REPORT - PRECOG RECRUITMENT TASK

- By Ashna Dua

DATA AND TASK DESCRIPTION:

The Development Data Lab (DDL) is an esteemed organization that employs the utilization of data analysis and visualization to illuminate policy and decision-making in developing nations. They are acclaimed for producing a plethora of data sets, consisting of information pertaining to the judiciaries of several countries, including but not limited to court case backlogs, judicial budgets, and the number of judges and court staff. The purpose of this data is to assist policymakers, researchers, and other stakeholders to comprehend and tackle issues related to the efficiency and functioning of judiciaries in developing countries. This data can be utilized to detect patterns and trends, pinpoint areas in need of amelioration, and monitor progress over time.

The Judicial Data produced by the Development Data Lab was provided for the purpose of exploration and analysis in the PreCog Recruitment Task. This task is divided into two parts, Analysis and Classification, which require the necessary preprocessing and plotting steps to be carried out.

PART 1: ANALYSIS

CASE LOAD:

In the legal field, a case load refers to the number of cases or clients that a lawyer, judge or legal organization is responsible for handling at a given time. A case load for a judge refers to the number of cases that a judge is responsible for hearing and making decisions on at a given time. This can include criminal cases, civil lawsuits, and other legal matters that come before the court. A judge's case load can vary depending on the type of court they preside over, the area in which they serve, and the overall workload of the court system.

A judge who has a heavy caseload may have a plethora of cases scheduled for trial or hearing at any one time, and may be required to proceed expeditiously to keep pace with the volume, whereas a judge with a lighter case load may have fewer cases to hear and may be able to devote more time to each case.

Also, in some jurisdictions, judges are assigned to specific types of cases, such as criminal, civil, or family law, and their caseload will reflect the type of cases they are assigned to hear. Upon conducting a thorough analysis of the data, the mean caseload per judge of each state was meticulously plotted, and the variation in this case load over the years was observed.

This examination revealed that judges in the states of Chandigarh, Delhi, and Punjab were faced with the most substantial case loads, whereas judges in the states of Sikkim, Meghalaya, Mizoram, and Tripura were comparatively less burdened.
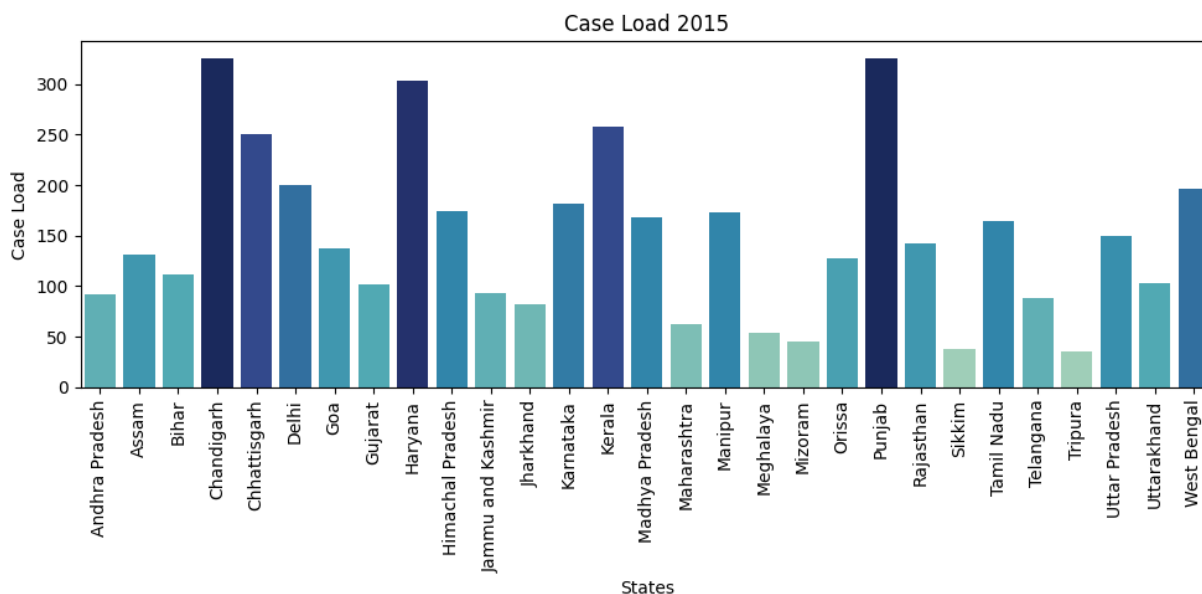

