

Precog Task - 1 Report

Ishan Kavathekar

ishan.kavathekar@research.iiit.ac.in

2022121003

1. INTRODUCTION

The dataset contains ~ 80 million Indian district court cases across various states. This data is obtained from the Development Data Lab. The dataset has been divided into cases, keys, acts, and judges. The cases subsection has details of court cases from 2010 to 2018 across various states and districts. The keys subsection consists of all the keys(state code, district code, display name, etc.) and their corresponding values. These keys are used in other sections. The act's section contains information about case ids, act IDs, section IDs, the number of sections filed under and criminal case status. The Judge section provides all the information about judges, like designation, the start date of tenure, the end date of tenure and gender across all the states. The following report is the exploratory data analysis of the given dataset. Exploratory Data Analysis is the critical process of performing initial investigations on data to discover patterns, spot anomalies, test hypotheses and check assumptions with the help of summary statistics and graphical representations. This analysis aims to get an insight into the data, identify the structure of the data and observe trends among variables if present.

2. Technologies and libraries used

- Python
- NumPy
- Pandas
- Matplotlib
- Seaborn
- Plotly
- Geopandas
- Shapely

3. Analysis

A. Judges

File: judges_clean.csv

Corresponding notebook: Precog_task_judge.ipynb

- 1) The data frame was standardized for categorical data. '0 nonfemale' was replaced with 'male', '1 female' was replaced with 'female', and '-9998 unclear' was replaced with 'unclear'. The missing values in the gender column were imputed with 'unclear'.
- 2) This data frame was merged with the cases_state_key.csv file to obtain the state name for all of the present tuples.
- 3) It was found that the number of male judges is significantly more than the number of female judges, as shown in fig 1
- 4) Maharashtra has the highest number of judges, followed by Uttar Pradesh, Gujrat, Karnataka and Madhya Pradesh. The number of judges per state can be seen in fig 2

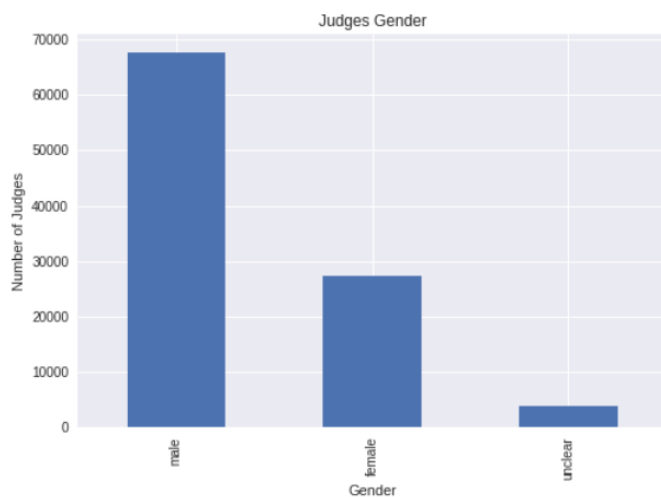


Fig 1 Gender-wise count of judges

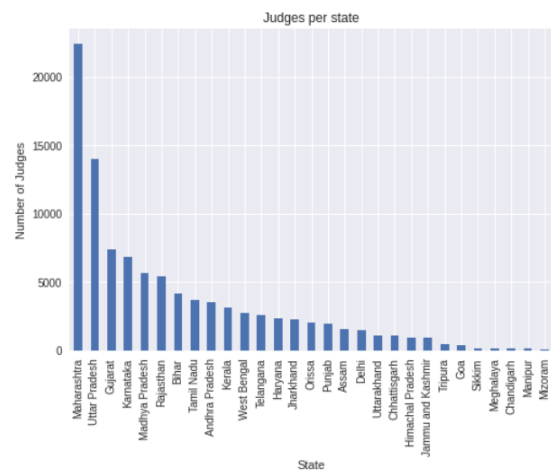


Fig 2 Number of judges per state

5) The analysis of judge designation was done in two settings: gender-wise and gender-neutral. Fig 3-5 shows the obtained results.

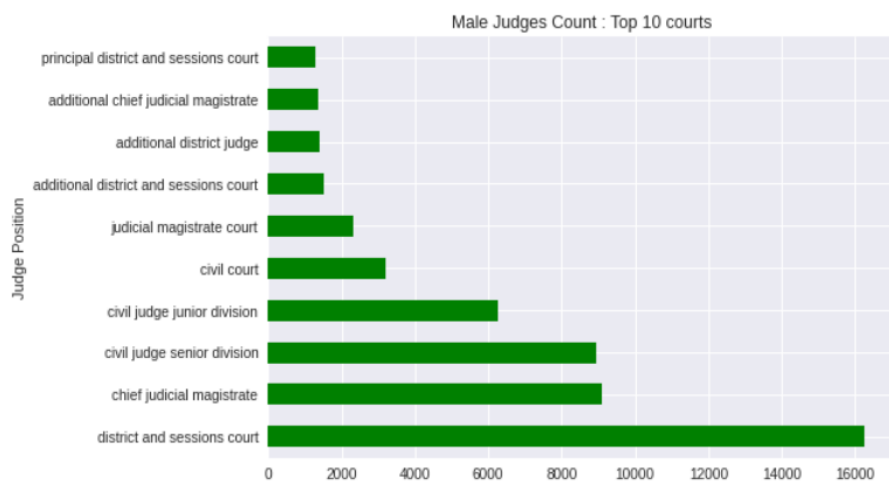


Fig 3 Count of male judges: designation

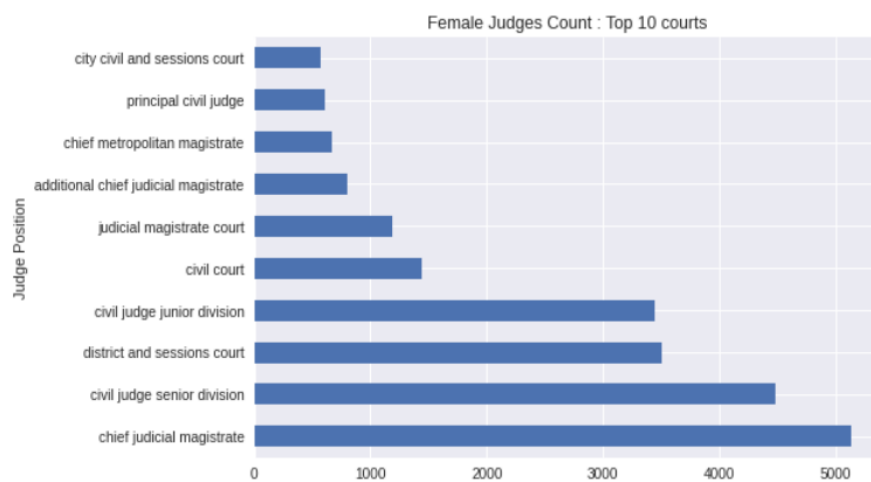


Fig 4 Count of male judges: designation

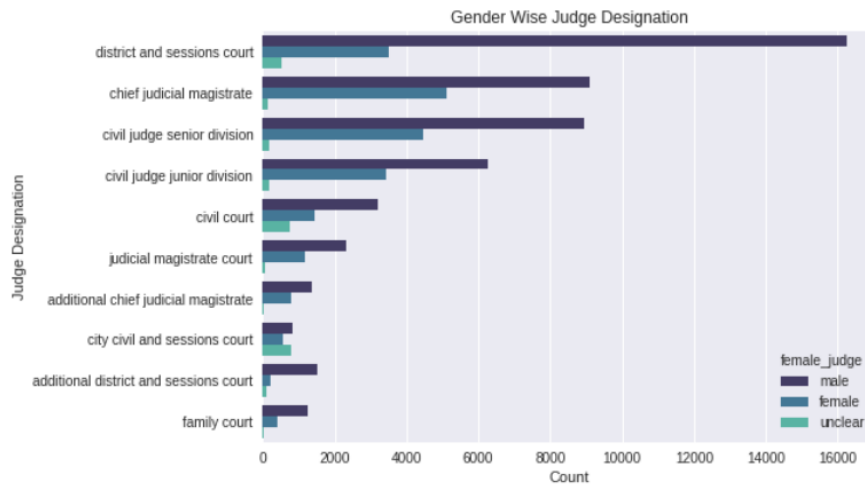


Fig 5 Gender-wise judge designation in top 10 courts

- 6) I plotted the count of judges according to the start date of their tenure over the years. It was observed that not many female judges had started their tenure in the last decade as compared to male judges. This can be seen in fig 6

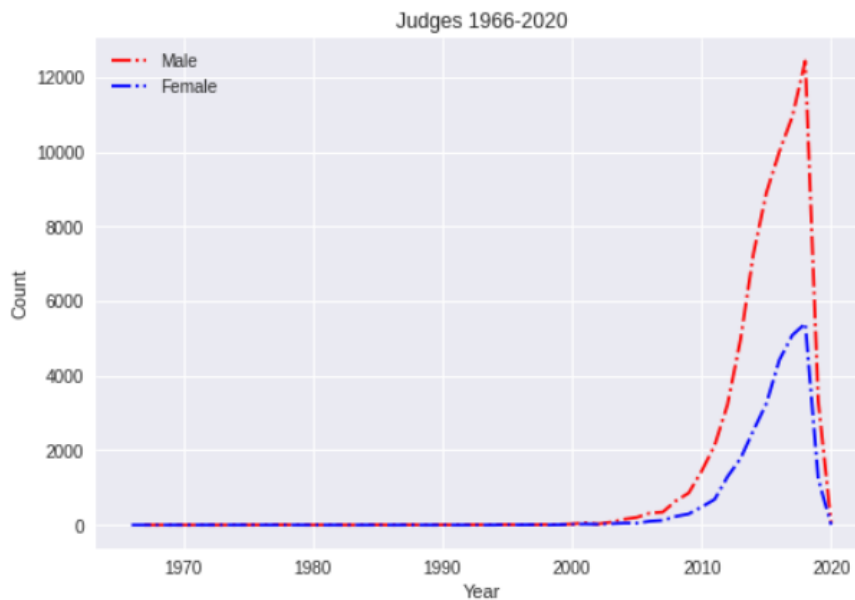


Fig 6 Number of judges over the years

- 7) The choropleth map of India helps visualize the number of judges per state and supports the data obtained in fig 2

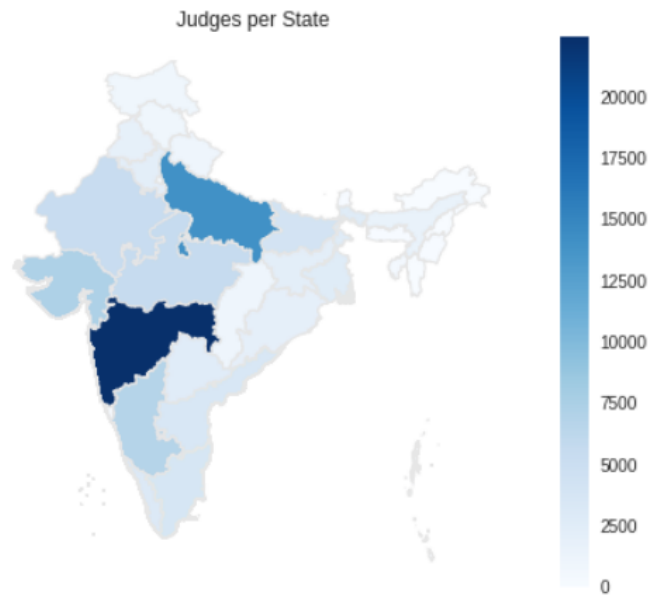


Fig 7 Judges per state

- 8) Gender-wise count density of court cases per state can be seen in fig 8 and fig 9. It can be seen that the female judge count in Gujarat ranks third.

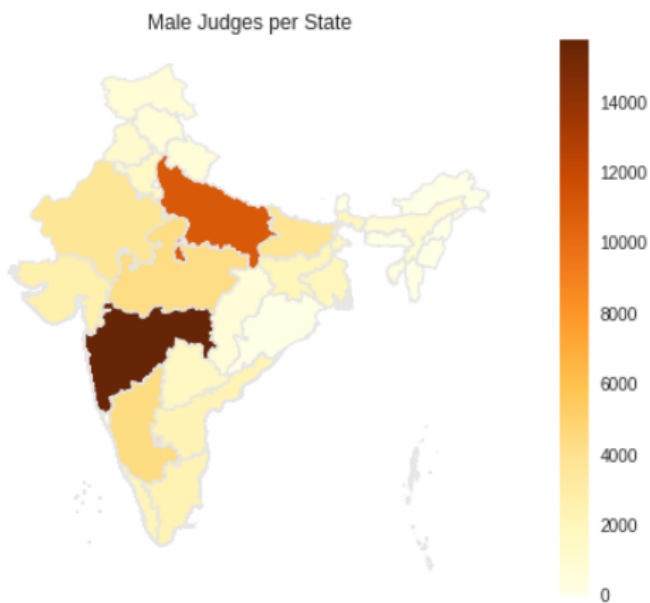


Fig 8 Male judges per state

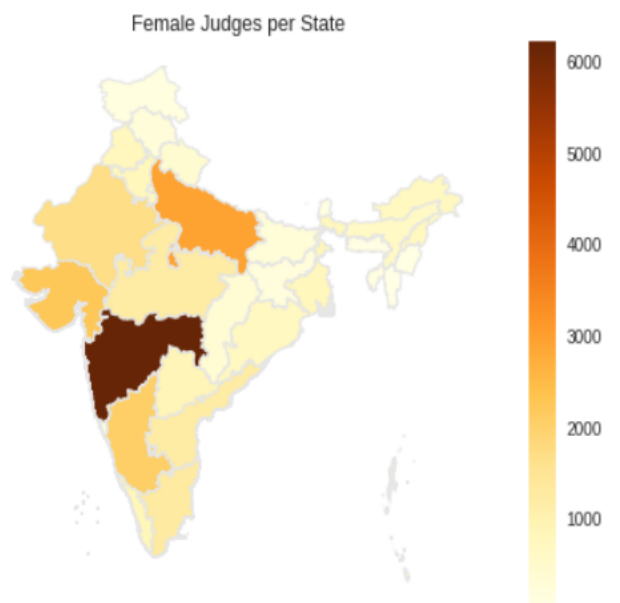


Fig 9 Female judges per state

- 9) The average tenure time per state was analyzed in two settings: gender-wise and gender-neutral. It is observed that the northern states have higher values for mean tenure time as compared to other states. Fig 10 - 12 shows the obtained results.

