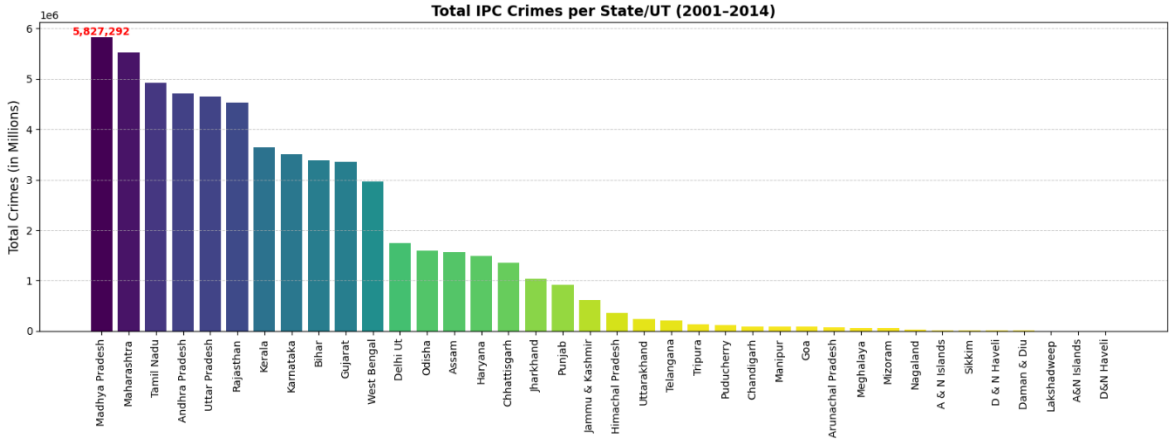
TIMED CHALLENGE 5

EDA Questions

1. **Determine the total number of crimes recorded across all districts and the average number of murders per district.**
   From 2002 to 2014, a total of approximately 58 million crimes were reported across various districts in India. On average, each district recorded around 88 cases of murder during this 13-year period. This highlights the significant scale of criminal activity as well as the persistent presence of violent crimes like murder across the country.

```
🔹 Total IPC Crimes in India (2001-2014): 58,894,630
🔹 Average Murders per District: 88.01
```

2. **Examine how crime distributions vary across different states and identify the top 5 districts with the highest total IPC crimes.**
   The distribution of IPC (Indian Penal Code) crimes varies significantly across different states in India. The bar chart below illustrates this variation for the years 2001 to 2024, with *Madhya Pradesh* recording the highest number of IPC crimes, followed by Maharashtra and Tamil Nadu.



   At the district level, the top five districts with the highest total IPC crimes during this period are:

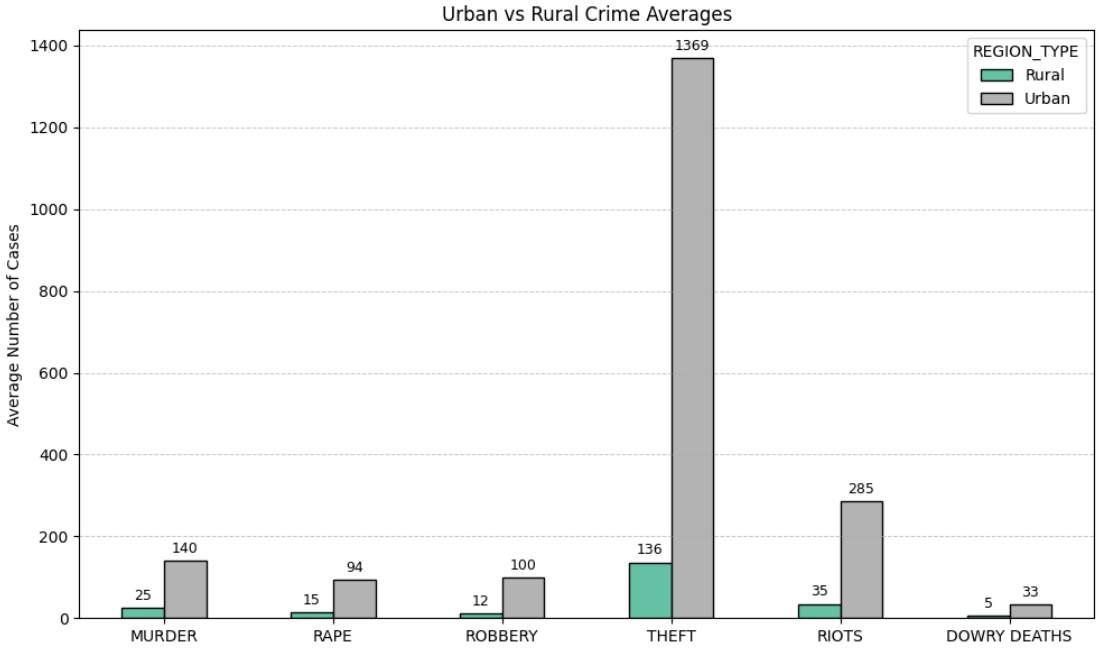
   1. Delhi
   2. Bangalore Commissionerate
   3. Mumbai Commissionerate
   4. Indore
   5. Ahmedabad Commissionerate

```
🔹 Top 5 Districts by Total IPC Crimes:
DISTRICT
Total              28814141
Delhi Ut Total       633174
Bangalore Commr.     380665
Mumbai Commr.        297871
Indore               250639
Ahmedabad Commr.     239263
```

