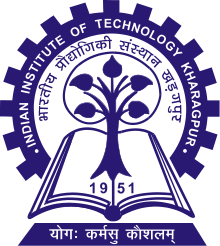
**Adhya Dagar**

**Summer Intern at CNERG Lab,IIT Kharagpur**

**15 May - 30 June,2019**

**Internship Report**

****

**PROJECT TITLE :** Pendency Analysis of Indian Court Cases using Legal Data Mining

**UNDER THE GUIDANCE OF**

**PROFESSOR SAPTARSHI GHOSH**

**Assistant Professor**

**Department of Computer Science and Engineering**

**Indian Institute of Technology Kharagpur**

***Abstract***

***In this project data has been scraped from National Judicial Data Grid Web Site and stored in a dictionary -key value format thus making it easily accessible.The key contains the header section a particular value comes under. After collection, analysis has been performed and feature selection takes place.Correlation between different features has been studied and dependency between various types of filed criminal cases determined.***

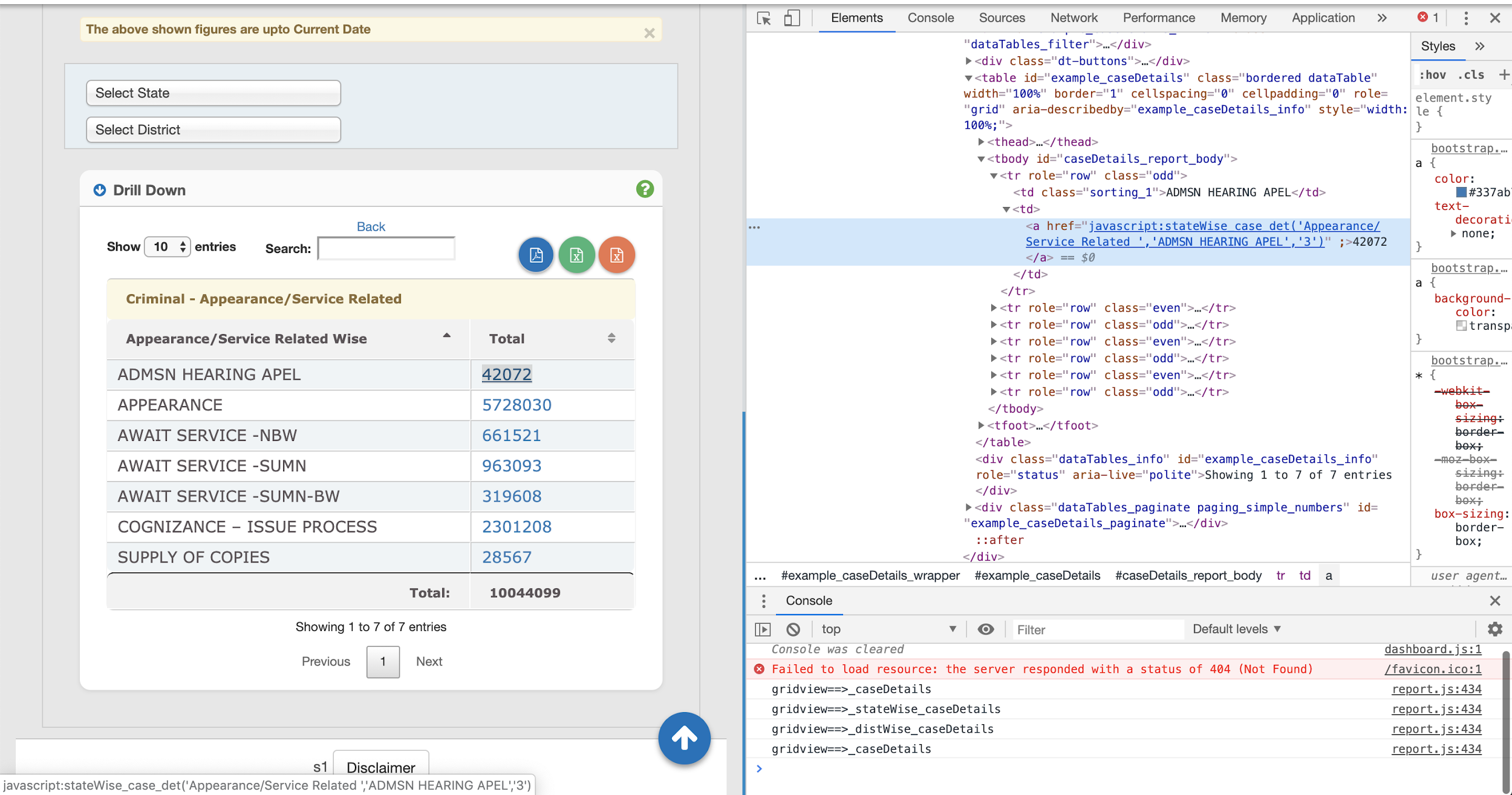
**Introduction**

The National Judicial Data Grid (NJDG) is a part of the on-going e-Courts Integrated Mission Mode Project. NJDG will work as a monitoring tool to identify, manage & reduce pendency of cases. It will also help to provide timely inputs for making policy decisions to reduce delay and arrears in the system, facilitate better monitoring of court performance and systemic bottlenecks, and, thus, facilitate better resource management.

**Methodology**

The structure of the National Judicial Data Grid Website was studied and different features and attributes analysed. Data scraping methodology was developed based on the architecture of website which is unique to different sub sections.

General approach included traversing the nested webpages through their respective xpath links. Components on every page had a different xpath so the code had to be revamped pertaining to the section that had to be crawled.



Sample xpath - //\*[@id="caseDetails\_report\_body"]/tr[1]/td[2]/a

Crawling needs to be performed till the point we have access to the Case Ids or Case Numbers which might be of the following format - CC / 400595 / 2013

With each loop, the key value of the dictionary is appended and the final result is in Key Value form so that dictionary can be indexed based on the section we want to search under.

**Sample Crawling Procedure for Pleadings section under Stage Wise Header -**

|  |  |  |
| --- | --- | --- |
| Iteration | Key | Value |
| 1 | 'Pleadings/Issues/Charge' | 1640597 |
| 2 | 'Pleadings/Issues/Charge - CHARGE-PLEA' | '1014699 |
| 3 | 'Pleadings/Issues/Charge - CHARGE-PLEA - Assam' | '932' |
| 4 | 'Pleadings/Issues/Charge - CHARGE-PLEA - Assam - Bongaigaon' | '4' |
| 5 | 'Pleadings/Issues/Charge - CHARGE-PLEA - Assam - Bongaigaon - Chief Judicial Magistrate' | '4' |
| 6 | 'Pleadings/Issues/Charge - CHARGE-PLEA - Assam - Bongaigaon - Chief Judicial Magistrate - 2018' | '4' |
| 7 | 'Pleadings/Issues/Charge - CHARGE-PLEA - Assam - Bongaigaon - Chief Judicial Magistrate - 2018 - C R Case ( Warrant) / 12 / 2018' | 'C R Case ( Warrant) / 12 / 2018' |

**Structure of the Web Site**

Data needed to be scraped from the following subsections -

1.Pending Cases

* 0-1 Years
* 1-3 Years
* 3-5 Years
* 5-10 Years
* 10-20 Years
* 20-30 Years
* Above 30 Years

2.Stage Wise

* Appearance/Service Related (10039957 records)
* Compliance/Steps/stay (1712916 records)
* Evidence/Argument/Judgement (7476854 records)
* Pleadings/Issues/Charge (1644448 records)

3.Senior Citizen

* Cases filed by Senior Citizens ( 478917 records)

4.Women

* Cases filed by Women

**Tools and Techniques Used**

Web Scraping - Web scraping is a technique for extracting information from the internet automatically using a software that simulates human web surfing.