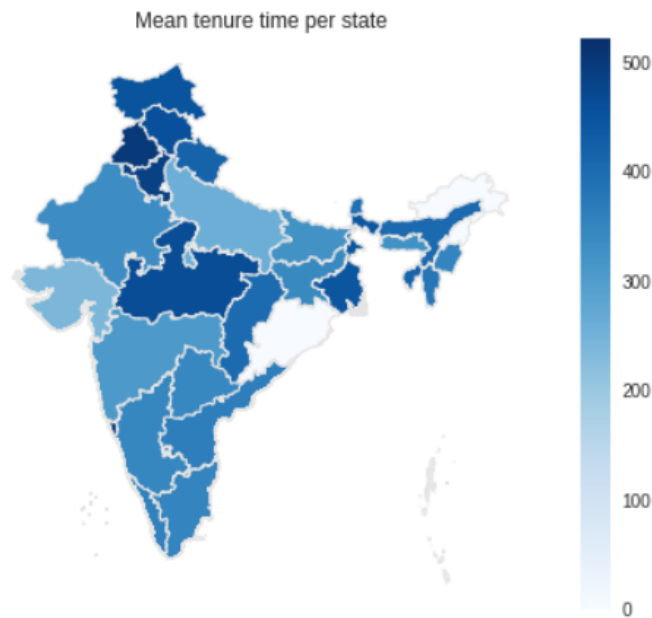


Fig 10 Average tenure time per state

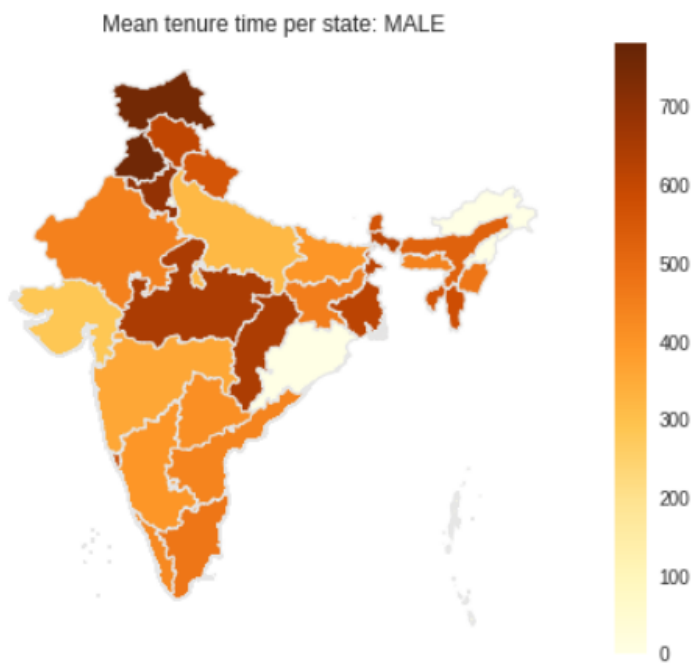


Fig 11 Average tenure for male judges per state

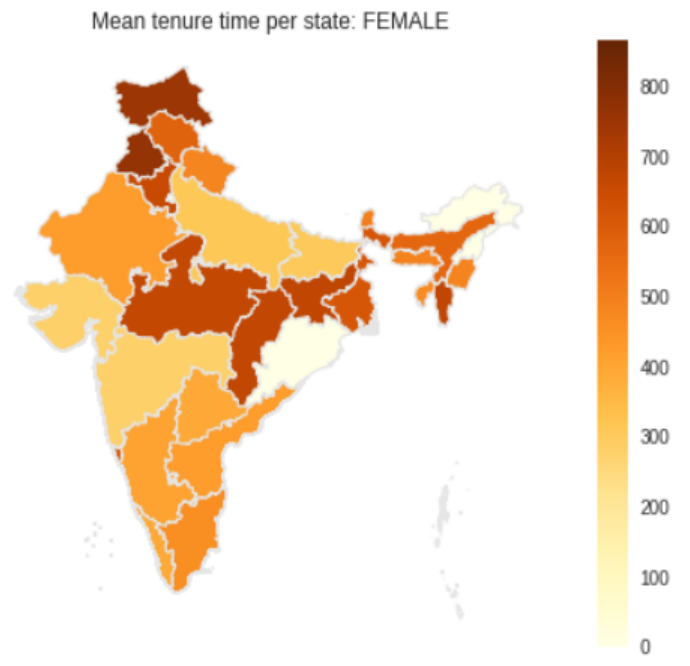


Fig 12 Average tenure for female judges per state

B. Display name and Court cases

File: disp_name_key.csv and cases_court_key.csv

Corresponding notebook: Precog task display_name_key and cases_court_key.ipynb

- 1) The count of all types of cases was plotted across the years (2010 - 2018). It can be seen that disposition cases increased exponentially over the years, as seen in fig 13.

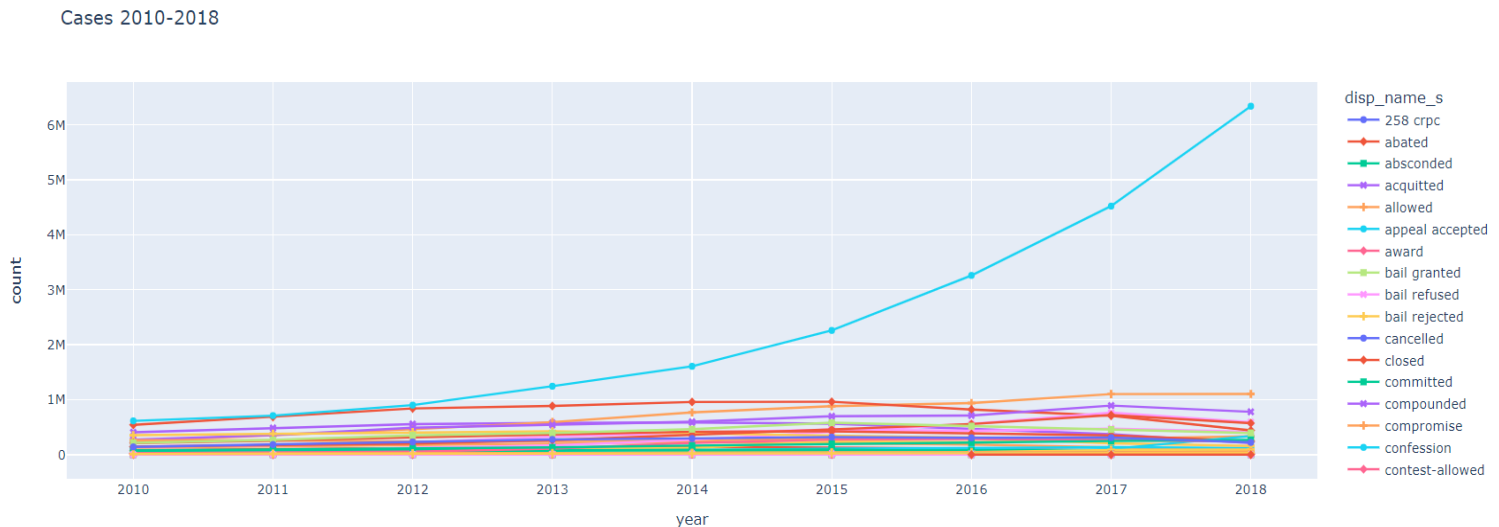


Fig 13 Count of different types of cases (2010 - 2018)

- 2) Fifteen types of cases were chosen to be studied. These were:

- | | | |
|--------------|--------------|--------------------------|
| ● Acquitted | ● Confession | ● Plead guilty |
| ● Award | ● Convicted | ● Referred to Lok Adalat |
| ● Closed | ● Decided | ● Reject |
| ● Committed | ● Dismissed | ● Transferred |
| ● Compromise | ● Judgement | ● Withdrawn |

We can see in fig 14 that the number of dismissed cases increased till 2015 and then decreased afterwards. Cases referred to Lok Adalat also increased over the years, except for 2018, where the count decreased. Acquitted cases also observed a dip since 2014. The number of cases with judgment has also been decreasing since 2013.

Cases 2010-2018

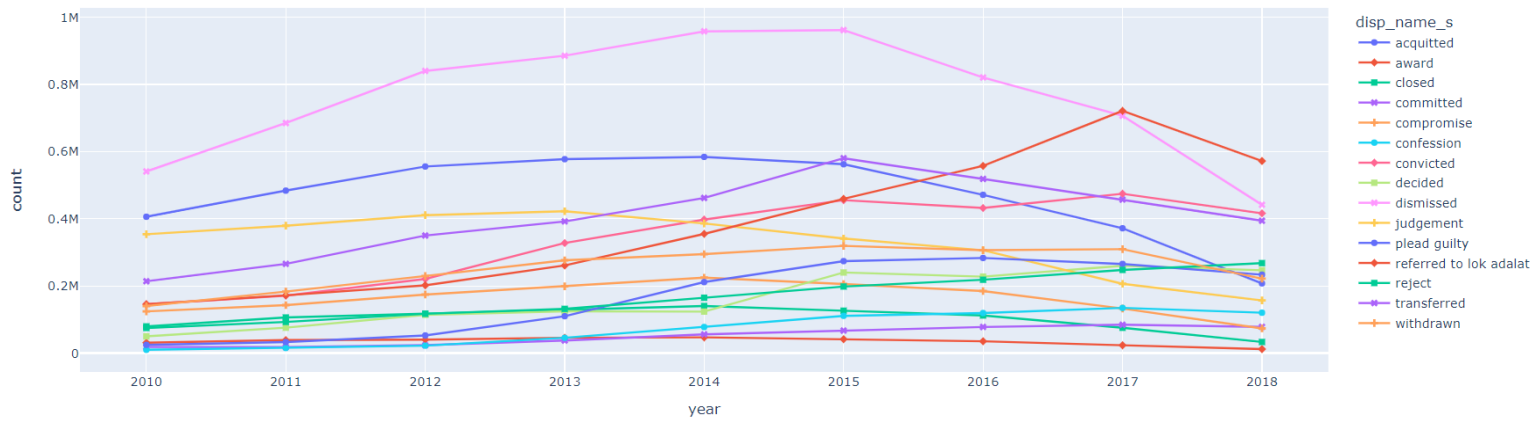


Fig 14 Count of selected 15 types of cases (2010 - 2018)

3) Rajasthan has the highest number of courts, followed by Tamil Nadu and Maharashtra. The number of courts over the years 2010 - 2018 is given in fig 15

Number of Courts : 2010-2018

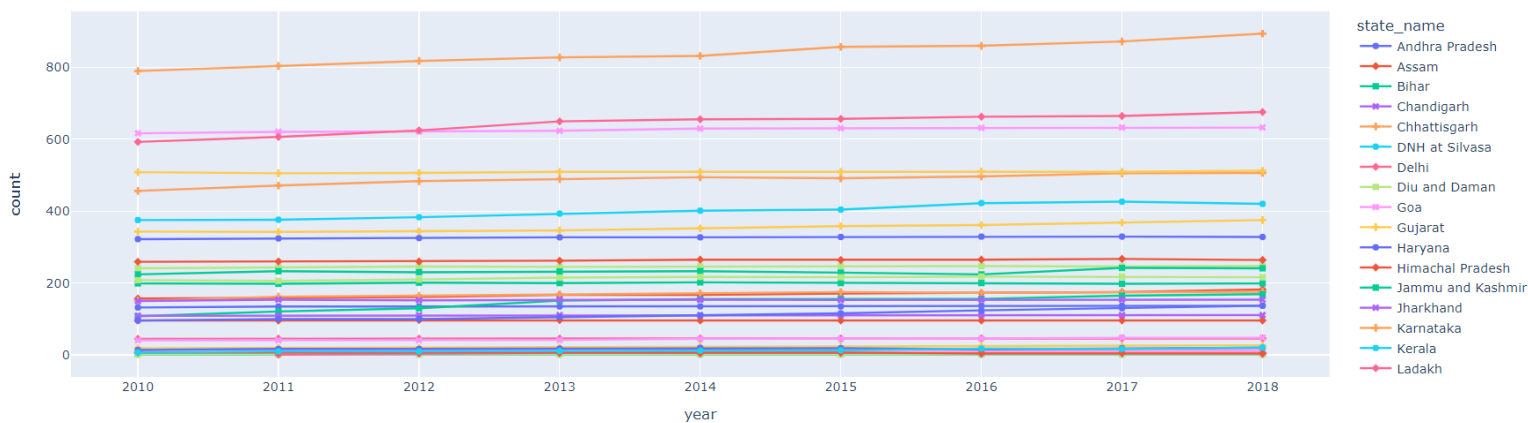


Fig 15 Number of courts per state

C. Cases

File: cases_2010.csv, cases_2011.csv and cases_2012.csv

Corresponding notebook: Precog task display_name_key and cases COURT_key.ipynb

- 1) The number of courts per state is highest for Maharashtra, followed by Uttar Pradesh and Karnataka. The increase in cases is visible in fig 16

