**PYTHON PROJECT REPORT**

(Project Semester: January-April 2025)

**Title of the Project:** **District and Taluk Court Cases India**

**Submitted by:**

**Ayush Kumar  
Registration No.: 12324270  
Programme and Section: B.Tech CSE (K23FD)  
Course Code: INT375**

**Under the Guidance of:  
Baljinder Kaur (UID : 27952)**

**Discipline of CSE/IT**  
**Lovely School of Computer Science & Engineering**  
**Lovely Professional University, Phagwara**

**DECLARATION**

I, **Ayush Kumar**, student of **Bachelor of Technology (B.Tech)** under CSE/IT Discipline at Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 12-April-2025

Signature:

Registration No.: 12324270  
Name of the Student: **Ayush Kumar**

# ****CERTIFICATE****

This is to certify that **Ayush Kumar** bearing Registration No. **12324270** has completed **INT375** project titled **“District and Taluk Court Cases”** under my guidance and supervision. To the best of my knowledge, the present work is the result of her original development, effort, and study.

Baljinder Kaur  
**School of Computer Science & Engineering**

**Lovely Professional University**  
**Phagwara, Punjab**

Date: **12-April-2025**

**ACKNOWLEDGMENT**

I would like to express my sincere gratitude to **Baljinder kaur ma’am**, my project guide, for their invaluable support, guidance, and encouragement throughout the development of this project. Their expert insights and constructive feedback have been instrumental in shaping the project's outcome.

I am also thankful to **Lovely Professional University** for providing a conducive learning environment and access to resources that made this project possible. Additionally, I extend my appreciation to my professors and peers for their continuous motivation and insightful discussions, which greatly enhanced my understanding of the subject.

Lastly, I would like to acknowledge the unwavering support of my family and friends, whose encouragement has been a source of inspiration throughout this journey.

# ****TABLE OF CONTENTS****

**Introduction**

**Source of Dataset**

**EDA Process**

**Analysis on Dataset (for each objective)**

* Introduction
* General Description
* Specific Requirements
* Analysis Results
* Visualization

**Conclusion**

**Future Scope**

**References**

# ****1. INTRODUCTION****

The Indian judiciary has long been burdened with a large volume of pending cases, resulting in delayed justice and inefficient resource allocation. This backlog, particularly at the District and Taluk court levels, can be attributed to a combination of procedural, administrative, and infrastructural issues. In this project, we aim to explore and visualize data related to court case pendency across various regions in India. By conducting exploratory data analysis (EDA), we intend to identify patterns and bottlenecks that contribute to delays in case resolution. This analysis leverages Python and its powerful libraries to draw insights from a comprehensive dataset containing information on the status, duration, type, and demographics related to court cases.

**2. SOURCE OF DATASET**

The dataset used for this analysis was sourced from the eCourts Services portal, which provides open-access data related to Indian district and taluk courts. The dataset includes information about:

* State and District names
* Types of court cases (civil, criminal, etc.)
* Case pendency duration categories (0-1 year, 1-3 years, etc.)
* Total pending cases
* Stage of each case (Appearance, Compliance, etc.)
* Institution and disposal statistics for recent months
* Special case indicators (e.g., cases filed by women and senior citizens)

This rich dataset enables multidimensional analysis to uncover insights on the court case backlog problem.

# ****3. EDA PROCESS****

**Exploratory Data Analysis (EDA) Process**

# Exploratory Data Analysis (EDA) is a critical phase in any data-driven project, especially when dealing with complex judicial datasets. The purpose of EDA is to better understand the structure, quality, and key patterns in the dataset before diving into deep analysis. Our EDA process was comprehensive and involved the following stages:

# Data Inspection and Loading: Initially, the dataset was loaded using the pandas library. We explored the structure of the dataset using functions such as df.head(), df.info(), and df.describe() to understand the data types, missing values, and basic statistics.

# Data Cleaning: This involved handling missing or null values using fillna() or dropna() functions, standardizing column names to ensure consistency, and filtering out any irrelevant or duplicate records. The cleaning step ensured data integrity for further analysis.

# Data Transformation and Grouping: We used grouping (groupby) and aggregation (sum, count) functions to restructure the dataset for our objectives. For instance, data was grouped by state, district, stage of the case, and case duration to compute meaningful summaries.

# Feature Engineering: Derived new features like total cases by year category and stage-wise counts to enrich the dataset. We also calculated percentages and ratios to enhance the analytical depth.

# Data Exploration Through Visualization: Various visualization techniques were employed to observe patterns, outliers, and relationships in the data. We used bar charts to compare states and stages, heatmaps to show intensity across districts, histograms to understand distribution, boxplots to identify outliers, and scatter plots to explore correlations.

# Consistency Checks: After transformation, we checked for consistency in the processed dataset by comparing totals and verifying that no important data had been inadvertently excluded.

# This detailed and iterative EDA process formed the foundation of our deeper analysis, helping us frame relevant hypotheses and select the right visual tools to represent the findings.

# ****4. ANALYSIS ON DATASET****

**Objective 1: Distribution of Pending Court Cases Across States and Districts**

**Introduction:**  
The geographical distribution of pending cases provides a clear understanding of workload imbalances and can guide resource allocation.

**General Description:**  
We analysed the number of pending cases in each state and district to find out which areas are under the most pressure.

**Specific Requirements:**

* Create bar plots for state-wise pending cases
* Generate heatmaps for district-level intensity

**Analysis Results:**

* Uttar Pradesh and Maharashtra emerged as high pendency states.
* Some districts in West Bengal and Bihar showed disproportionately high caseloads.

**Visualization:**

* Bar plots for top 10 states by pending cases
* Clean, uncluttered heatmap showing district-level case distribution

**Objective 2: Duration-Wise Trend of Pending Cases**

**Introduction:**  
Studying the duration of pending cases reveals aging trends and highlights long-term inefficiencies.

**General Description:**  
We categorized cases into buckets (0-1, 1-3, 3-5, 5-10, 10-20, 20-30, and 30+ years) and analysed their distributions.

**Specific Requirements:**

* Use stacked bar charts to compare duration categories
* Generate histograms and boxplots to identify frequency and outliers

**Analysis Results:**

* Most cases fall in the 1–5-year bracket
* Outliers exist where cases have remained pending for over 30 years

**Visualization:**

* Stacked bar charts per state
* Histograms showing case frequency across time spans
* Refined boxplot revealing long-pending outliers and regional variation

**Objective 3: Stage-Wise Status of Pending Cases**

**Introduction:**  
Each case goes through several stages; identifying bottlenecks in these can help speed up the process.