Selenium - It is a python binding for the API of Selenium Web Drivers. One would be able to conveniently access the API of Selenium Web drivers like [Firefox](https://developer.mozilla.org/en-US/docs/Mozilla/QA/Marionette/WebDriver), [Chrome](https://sites.google.com/a/chromium.org/chromedriver/downloads), [PhantomJS](http://phantomjs.org/) etc. This web driver API can be used to simulate all sorts of actions that can be performed on a web browser i.e. click on buttons of websites, scroll and navigate through pages, type something in input boxes, submit forms, use proxies.

Code has been compiled in PyCharm Python IDE.

**Sample Web Scraping Code to Crawl data under the Stage Wise Section**

**from selenium import webdriver**

**import time**

**from selenium.webdriver.support import expected\_conditions as EC**

**from selenium.webdriver.common.by import By**

**from selenium.webdriver.support.ui import WebDriverWait**

**from selenium.webdriver.support.ui import Select**

**from selenium.webdriver.chrome.options import Options**

**import pickle**

**driver = webdriver.Chrome("/Users/adhyadagar/PycharmProjects/WebCrawler1/Drivers/chromedriver")**

**site = driver.get("https://njdg.ecourts.gov.in/njdgnew/index.php")**

**title = driver.title**

**driver.execute\_script("window.scrollTo(0, 0)")**

**print(title)**

**my\_url = "https://njdg.ecourts.gov.in/njdgnew/index.php"**

**driver.maximize\_window()**

**dict = {}**

**pickle\_out = open("SW1\_7-8.pickle","wb")**

**class wait\_for\_page\_load(object):**

**def \_\_init\_\_(self, browser):**

**self.browser = browser**

**def \_\_enter\_\_(self):**

**self.old\_page = self.browser.find\_element\_by\_tag\_name('html')**

**def page\_has\_loaded(self):**

**new\_page = self.browser.find\_element\_by\_tag\_name('html')**

**return new\_page.id != self.old\_page.id**

**def \_\_exit\_\_(self, \*\_):**

**wait\_for\_page\_load(self.page\_has\_loaded)**

**for i in range(17, 18):**

**try:**

**#WebDriverWait(driver, 10).until(EC.presence\_of\_element\_located((By.XPATH, '//\*[@id="report\_body"]/tr[' + str(i) + ']/td[1]')))**

**element = driver.find\_element\_by\_xpath('//\*[@id="report\_body"]/tr[' + str(i) + ']/td[3]/a')**

**driver.execute\_script("arguments[0].scrollIntoView(true);", element)**

**key = driver.find\_element\_by\_xpath('//\*[@id="report\_body"]/tr[' + str(i) + ']/td[1]').text**

**value = driver.find\_element\_by\_xpath('//\*[@id="report\_body"]/tr[' + str(i) + ']/td[3]/a').text**

**time.sleep(3)**

**dict[key]=value**

**print(dict)**

**#WebDriverWait(driver, 10).until(EC.presence\_of\_element\_located((By.XPATH, '//\*[@id="report\_body"]/tr[' + str(i) + ']/td[1]')))**

**driver.find\_element\_by\_xpath('//\*[@id="report\_body"]/tr[' + str(i) + ']/td[3]/a').click()**

**except Exception as e:**

**break**

**for j in range(7, 8):**

**try:**

**dict1 = {}**

**time.sleep(3)**

**#driver.execute\_script("window.scrollTo(0, 0)")**

**#element = driver.find\_element\_by\_xpath('//\*[@id="caseDetails\_report\_body"]/tr['+ str(j) +']/td[1]')**

**#driver.execute\_script("arguments[0].scrollIntoView(true);", element)**

**select = Select(driver.find\_element\_by\_xpath('//\*[@id="example\_caseDetails\_length"]/label/select'))**

**select.select\_by\_value('50')**

**key1 = driver.find\_element\_by\_xpath('//\*[@id="caseDetails\_report\_body"]/tr[' + str(j) + ']/td[1]').text**

**value1 = driver.find\_element\_by\_xpath('//\*[@id="caseDetails\_report\_body"]/tr[' + str(j) + ']/td[2]/a').text**

**Key1 = key + ' - ' + key1**

**dict1[Key1] = value1**

**dict.update(dict1)**

**print(dict)**

**driver.find\_element\_by\_xpath('//\*[@id="caseDetails\_report\_body"]/tr[' + str(j) + ']/td[2]/a').click()**

**except Exception as e:**

**driver.execute\_script("window.scrollTo(0, 0)")**

**with wait\_for\_page\_load(driver):**

**# time.sleep(10)**

**driver.find\_element\_by\_link\_text("Back")**

**driver.find\_element\_by\_link\_text("Back").click()**

**print("HERE")**

**break**

**for k in range(1, 51):**

**try:**

**dict2={}**

**time.sleep(3)**

**#element = driver.find\_element\_by\_xpath('//\*[@id="stateWise\_caseDetails\_report\_body"]/tr[' + str(k) + ']/td[2]/a')**

**#driver.execute\_script("arguments[0].scrollIntoView(true);", element)**

**select = Select(driver.find\_element\_by\_xpath('//\*[@id="example\_stateWise\_caseDetails\_length"]/label/select'))**

**select.select\_by\_value('50')**

**key2 = driver.find\_element\_by\_xpath('//\*[@id="stateWise\_caseDetails\_report\_body"]/tr[' + str(k) + ']/td[1]').text**

**value2 = driver.find\_element\_by\_xpath('//\*[@id="stateWise\_caseDetails\_report\_body"]/tr[' + str(k) + ']/td[2]/a').text**

**Key2 = Key1 + ' - ' + key2**

**dict2[Key2] = value2**

**dict.update(dict2)**

**print(dict)**

**#time.sleep(10)**

**driver.find\_element\_by\_xpath('//\*[@id="stateWise\_caseDetails\_report\_body"]/tr[' + str(k) + ']/td[2]/a').click()**