Fig 1. Case Load in 2015

METHOD:

The Pandas Library of Python, a powerful tool for data manipulation and analysis, was utilized to scrutinize the data of the cases over the years, as well as the data of the judges. The calculation of the caseload was executed in the following manner:

$$Caseload\ in\ a\ given\ year = \frac{Number\ of\ cases\ in\ the\ state\ for\ a\ given\ year}{Number\ of\ judges\ in\ the\ state\ for\ a\ given\ year}$$

The caseload for each state was calculated and plotted with precision, and the trend was studied with great attention from 2010 to 2018.

Link to view the trend with the help of a GIF: Case Load

GENDER BIAS:

Gender bias refers to the preferential treatment or prejudice towards individuals based on their gender. This can manifest in a variety of forms, including discrimination, stereotypes, and societal expectations. Gender bias can be both conscious and unconscious, and it can affect both men and women, although it is more commonly directed towards women. It can occur in various settings such as the workplace, education, and healthcare, including the judicial system.

Gender bias can have negative consequences on individuals, organizations, and society as a whole, by limiting opportunities, hindering performance and productivity, and promoting inequality.

An examination of the number of criminal and civil cases presided over by judges revealed that the proportion of cases handled by women is significantly less than 50%, indicating a pervasive and persistent gender bias within the legal system. While the percentage of women handling criminal and civil cases has been gradually increasing over the years, it will take a considerable amount of time at this rate to achieve a state of parity.
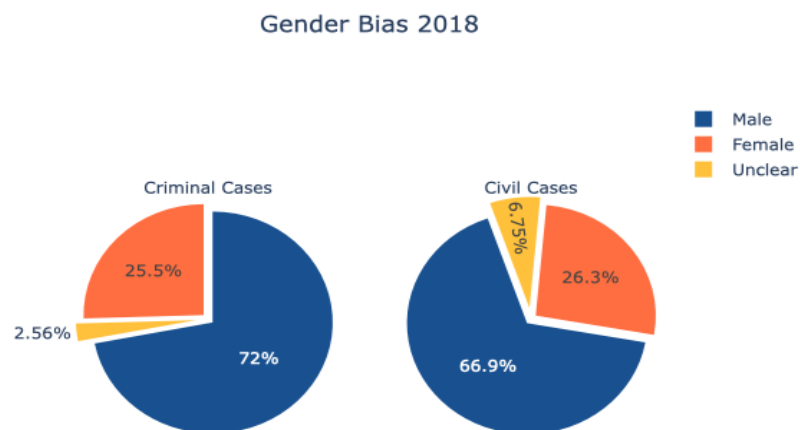

Fig 2. Gender Bias in 2018

METHOD:

In order to demonstrate the presence of gender bias among judges handling criminal and civil cases, the Pandas and Plotly libraries of Python were employed. The datasets of acts and sections were seamlessly integrated with the cases dataset and the judge dataset to facilitate the identification of the genders of the judges handling said cases.

Utilizing the Plotly library, these results were graphically represented through the use of pie charts, depicting the proportion of men and women handling these cases respectively, thus offering a clear and concise visual representation of the gender bias present in the legal system.

Link to view the trend with the help of a GIF: Gender Bias

PENDING CASES:

Pending cases in the Indian judicial system refer to legal proceedings that have been initiated but have not yet been resolved or decided upon by a judge or a bench of judges. These cases include a wide range of legal disputes such as criminal trials, civil lawsuits, appeals and more. The problem of pending cases has been a chronic issue for the Indian judiciary for a long time now, which can be attributed to several reasons such as an inadequate number of judges, insufficient resources and an overwhelming backlog of cases.

