**DATA SCIENCE**

**PROJECT REPORT**

(Project Semester January-April 2025)

**Exploratory Data Analysis on** **Court Case Backlogs Across India**

Submitted by

Maneet Kumar

Registration No: 12305153

Programme and Section: B.Tech CSE K23EU

Course Code: INT375

Under the Guidance of

Dr. Tanima Thakur

Discipline of CSE/IT

**Lovely School of Computer Science and Engineering**

**Lovely Professional University, Phagwara**

# DECLARATION

I, Maneet Kumar, student of 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: 2025-04-11

Signature Maneet Kumar

Registration No. 12305153  
  
Name of the student: Maneet Kumar

# CERTIFICATE

This is to certify that Mr. Maneet Kumar bearing Registration No. 12305153 has completed INT375 project titled, “Exploratory Data Analysis on Court Case Backlogs Across India” under my guidance and supervision. To the best of my knowledge, the present work is the result of his original development, effort and study.

Signature and Name of the Supervisor  
Dr. Tanima Thakur  
  
Designation of the Supervisor  
School of Computer Science and Engineering  
Lovely Professional University  
Phagwara, Punjab.

Date: 2025-04-13

# ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my mentor Dr. Tanima Thakur for her valuable guidance throughout this project. I also thank Lovely Professional University for providing the necessary resources and support.

# TABLE OF CONTENTS

1. Introduction

2. Source of Dataset

3. EDA Process

4. Analysis on Dataset

* General Description
* Specific Requirements, Functions and Formulas
* Analysis Results
* Visualizations

5. Conclusion

6. Future Scope

7. References

# Description

This project is about studying a dataset that shows how many court cases are still waiting to be solved in different parts of India. My main aim was to find out how many cases are pending, where they are happening the most, and if there are any interesting patterns. I also made some nice graphs to show things like how long the cases have been stuck and which states have the most pending cases.

I cleaned the data first by fixing things like wrong column names and missing information. Then I did some basic math to understand the data better. I made different kinds of graphs like bar charts, pie charts, histograms, box plots, and scatter plots. I also tried to see how delayed cases are related to the total pending cases.

**Challenges Faced:**

Some parts of the project were a bit hard! Cleaning the data took time because some names had extra spaces and some information was missing. Choosing the right type of graph was also tricky—I had to think carefully to make sure the graphs looked good and explained the data well. Also, understanding how delayed cases and pending cases are connected was confusing at first, but I learned a lot by working through it.

# Introduction

Today, data is very helpful for making smart decisions. By knowing how many court cases are still not solved, we can help improve the court system. In this project, I used EDA (Exploratory Data Analysis) to study a dataset about pending court cases in different states and districts of India. I used Python and some cool tools to look at the data and find patterns. This helped me see which places have more pending cases and spot trends that can help judges and leaders. This project also taught me how to work with real data and find useful information from it.

# Source of Dataset

The data I used in this project came from a trusted website called the National Data and Analytics Platform (<https://ndap.niti.gov.in/>). It has a lot of information about pending court cases in different states and districts of India.

You can check out the same dataset I used by going to this link: <https://ndap.niti.gov.in/dataset/7150?filter_id=3609>.