**except Exception as e:**

**driver.execute\_script("window.scrollTo(0, 0)")**

**with wait\_for\_page\_load(driver):**

**# time.sleep(10)**

**driver.find\_element\_by\_link\_text("Back").click()**

**#driver.find\_element\_by\_link\_text("Back").click()**

**print("HERE2")**

**break**

**for l in range(1 ,51):**

**try:**

**dict3 = {}**

**time.sleep(3)**

**select = Select(driver.find\_element\_by\_xpath(**

**'//\*[@id="example\_distWise\_caseDetails\_length"]/label/select'))**

**select.select\_by\_value('50')**

**key3 = driver.find\_element\_by\_xpath(**

**'//\*[@id="distWise\_caseDetails\_report\_body"]/tr[' + str(l) + ']/td[1]').text**

**value3 = driver.find\_element\_by\_xpath(**

**'//\*[@id="distWise\_caseDetails\_report\_body"]/tr[' + str(l) + ']/td[2]/a').text**

**Key3 = Key2 + ' - ' + key3**

**dict3[Key3] = value3**

**dict.update(dict3)**

**print(dict)**

**#time.sleep(10)**

**driver.find\_element\_by\_xpath('//\*[@id="distWise\_caseDetails\_report\_body"]/tr[' + str(l) + ']/td[2]/a').click()**

**except Exception as e:**

**driver.execute\_script("window.scrollTo(0, 0)")**

**with wait\_for\_page\_load(driver):**

**# time.sleep(10)**

**driver.find\_element\_by\_link\_text("Back").click()**

**#driver.find\_element\_by\_link\_text("Back").click()**

**print("HERE3")**

**break**

**for m in range(1,51):**

**try:**

**dict4 = {}**

**time.sleep(3)**

**select = Select(driver.find\_element\_by\_xpath('//\*[@id="example\_estCodeWise\_caseDetails\_length"]/label/select'))**

**select.select\_by\_value('50')**

**key4 = driver.find\_element\_by\_xpath('//\*[@id="estCodeWise\_caseDetails\_report\_body"]/tr[' + str(m) + ']/td[1]').text**

**value4 = driver.find\_element\_by\_xpath('//\*[@id="estCodeWise\_caseDetails\_report\_body"]/tr[' + str(m) + ']/td[2]/a').text**

**Key4 = Key3 + ' - ' + key4**

**dict4[Key4] = value4**

**dict.update(dict4)**

**print(dict)**

**#time.sleep(10)**

**driver.find\_element\_by\_xpath('//\*[@id="estCodeWise\_caseDetails\_report\_body"]/tr[' + str(m) + ']/td[2]/a').click()**

**except Exception as e:**

**driver.execute\_script("window.scrollTo(0, 0)")**

**with wait\_for\_page\_load(driver):**

**# time.sleep(10)**

**driver.find\_element\_by\_link\_text("Back").click()**

**#driver.find\_element\_by\_link\_text("Back").click()**

**break**

**for t in range(1,51):**

**try:**

**dict5 = {}**

**#print(t)**

**time.sleep(3)**

**with wait\_for\_page\_load(driver):**

**driver.find\_element\_by\_xpath('//\*[@id="example\_regYearWise\_caseDetails\_length"]/label/select')**

**select = Select(driver.find\_element\_by\_xpath('//\*[@id="example\_regYearWise\_caseDetails\_length"]/label/select'))**

**select.select\_by\_value('50')**

**with wait\_for\_page\_load(driver):**

**driver.find\_element\_by\_xpath(**

**'//\*[@id="regYearWise\_caseDetails\_report\_body"]/tr[' + str(t) + ']/td[1]')**

**key5 = driver.find\_element\_by\_xpath('//\*[@id="regYearWise\_caseDetails\_report\_body"]/tr[' + str(t) + ']/td[1]').text**

**value5 = driver.find\_element\_by\_xpath('//\*[@id="regYearWise\_caseDetails\_report\_body"]/tr[' + str(t) + ']/td[2]/a').text**

**Key5 = Key4 + ' - ' + key5**

**dict5[Key5] = value5**

**dict.update(dict5)**

**print(dict)**

**time.sleep(5)**

**z = driver.find\_element\_by\_xpath('//\*[@id="regYearWise\_caseDetails\_report\_body"]/tr[' + str(t) + ']/td[2]/a').text**

**driver.find\_element\_by\_xpath('//\*[@id="regYearWise\_caseDetails\_report\_body"]/tr[' + str(t) +**

**']/td[2]/a').click()**

**z = str(z)**

**z = int(z)**

**y = int(z / 100)**

**if z > 0:**

**for o in range(1, y + 2):**

**for p in range(1, 101):**

**try:**

**dict6 = {}**

**time.sleep(3)**

**select = Select(driver.find\_element\_by\_xpath('//\*[@id="example\_regnoWise\_caseDetails\_length"]/label/select'))**

**select.select\_by\_value('100')**

**driver.find\_element\_by\_xpath('// \*[ @ id = "example\_regnoWise\_caseDetails\_next"]').click()**

**key6 = driver.find\_element\_by\_xpath('//\*[@id="regnoWise\_caseDetails\_report\_body"]/tr[' + str(p) + ']/td/a').text**

**value6 = driver.find\_element\_by\_xpath('//\*[@id="regnoWise\_caseDetails\_report\_body"]/tr[' + str(p) + ']/td/a').text**

**Key6 = Key5 + key6**

**dict6[Key6] = value6**

**dict.update(dict6)**

**print(dict)**

**# time.sleep(10)**

**except Exception as e:**

**driver.execute\_script("window.scrollTo(0, 0)")**

**with wait\_for\_page\_load(driver):**

**# time.sleep(10)**

**driver.find\_element\_by\_link\_text("Back")**

**driver.find\_element\_by\_link\_text("Back").click()**

**break**

**with wait\_for\_page\_load(driver):**

**driver.find\_element\_by\_xpath('//\*[@id="example\_regnoWise\_caseDetails\_next"]').click()**

**else:**

**for p in range(1, 101):**

**try:**

**dict6 = {}**

**time.sleep(3)**

**select = Select(driver.find\_element\_by\_xpath('//\*[@id="example\_regnoWise\_caseDetails\_length"]/label/select'))**

**select.select\_by\_value('100')**

**driver.find\_element\_by\_xpath('// \*[ @ id = "example\_regnoWise\_caseDetails\_next"]').click()**

**key6 = driver.find\_element\_by\_xpath('//\*[@id="regnoWise\_caseDetails\_report\_body"]/tr[' + str(p) + ']/td/a').text**

**value6 = driver.find\_element\_by\_xpath('//\*[@id="regnoWise\_caseDetails\_report\_body"]/tr[' + str(p) + ']/td/a').text**

**Key6 = Key5 + key6**

**dict6[Key6] = value6**

**dict.update(dict6)**

**print(dict)**

**# time.sleep(10)**

**except Exception as e:**

**#time.sleep(10)**

**driver.execute\_script("window.scrollTo(0, 0)")**

**with wait\_for\_page\_load(driver):**

**#time.sleep(10)**

**driver.find\_element\_by\_xpath('//\*[@id="regno\_Wise\_caseDetails\_tab"]/a')**

**driver.find\_element\_by\_xpath('//\*[@id="regno\_Wise\_caseDetails\_tab"]/a').click()**

**break**

**except Exception as e:**

**driver.execute\_script("window.scrollTo(0, 0)")**

**with wait\_for\_page\_load(driver):**

**driver.find\_element\_by\_link\_text("Back")**

**driver.find\_element\_by\_link\_text("Back").click()**

**break**

**pickle.dump(dict, pickle\_out)**

**pickle\_out.close()**

**Similarly code has been written to scrape data under the pending cases,women and senior citizen sections.**

**Data Type -**

Data collected has been stored in a dictionary format with the key value containing the sub heading the numeric value comes under.