**General Description:**  
We focused on stages such as Appearance, Compliance, Evidence, and Pleadings to determine where most delays occur.

**Specific Requirements:**

* Bar plots and pie charts for stage distributions
* Heatmap showing stage vs district concentration

**Analysis Results:**

* Appearance and Compliance stages showed the highest case accumulation
* Pleading and Evidence stages had comparatively fewer delays

**Visualization:**

* Pie chart with labelled proportions for stages
* Bar plot comparing case counts at each stage
* Heatmap showing distribution of stages across districts

**Objective 4: Monthly Trends – Cases Instituted vs Disposed**

**Introduction:**  
Tracking how many cases are filed versus resolved over time indicates whether the backlog is increasing or being reduced.

**General Description:**  
We analysed monthly trends to see the net change in caseloads.

**Specific Requirements:**

* Line chart and grouped bar plots for comparative analysis
* Scatter plot to identify correlations

**Analysis Results:**

* In many regions, instituted cases consistently outnumber disposals
* Some districts maintained a healthy balance, reflecting better efficiency

**Visualization:**

* Grouped bar plots for monthly trends
* Scatter plots showing correlation between filing and disposal rates

**Objective 5: Special Case Categories (Women, Senior Citizens, Delayed Cases)**

**Introduction:**  
Cases involving women and senior citizens require special attention due to their socio-legal implications.

**General Description:**  
We analysed the number and distribution of such cases and also examined delayed cases for strategic intervention.

**Specific Requirements:**

* Compare women vs senior citizen case counts
* Highlight delayed cases using heatmaps and boxplots

**Analysis Results:**

* States like Tamil Nadu and Kerala showed higher filings from senior citizens and women
* Significant delays were observed in some northern districts

**Visualization:**

* Bar plots comparing vulnerable group case counts
* Heatmap showing high-delay districts
* Boxplot indicating delay spread across regions

**🔹 Formula Used and Information Explained to the User**

In the course of this analysis, several derived metrics and calculations were used to interpret the dataset more meaningfully and communicate insights clearly to the user. Below are the key formulas and the information they convey:

**1. Total Pending Cases**

**Formula:**  
Total Pending Cases = Sum of all pending duration categories per district/state  
**Purpose:**  
Helps users understand the overall burden of pending cases in a region, highlighting where intervention may be needed.

**2. Percentage of Cases at Each Stage**

**Formula:**  
Stage % = (Cases at Specific Stage / Total Pending Cases) × 100  
**Purpose:**  
This metric reveals bottlenecks in case progress by showing which procedural stages are causing delays (e.g., Appearance, Compliance).

**3. Case Age Distribution**

**Formula:**  
Category-wise Case Count (e.g., 0–1 year, 1–3 years, etc.)  
**Purpose:**  
This categorization helps users assess how long cases remain unresolved and whether courts are effective in clearing recent vs. older cases.

**4. Filing vs. Disposal Ratio**

**Formula:**  
Ratio = Total Cases Filed / Total Cases Disposed  
**Purpose:**  
Indicates whether courts are keeping up with incoming cases or adding to the backlog over time.

**5. Special Case Proportion**

**Formula:**  
(Special Category Cases / Total Cases) × 100  
Where special category includes:

* Women Litigants
* Senior Citizens  
  **Purpose:**  
  Measures the inclusiveness and social impact of court operations, identifying districts where vulnerable populations are more engaged in the legal system.

**6. District-Wise Delay Score**

**Formula (Derived):**  
Delay Score = Weighted Sum of Long-Pending Cases (e.g., 10–20 years, >20 years)  
**Purpose:**  
Highlights areas where prolonged delays are most severe, guiding reform priorities or staff allocations.

These formulas and calculations were integrated into the visualizations (bar charts, heatmaps, scatter plots, etc.) to make the findings more accessible. Information was clearly labeled in each chart so users could understand the context and implications without needing technical expertise. The goal was to make the report interpretable for both judicial experts and the general public.

**5. CONCLUSION**

This project provided a comprehensive and data-driven overview of the challenges faced by the Indian judiciary, specifically at the District and Taluk court levels. Through rigorous exploratory data analysis (EDA), we have unearthed critical insights that highlight both systemic issues and opportunities for improvement.

The analysis of pending court cases revealed significant disparities in case load across different states and districts. Heavily populated or litigation-prone states like Uttar Pradesh, Maharashtra, and Bihar bear the highest burdens, indicating the need for region-specific judicial reforms. At the district level, several smaller regions exhibited disproportionately high case pendency, possibly due to limited judicial infrastructure, staff shortages, or inefficiencies in process management.

Our duration-wise analysis showed a concerning number of cases pending for more than 5 years — some extending beyond three decades. This signals deeply rooted procedural delays and the need for stronger mechanisms to prioritize and fast-track older cases.

Furthermore, the stage-wise breakdown of case pendency helped pinpoint where delays tend to accumulate. The fact that a majority of cases are stuck at early procedural stages like "Appearance" and "Compliance" suggests that basic operational hurdles — such as serving notices, verifying documents, or recording initial proceedings — contribute heavily to the overall backlog. This observation underscores the importance of administrative reform and better case management systems.

By comparing cases instituted versus those disposed of on a monthly basis, we observed that in most regions, the rate of case filings exceeds the rate of disposals. This growing imbalance implies that the backlog is not only persistent but is also likely to worsen unless measures are taken to improve judicial throughput.

An important social dimension was captured by analyzing special case categories, particularly those involving women and senior citizens. The data revealed that these demographics are active litigants in certain districts, highlighting the judiciary's relevance to vulnerable populations. However, the delays in these sensitive cases also signal a need for greater sensitivity, prioritization, and fast-track procedures.

Our visualizations served as a powerful tool in identifying hidden patterns and communicating findings effectively. They made the analysis accessible not only to technical audiences but also to stakeholders in the legal and policy domains.

In essence, this project doesn’t just quantify the backlog — it dissects it across time, geography, stage, and social impact. The findings suggest that addressing this crisis will require a multi-pronged approach involving judicial appointments, infrastructure expansion, digitization, and administrative streamlining.

This work provides a strong foundation for further predictive modeling and decision support systems aimed at improving case clearance rates and delivering timely justice.