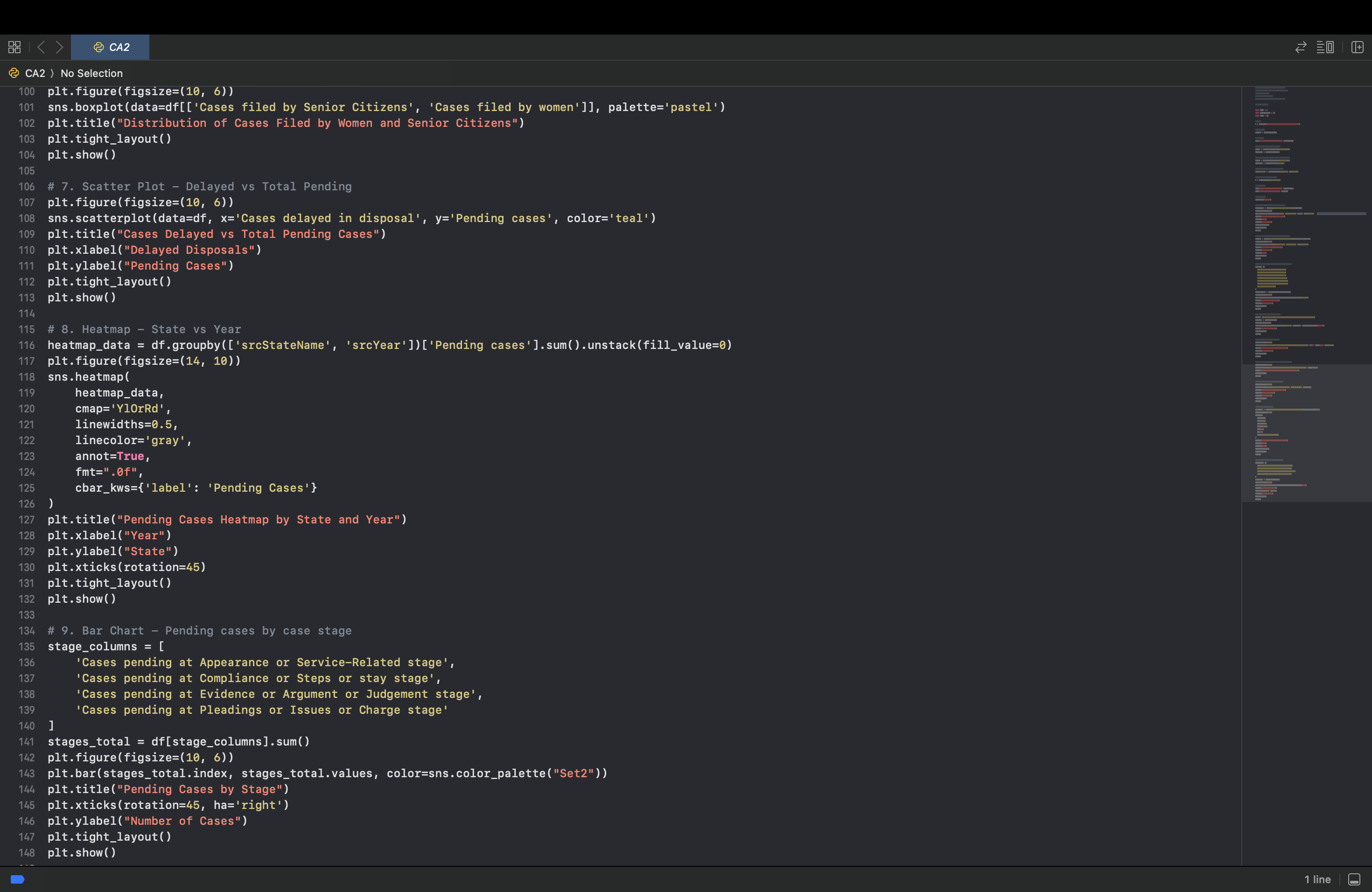
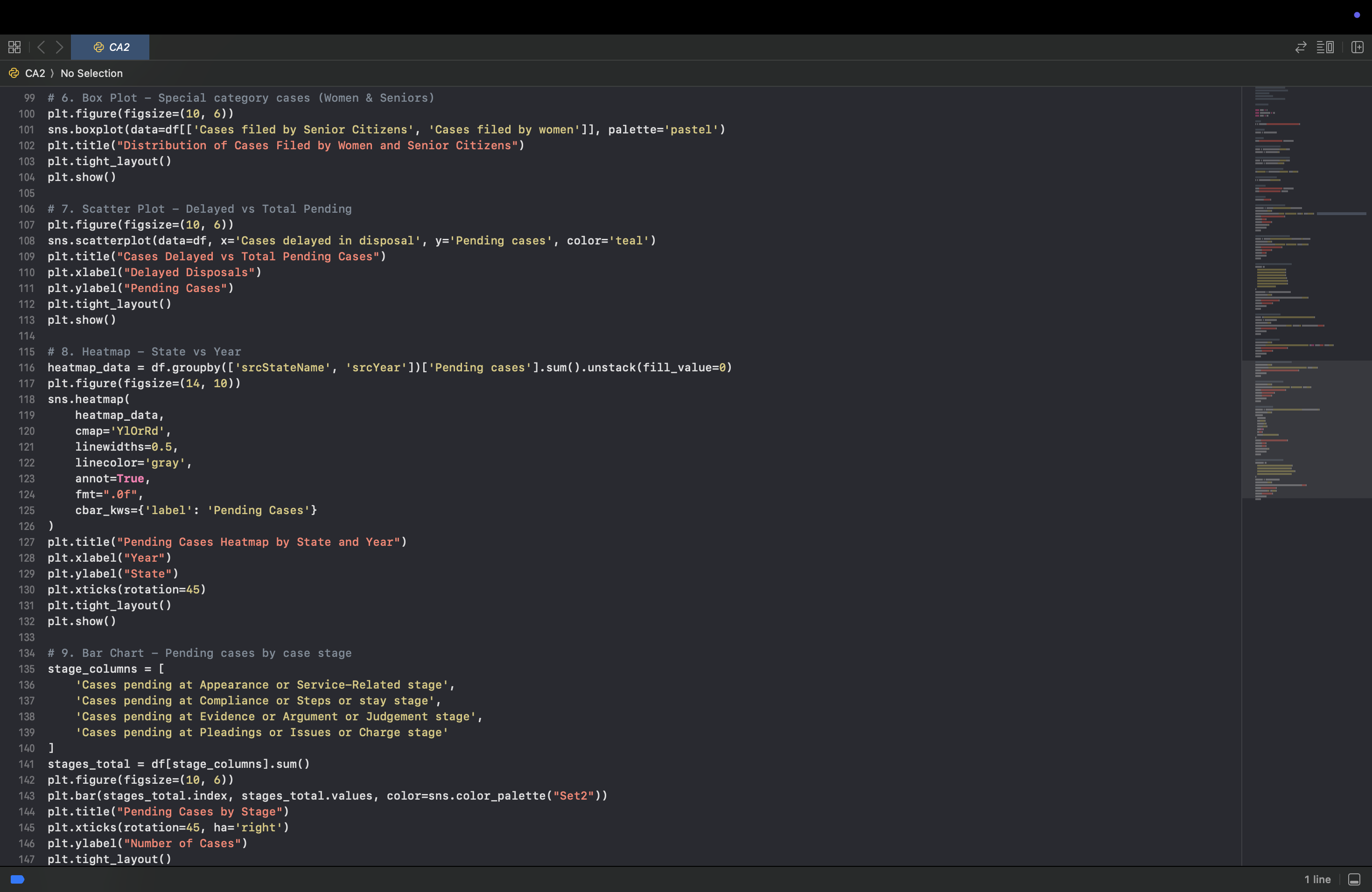
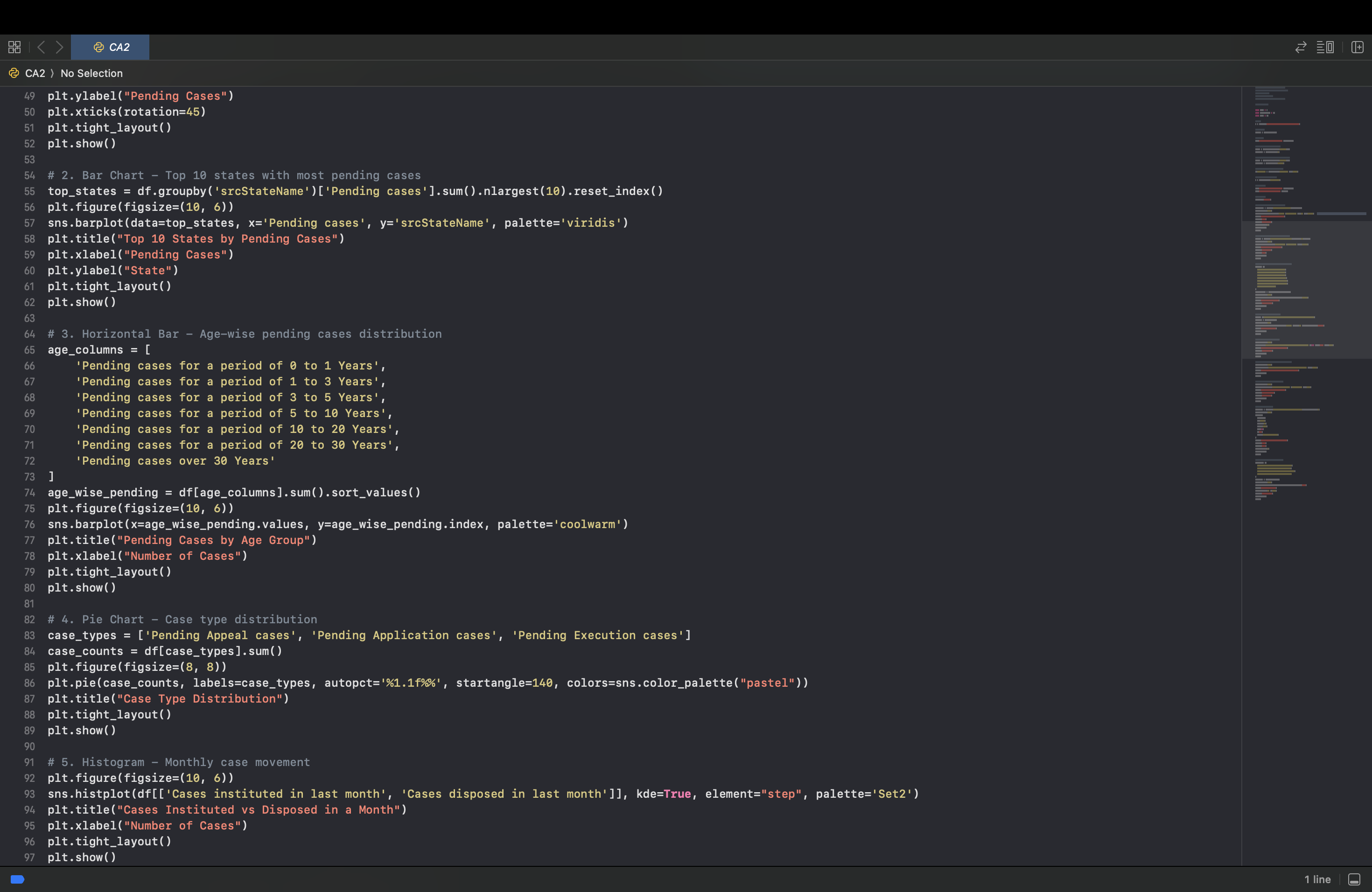
