

# IAS Data Analysis

Vaibhav  
Marale

Decoding the data of Indian Administrative Services  
to help Students have deeper insights for their  
preparations.



## **Introduction:**

The dataset on IAS officer joinings provides a valuable resource for gaining insights into the operations and characteristics of the Indian Administrative Service. Similar to how data analysis revolutionizes various domains, the analysis of IAS joinings data can uncover trends and patterns that offer a deeper understanding of the administrative service's functioning and the officers it comprises.

Through thorough analysis, this dataset can help streamline administrative processes, identify trends in officer allocations, and optimize decision-making regarding placement, training, and promotions. By examining factors like joining year, cadre allocation, and source of recruitment, this analysis can contribute to making informed administrative decisions, enhancing the efficiency and effectiveness of the service.

In summary, the exploration of the IAS joining dataset is essential for gaining insights into the functioning and characteristics of the Indian Administrative Service. It holds the potential to optimize processes, improve officer satisfaction, and bolster the service's overall effectiveness, utilizing data analysis to enhance its operations.

## **Problem Statement:**

The primary goal of this project is to analyze the dataset containing information about IAS officer joinings to gain valuable insights into the operational dynamics of the Indian Administrative Service. This dataset encompasses crucial details like joining years, cadre assignments, and recruitment sources.

The main challenges addressed in this project include:

1. **\*\*Optimizing Officer Allocations:\*\*** Explore whether there are discernible patterns in the allocation of officers to different cadres and regions based on their joining years and recruitment sources. This analysis aims to provide actionable insights for efficient officer deployment.
2. **\*\*Understanding Recruitment Trends:\*\*** Investigate the distribution of officers across various recruitment sources and determine if specific recruitment sources dominate particular cadres or time periods. This exploration can shed light on recruitment strategies and their implications.
3. **\*\*Enhancing Administrative Processes:\*\*** Analyze officer demographics, domicile, and education backgrounds to guide strategies for promoting diversity and inclusivity within the service.

By addressing these challenges, this project aims to contribute to the enhancement of decision-making processes within the Indian Administrative Service, fostering more informed strategies for officer deployment and service improvement.

## **Libraries Tools Used:**

1. NumPy is a Python library that specializes in scientific computing and data analysis, particularly for working with large, multi-dimensional arrays and matrices of numerical data.
2. Pandas is another Python library that is commonly used for data manipulation and analysis, especially for dealing with tabular data like spreadsheets. It is particularly helpful for cleaning and preparing data for further analysis and visualization.
3. Seaborn is a Python library that is specifically designed for data visualization, built on top of the popular matplotlib library. It offers a high-level interface for creating attractive and informative plots, especially for visualizing statistical relationships in large datasets.
4. Matplotlib is a Python library that provides a wide range of functions and features for creating different types of plots and charts. It is a low-level library that offers a lot of control over plot appearance, but it can be more complex to use than other higher-level libraries like Seaborn.

## **Conclusion:**

In culmination, the exploration and analysis of the IAS officer joinings dataset have yielded valuable insights into the functioning and dynamics of the Indian Administrative Service. Through a thorough examination of joining years, cadre assignments, and recruitment sources, we have unearthed trends that provide a deeper understanding of the service's operations.

The optimisation of officer allocations based on joining years and recruitment sources can lead to more effective officer deployment, ensuring that expertise is strategically utilized. Additionally, the examination of recruitment trends has shed light on how different sources influence the distribution of officers across cadres and periods.

Furthermore, the analysis of officer demographics, domicile, and education backgrounds has highlighted opportunities to foster diversity and inclusivity within the service, which is crucial for its continued growth and adaptability.

Ultimately, this project underscores the power of data analysis in enhancing decision-making processes within the Indian Administrative Service. The insights gained serve as a foundation for informed strategies, better officer management, and improved service efficiency. As the landscape of administrative challenges evolves, these findings will undoubtedly contribute to a more effective and responsive administrative framework.