**6. FUTURE SCOPE**

While the current project successfully explored the state of pending court cases across Indian district and taluk courts, there is substantial potential to enhance this analysis and extend its real-world utility. The following future enhancements and research directions can provide deeper insights and practical tools for judicial reforms and policy making:

**1. Predictive Modelling and Forecasting**

* Implement machine learning models such as time series forecasting (e.g., ARIMA, Prophet) or regression models to predict the future pendency of cases based on historical trends.
* Classification models could also predict the likelihood of delay for a newly instituted case based on attributes like court type, district, case category, and stage.

**2. Real-Time Dashboard for Monitoring**

* Develop an interactive dashboard using tools like Dash, Tableau, or Power BI to provide real-time visualization of pendency statistics.
* Such a dashboard can be integrated with official systems to track changes in filing, disposal, and bottlenecks dynamically, allowing court administrators to make timely interventions.

**3. Time-Series and Trend Analysis**

* With extended temporal data (monthly or quarterly over several years), one could apply more advanced time-series analyses to detect cyclical patterns, seasonal backlogs, and long-term judicial trends.
* This would help identify specific months or periods when courts experience high or low throughput.

**4. Geospatial and Infrastructure Mapping**

* Integrate court case data with geographical data (GIS mapping) to visualize not just case loads but also correlate them with factors such as court infrastructure, population density, and distance to nearest court.
* This can guide the establishment of new courts or redistribution of case load to balance regional pressures.

**5. Policy Simulation Tools**

* Develop simulation models that allow policy makers to test "what-if" scenarios. For example:
  + What if 10 more judges are added in a district?
  + What if digital filing is introduced in rural courts?
  + These simulations can estimate the impact of interventions before implementation.

**6. Integration with Legal NLP (Natural Language Processing)**

* Use NLP techniques to analyse case documents, categorize case types more accurately, and extract relevant details (e.g., delays, adjournments) from unstructured legal texts.
* Sentiment analysis on case outcomes or judge remarks could also provide deeper judicial behaviour insights.

**7. Citizen-Centric Applications**

* Build mobile/web applications that allow citizens to track the progress of their cases, understand delays, and view historical trends in their regions.
* These apps could also include AI-driven insights for litigants regarding expected timelines and next possible actions.

**8. Collaboration with Government and Judicial Bodies**

* Extend this project into a collaborative pilot with district courts, judicial academies, or Ministry of Law & Justice to validate findings on the ground and implement data-driven improvements.