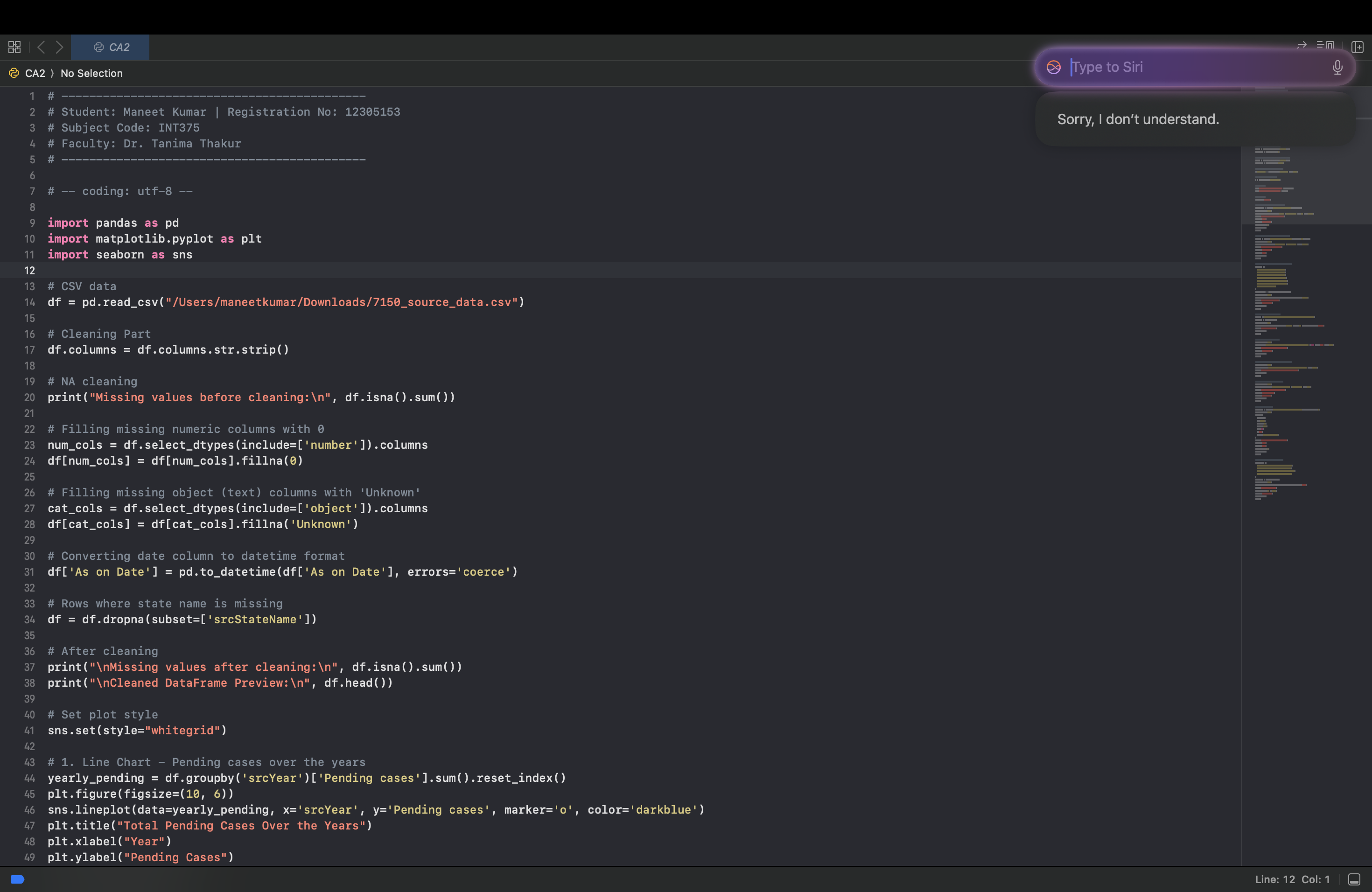
This website is really cool because it has many types of data that can be used for academic and research. The dataset was in CSV format, and I used the pandas library in Python to open it and study it.