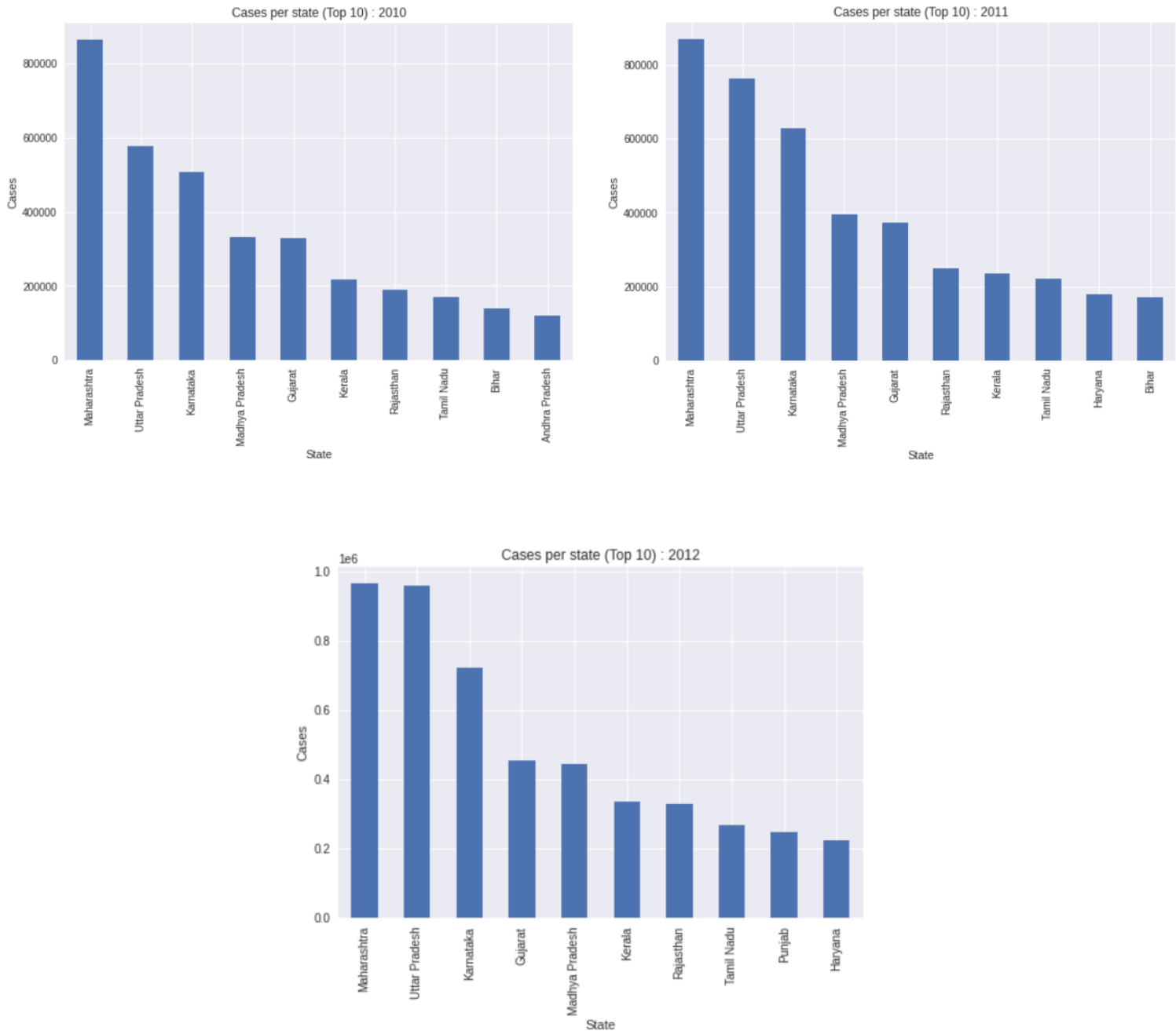


Fig 16 Cases per state (Top 10): 2010 - 2012

2) Defendant gender: Males are the predominant class, followed by unclear and female

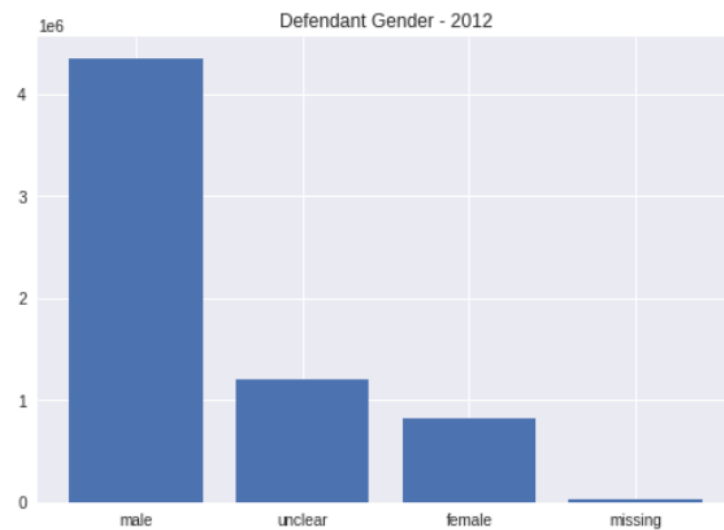
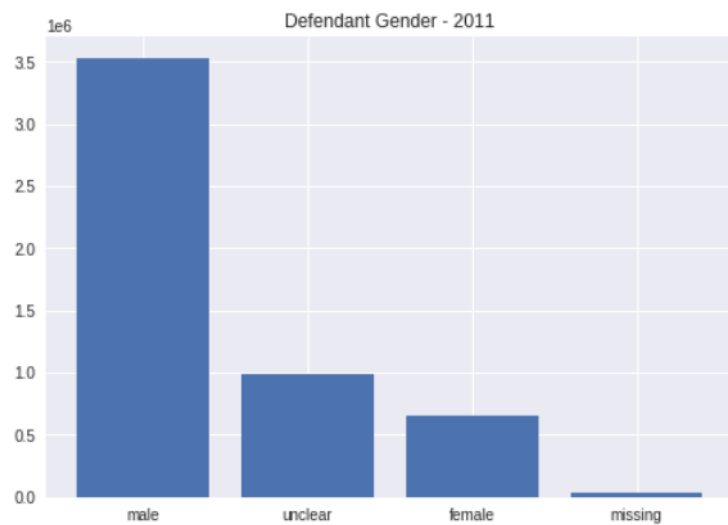
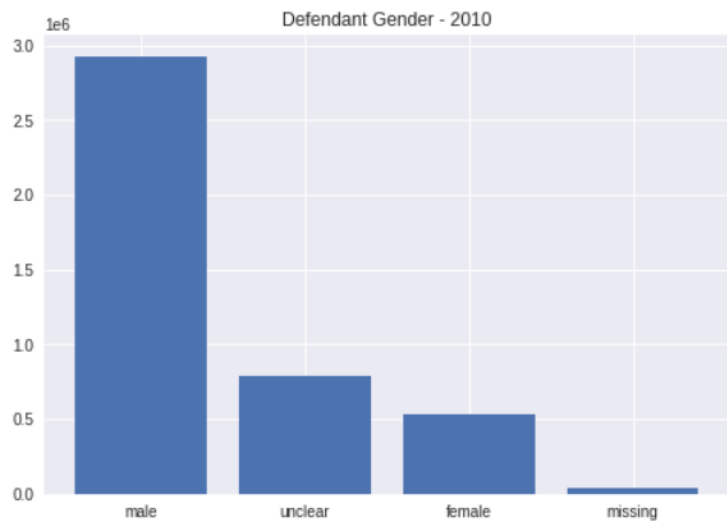
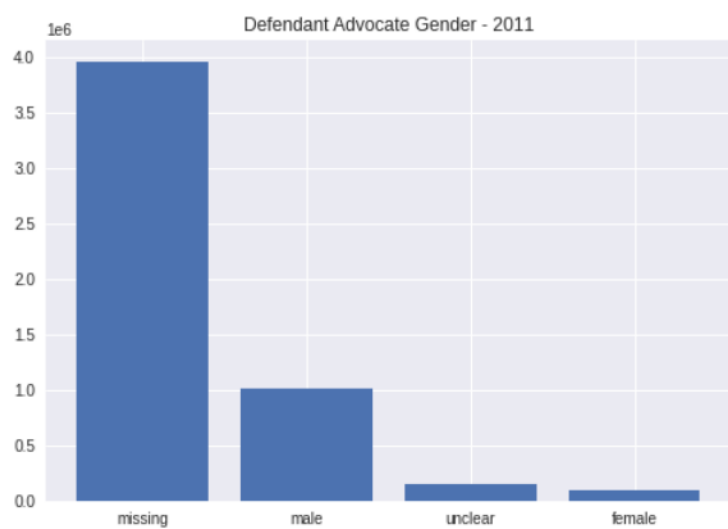
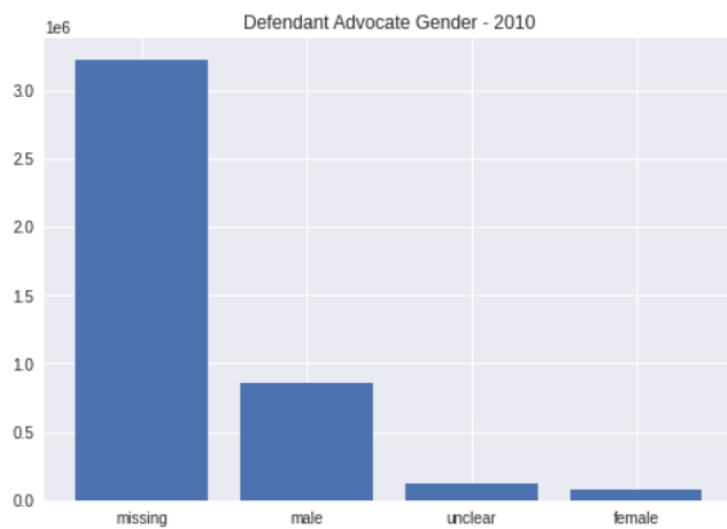


Fig 17 Defendant gender: 2010 - 2012

3) Defendant advocate gender



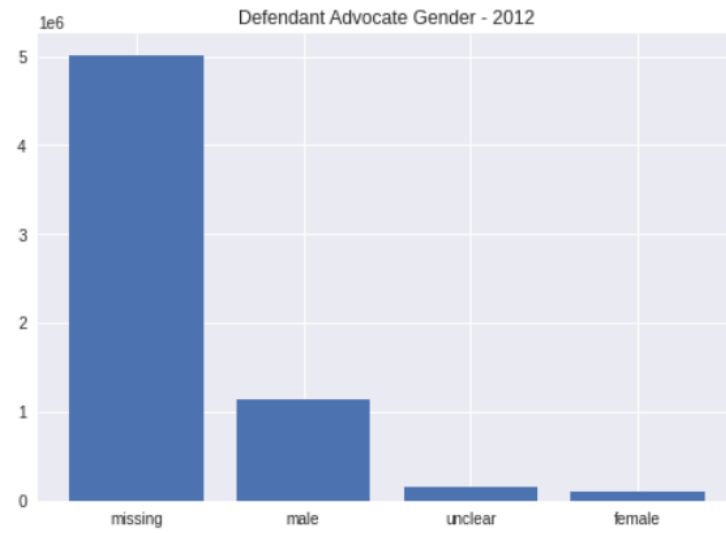


Fig 18 Defendant advocate gender: 2010 - 2012

4) Petitioner gender

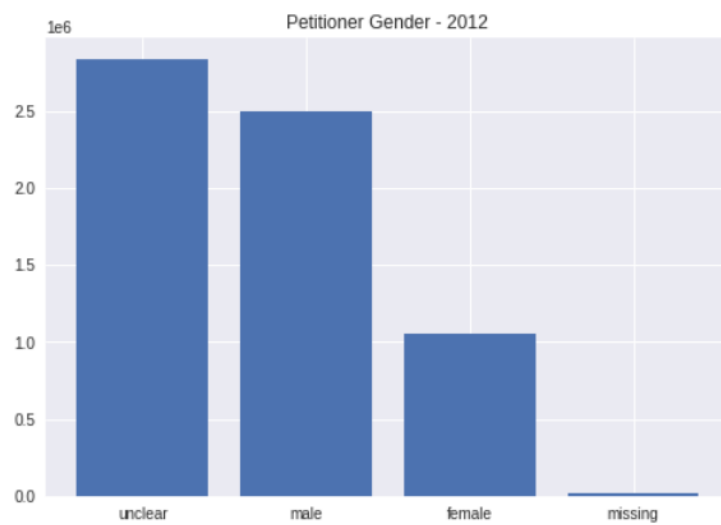
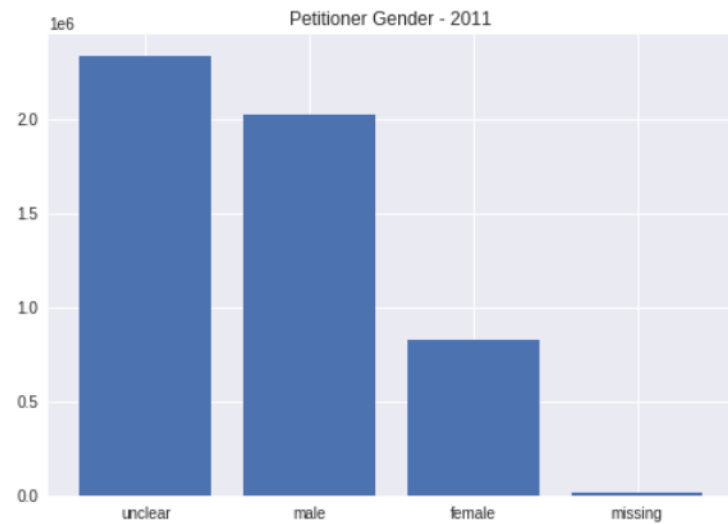
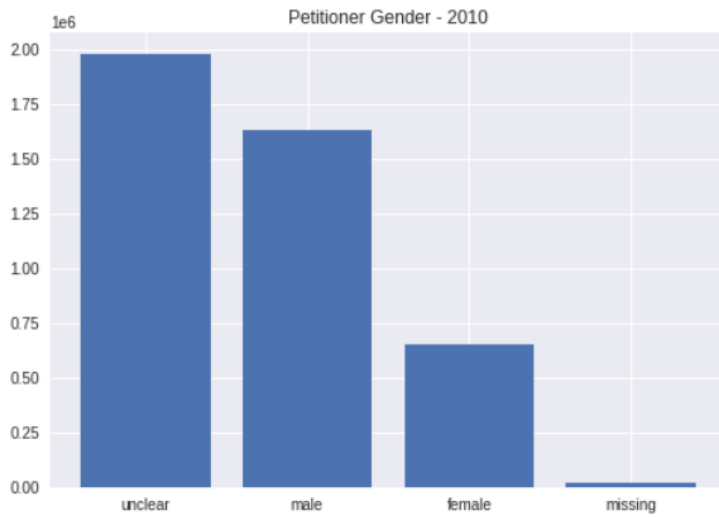