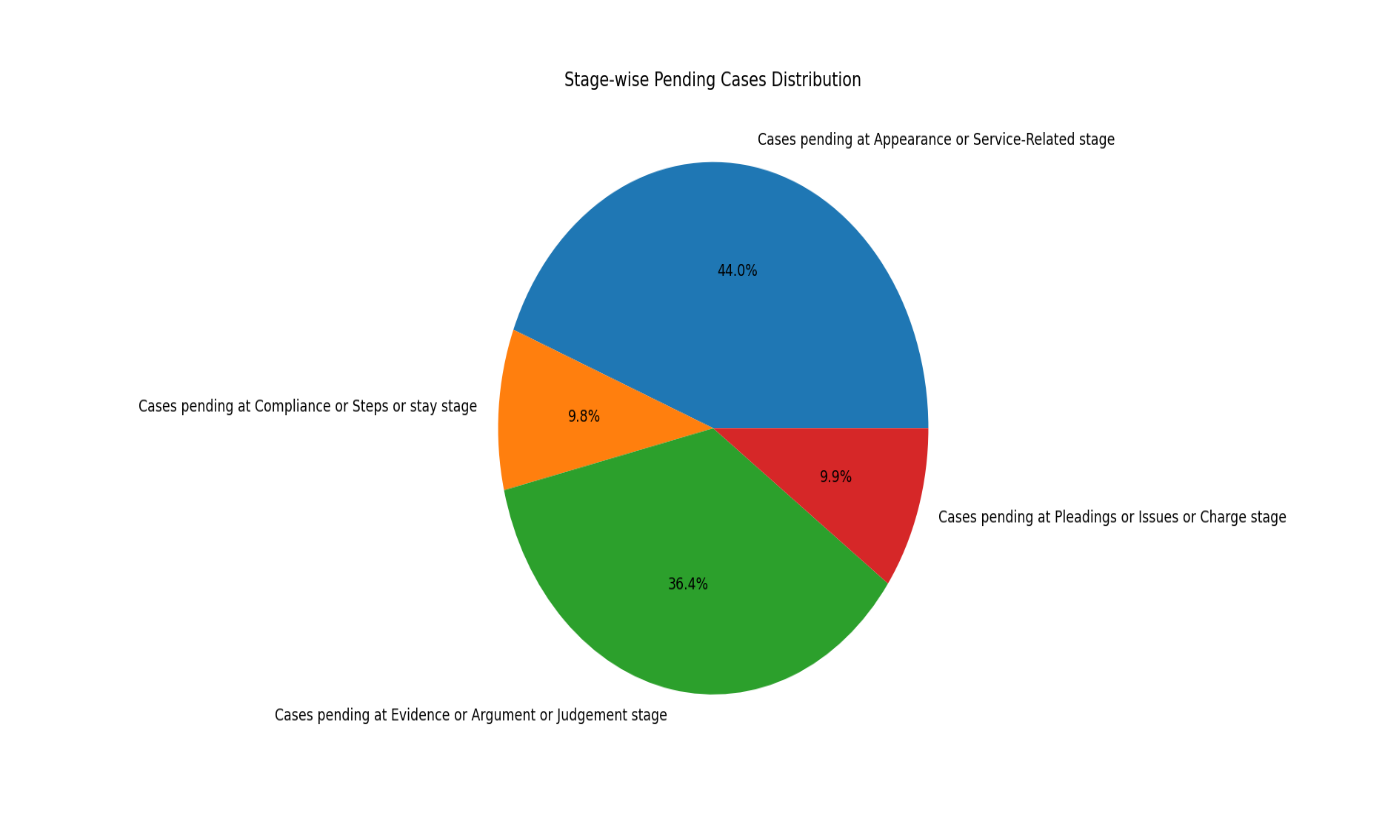
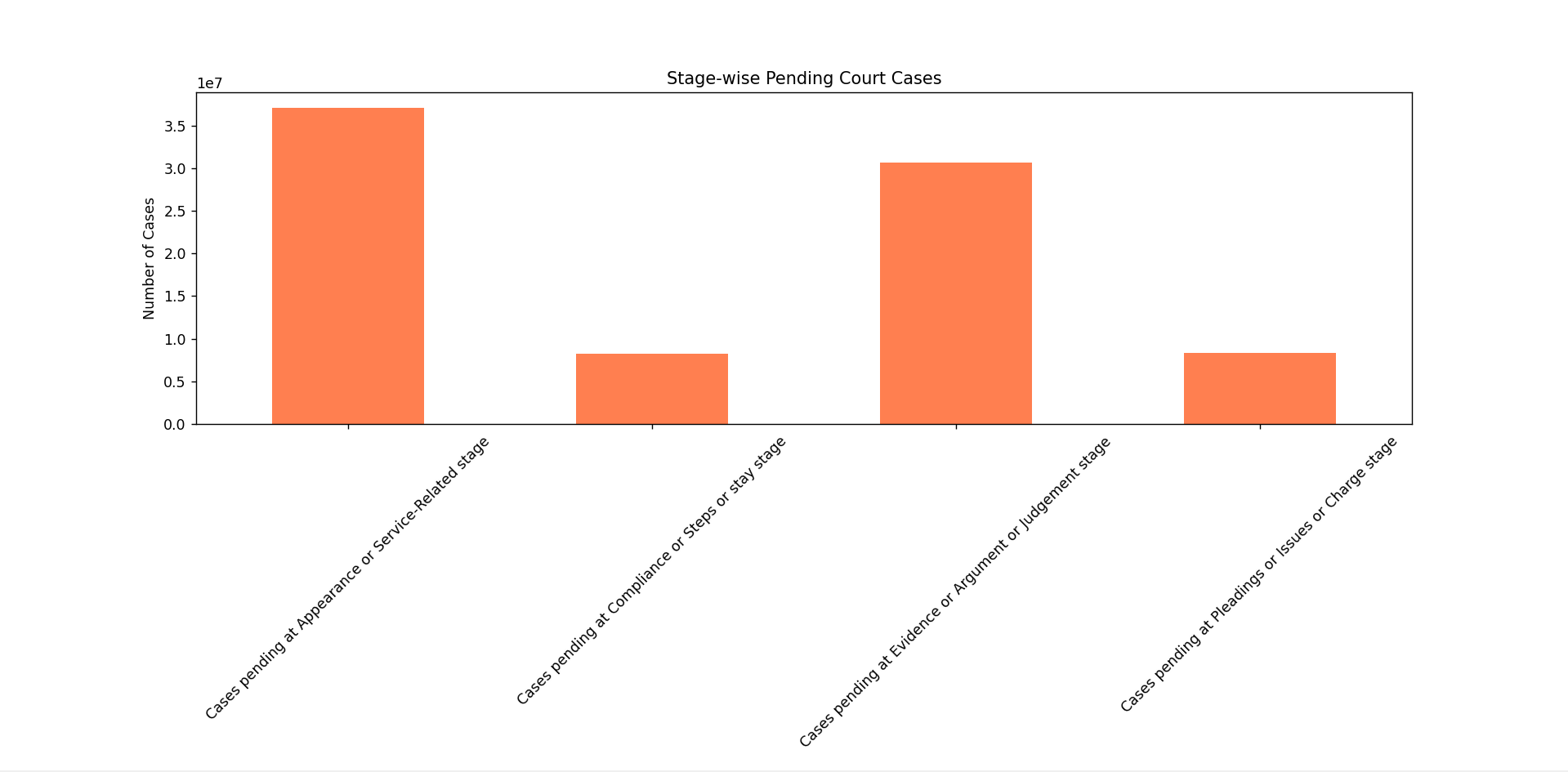
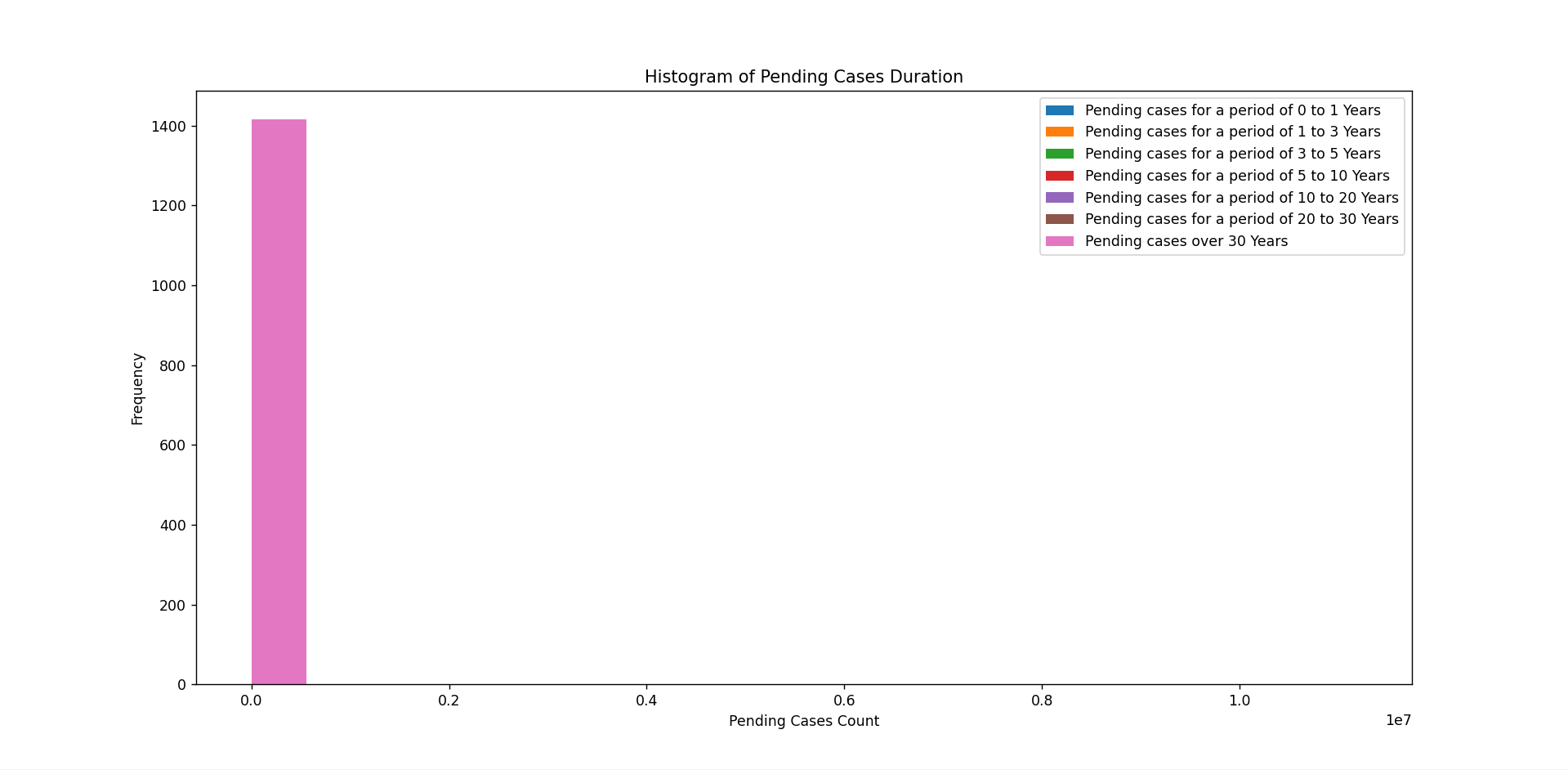