# EDA Process

I did Exploratory Data Analysis (EDA) step-by-step to understand the court cases dataset and find some interesting trends. Here’s what I did:

* **Loading the Dataset**: I opened the CSV file in Python using pd.read\_csv() so I could work with the data easily.
* **First Look at the Data**: I used .shape to see how big the dataset was, .info() to check the columns, and .head() to look at a few sample rows.
* **Cleaning the Data**: I fixed the column names by removing extra spaces using .str.strip(). I made sure the state and district names were easy to read. I also checked for missing values. If a number was missing, I filled it with 0. If text was missing, I filled it with "Unknown" so everything stayed neat.
* **Making Graphs**: I made many types of graphs like line charts, bar charts, pie charts, histograms, box plots, scatter plots, and even a heatmap. I used Seaborn and Matplotlib libraries to do this and find patterns in the data.
* **Looking for Patterns**: I compared delayed cases with pending cases using scatter plots and other graphs.
* **What I Learned**: I found out which states had the most pending cases, how long the cases have been waiting, and what stage they’re stuck at.

Each step helped me turn messy data into clear and useful information about court cases!



# 4.Analysis on Dataset

I. General Description

The dataset has information about pending court cases in different states and districts of India. It shows if the cases are civil or criminal, how long they’ve been waiting (like 0–1 years or more than 30 years), the type of cases (like appeals or applications), and what stage they are stuck at (like evidence or arguments). By looking at this, we can understand which courts are very busy and why cases are getting delayed.

II. Specific Requirements, Functions and Formulas

**Libraries Used**: pandas, matplotlib, and seaborn

**Some Python Functions I Used**:

* .str.strip() – to clean extra spaces in column names
* .isna().sum() – to find missing data
* .fillna() – to fill in missing data
* .to\_datetime() – to fix date columns
* .groupby() – to group data by state or year
* Plotting functions like sns.lineplot(), sns.barplot(), plt.pie(), sns.histplot(), sns.boxplot(), sns.scatterplot(), and sns.heatmap() for making graphs

III. Analysis Results

* **Line charts** showed that the number of pending cases changes over time—some years had more, some had less.
* **States like Uttar Pradesh and Maharashtra** had the most pending cases.