Eg.

'Pleadings/Issues/Charge - ISSUES - Andhra Pradesh - Ananthapur - District Courts, Ananthapur - 2018 - SC / 289 / 2018': 'SC / 289 / 2018'

The data has been stored under the following labels and uploaded on the drive.

**Analysis Carried out on Data**

**Feature Selection -**

A total of 38 features we selected to the dependency between pending cases.

6 sample states based on population and geo-political importance were chosen to study the correlation between different factors, these states are -

1.Bihar

2.Delhi

3.Maharashtra

4.Tamil Nadu

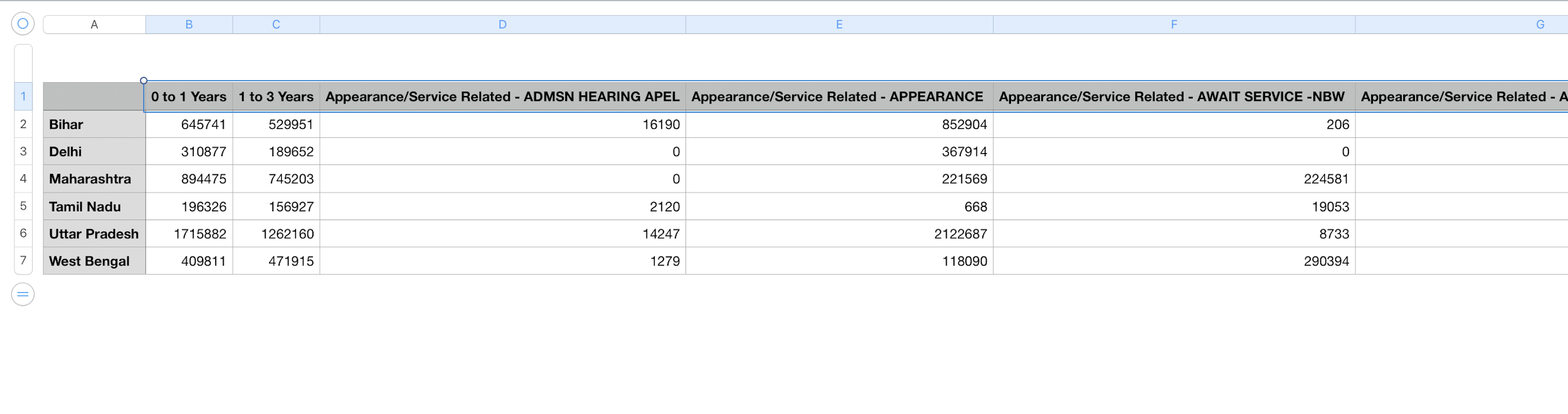
5.Uttar Pradesh

6.West Bengal

**Results**

**1.Feature Vector**

A table of 6 rows and 38 columns were made number of criminal pending cases of a particular state

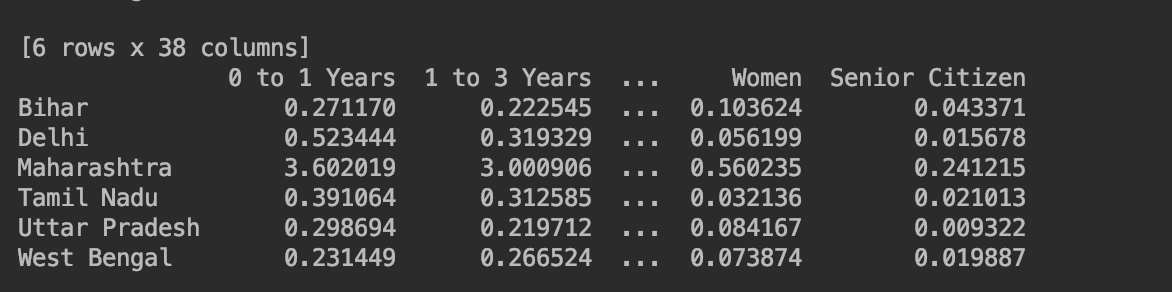
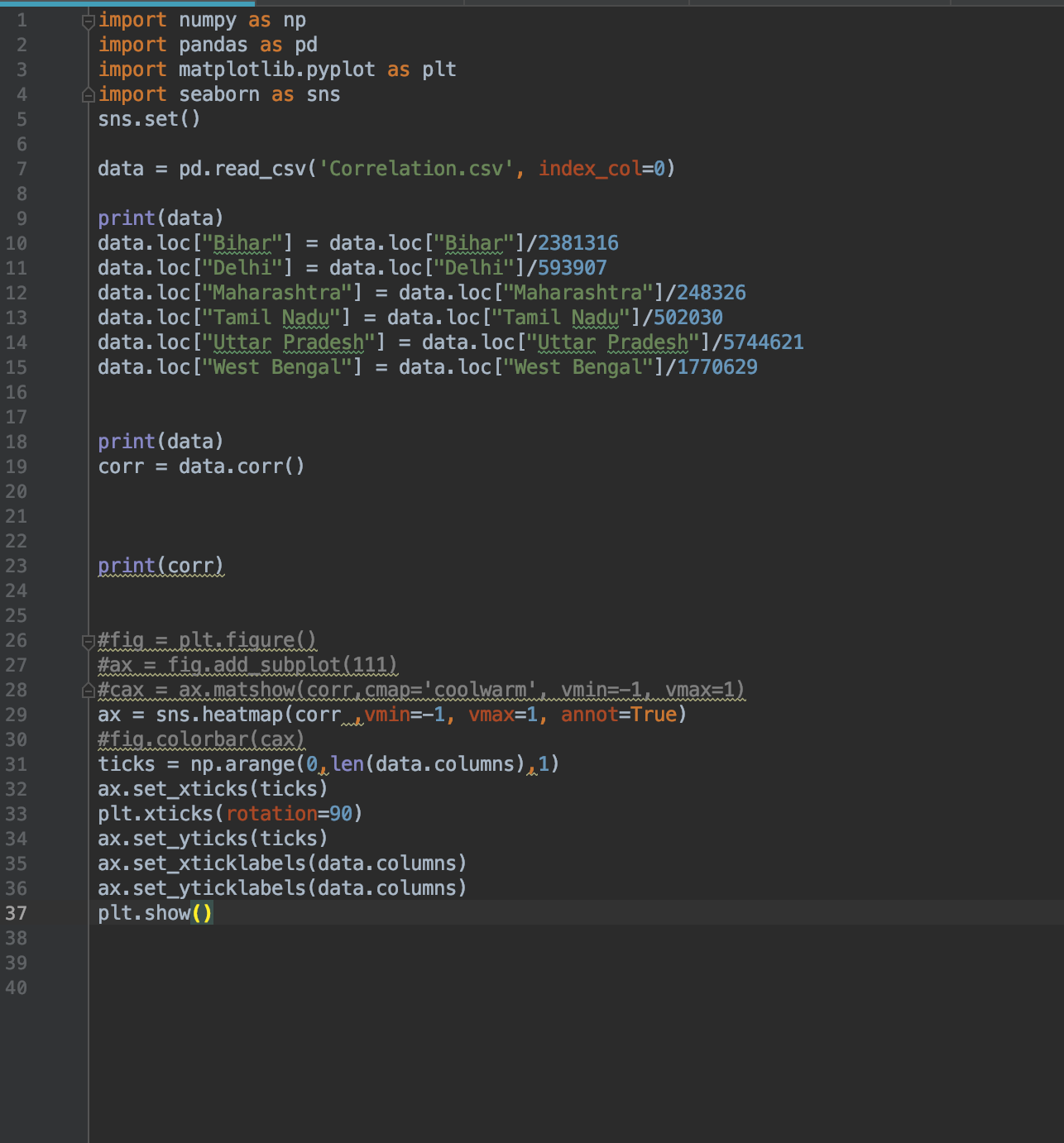
****