   These findings indicate that metropolitan and urban regions tend to report higher volumes of IPC crimes, which could be attributed to higher population density and various socio-economic factors.

3. **Further, analyze how crime patterns differ across various crime categories in urban vs. rural districts (or using a proxy like population if urban/rural data is unavailable) and investigate whether there is a correlation between different crime types such as murder and theft.**

To understand the variation in crime patterns, we analyzed average crime rates across urban and rural regions by grouping the data based on region type for each crime category. The bar chart below presents the comparative averages for different crime categories, and the accompanying table summarizes the findings.



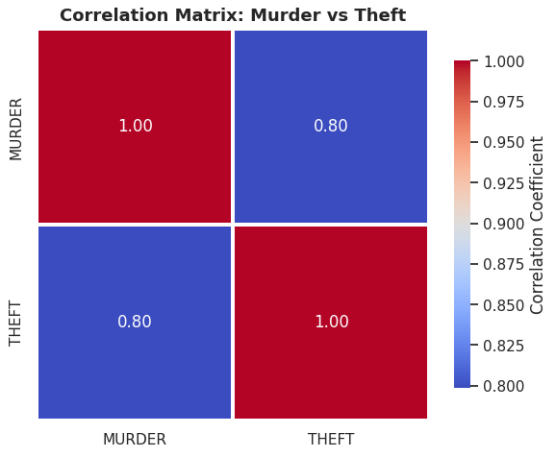
The numbers clearly show that urban areas experience much higher rates of all major crimes compared to rural regions, especially in categories like theft, robbery, riots, and crimes against women. This could be due to factors like higher population density and more complex socio-economic conditions in cities.

	REGION_TYPE	MURDER	RAPE	ROBBERY	THEFT	RIOTS	DOWRY DEATHS
0	Rural	25.29	15.37	11.71	136.15	34.75	5.30
1	Urban	140.32	94.19	99.57	1368.96	284.74	32.59

Interestingly, while investigating we did found correlation between ***murder*** **and** ***theft*** potentially hinting that *districts with high levels of one type of crime often see spikes in others too.*

This relationship is also visualized in the adjacent heatmap, which highlights the strength of correlation between various crime types.

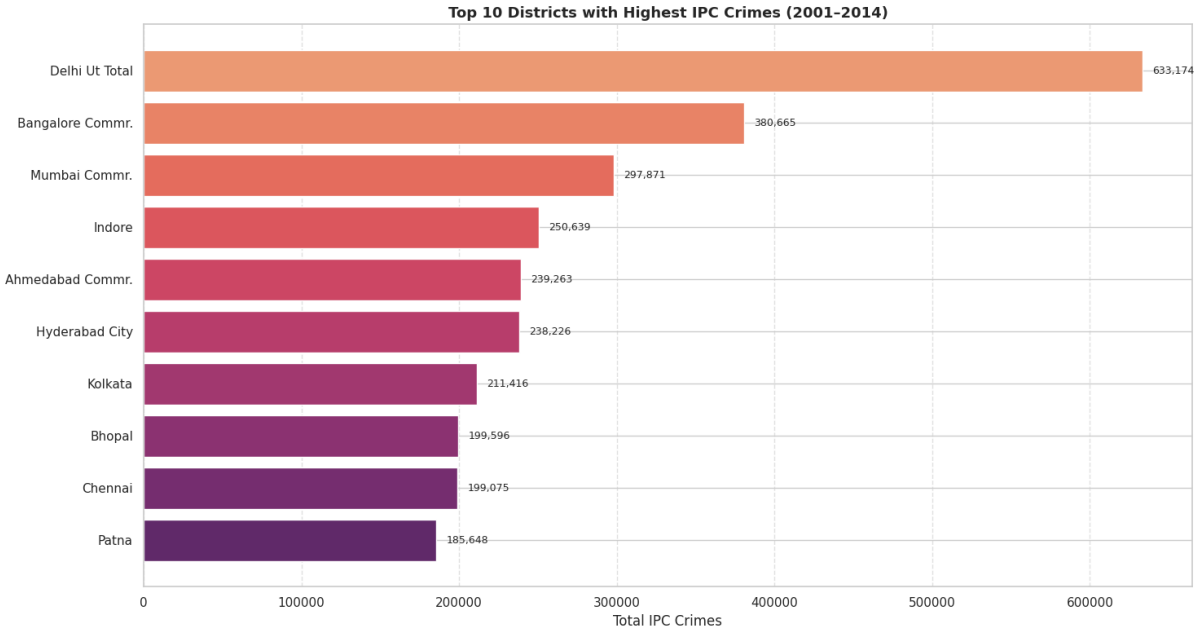
```
===== Correlation Summary =====
Pearson Correlation (linear):    0.80
Spearman Correlation (ranked): 0.64
================================
```