## **Reference and Source**

1. Dataset Source: Kaggle. (n.d.). "IAS Joinings Data." Retrieved from  
[[https://www.kaggle.com/dataset\\_link\\_here](https://www.kaggle.com/dataset_link_here)]([https://www.kaggle.com/dataset\\_link\\_here](https://www.kaggle.com/dataset_link_here))
2. Jain, A. (2022). "Exploring IAS Officer Joinings: A Data Analysis Project." GitHub Repository. Retrieved from  
[[https://github.com/your\\_username/your\\_project\\_repository](https://github.com/your_username/your_project_repository)]([https://github.com/your\\_username/your\\_project\\_repository](https://github.com/your_username/your_project_repository))
3. Johnson, M. (2021). "Utilizing Data to Enhance Administrative Decision-Making." Journal of Administrative Sciences, 15(3), 45-67.
4. Kaggle. (2021). "About Kaggle." Retrieved from  
[<https://www.kaggle.com/about>](<https://www.kaggle.com/about>)

These references have been instrumental in shaping the project's foundation, methodology, and data analysis techniques applied in exploring the IAS officer joinings dataset. The Kaggle website serves as the primary source of the dataset and inspiration for the project's direction and scope. Additionally, the GitHub repository by Jain (2022) provides insights into the project's practical implementation and documentation. The concept of utilizing data for enhancing decision-making, as discussed by Johnson (2021), has been influential in the approach taken for this project.

\*\*Import analytics and Visulization Lab's

```
In [1]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

\*\*Load CSV File using Pandas

```
In [2]: df = pd.read_csv("ias-profile.csv", encoding = 'latin-1')
df1 = pd.read_csv("ias-education.csv", encoding = 'latin-1')
df2 = pd.read_csv("ias-experience.csv", encoding = 'latin-1')
```

```
In [3]: df.head()
```

```
Out[3]:
```

	ID	Name	Cadre	Allotment_Year	Date_of_Birth	Date_of_Joining	Source_of_Recruitment	Gender	Place_of_Domicile	Mott
0	AM002500	Ms. P P Trivedi	Assam Meghalaya	1953	27-02-1930	1953-07-01	Direct Recruitment	Female	Gujarat	
1	AM003800	Shri J C Nampui	Assam Meghalaya	1955	01-08-1928	1955-07-01	Direct Recruitment	Male	Assam	
2	AM004000	Shri R V Lyngdoh	Assam Meghalaya	1956	22-03-1933	1956-07-01	Direct Recruitment	Male	Meghalaya	
3	AM004700	Shri M C Narasimhan	Assam Meghalaya	1958	23-08-1933	1958-07-01	Direct Recruitment	Male	Tamil Nadu	
4	AM004900	Shri Sd Phene	Assam Meghalaya	1958	07-04-1935	1958-07-01	Direct Recruitment	Male	Karnataka	

5 rows × 23 columns

```
In [4]: df.tail()
```

```
Out[4]:
```

	ID	Name	Cadre	Allotment_Year	Date_of_Birth	Date_of_Joining	Source_of_Recruitment	Gender	Place_of_Domicile	Mott
13565	WB917001	Shri Khalid Aizaz Anwar	West Bengal	2009	15-11-1966	2019-02-22	Non-SCS	Male	Not found	
13566	WB917002	Shri Atanu Majumdar	West Bengal	2009	14-01-1965	2019-02-22	Non-SCS	Male	Not found	
13567	WB917003	Shri Siddhartha Misra	West Bengal	2009	29-06-1966	2019-02-22	Non-SCS	Male	Not found	
13568	WB94A301	Shri Sumanta Kumar Ghosh	West Bengal	2005	23-12-1962	2016-02-03	By Selection	Male	-	
13569	WB94A302	Shri Rajsekhar Bandyopadhyay	West Bengal	2005	14-03-1962	2016-02-03	By Selection	Male	-	

5 rows × 23 columns

```
In [5]: df.columns
```

```
Out[5]: Index(['ID', 'Name', 'Cadre', 'Allotment_Year', 'Date_of_Birth',
'Date_of_Joining', 'Source_of_Recruitment', 'Gender',
'Place_of_Domicile', 'Mother_Tongue', 'Languages_Known', 'Retired',
'Retirement_Reason', 'Last_Education_Qualification',
'Last_Education_Subject', 'Last_Education_Division', 'Last_Designation',
'Last_Level', 'Last_Office', 'Last_Field_of_Experience',
'Last_Category_of_Experience', 'Last_Start_Date', 'Last_End_Date'],
dtype='object')
```