Fig 19 Petitioner gender: 2010 - 2012

5) Petitioner advocate gender

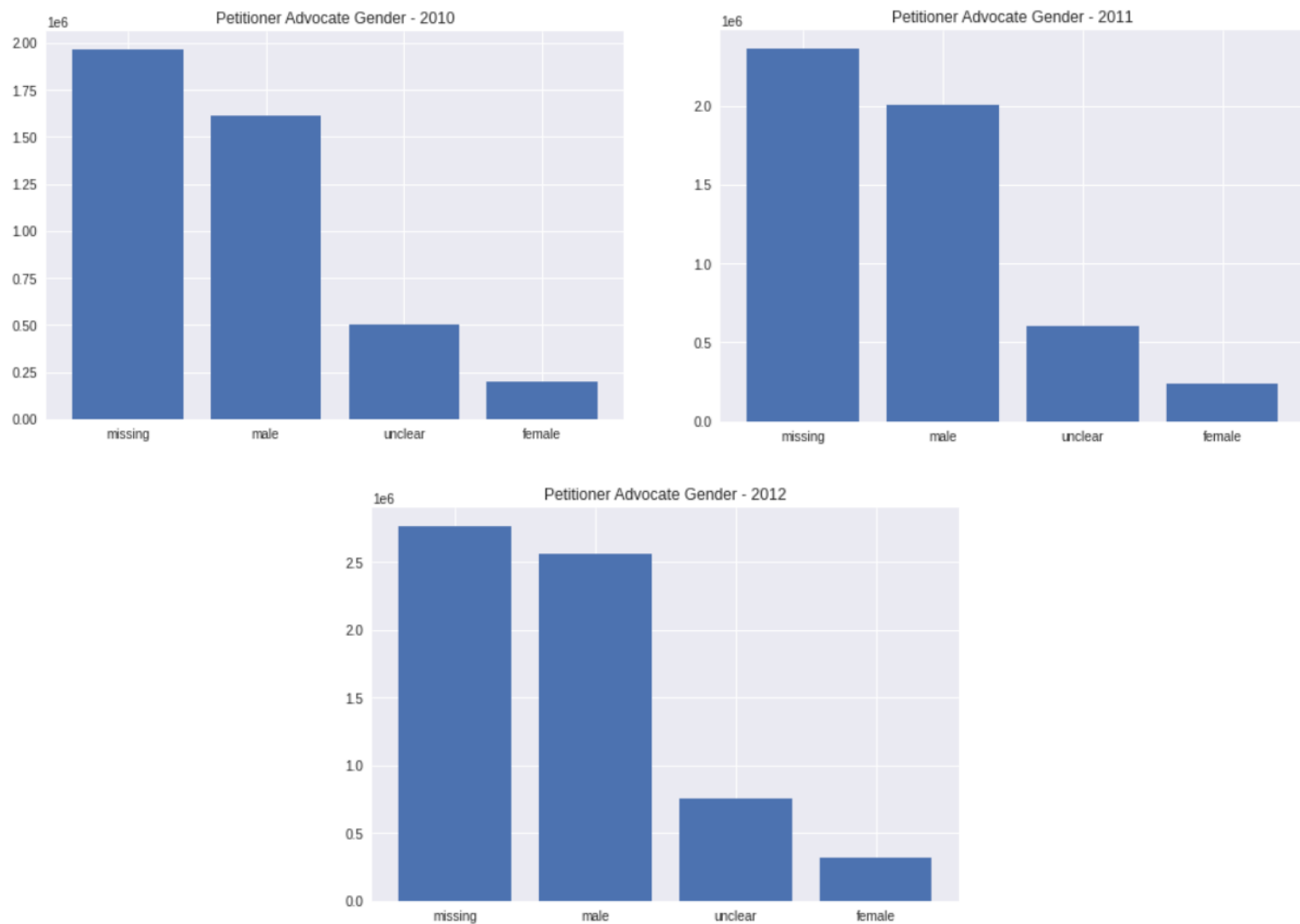
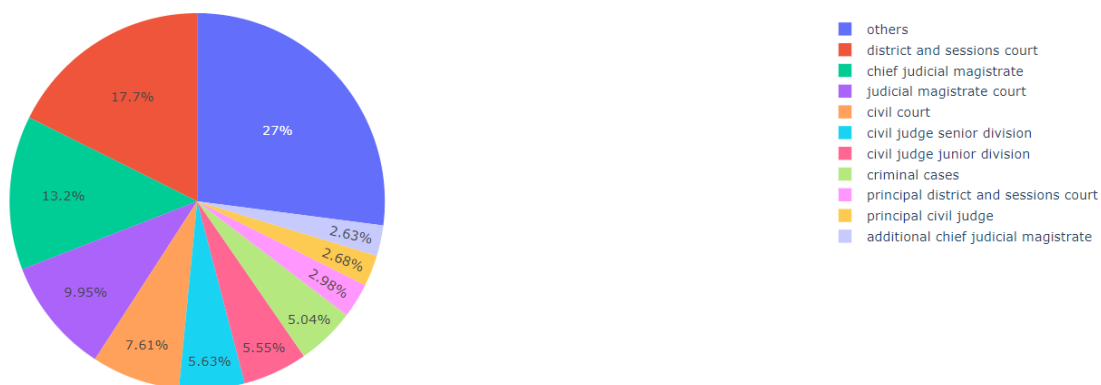


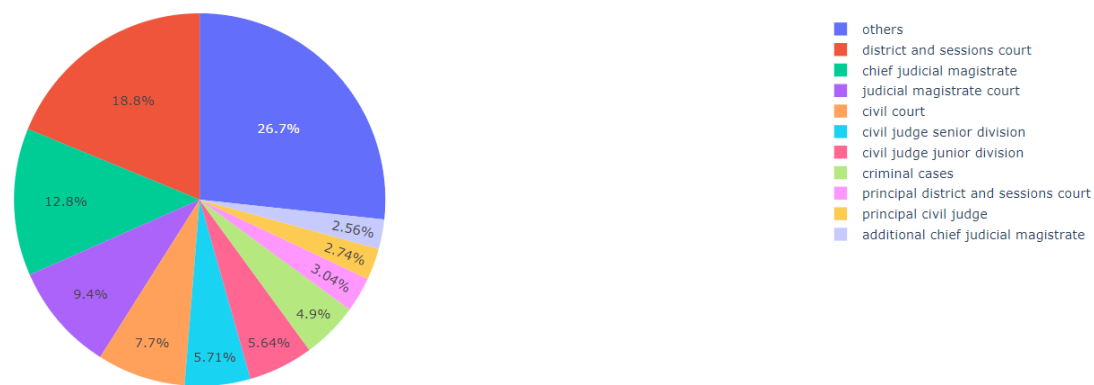
Fig 20 Petitioner advocate gender: 2010 - 2012

6) Fig 21 shows that district and session court judges were highest in number, followed by the chief judicial magistrate and judicial magistrate court.

Judge designation - 2010



Judge designation - 2011



Judge designation - 2012

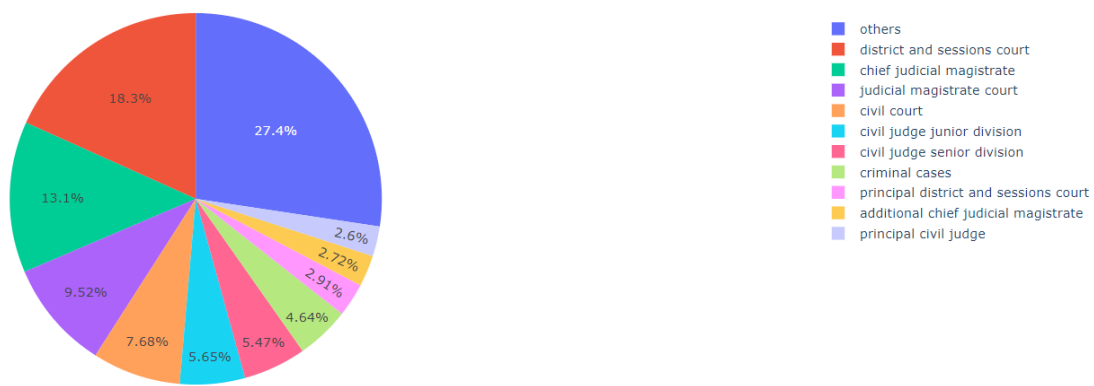
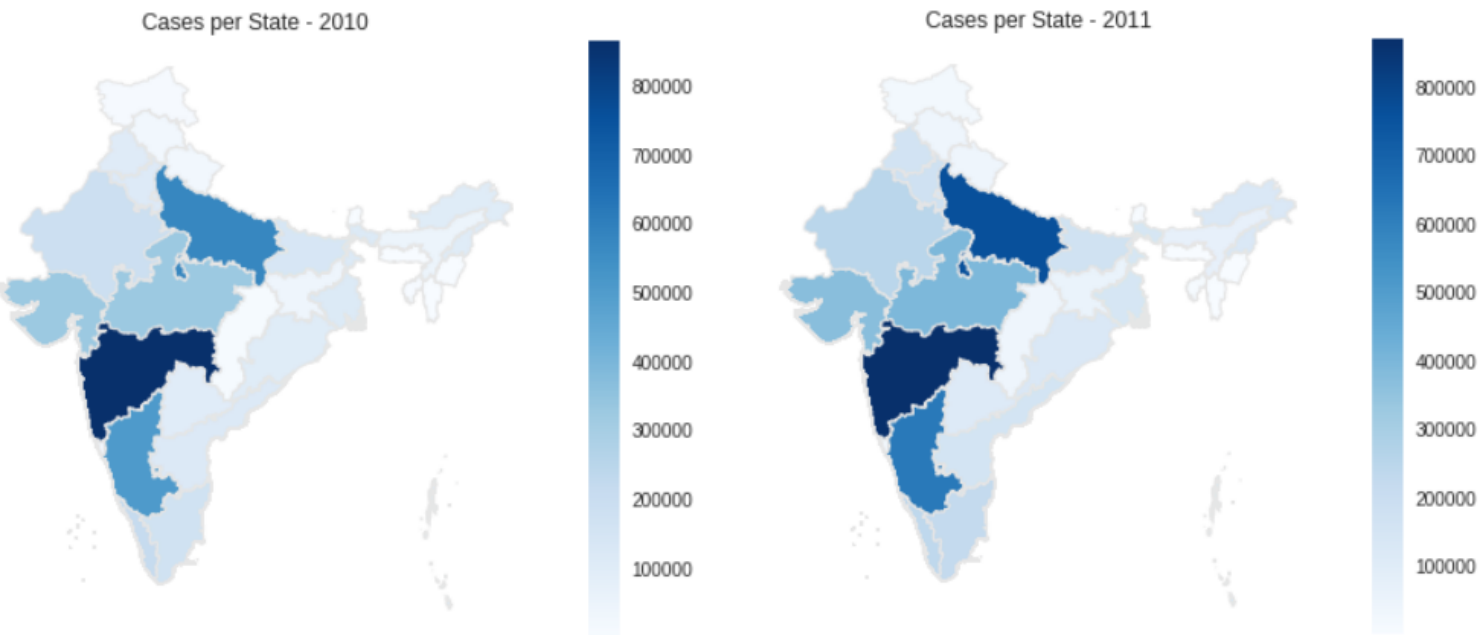


Fig 21 Judge designation: 2010 - 2012

7) Fig 22, showing cases per state during 2010 - 2012, confirms that the state of Maharashtra has the highest number of cases.



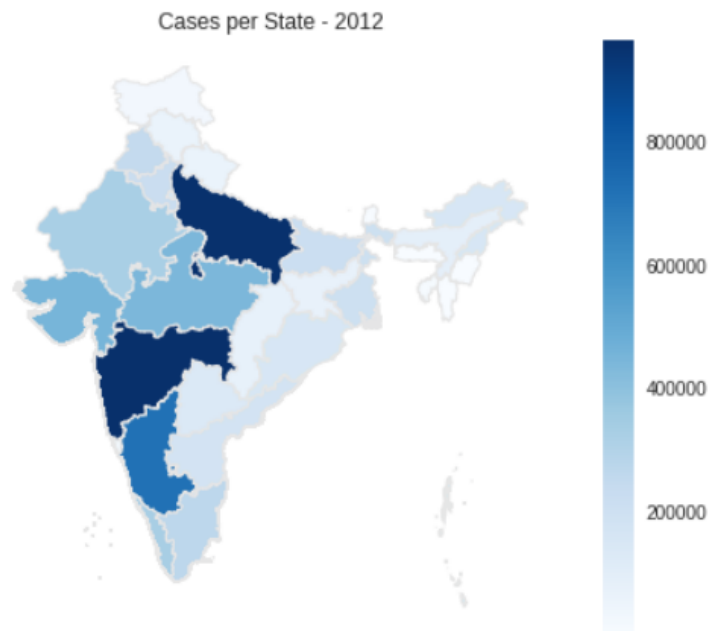
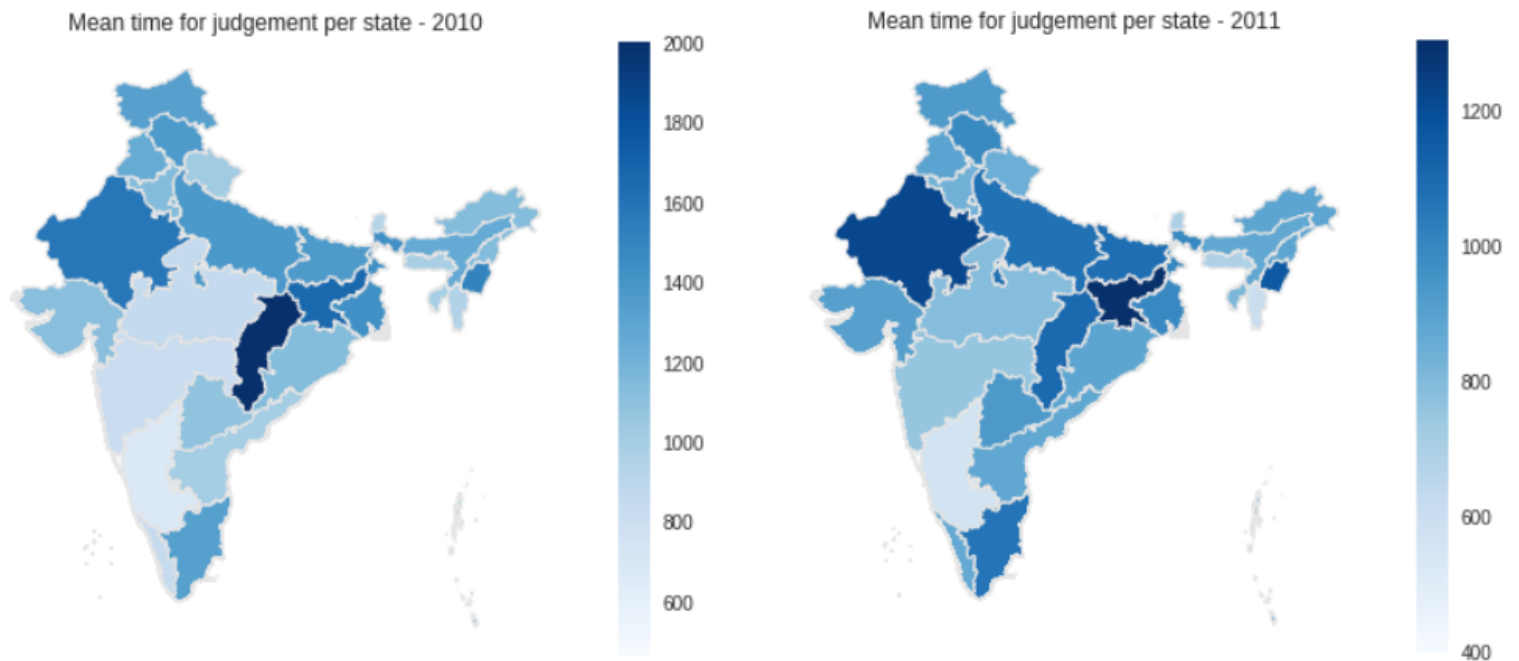


Fig 22 Cases per state: 2010 - 2012

- 8) Fig 23 shows an analysis of the meantime for judgment, which is the days between the filing date and the decision date. The cases with no date of decision are still pending and hence are not considered in these choropleth maps.