Visualization Questions

1. **How can visualizations be used to explore crime patterns in India by identifying the top 10 districts with the highest crime rates, understanding the overall distribution of total IPC crimes, analyzing crime density across different states, and comparing trends in violent crimes such as murder and rape across various districts?**

1.1 Top 10 districts with the highest crime rates:



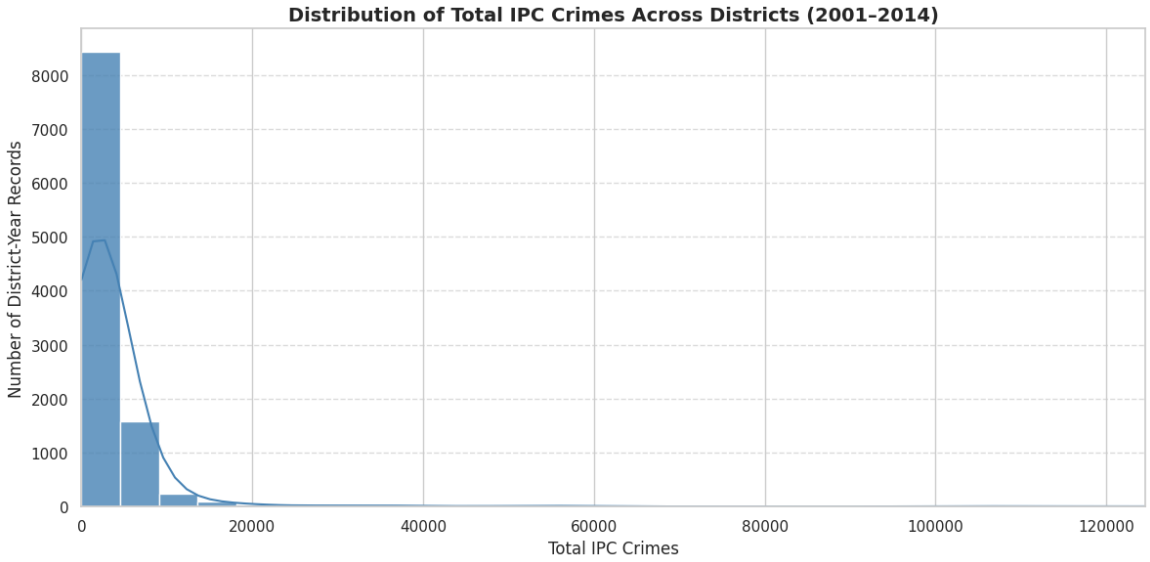
1.2 Crime density across different states:

The data for the top 10 districts is visualized with a horizontal bar chart that displays each district along with its total crime count. This allows for quick comparison of crime rates between districts.