**2. Normalisation Results**

Normalisation results were obtained by dividing value in every cell of the feature matrix by total number of criminal cases pending in that particular state.

|  |  |
| --- | --- |
| **State** | **Total Number of Criminal Cases Pending** |
| Bihar | 2381316 |
| Delhi | 593907 |
| Maharashtra | 2498327 |
| Tamil Nadu | 502030 |
| Uttar Pradesh | 5744621 |
| West Bengal | 1770629 |

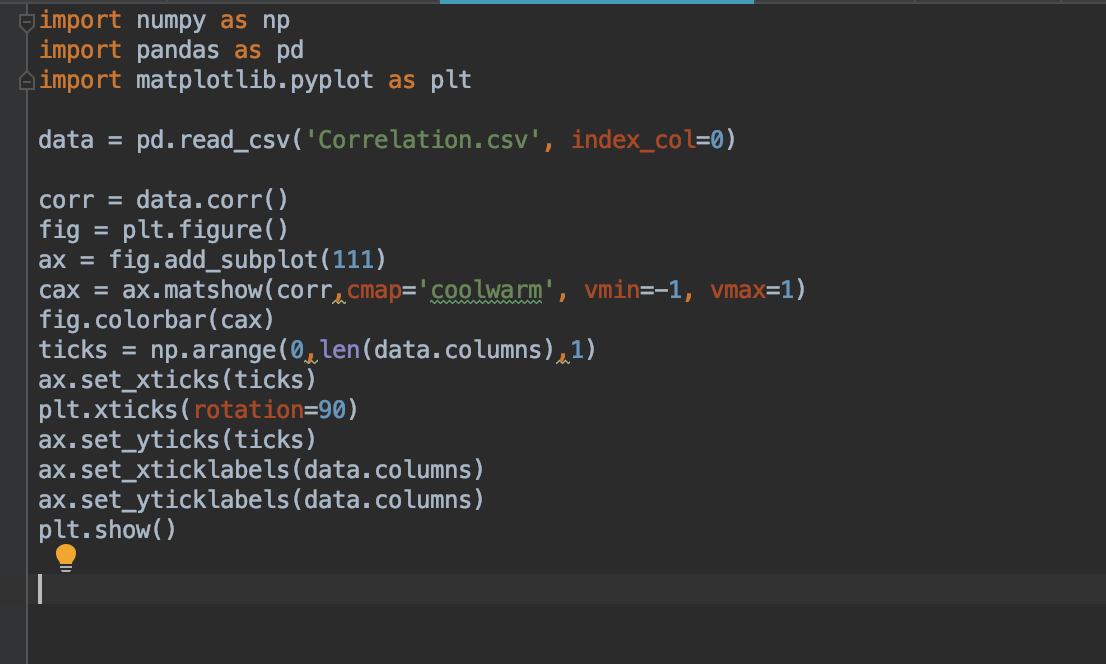
**Code -**

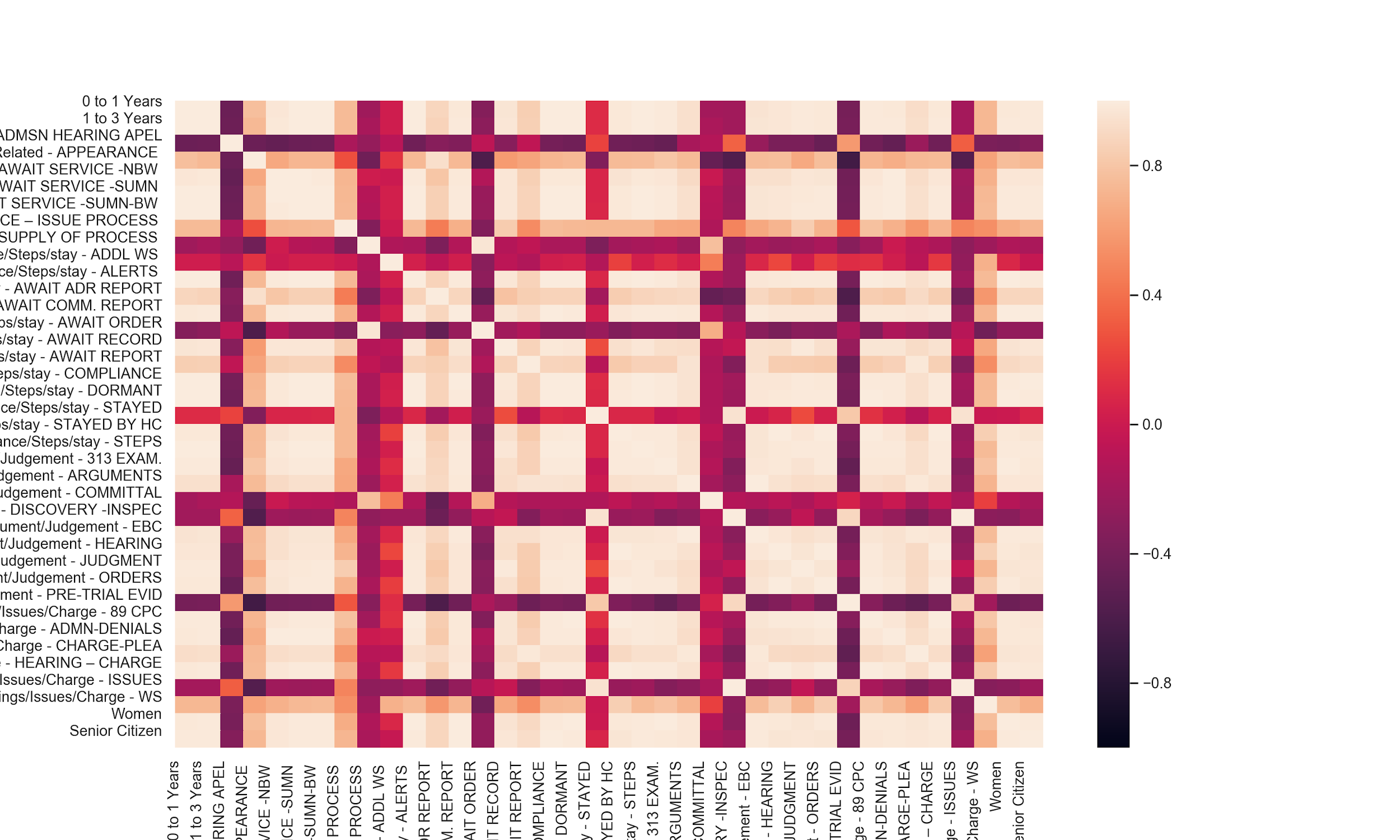
**Result -**

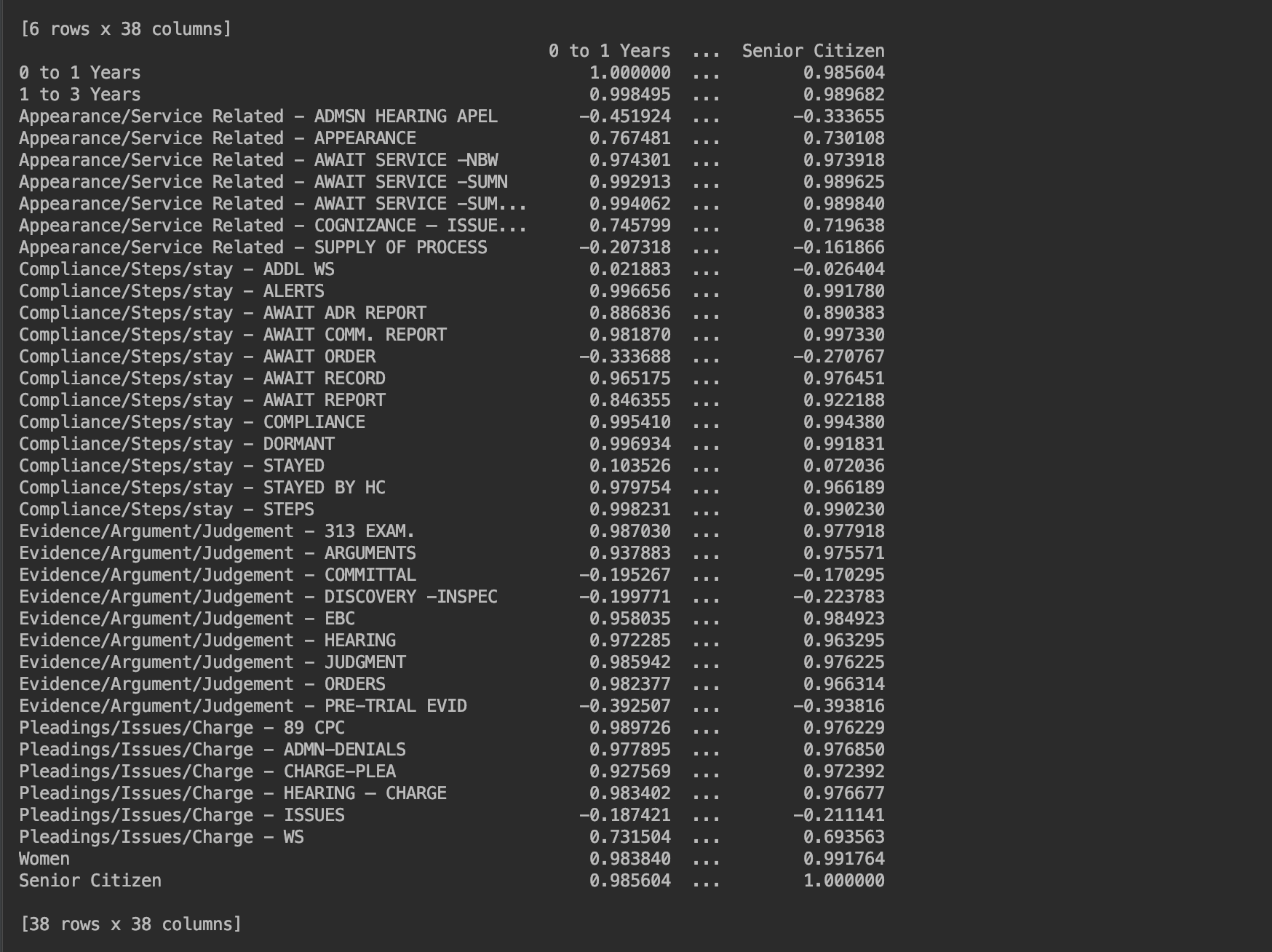
**3.Correlation Results on Normalised Matrix**