These factors have resulted in an overburdened legal system, leading to delays and inefficiencies in the administration of justice. Through an analysis of the decision dates of cases, a map was created to give a comprehensive overview of the number of pending cases in each state.

The findings of this analysis reveal that over the years, the maximum number of pending cases are observed in the state of Uttar Pradesh, indicating a significant backlog of unresolved legal matters in this region. On the other hand, the least number of pending cases are observed in the states of East India, indicating a more efficient and streamlined legal system in these areas. This map serves as a valuable tool for identifying areas in need of improvement and for implementing targeted measures to address the issue of pending cases.
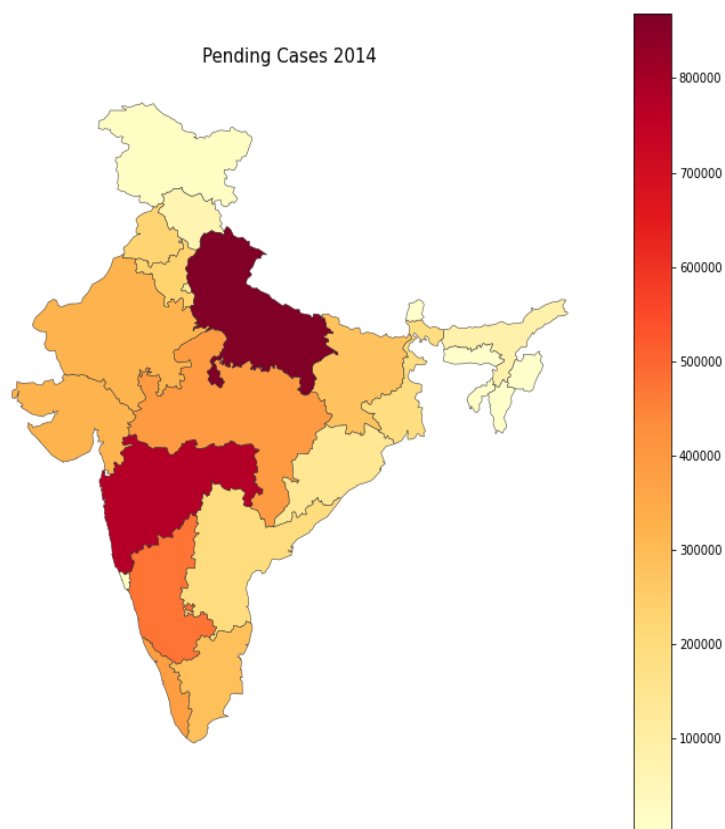

Fig 3. Pending Cases in 2014

METHOD:

In order to provide a visual representation of the number of pending cases in each state, the Pandas and GeoPandas libraries of Python were employed. The number of pending cases were determined through a calculation that took into account the number of cases in a given year, whose decision date had not yet been reached by December 31st of that year.

Utilizing the GeoPandas library, these results were plotted on a map of India, with each state being represented by a color corresponding to a scale that indicated the number of pending cases. This graphical representation offered a clear and concise understanding of the distribution of pending cases across the states of India.

Link to view the trend with the help of a GIF: Pending Case

DISPOSITIONS:

In the legal system, a disposition refers to the final outcome or ruling of a case. It is the last act of determining the result of a legal matter, made by a court or other judicial body with the authority to decide legal matters. Dispositions can include a verdict of guilt or innocence in criminal trials, an award of compensation in civil lawsuits, or an order for a specific action in legal disputes.

These dispositions are the official conclusion of the legal process and bring closure to all parties involved. They are typically recorded in a written document called a judgment or opinion, which includes the rationale and legal reasoning behind the decision, and serves as a legal precedent for future cases.