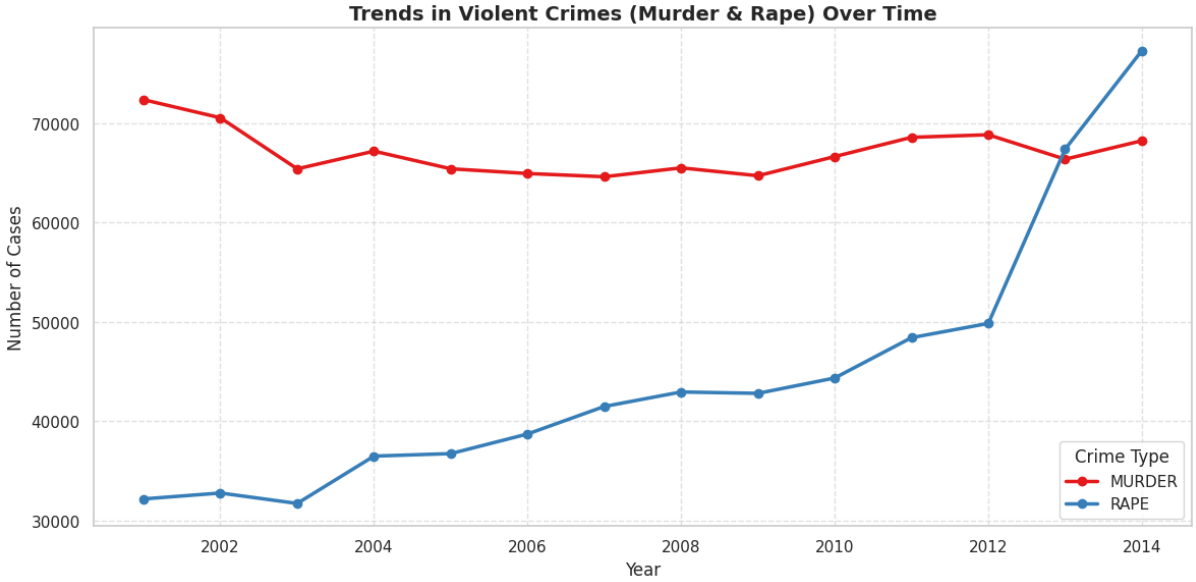
DISTRICT	
Delhi Ut Total	633174
Bangalore Commr.	380665
Mumbai Commr.	297871
Indore	250639
Ahmedabad Commr.	239263
Hyderabad City	238226
Kolkata	211416
Bhopal	199596
Chennai	199075
Patna	185648

1.3 The overall distribution of total IPC crimes:

This plot shows the distribution of total IPC crimes across districts from 2001 to 2014. The histogram represents the frequency of different crime totals, while the KDE (kernel density estimate) curve provides a smooth approximation of the distribution. The x-axis shows the total number of IPC crimes, and the y-axis indicates how often each total appears across district-year records. This gives us the overall trend and spread of crimes, identifying potential patterns in the data.