Correlation Matrix Results

**Code -**

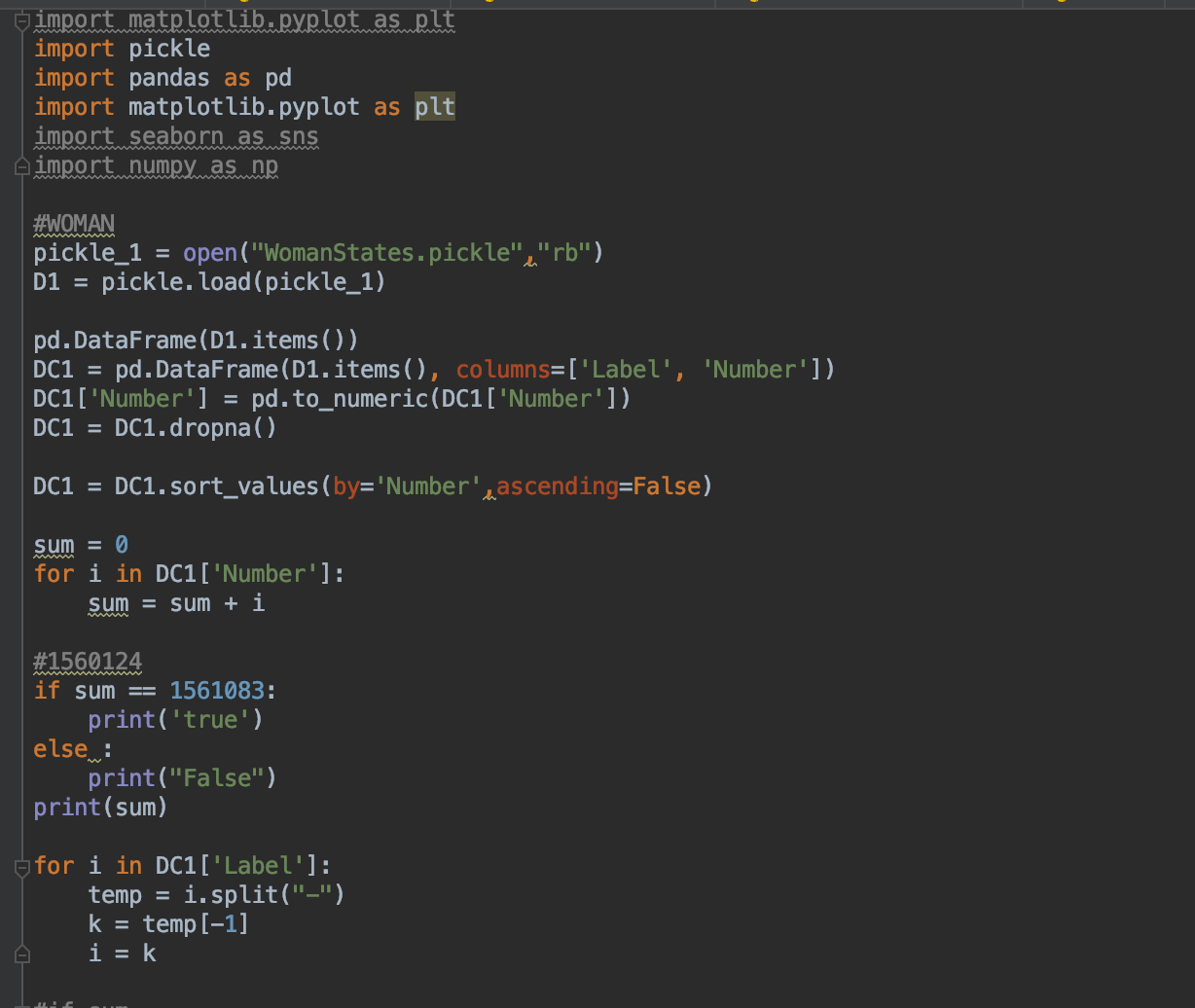
****

**Result** 



**4. States with maximum cases filed by Women**