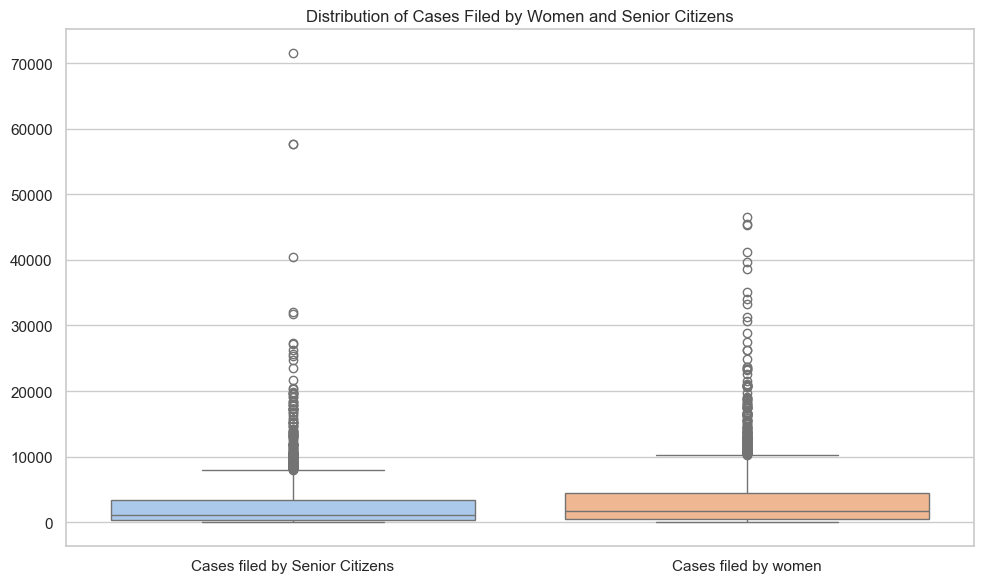
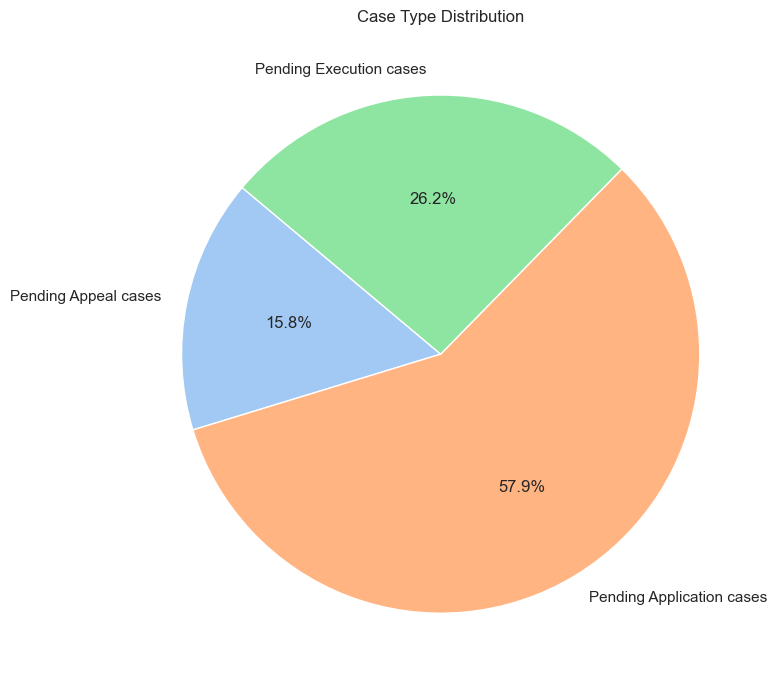
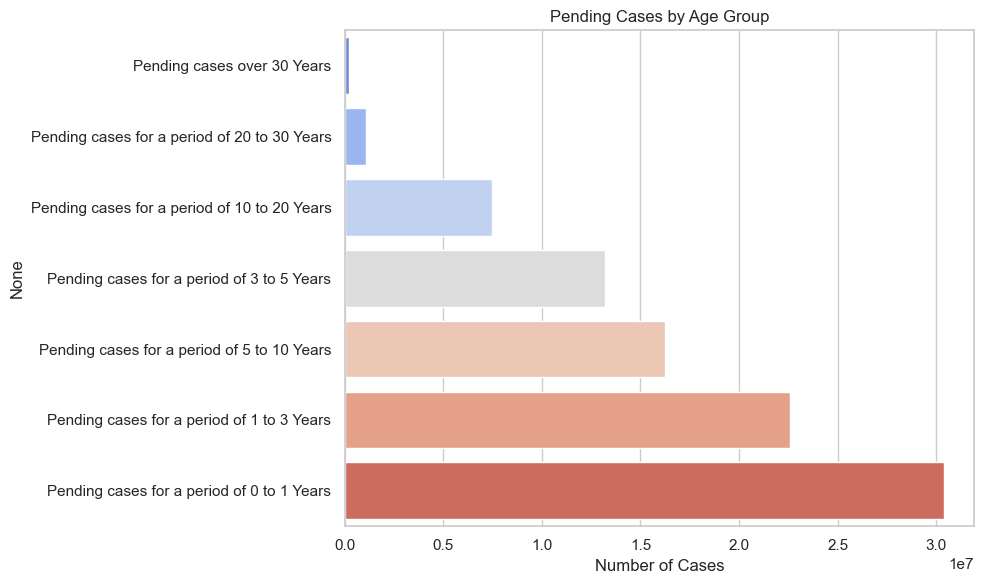
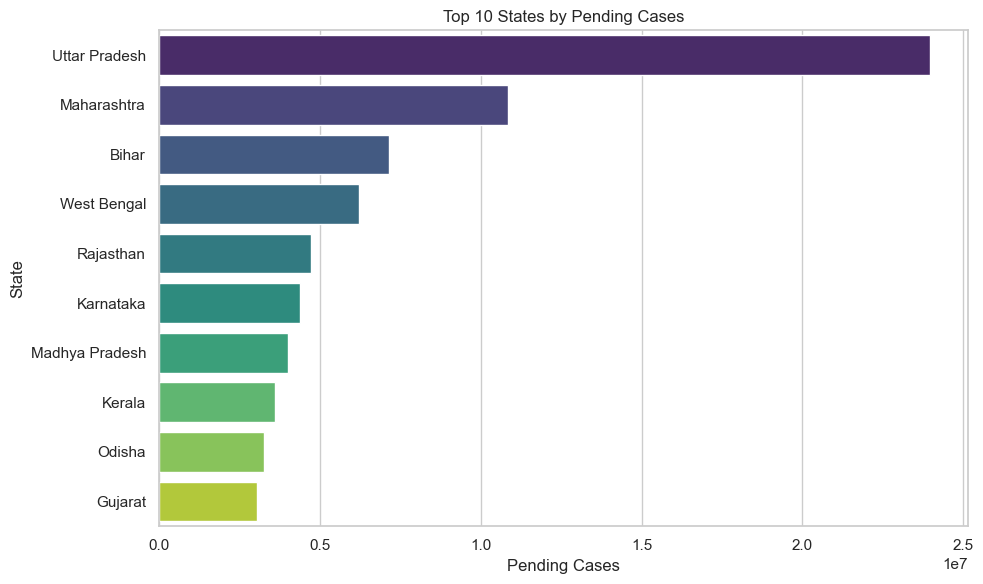
Places with more pending cases also had more delayed ones. That means both are connected.