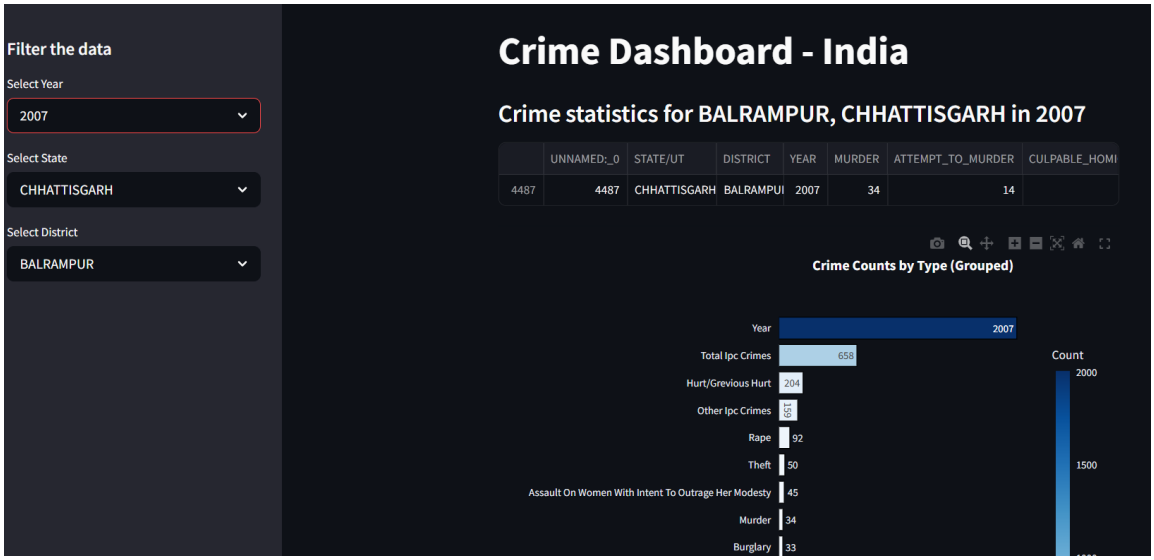


1.4 Trends for violent crimes across various districts



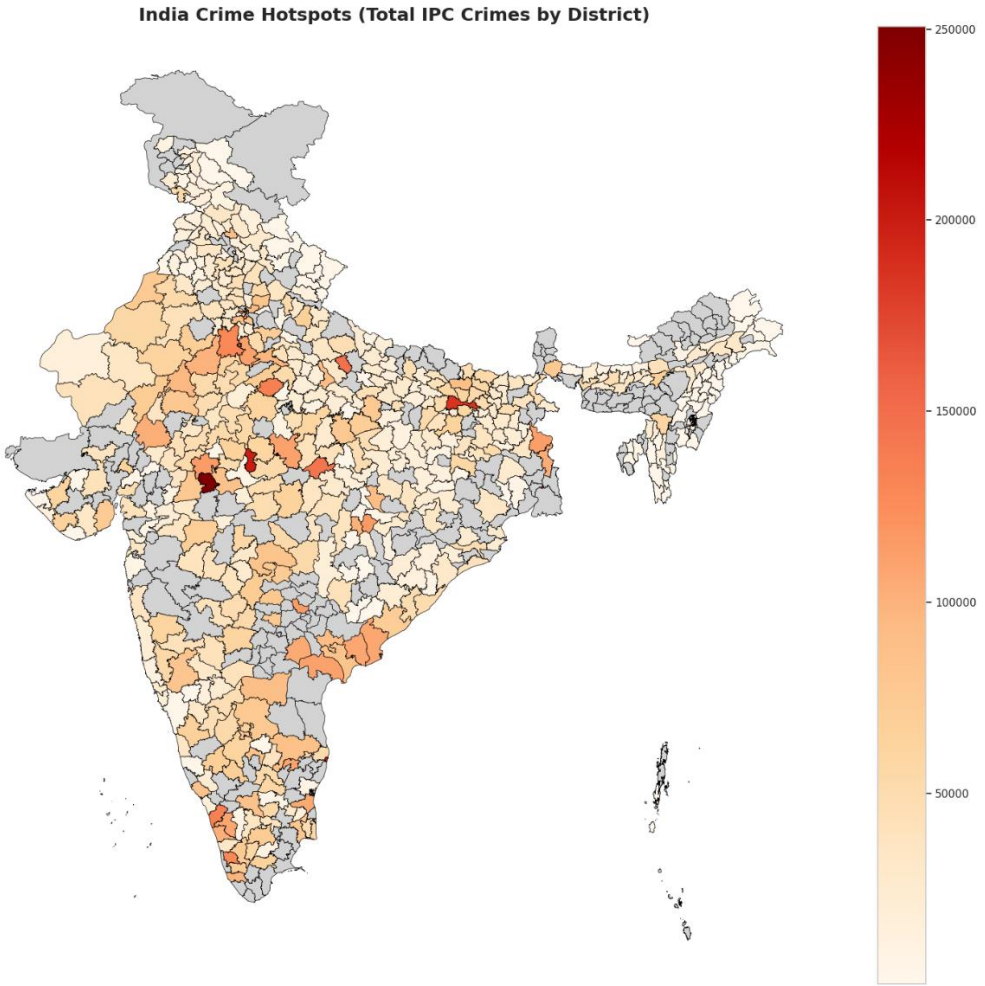
The graph illustrates the trends for murder and rape crimes, from 2001 to 2014. While murder cases remained relatively stable with a slight overall decline, rape cases showed a significant and consistent increase over the years. Notably, from 2012 to 2014, rape cases surged sharply, surpassing murder cases by 2013. This trend indicates a growing concern around rape incidents despite stable murder rates.

2. **Create an interactive dashboard that allows users to filter crime data by year, state, and district.**
   https://29a5-35-194-254-26.ngrok-free.app/ (click visit page in-case dashboard does not show up)
   preview:



An interactive dashboard was developed using Streamlit, allowing users to filter and explore crime statistics based on year, state, and district. This enables a dynamic view of crime trends, letting users focus on specific regions and timeframes to gain insights into the crime patterns of a particular district within a selected state and year.

3. Use a geospatial map to visualize crime hot spots across India. (Matplotlib)



The above Geopandas generated map of India shows the crime hotspots across the country.

Advanced Questions

1. **Identify the state with the lowest crime rate and analyze why it might be lower than others.**
   The state with the lowest total IPC (Indian Penal Code) crimes from 2001 to 2014 is *Dadra and Nagar Haveli*, with just 554 reported cases over the entire period.

   This notably low crime rate could be attributed to several factors:
   - Small population and limited geographic area, which naturally reduces the likelihood of large-scale criminal activity.
   - Low levels of urbanization and industrial activity, leading to fewer socio-economic pressures often associated with crime.
   - Potential underreporting or limited administrative reach, especially in less densely populated regions.
   - A higher police-to-population ratio compared to urban metros, possibly contributing to better deterrence and quicker law enforcement response.