An analysis of legal cases revealed a pattern in which dispositions of "disposed-otherwise", "allowed", and "acquitted" were the most frequent outcomes. This trend was found across the entire dataset and can be useful for predicting future dispositions and formulating legal strategies, as well as providing insight into the judicial system.
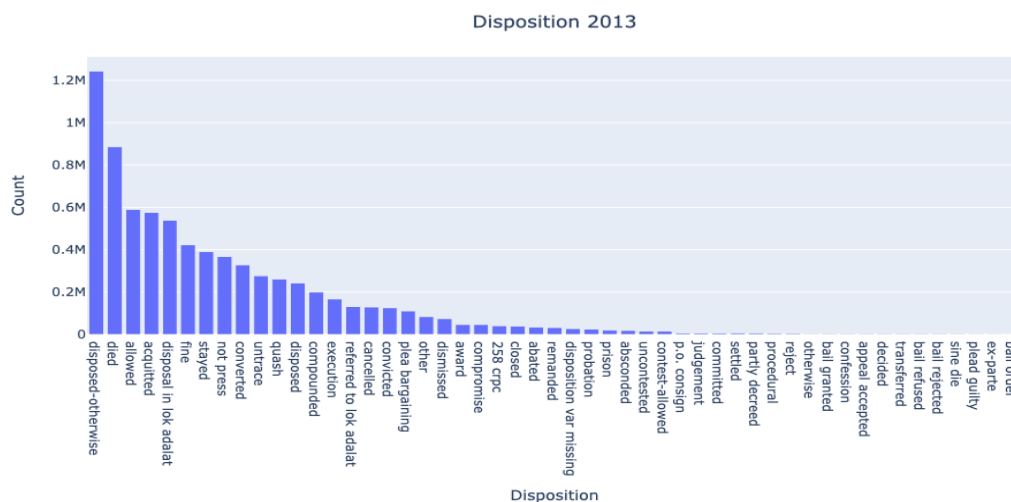

Fig 4. Disposition Distribution in 2013

METHOD:

In order to chart the aggregate number of dispositions in each year, the Plotly library was employed to create a bar graph, wherein each bar represented a distinct disposition. This graph offered a valuable perspective into the ultimate determination of the majority of cases, by providing a visual representation of the frequency of each disposition over time. This approach was particularly useful in identifying any temporal variations in the distribution of dispositions, thereby allowing for a more informed understanding of the judicial process.

Link to view the trend with the help of a GIF: Disposition Trend

PART 2: CLASSIFICATION

The aim of this classification problem is to determine whether a criminal case is eligible for bail or not, by analyzing the features of the case such as the Act, Section, Type of Case, Disposition, and Purpose of the case. These features were selected in order to predict the bailability of a criminal case.
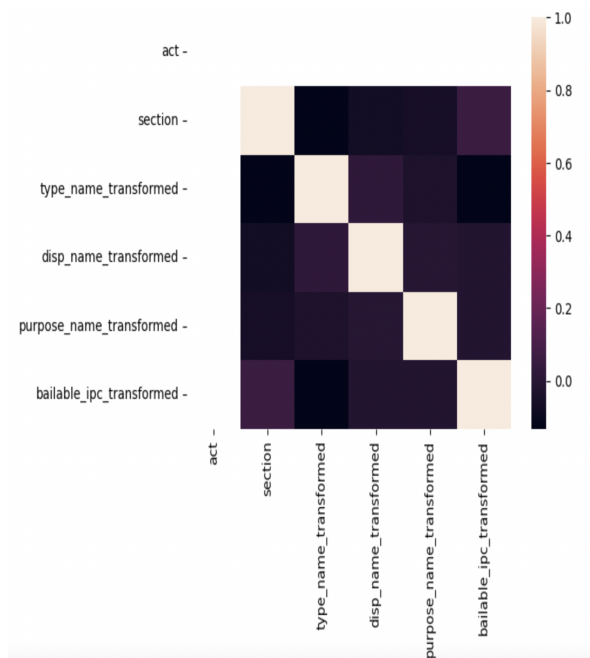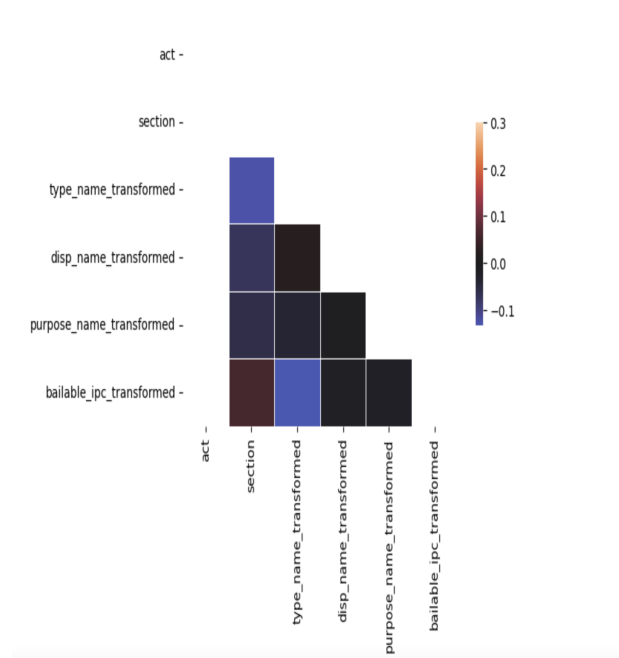