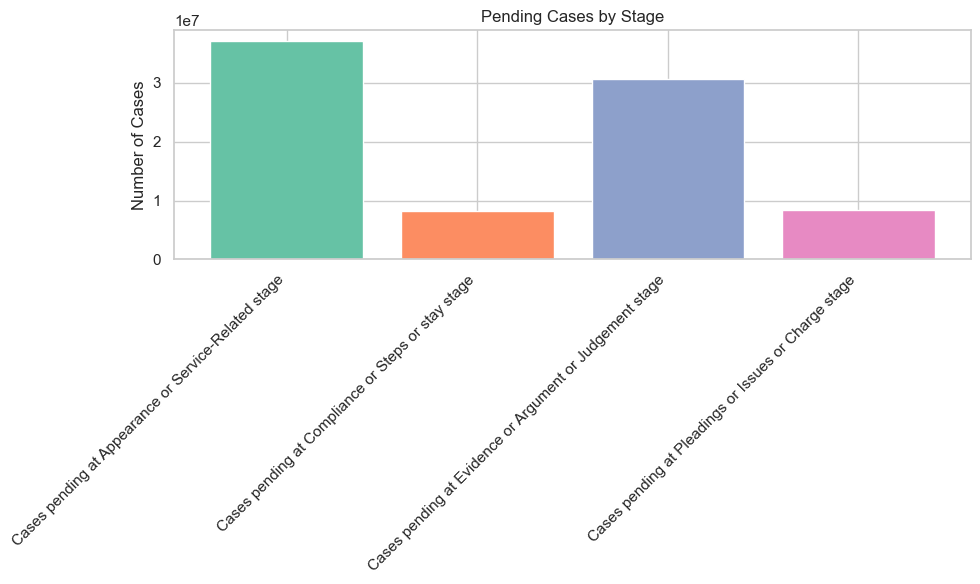
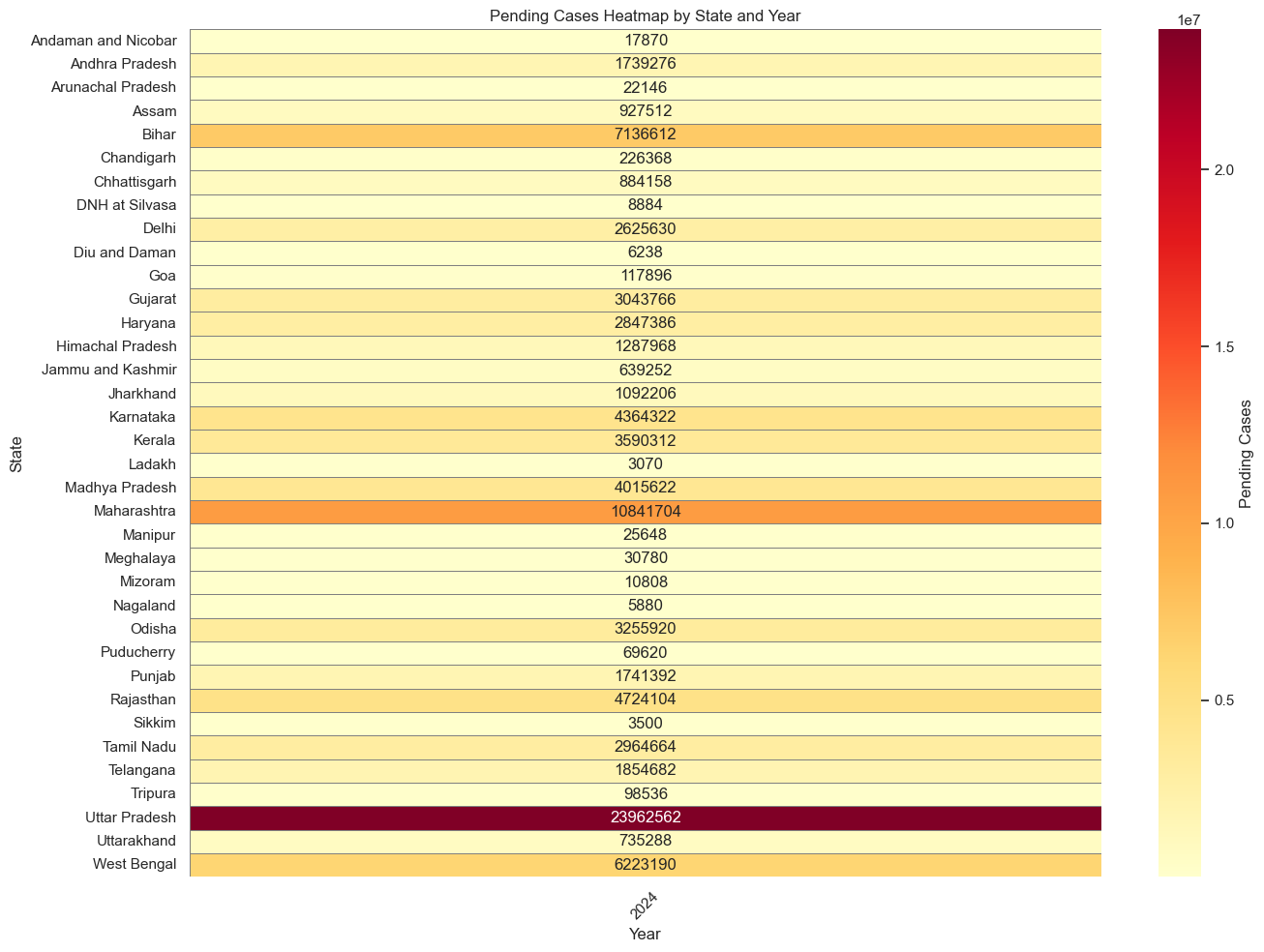
* **Most cases are new** (0–1 years or 1–3 years), but some are **very old** (more than 30 years!).