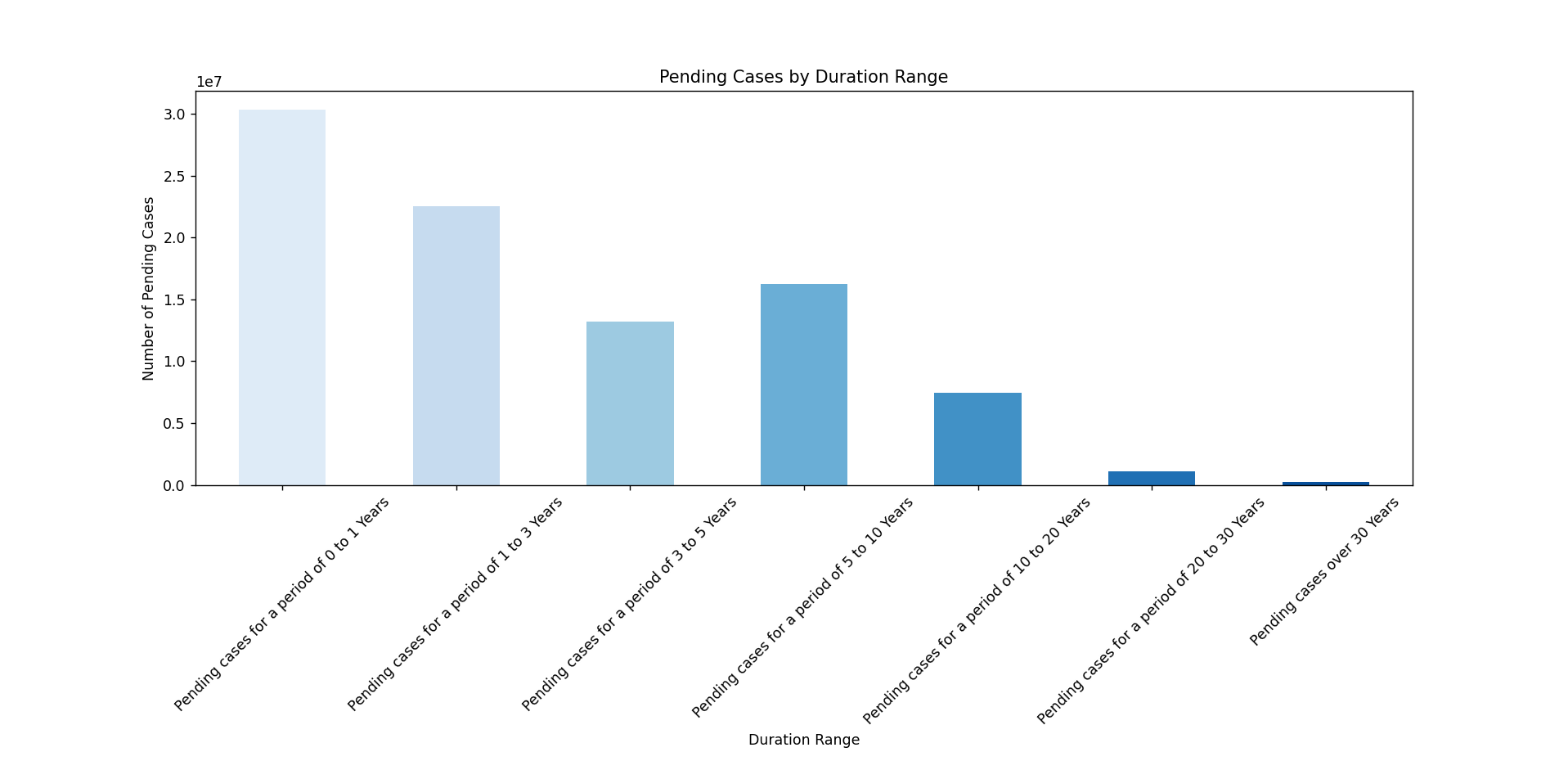
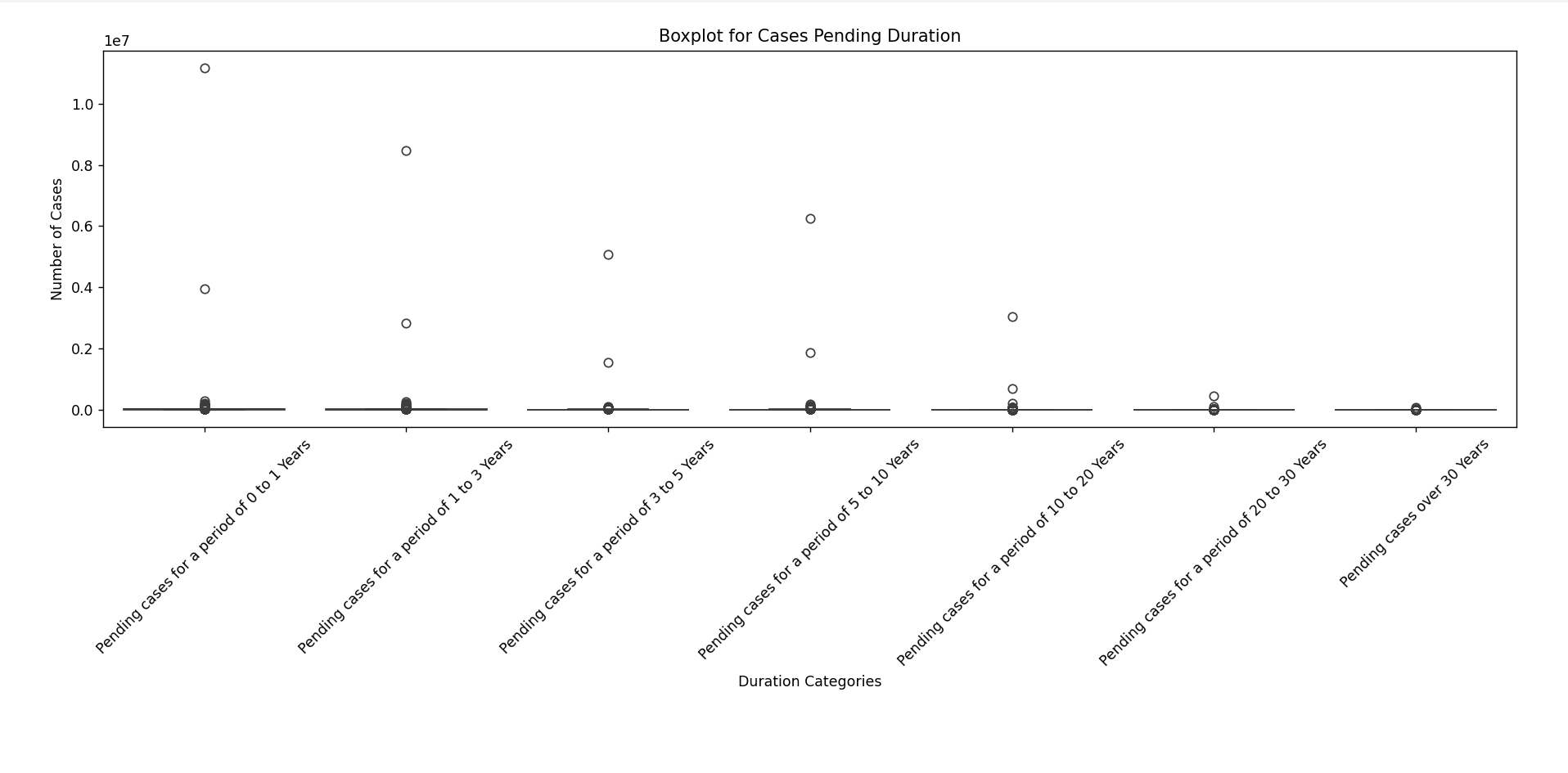
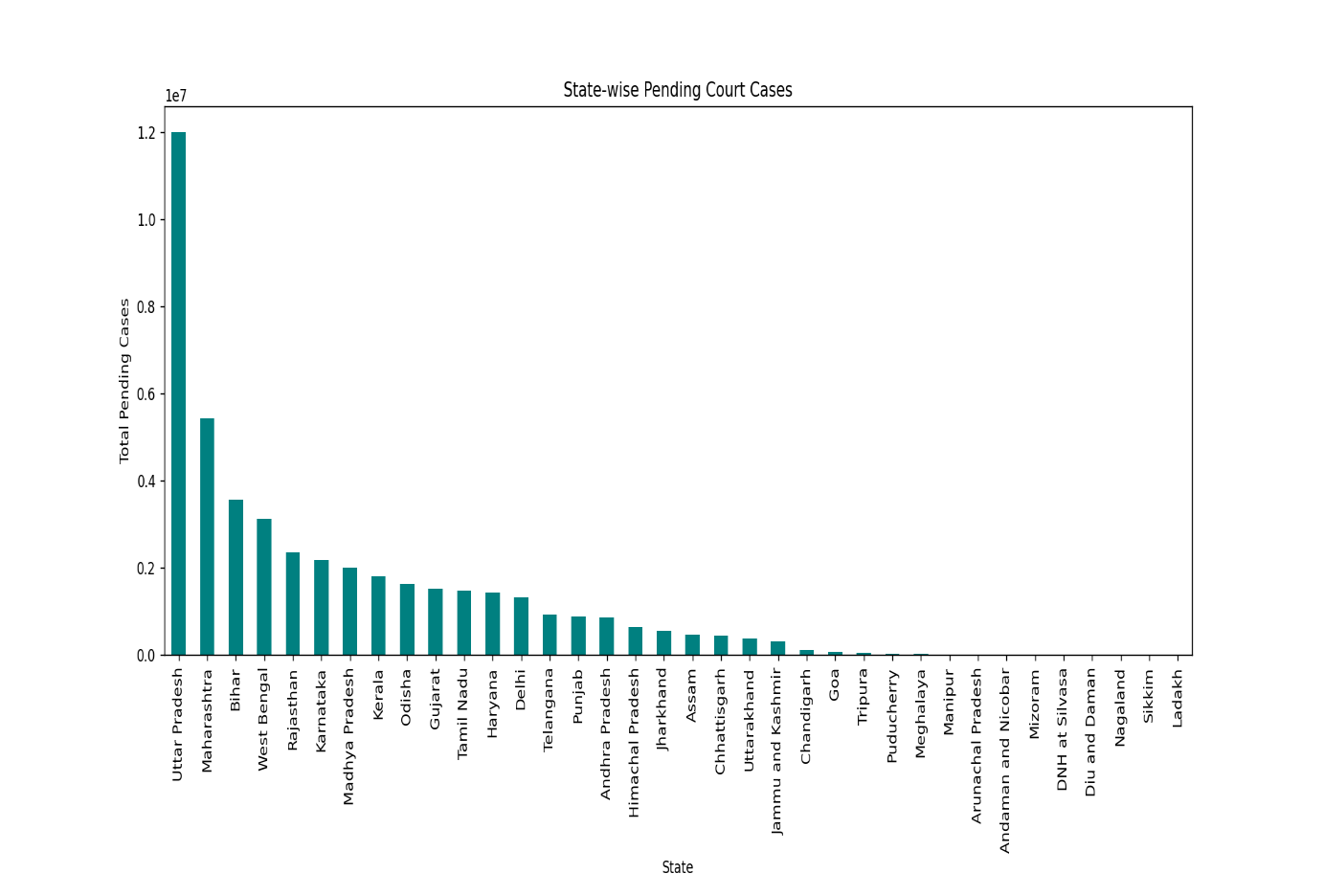
**9. Incorporate Additional Variables**