Fig 5. Correlation between various features

Fig 6. Correlation between various features

DECISION TREE CLASSIFIER:

A decision tree classifier is a type of supervised machine learning algorithm used for classification problems. It is represented graphically as a tree-like model where each internal node represents a feature(or attribute), each branch represents a decision based on that feature and each leaf node represents the outcome.

The performance report achieved by using this classifier is as follows:

```
                          precision    recall  f1-score   support

               bailable       1.00      1.00      1.00   2388513
                depends       1.00      1.00      1.00    147416
           non-bailable       1.00      1.00      1.00   1786802
 not applicable to section    1.00      1.00      1.00    172870
   unclear, need to check     1.00      1.00      1.00      5710

               accuracy                           1.00   4501311
              macro avg       1.00      1.00      1.00   4501311
           weighted avg       1.00      1.00      1.00   4501311
```

INSIGHTS:

1. IMPORTANCE OF FEATURES

The importance of features while training the Decision Tree is as follows:

Feature	Importance
Act	0.00000000e+00
Section	9.98864805e-01
Type	4.82389794e-04
Disposition	6.51171779e-04
Purpose	1.63318130e-06

Analysis indicates that the feature "Act" did not contribute significantly (at all) towards forming the decision tree. In contrast, the feature "Section" was found to be the most important feature, followed by "Disposition".

2. SCORE: 0.999998667054998

The decision tree classifier achieved a high score of 0.999998667054998 when tested. This high score can likely be attributed to the large size of the dataset used, as the decision tree had already been trained on a significant amount of data. This prior knowledge allowed the decision tree to accurately predict the outcome when tested on similar data, resulting in the high score.

3. LARGE SIZED DECISION TREE

The decision tree formed from the classifier's training on the dataset of all the years has a large number of branches and leaf nodes. This does not guarantee that the model will perform well when evaluated on completely unseen data. The classifier's high performance is likely due to the fact that the decision tree has a separate branch for each unique value of the features, allowing it to accurately predict the outcome based on the specific feature values. However, having a large number of branches and leaf nodes can also indicate overfitting, meaning the model has learned the training data too well and may not generalize well to new unseen data. The tree can be viewed in a graphical as well as a text representation.

CONCLUSION

In summation, the present study aimed to classify the bailability of criminal cases through an examination of various features such as the Act, Section, Type of Case, Disposition, and Purpose of the case. A decision tree classifier was implemented to perform this analysis, and the results demonstrated that the feature of "Section" was the most salient in determining the bailability of a case, followed by the "Disposition" feature. The decision tree classifier exhibited an impressive score of 0.999998667054998 when tested on the dataset. This can be attributed to the extensive size of the dataset utilized during the training phase. It must be noted, however, that this high score does not necessarily imply a robust performance on unseen data. Therefore, it is imperative to employ methods such as pruning to avoid any potential overfitting.