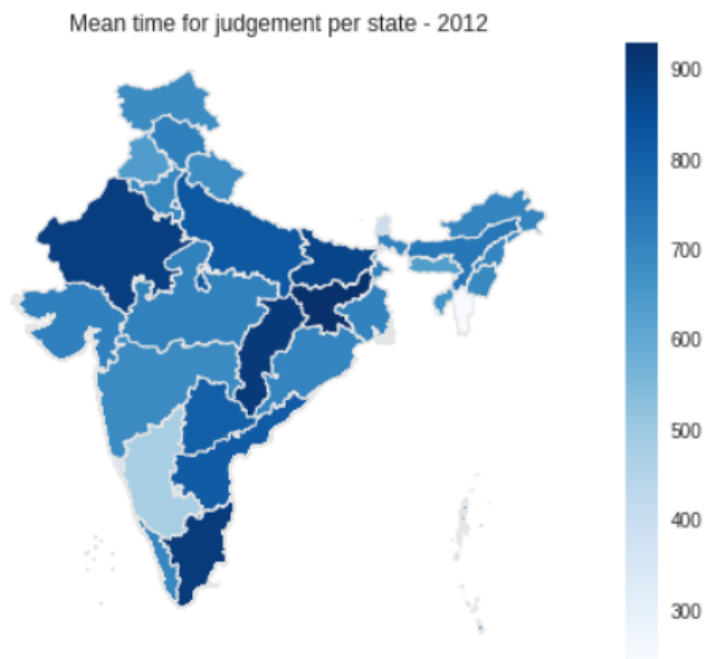
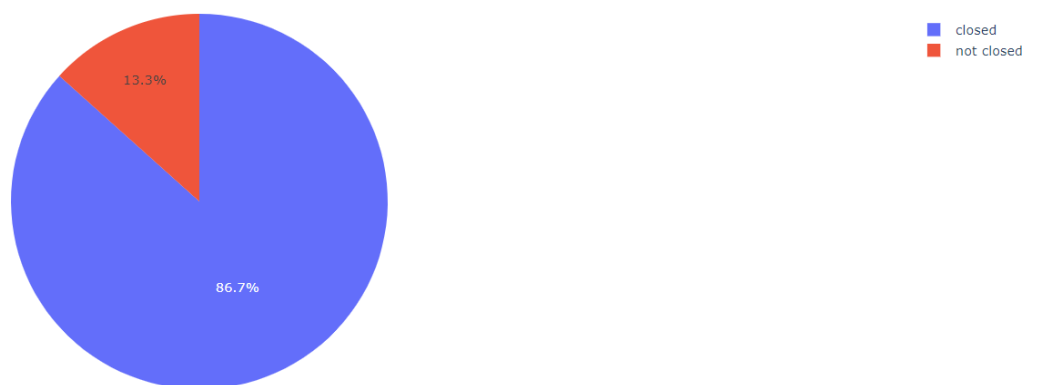


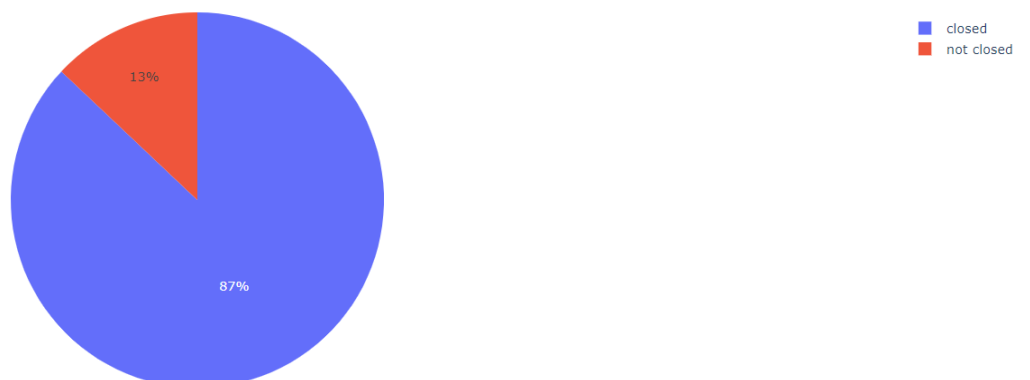
Fig 23 Average time for judgment per state: 2010 - 2012

- 9) It is observed that many cases are closed out of the total number of cases, but the number of new cases are increasing at a very high rate.

Decision of Cases - 2010



Decision of Cases - 2011



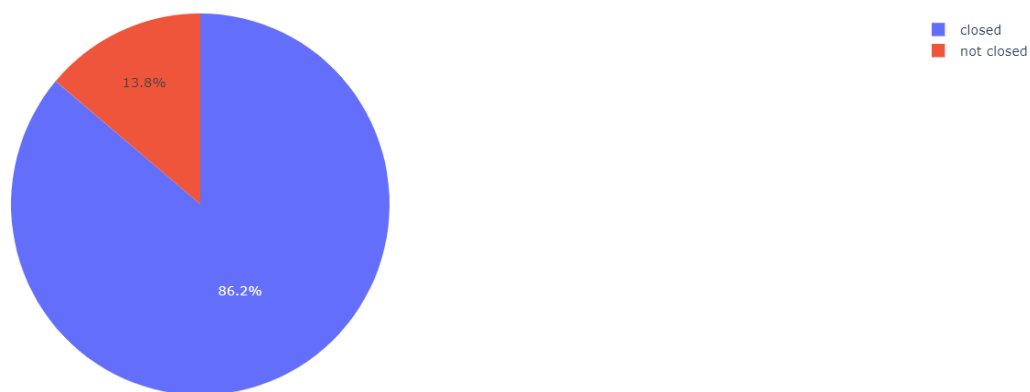


Fig 24 Average time for judgment per state: 2010 - 2012

10) Maharashtra has the maximum number of closed cases, followed by Karnataka, Uttar Pradesh and Madhya Pradesh.

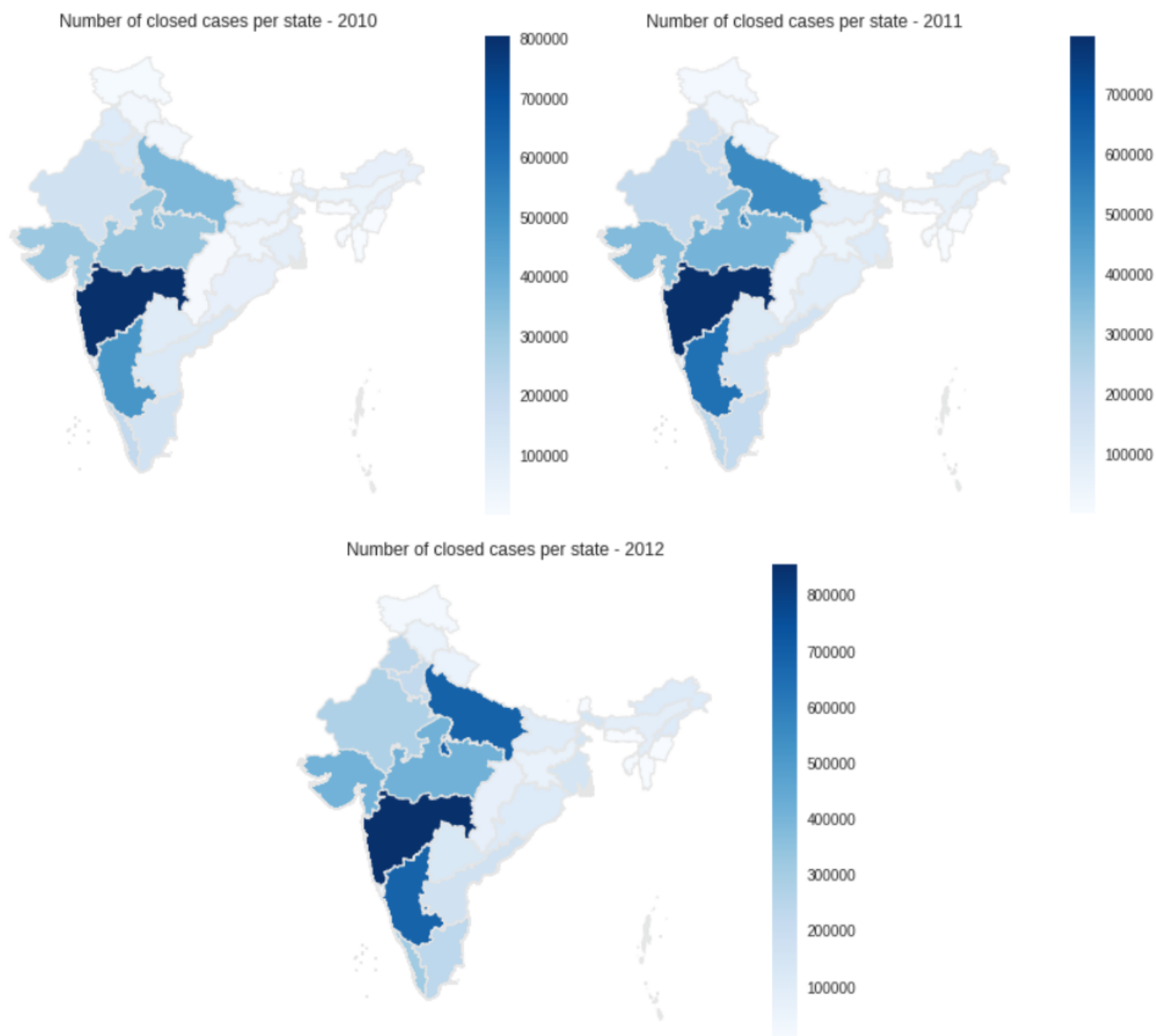


Fig 25 Number of closed cases per state: 2010 - 2012

11) Madhya Pradesh has the most number of unclosed cases, which is persistent over 2010 - 2012

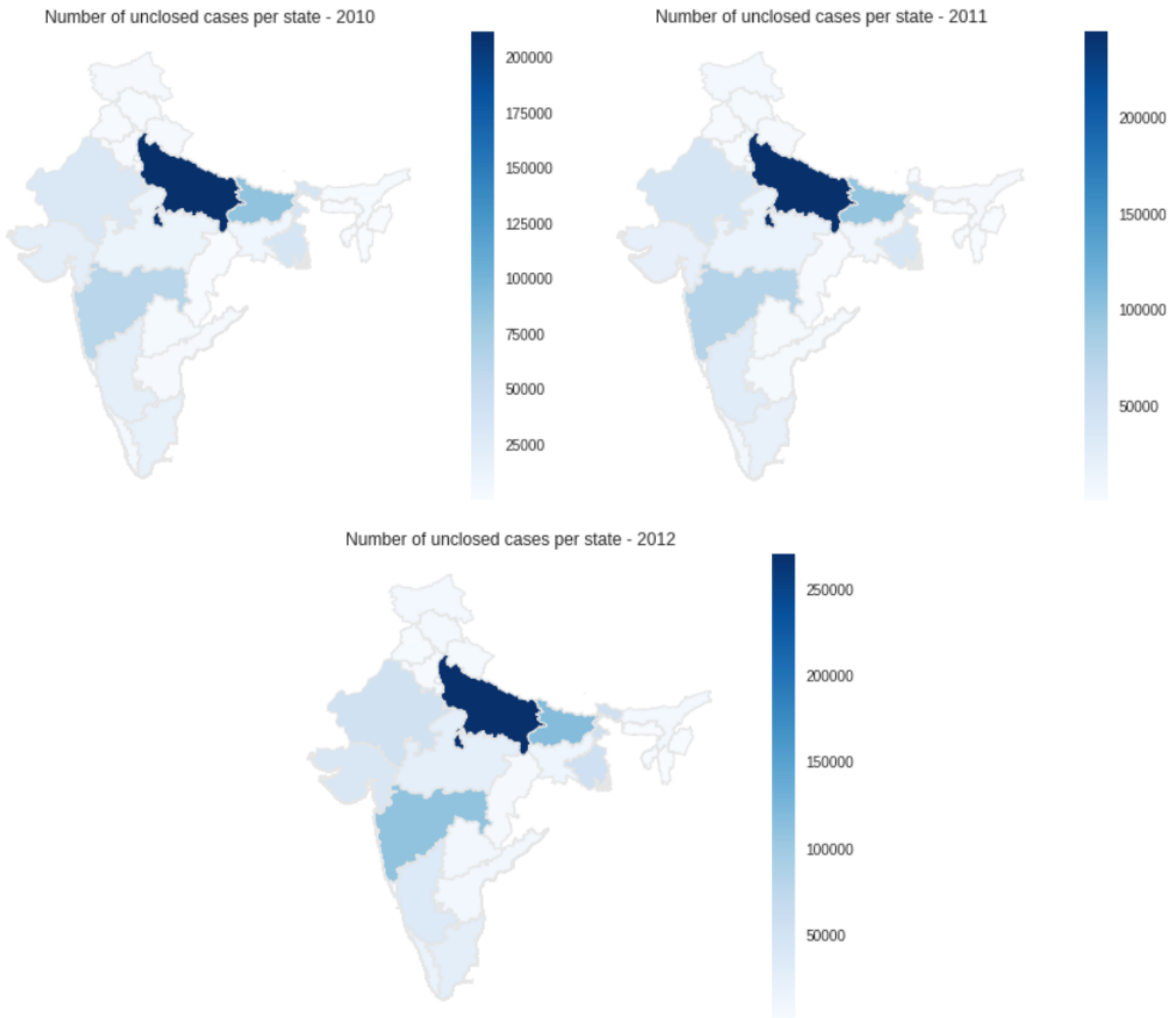


Fig 25 Number of unclosed cases per state: 2010 - 2012

12) It was observed that the number of courts had not increased over 2010-2012 significantly.