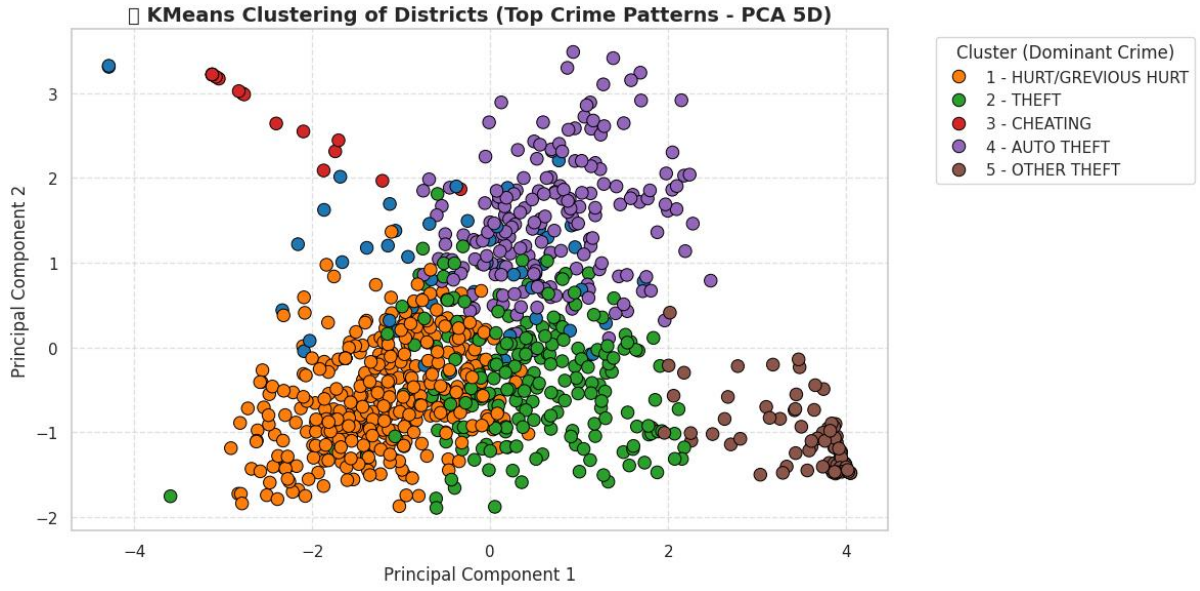
```
📉 State/UT with the Lowest Total IPC Crimes (2001–2014):
  STATE/UT  TOTAL IPC CRIMES
D&N Haveli               554

Possible reasons for low crime:
- Small population and geographic size
- Limited urbanization and industrial zones
- Lower reporting or administrative coverage
- High police-to-population ratio relative to urban metros
```

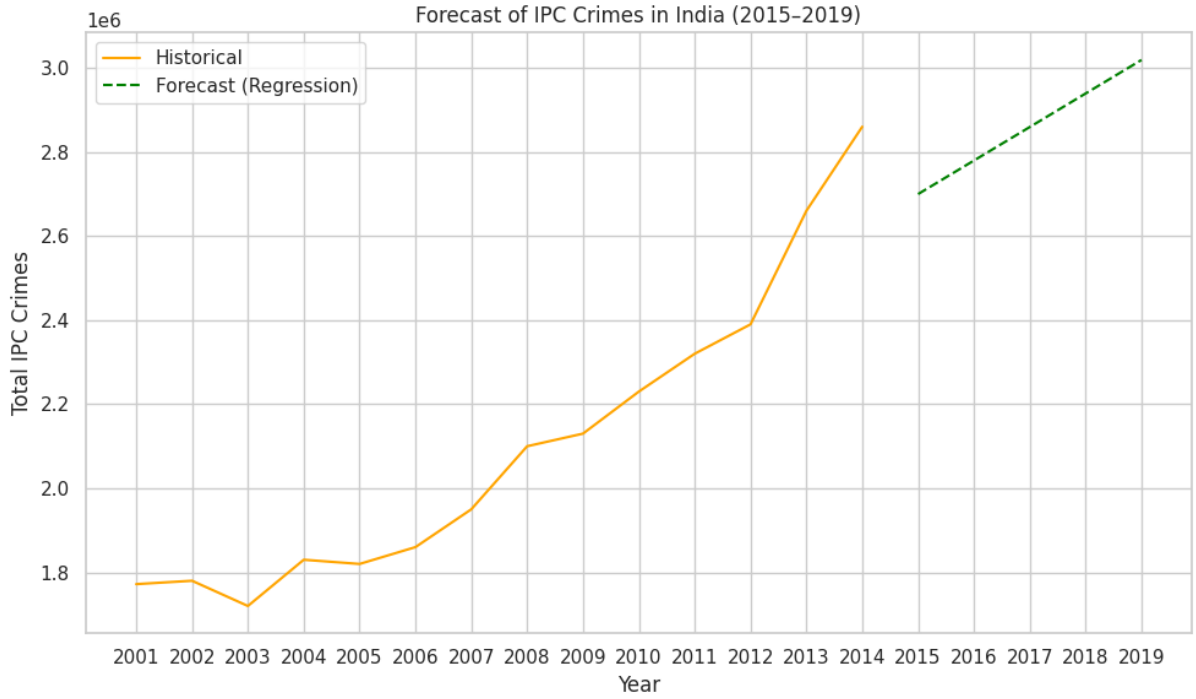
2. **Find the most common type of crime committed in each district.**
   Below are few districts with their most committed crime:

	STATE/UT	DISTRICT	MOST COMMON CRIME	CRIME COUNT
9840	Andhra Pradesh	Anantapur	THEFT	3800
9841	Andhra Pradesh	Chittoor	CAUSING DEATH BY NEGLIGENCE	2567
9842	Andhra Pradesh	Cuddapah	THEFT	2604
9843	Andhra Pradesh	East Godavari	THEFT	3791
9844	Andhra Pradesh	Guntakal Railway	THEFT	413
9845	Andhra Pradesh	Guntur	THEFT	2606
9846	Andhra Pradesh	Guntur Urban	THEFT	1740
9847	Andhra Pradesh	Krishna	CRUELTY BY HUSBAND OR HIS RELATIVES	3942
9848	Andhra Pradesh	Kurnool	CHEATING	3522
9849	Andhra Pradesh	Nellore	THEFT	3037
9850	Andhra Pradesh	Prakasham	THEFT	2687
9851	Andhra Pradesh	Rajahmundry	THEFT	657

3. Apply clustering algorithms (e.g., K-Means) to group districts based on crime patterns.



4. Predict future crime trends using regression analysis.



To analyze and forecast the trend of IPC (Indian Penal Code) crimes in India, a simple linear regression model was used based on historical data from 2001 to 2014. The model was trained using LinearRegression to capture the trend of increasing crime rates over the years. Using the trained model, IPC crime figures were predicted for the years 2015 to 2019. The results showed a steady upward trend, consistent with past data.

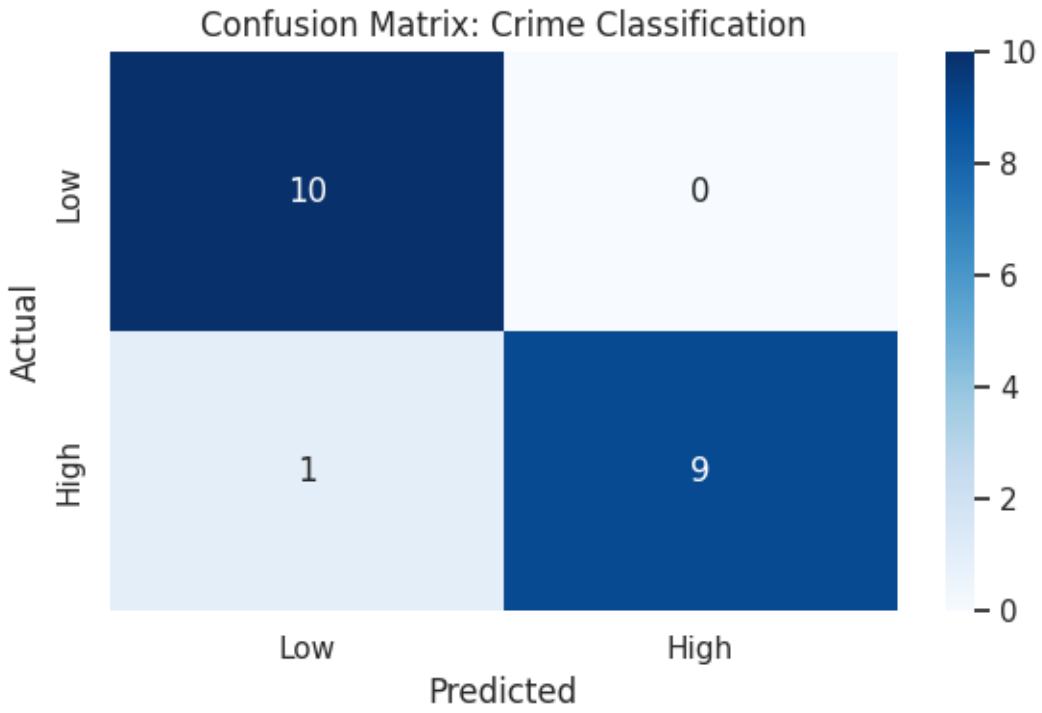
The forecasted values are summarized below:

Forecasted IPC Crimes for 2015–2019:

	Year	Predicted IPC Crimes
0	2015	2699824
1	2016	2779591
2	2017	2859358
3	2018	2939125
4	2019	3018892