**Code-**

**Result-**

Top 6 States -

1.Andhra Pradesh

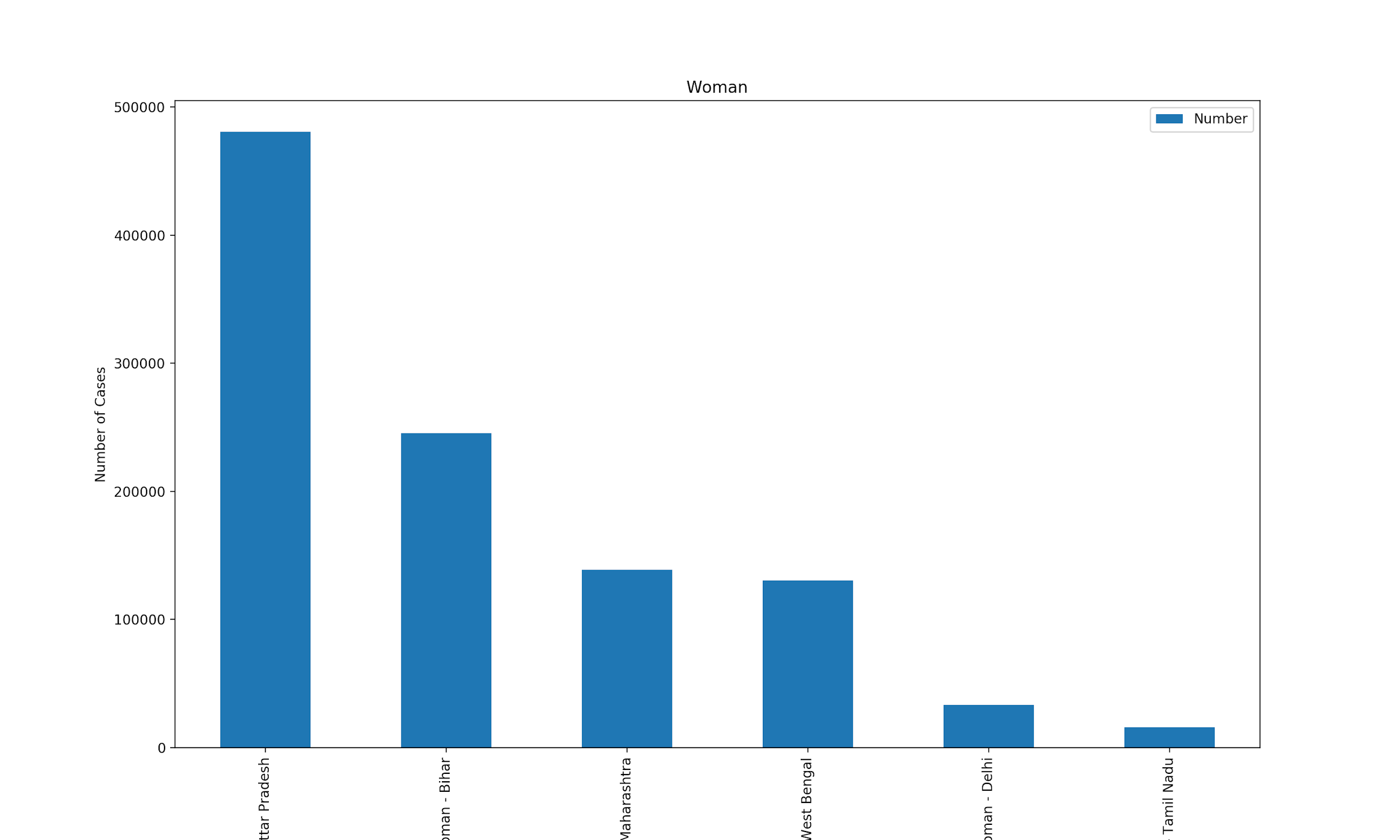
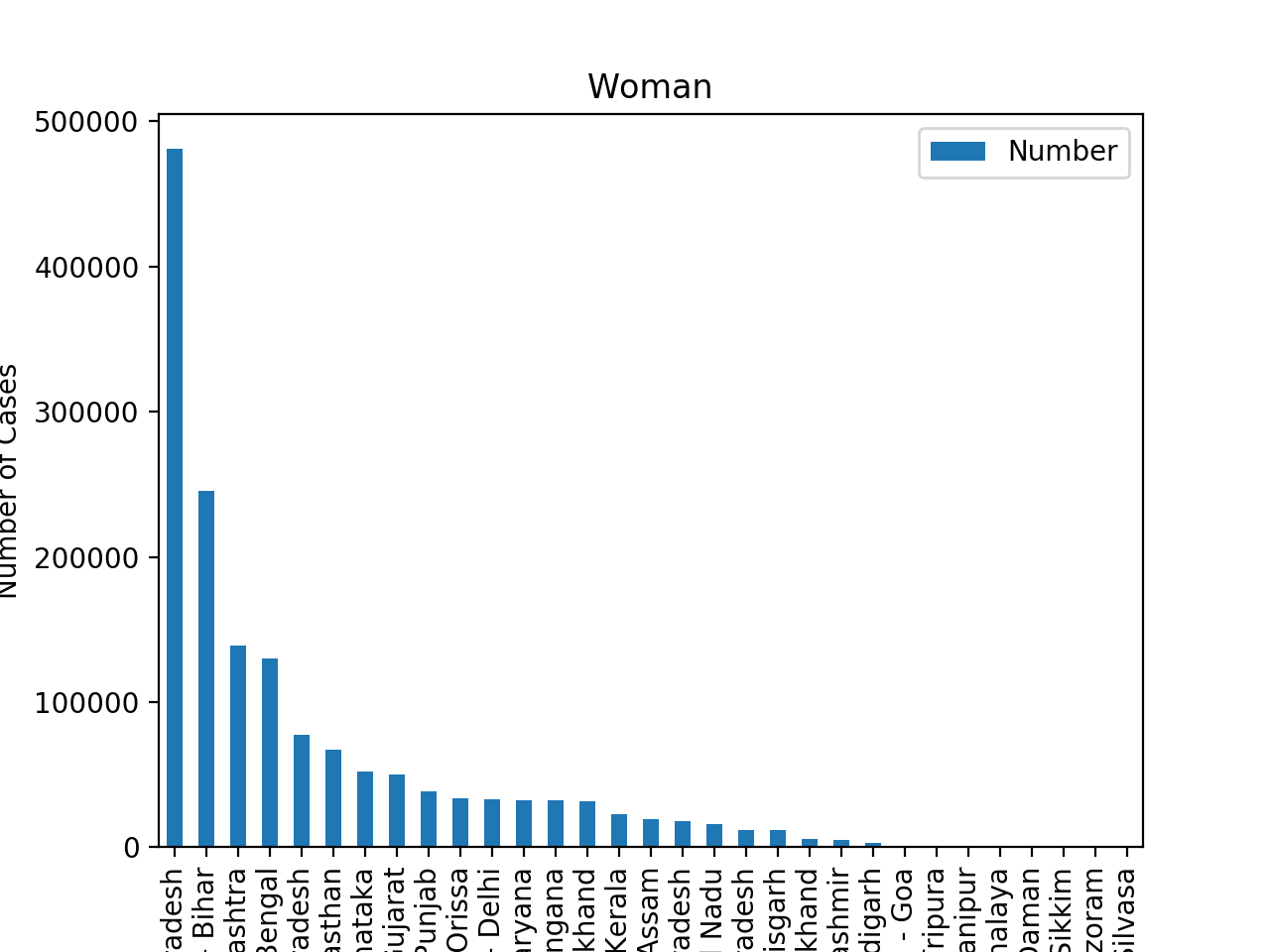
2.Bihar