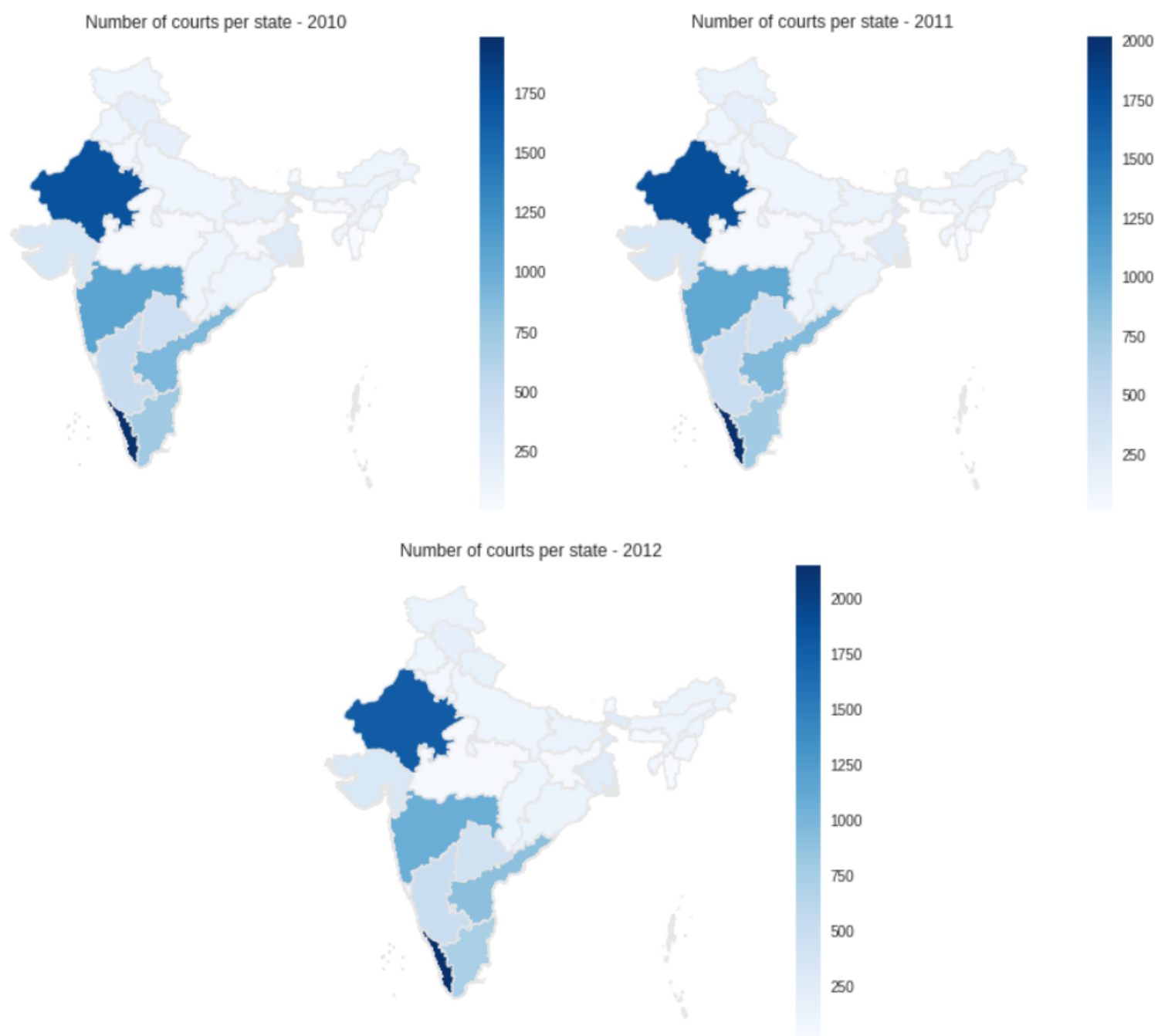
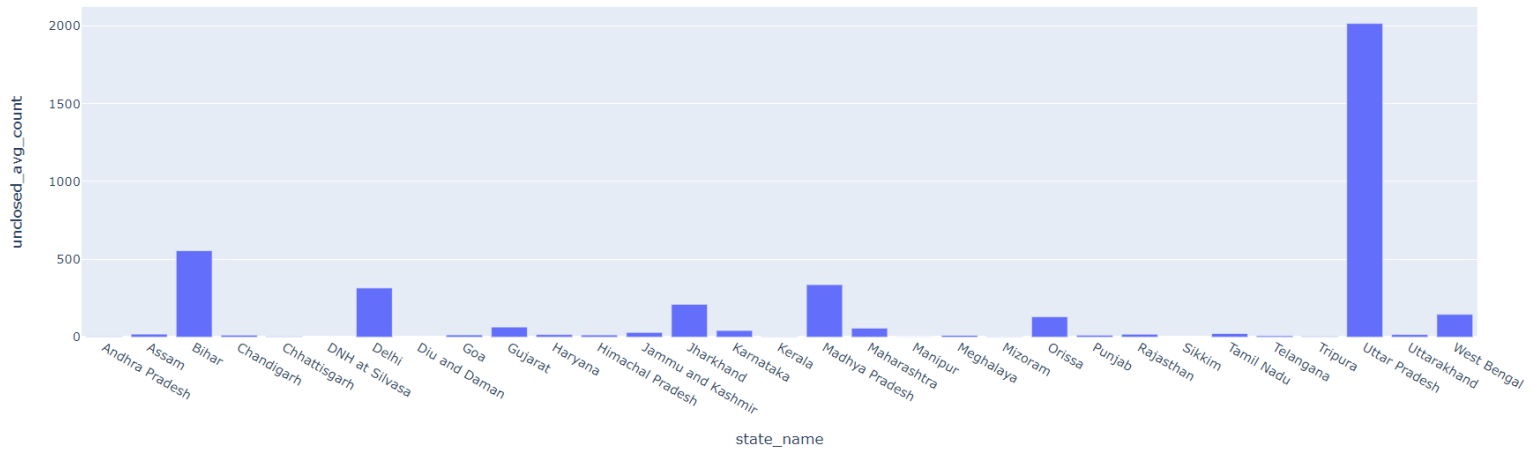


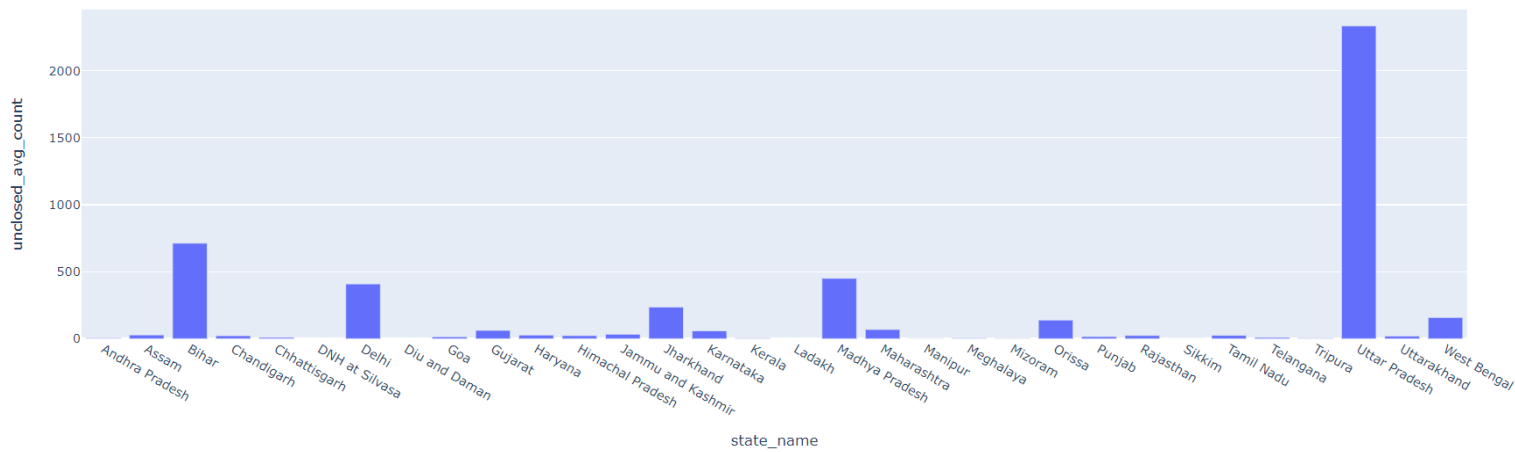
Fig 26 Number of courts per state: 2010 - 2012

13) Number of closed cases per court per state

Average unclosed cases per court per state - 2010



Average unclosed cases per court per state - 2011



Average unclosed cases per court per state - 2012

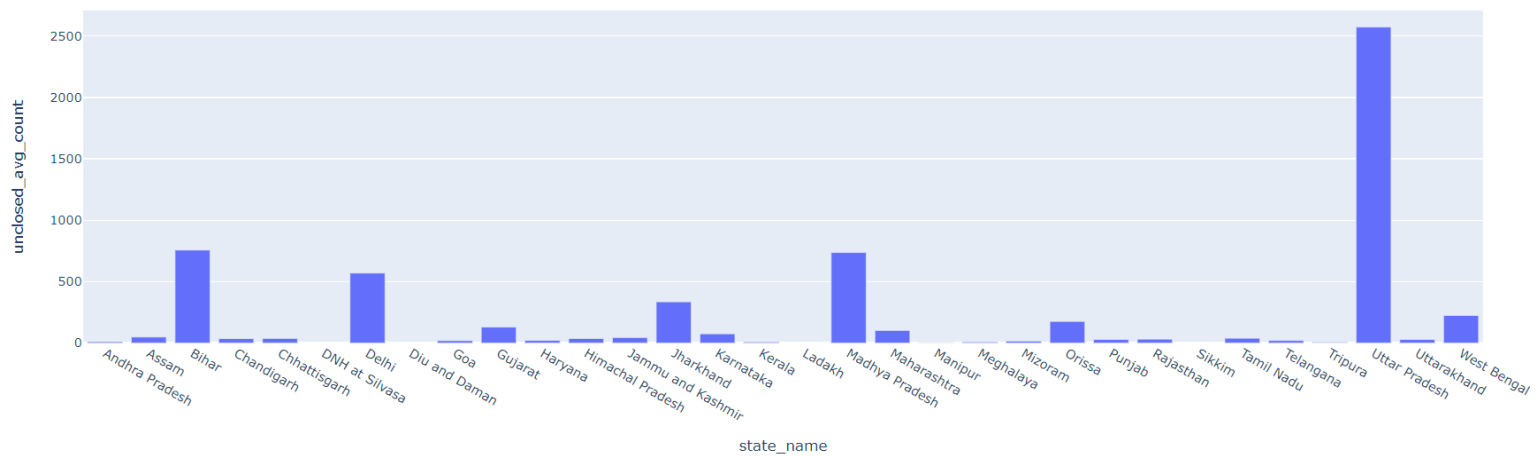
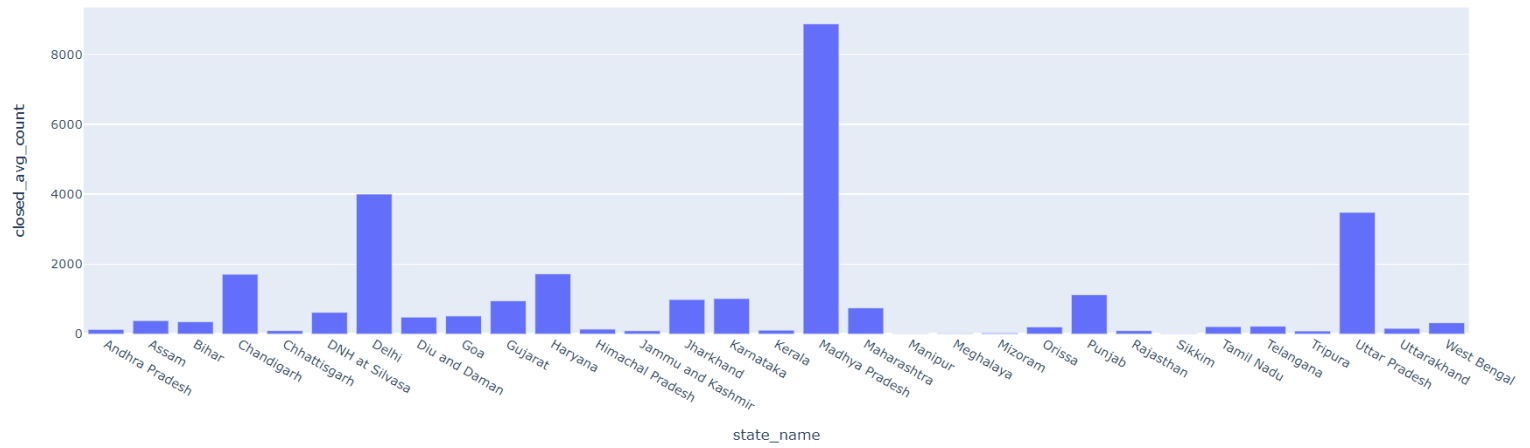