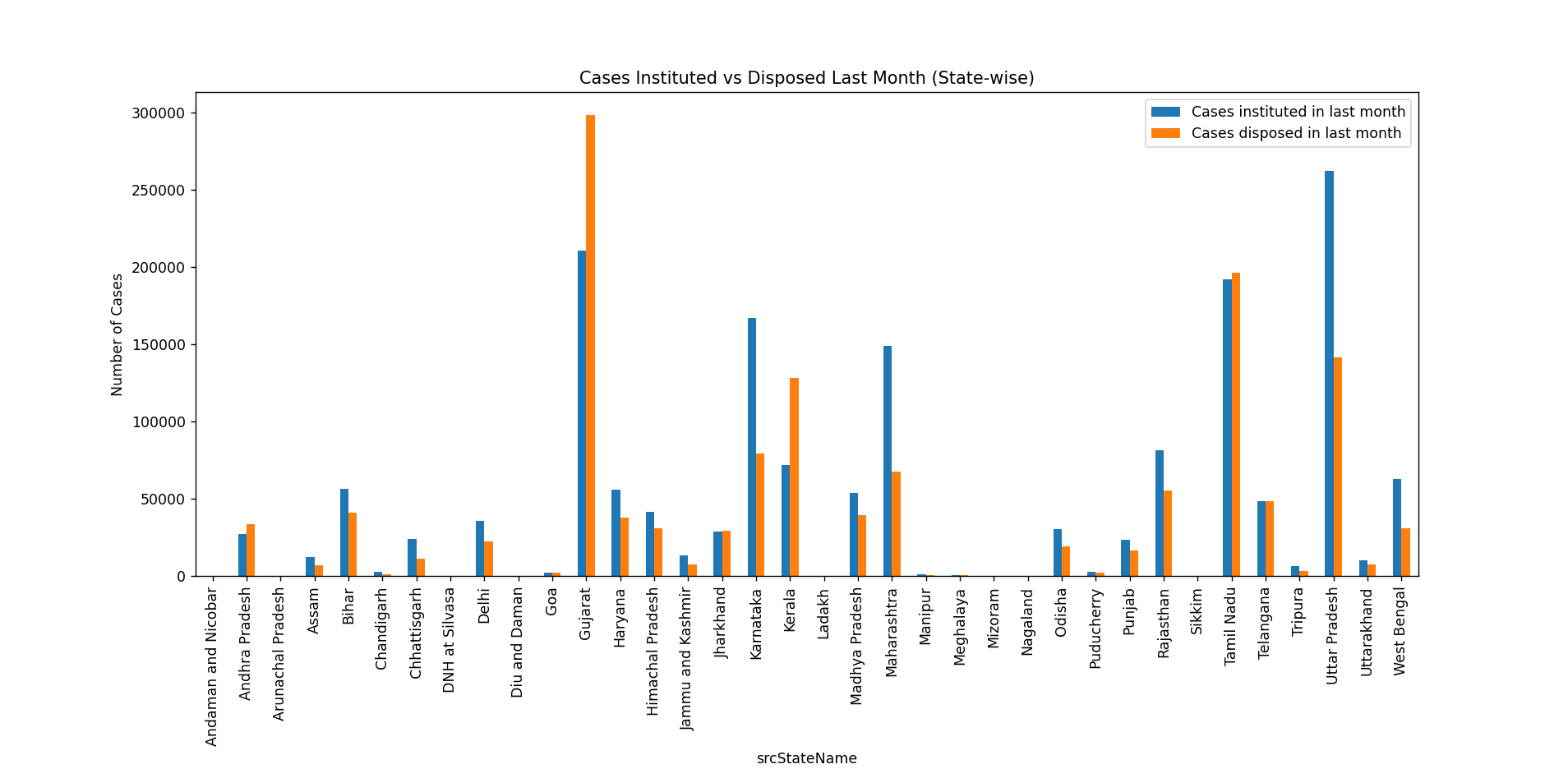
* Future datasets could include variables like number of judges per district, number of adjournments, average hearing time, or digitization level of the court — enabling a more holistic view of systemic delays.

