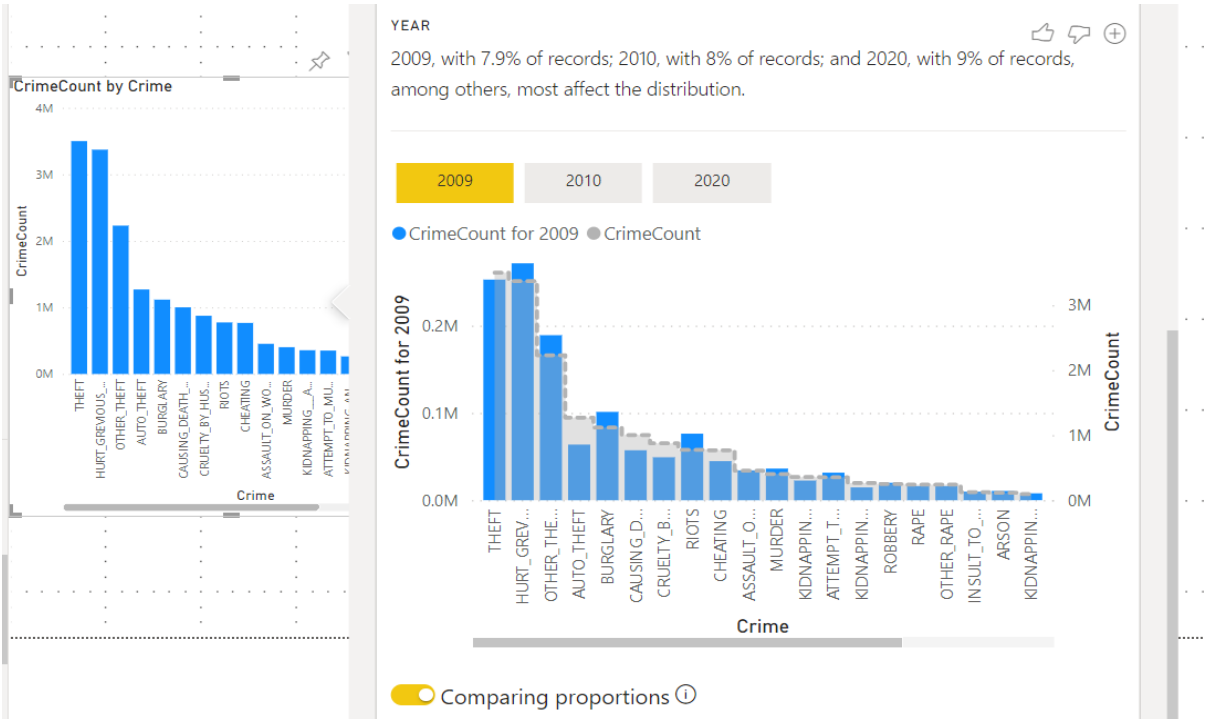


Fig. 5 Which years affect the distributions most

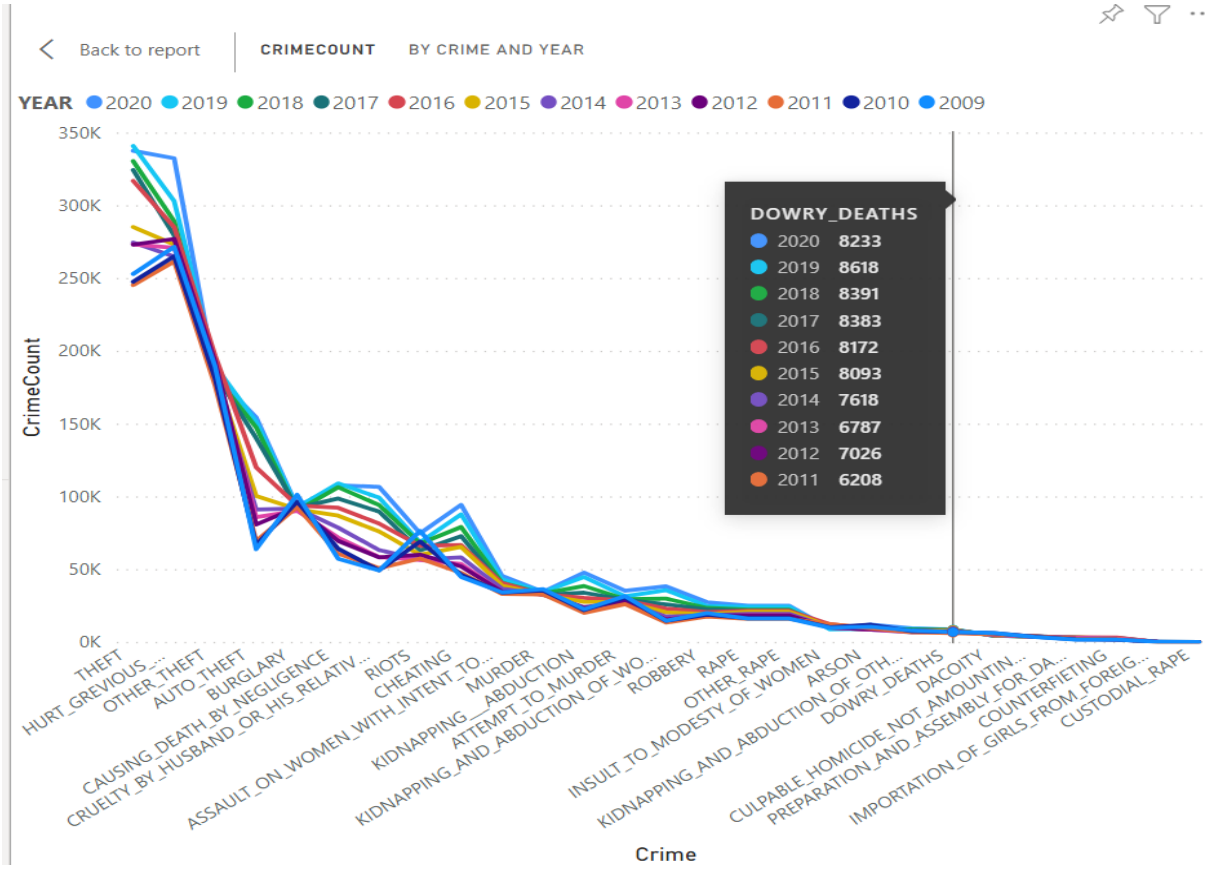


Fig. 5 Which crime has more crime count and in which year it happens in more numbers

In figure 5, we can see that the THEFT is more likely to happen in 2019. By pointing out every time we can see that Crime counts and in which year it is most likely to happen as

we see in the figure by pointing out the in the DOWRY_DEATH that in 2019 that crime count is 8618 and it is high in that year and in 2011 it is 6208 which is less in that year.