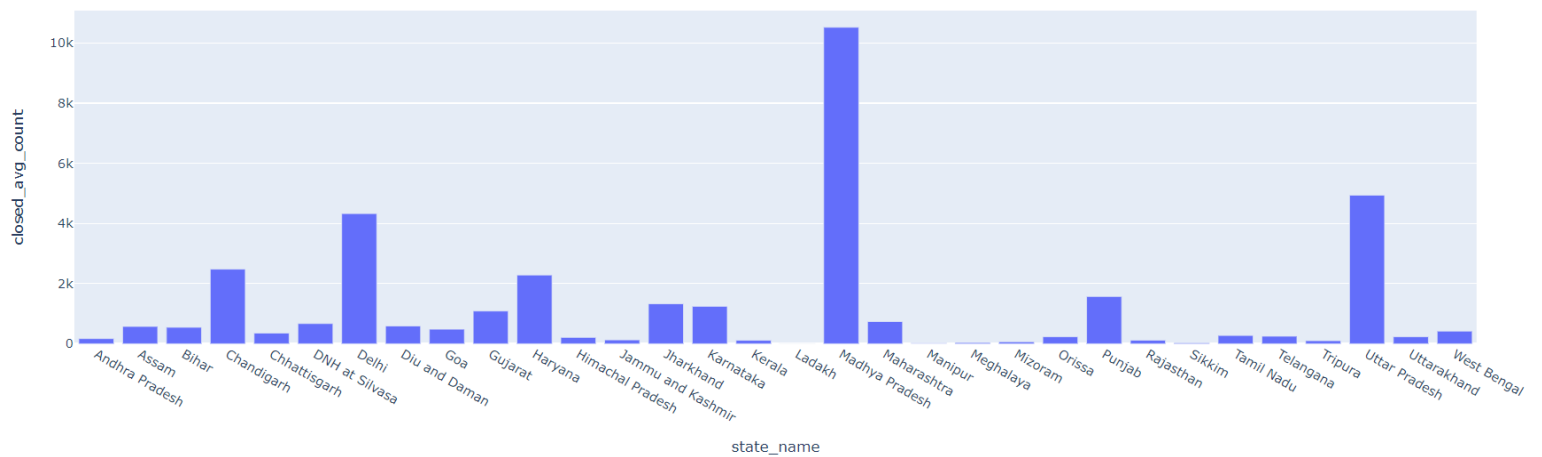
Fig 27 Number of unclosed cases per court per state: 2010 - 2012

14) Number of unclosed cases per court per state

Average closed cases per court per state - 2010



Average closed cases per court per state - 2011



Average closed cases per court per state - 2012

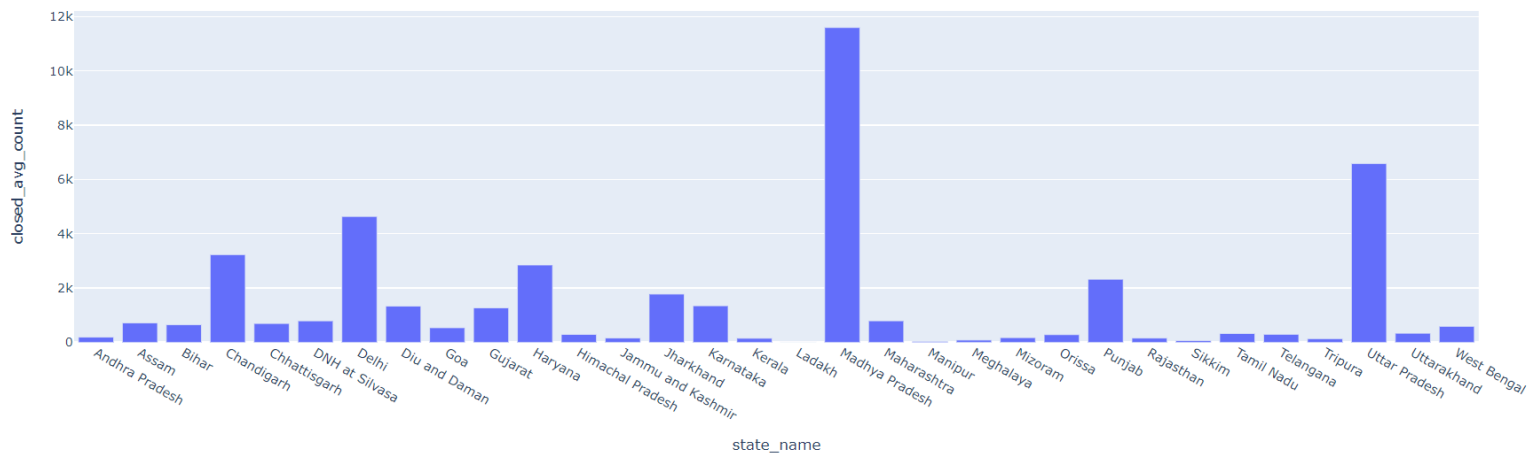


Fig 28 Number of closed cases per court per state: 2010 - 2012

D. All cases

File: cases_2010.csv to cases_2018.csv

Corresponding notebook: Precog task cases_all.ipynb

- 1) The majority of the defendants are male, according to the given data. The female defendants are just equal to a small portion of the male count and nowhere comparable.

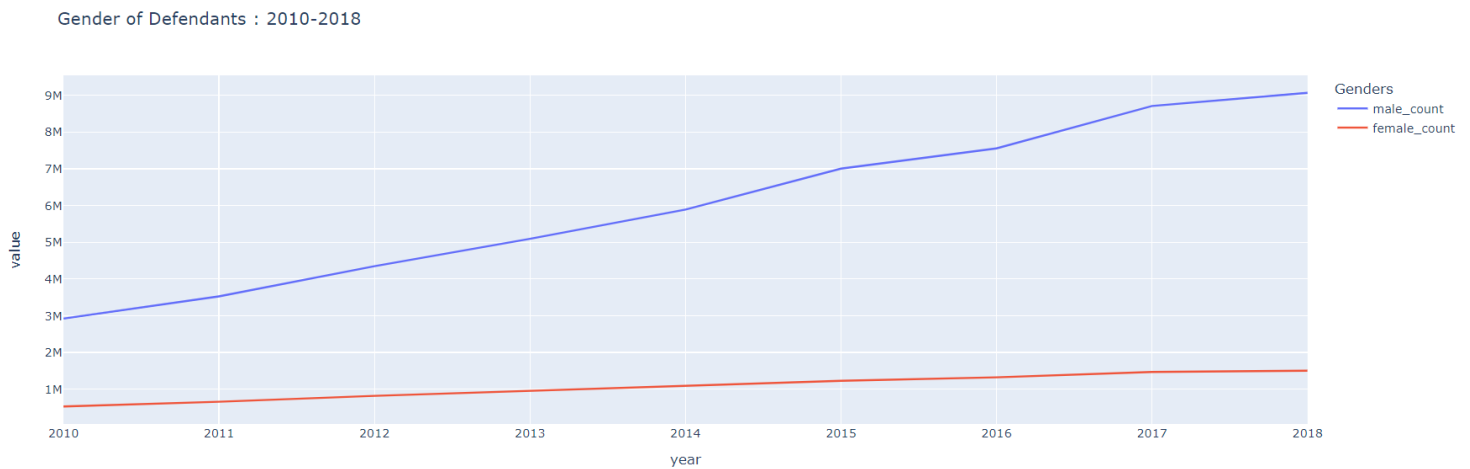
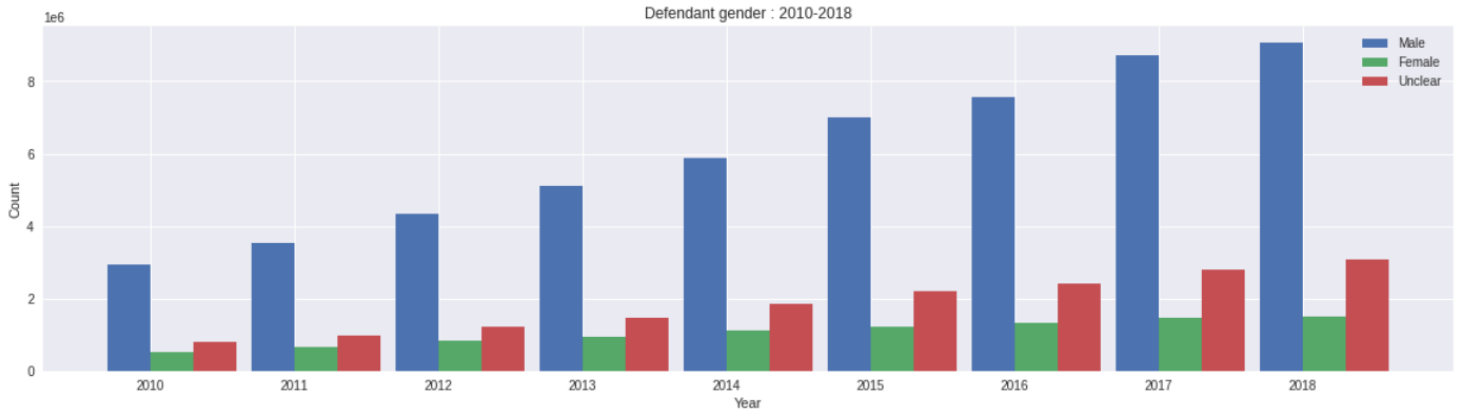
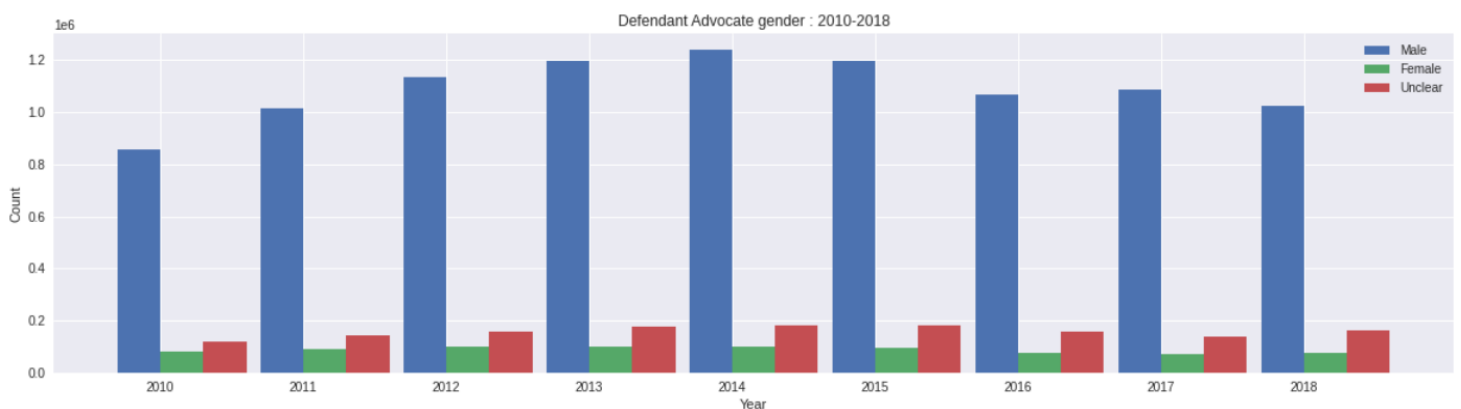


Fig 29 Defendant gender: 2010-2018

- 2) The defendant's advocate is male to a large extent.



Gender of Defendant's Advocate : 2010-2018

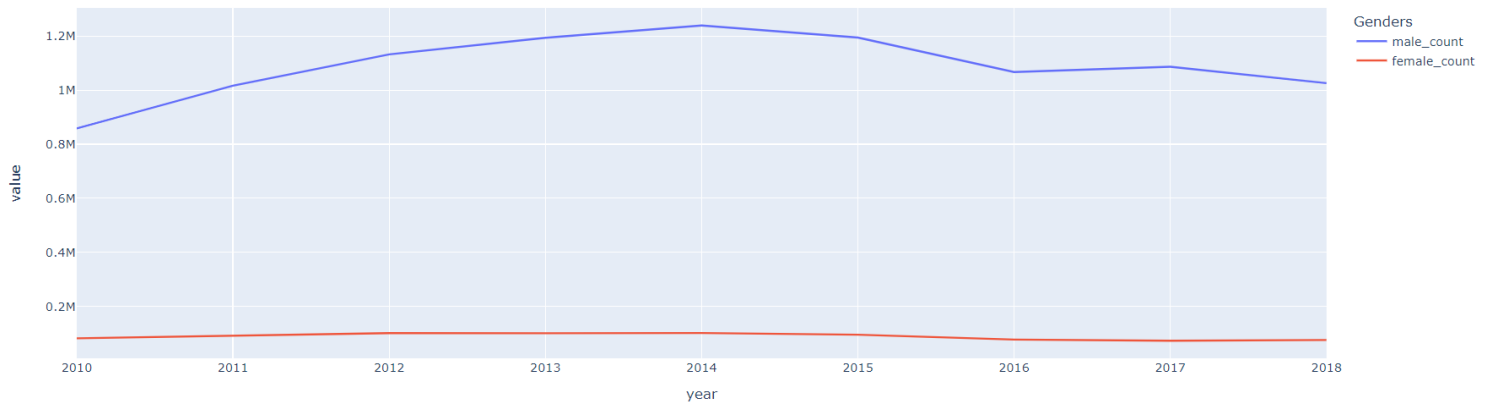
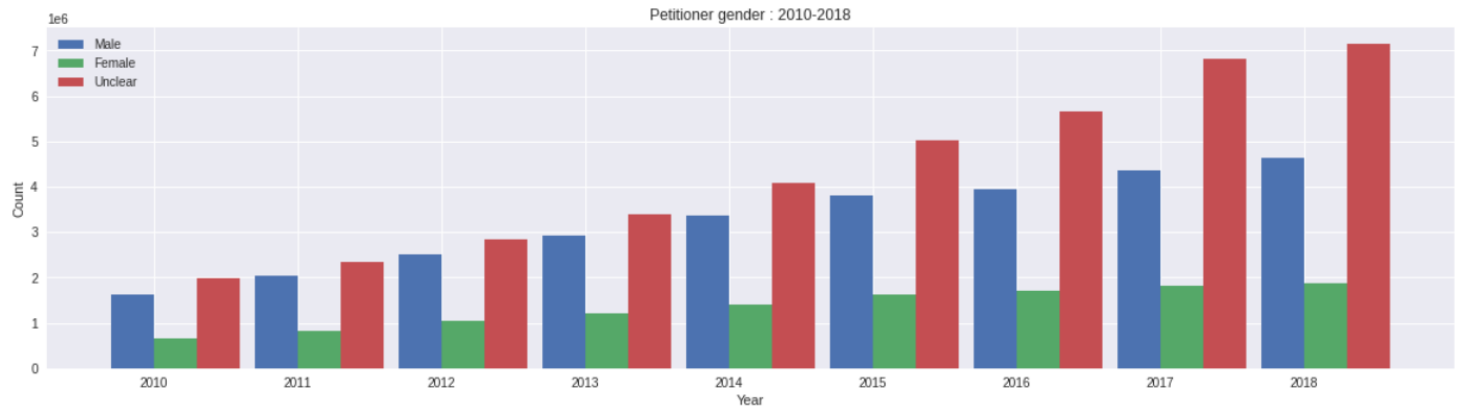


Fig 30 Defendant advocate gender: 2010-2018

- 3) According to the given data, the gender of the petitioner was unclear in many cases, and this trend has been consistent over the years. Comparing the male and female counts yields the result that the male count is more than the female count.