**10. Ethical and Social Impact Analysis**

* Study the impact of court delays on vulnerable groups (e.g., women, elderly, marginalized communities) to design more inclusive and equitable reforms.
* Collaborate with NGOs or legal aid groups to better understand and respond to the lived experiences behind the data.

1. **Screenshot of Objective**





A screenshot of a graph

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A close-up of a chart

AI-generated content may be incorrect.

**Python Code:**

import pandas as pd

df = pd.read\_csv('cleaned\_data.csv')

print(df.head())

print(df.info())

print(df.isnull().sum())

print(df.columns)

print(df.describe())

#Total Pending Cases State-wise

state\_cases = df.groupby('srcStateName')['Pending cases'].sum().reset\_index()

print(state\_cases)

#Total Pending Cases District-wise

district\_cases = df.groupby('srcDistrictName')['Pending cases'].sum().reset\_index()

print(district\_cases)

#Analyze Pending Cases by Year Range

year\_range\_cols = [

'Pending cases for a period of 0 to 1 Years',

'Pending cases for a period of 1 to 3 Years',

'Pending cases for a period of 3 to 5 Years',

'Pending cases for a period of 5 to 10 Years',

'Pending cases for a period of 10 to 20 Years',

'Pending cases for a period of 20 to 30 Years',

'Pending cases over 30 Years'

]

import matplotlib.pyplot as plt

import seaborn as sns

#Pending Cases Across States & Districts

# State-wise Pending Cases - Bar Plot

plt.figure(figsize=(12,6))

state\_cases = df.groupby('srcStateName')['Pending cases'].sum().sort\_values(ascending=False)

state\_cases.plot(kind='bar', color='teal')

plt.title('State-wise Pending Court Cases')

plt.xlabel('State')

plt.ylabel('Total Pending Cases')

plt.xticks(rotation=90)

plt.show()

pending\_by\_years = df[year\_range\_cols].sum()

print(pending\_by\_years)

#Analyze Cases by Stage

stage\_cols = [

'Cases pending at Appearance or Service-Related stage',

'Cases pending at Compliance or Steps or stay stage',

'Cases pending at Evidence or Argument or Judgement stage',

'Cases pending at Pleadings or Issues or Charge stage'