3.Maharshtra

3.West Bengal

4.Delhi

5.Tamil Nadu

****

**4.States with maximum cases filed by Senior Citizen**

**Code-**

****

**Result-**

Top 6 states -

1.Bihar

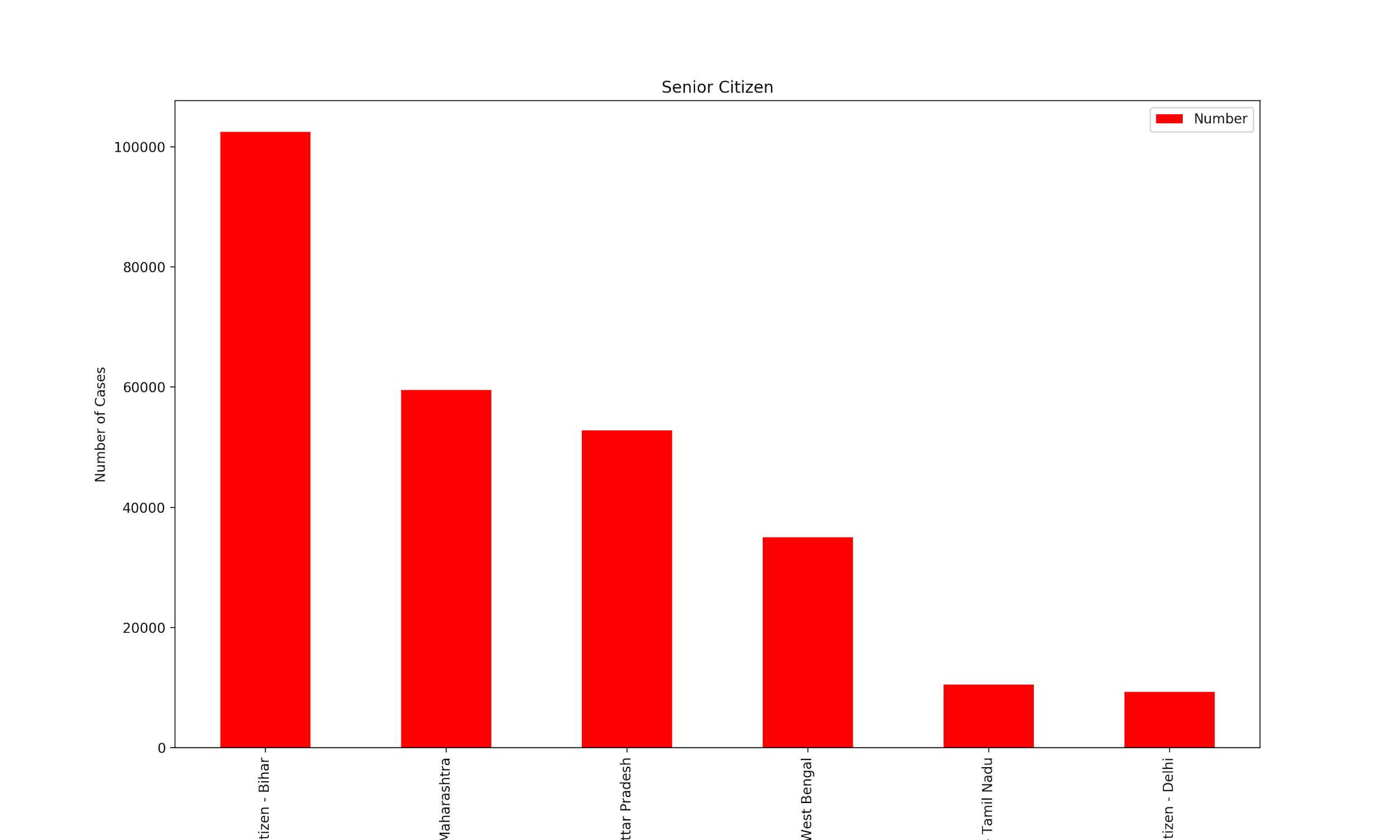
2.Maharshtra

3.Uttar Pradesh

4.West Bengal

5.Tamil Nadu

6.Delhi

****

**Conclusion**

The following data has been collected - Pending Cases (0-1 years,20-30 years , above 30 years); Stage Wise (Appearance Related,Compliance (half),Evidence, Pleadings),Senior Citizens,Women (half - till districts).The rest of the slots are under progress.Analysis was performed on the collected data to study unique features about different types of cases from different states.Correlation analysis showed us the dependency between 2 different features and how they can give us information about how long a case will be pending in the court system.Results have been published and uploaded on the drive.