Gender of Petitioner : 2010-2018

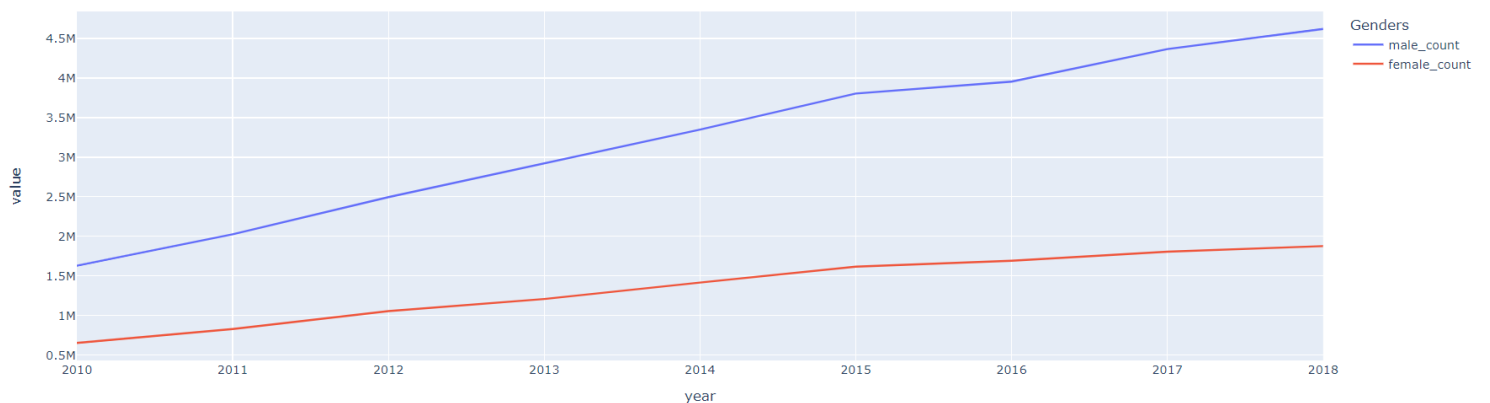
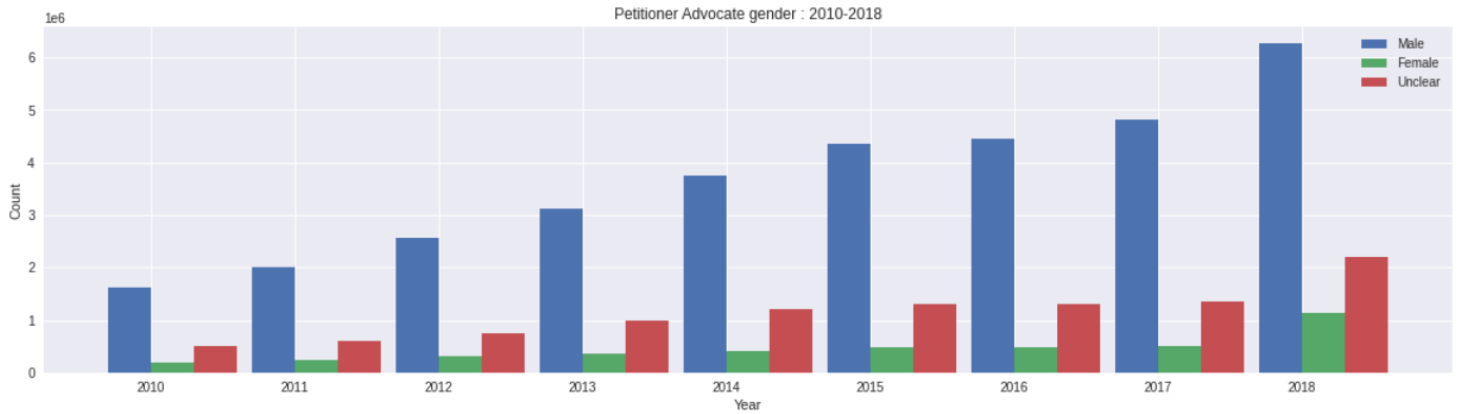


Fig 31 Petitioner gender: 2010-2018

4) The petitioner's advocate is male to a large extent.



Gender of Petitioner's Advocate : 2010-2018

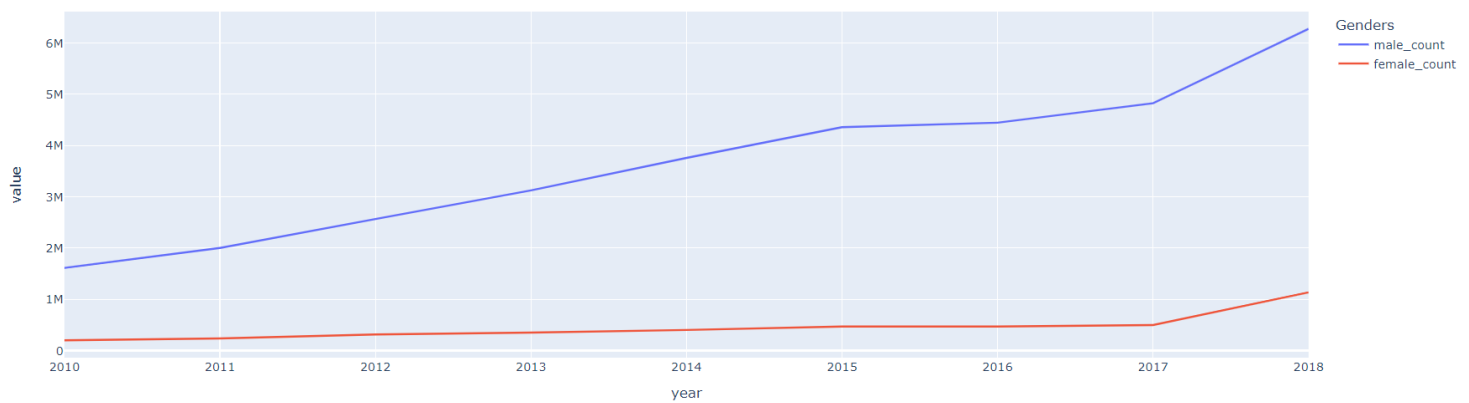


Fig 32 Petitioner advocate gender: 2010-2018

5) By plotting the number of open and closed cases over the years, we notice that the cases are closing linearly, but the number of open cases is increasing significantly. This will cause a backlog of cases and a burden for the courts.

Open and Closed Cases : 2010-2018

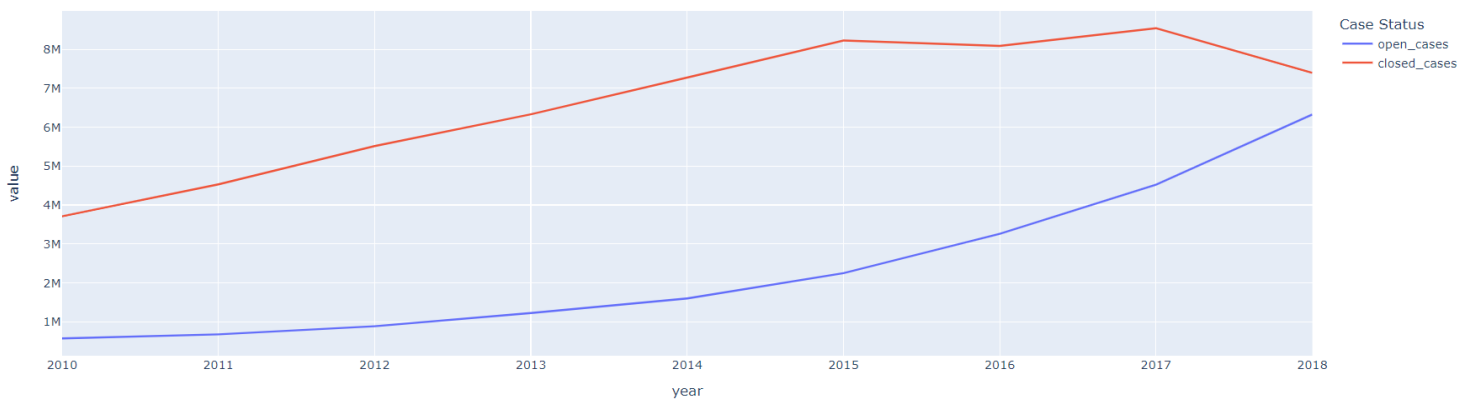


Fig 33 Open and closed cases: 2010-2018

- 6) The number of cases per year is growing linearly. This number considers only new cases opened that particular year and not the pending cases of the previous year also.

Court Cases : 2010-2018

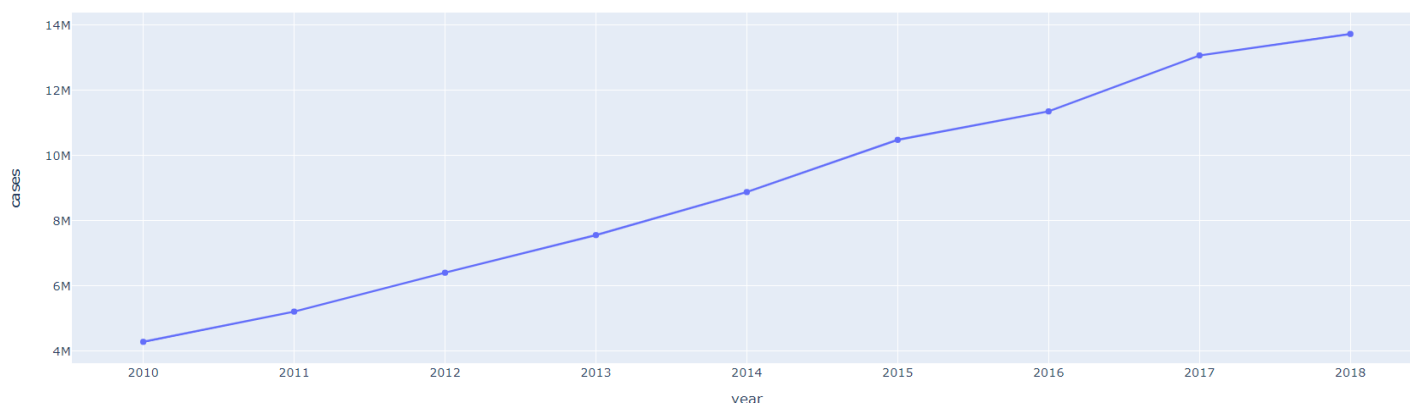


Fig 34 Court cases: 2010-2018

4. Future Work and Conclusion

In this paper, I performed the EDA (exploratory data analysis) of the given dataset to find some structure and trends. From the analysis, we can conclude that the number of cases and judges in Maharashtra is the highest. The number of female judges who started their tenure is significantly less compared to their male counterparts. The number of cases being closed per year is increasing linearly while the number of cases pending is increasing exponentially. This can cause a burden on the courts because the number of courts is not increasing significantly, and the cases are piling up. However, the total number of new cases in a year is increasing linearly. I plan to expand this report district-wise to gain insights on a much more profound level.

5. References

The following articles and websites were referred to while conducting the analysis and preparing this report:

1. <https://towardsdatascience.com/plot-choropleth-maps-with-shapefiles-using-geopandas-a6bf6ade0a49>
2. <https://towardsdatascience.com/a-beginners-guide-to-create-a-choropleth-map-in-python-using-geopandas-and-matplotlib-9cc4175ab630>
3. <https://practicaldatascience.co.uk/data-science/how-to-calculate-the-time-difference-between-two-dates-in-pandas#:~:text=There%20are%20several%20ways%20to,second%20between%20the%20two%20dates.>
4. <https://towardsdatascience.com/11-simple-code-blocks-for-complete-exploratory-data-analysis-eda-67c2817f56cd>
5. <https://towardsdatascience.com/5-advanced-visualisation-for-exploratory-data-analysis-eda-c8eafeb0b8cb>
6. <https://towardsdatascience.com/data-cleaning-in-python-the-ultimate-guide-2020-c63b88bf0a0d>
7. <https://plotly.com/python/>

8. https://www.analyticsvidhya.com/blog/2021/10/handling-missing-value/#:~:text=You%20can%20use%20the%20'fillna','%2C%20and%20'Self_Employed'.&text=Median%20is%20the%20middlemost%20value,Loan_Amount_Term'%20with%20the%20median%20value.
9. <https://towardsdatascience.com/7-ways-to-handle-missing-values-in-machine-learning-1a6326adf79e>
10. <https://medium.com/@DianaUrban/how-to-create-an-animated-gif-with-fading-effect-s-in-photoshop-d36b30e0d934>
11. <https://www.statology.org/pandas-select-rows-based-on-column-values/>
12. <https://towardsdatascience.com/handling-imbalanced-datasets-in-machine-learning-7a0e84220f28>
13. <https://towardsdatascience.com/dealing-with-imbalanced-classes-in-machine-learning-d43d6fa19d2>
14. <https://towardsdatascience.com/methods-for-dealing-with-imbalanced-data-5b761be45a18>
15. <https://towardsdatascience.com/having-an-imbalanced-dataset-here-is-how-you-can-solve-it-1640568947eb>
16. https://matplotlib.org/3.5.3/api/_as_gen/matplotlib.pyplot.html
17. <https://www.kdnuggets.com/2017/06/7-techniques-handle-imbalanced-data.html>