]

cases\_by\_stage = df[stage\_cols].sum()

print(cases\_by\_stage)

#Total Cases Filed vs Disposed in Last Month

filed\_vs\_disposed = df[['Cases instituted in last month', 'Cases disposed in last month']].sum()

print(filed\_vs\_disposed)

#Total Cases Filed by Senior Citizens and Women

special\_cases = df[['Cases filed by Senior Citizens', 'Cases filed by women']].sum()

print(special\_cases)

# Objective 2: Duration-wise Pending Cases Analysis

duration\_cols = [

'Pending cases for a period of 0 to 1 Years',

'Pending cases for a period of 1 to 3 Years',

'Pending cases for a period of 3 to 5 Years',

'Pending cases for a period of 5 to 10 Years',

'Pending cases for a period of 10 to 20 Years',

'Pending cases for a period of 20 to 30 Years',

'Pending cases over 30 Years'

]

#Boxplot for Cases Pending Duration

plt.figure(figsize=(10, 6))

sns.boxplot(data=df[duration\_cols])

plt.title('Boxplot for Cases Pending Duration')

plt.xlabel('Duration Categories')

plt.ylabel('Number of Cases')

plt.xticks(rotation=45)

plt.tight\_layout()

plt.show()

# Stacked Bar Chart

df\_duration\_sum = df[duration\_cols].sum()

df\_duration\_sum.plot(kind='bar', stacked=True, color=sns.color\_palette("Blues", 7))

plt.title('Pending Cases by Duration Range')

plt.xlabel('Duration Range')

plt.ylabel('Number of Pending Cases')

plt.xticks(rotation=45)

plt.show()

# Histogram

df[duration\_cols].plot.hist(bins=20, figsize=(10,6))

plt.title('Histogram of Pending Cases Duration')

plt.xlabel('Pending Cases Count')

plt.show()

# Objective 3: Stage-wise Pending Cases Analysis

stage\_cols = [

'Cases pending at Appearance or Service-Related stage',

'Cases pending at Compliance or Steps or stay stage',

'Cases pending at Evidence or Argument or Judgement stage',

'Cases pending at Pleadings or Issues or Charge stage'

]

# Bar Plot

df\_stage\_sum = df[stage\_cols].sum()

df\_stage\_sum.plot(kind='bar', color='coral')

plt.title('Stage-wise Pending Court Cases')

plt.xticks(rotation=45)

plt.ylabel('Number of Cases')

plt.show()

# Pie Chart

df\_stage\_sum.plot(kind='pie', autopct='%1.1f%%', figsize=(7,7))

plt.title('Stage-wise Pending Cases Distribution')

plt.ylabel('')

plt.show()

# Objective 4: Cases Instituted vs Disposed Last Month

plt.figure(figsize=(12,6))

df.groupby('srcStateName')[['Cases instituted in last month', 'Cases disposed in last month']].sum().plot(kind='bar')

plt.title('Cases Instituted vs Disposed Last Month (State-wise)')

plt.ylabel('Number of Cases')

plt.xticks(rotation=90)

plt.show()

# Scatter Plot: Correlation

plt.figure(figsize=(8,6))

sns.scatterplot(x='Cases instituted in last month', y='Cases disposed in last month', data=df)

plt.title('Correlation: Cases Instituted vs Disposed')

plt.xlabel('Cases Instituted')

plt.ylabel('Cases Disposed')

plt.show()

# -----------------------------------------------

# Objective 5: Special Case Categories (Senior Citizens, Women, Delayed)

# Bar Plot: Senior Citizens vs Women Cases

plt.figure(figsize=(10,6))

df[['Cases filed by Senior Citizens', 'Cases filed by women']].sum().plot(kind='bar', color=['orange', 'purple'])

plt.title('Cases Filed by Senior Citizens vs Women')

plt.ylabel('Total Cases')

plt.xticks(rotation=0)

plt.show()

# Group by State and Sum Duration-wise Pending Cases

state\_duration = df.groupby('srcStateName')[duration\_cols].sum()

# Plot Heatmap

plt.figure(figsize=(14, 8))

sns.heatmap(state\_duration, annot=True, fmt='.0f', cmap='YlGnBu', linewidths=0.5)

plt.title('Heatmap: State vs Duration-wise Pending Cases')

plt.xlabel('Case Pending Duration')

plt.ylabel('State')

plt.show()

# ****9.REFERENCES****

1.  **National Judicial Data Grid (NJDG)** –  
   https://njdg.ecourts.gov.in/njdgnew/  
   *(Official source for data on pending and disposed cases in Indian courts.)*
2.  **Department of Justice, Government of India** –  
   <https://doj.gov.in/>  
   *(Provides updates on judicial reforms, court infrastructure, and justice delivery.)*
3.  **eCourts Mission Mode Project** –  
   <https://ecourts.gov.in/>  
   *(A Government of India initiative for digital transformation of courts.)*
4.  **Law Commission of India Reports** –  
   https://lawcommissionofindia.nic.in/reports.htm  
   *(Contains valuable studies and recommendations on judicial delays and case backlogs.)*
5.  **KPMG Report on Judicial System Reforms** –  
   *KPMG (2020), “Reimagining the Indian Judiciary through Digitization”*  
   *(Available through KPMG India insights.)*
6.  **NITI Aayog Reports on Justice Delivery** –  
   <https://www.niti.gov.in/>  
   *(Strategic insights on improving governance and judicial infrastructure in India.)*
7.  **Academic Journals and Articles** –
8. Singh, A. (2021). *Delays in the Indian Judiciary: Causes and Remedies*. Indian Journal of Law and Governance.
9. Mehta, P. B. (2019). *Judicial Backlogs and Institutional Reforms in India*. Journal of Constitutional Studies.
10.  **Python and Data Analysis Tools** –
11. McKinney, W. (2017). *Python for Data Analysis*. O'Reilly Media.
12. Official documentation of Pandas, Matplotlib, and Seaborn libraries.  
    <https://pandas.pydata.org/>,  
    <https://matplotlib.org/>,  
    <https://seaborn.pydata.org/>

Linkedin :

<https://www.linkedin.com/posts/ayush-kumar-6a16132a3_python-datascience-datavisualization-activity-7316756396799090689-4tTN?utm_source=share&utm_medium=member_desktop&rcm=ACoAAEk9YRkB9uS-1dsbwAYNs4YcWAy_ETSQH9w>

Github:

<https://github.com/Ayushk-07/python-project>