5. **Use a machine learning model to classify high-crime and low-crime districts.**

	precision	recall	f1-score	support
0	0.97	0.94	0.95	1573
1	0.95	0.97	0.96	1631
accuracy			0.96	3204
macro avg	0.96	0.96	0.96	3204
weighted avg	0.96	0.96	0.96	3204



Sample of classified districts:

	DISTRICT	TOTAL IPC CRIMES	CRIME_LABEL	PREDICTED_LABEL
0	District_1	172	1	1
1	District_2	161	0	0
2	District_3	162	0	0
3	District_4	149	0	0
4	District_5	154	0	0
5	District_6	156	0	0
6	District_7	155	0	0
7	District_8	167	1	1
8	District_9	155	0	0
9	District_10	142	0	0

6. **Develop a crime risk index for districts based on historical data.**

	DISTRICT	STATE/UT	CRIME_RISK_INDEX
10245	Total	Madhya Pradesh	1.000000
10292	Total	Maharashtra	0.917081
10596	Total	Uttar Pradesh	0.882726
9472	Zz Total	Maharashtra	0.860372
9426	Zz Total	Madhya Pradesh	0.834926

Bonus Questions

1. **What percentage of crimes are committed against women?**

```
Percentage of crimes against women: 7.77%
```

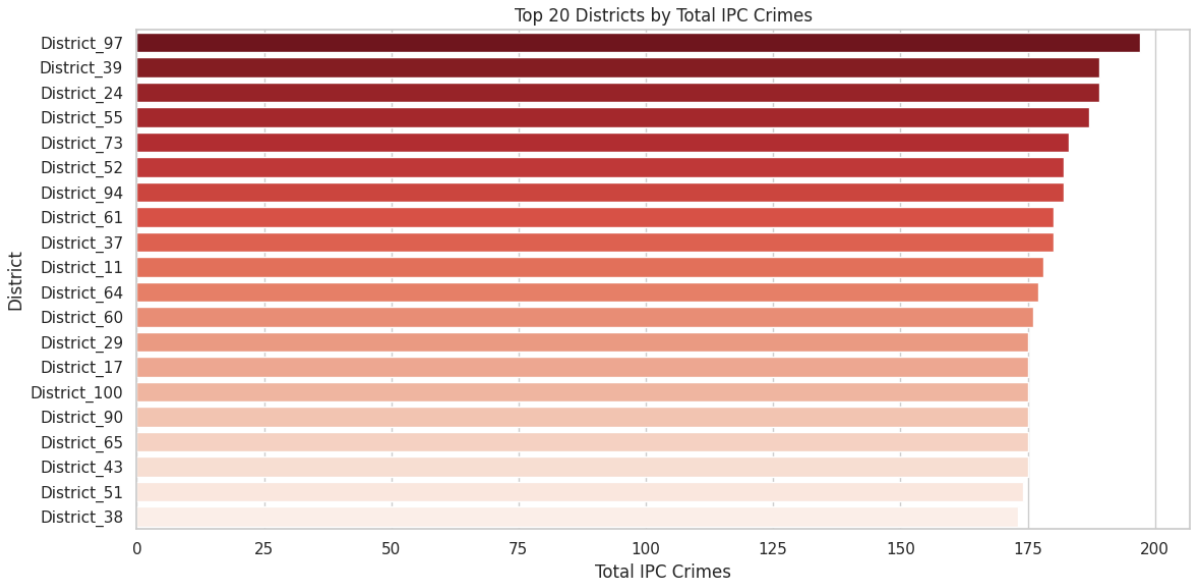
2. **Identify the state with the highest number of dowry deaths.**

```
State with Highest Dowry Deaths:
STATE/UT
Uttar Pradesh    57256
Name: DOWRY DEATHS, dtype: int64
```

3. **Analyze seasonal variations in crime trends (e.g., do crimes increase during certain months?).**
   Observation: The dataset does not have any information related to months.

4. **Examine if there is a link between cities and crime rates.**

   Insight: Cities with higher urbanization, population, or economic activity tend to report more crimes. Capital cities or metro districts (e.g., Delhi, Mumbai, Bangalore) usually rank high. Visualization helps pinpoint clusters of concern.



5. **Build a time-series model to forecast crime rates for the next five years.**
   To improve the forecasting of IPC crimes, we also used the ARIMA model, which is more suitable for time series data. Unlike linear regression, which assumes a straight-line trend, ARIMA considers past values and moving averages, making it more adaptive to real-world fluctuations. As seen in the plot, ARIMA provides a smoother and more realistic forecast by capturing underlying patterns in the data. This makes it a better choice for predicting crime trends over time compared to simple regression.