IV. Visualizations

* **Line Chart**: Total pending cases over the years.
* **Bar Chart**: Top 10 states with the most pending cases.
* **Horizontal Bar Chart**: Cases pending for different time periods (like 0-1 years or over 30 years).
* **Pie Chart**: Types of cases (appeal, application, execution).
* **Histogram**: Cases filed vs. cases solved in a month.
* **Box Plot**: Cases filed by women and senior citizens.
* **Scatter Plot**: Delayed cases vs. total pending cases.
* **Heatmap**: Pending cases by state and year.

A graph with numbers and a dot

AI-generated content may be incorrect.A graph with blue dots

AI-generated content may be incorrect.

# 5. Conclusion

This project gave me a deep look at how court cases are piling up in India. I used data analysis and cool graphs to understand the story behind the numbers. By studying the dataset, I found out which states have the most pending cases and saw how delayed cases are connected to the total number of cases. These patterns helped me understand where courts are very busy and what might be causing the delays.

But this project wasn’t just about the results—it was also a big learning journey for me! I got better at cleaning messy data, making clear graphs, and finding patterns in real-world information. Sometimes the data was hard to work with, but I learned how to fix problems, focus on the important parts, and show things in a simple way using graphs. Overall, this project made me much more confident in working with data and helped me learn how to find useful things hidden in it.

# 6. Future Scope

There’s still a lot more I can do to make this court cases project even better! Here are some cool ideas for the future:

* I can **build prediction models** to guess how many court cases might pile up in the coming years.
* I can **use maps** to show where pending cases are increasing across different parts of India.
* If I get older court data, I can **check trends over time** to see how things are changing.
* I can also **group states or districts** with similar case patterns using something called "clustering," which helps spot areas that are facing the same kind of issues.

# 7. References

[1] Wes McKinney, *Python for Data Analysis*, O’Reilly Media  
[2] pandas documentation: <https://pandas.pydata.org/>  
[3] seaborn documentation: <https://seaborn.pydata.org/>  
[4] matplotlib documentation: <https://matplotlib.org/>

# 8. Social Media links:

LinkedIn:<https://www.linkedin.com/posts/maneet-kumar-423857298_dataforgood-datawithpurpose-judicialreform-activity-7317118283701186561-pG83?utm_source=share&utm_medium=member_desktop&rcm=ACoAAEgHZpsBQcCBWSMyG7CE0AJu2goDaxwizXo>