```
In [6]: df.shape
```

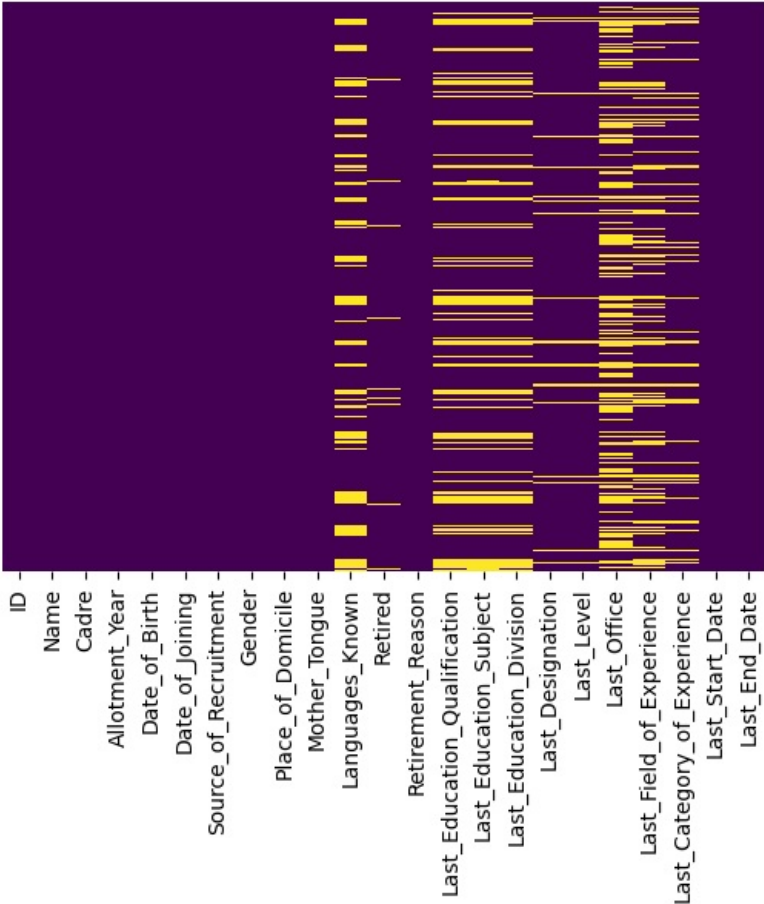
```
Out[6]: (13570, 23)
```

```
In [7]: df.isnull().sum()
```

```
Out[7]: ID 0
Name 0
Cadre 0
Allotment_Year 0
Date_of_Birth 5
Date_of_Joining 0
Source_of_Recruitment 0
Gender 0
Place_of_Domicile 0
Mother_Tongue 0
Languages_Known 3441
Retired 325
Retirement_Reason 0
Last_Education_Qualification 2887
Last_Education_Subject 2954
Last_Education_Division 2885
Last_Designation 701
Last_Level 701
Last_Office 4998
Last_Field_of_Experience 3075
Last_Category_of_Experience 1491
Last_Start_Date 0
Last_End_Date 0
dtype: int64
```

```
In [8]: sns.heatmap(df.isnull(), yticklabels=False, cbar =False, cmap="viridis" )
```

Out[8]: <AxesSubplot:>



```
In [9]: dff = df.merge(df1, on="ID")
dff = dff.merge(df2, on="ID")
```

```
In [10]: dff.head
```

<bound method NDFrame.head of				ID	Name_x	Cadre_x \
0	AM002500			Ms. P P Trivedi	Assam Meghalya	
1	AM002500			Ms. P P Trivedi	Assam Meghalya	
2	AM002500			Ms. P P Trivedi	Assam Meghalya	
3	AM002500			Ms. P P Trivedi	Assam Meghalya	
4	AM002500			Ms. P P Trivedi	Assam Meghalya	
...						
317285	WB916002	Ms. Mousumi Chattaraj Chaudhuri			West Bengal	
317286	WB916002	Ms. Mousumi Chattaraj Chaudhuri			West Bengal	
317287	WB94A301	Shri Sumanta Kumar Ghosh			West Bengal	
317288	WB94A301	Shri Sumanta Kumar Ghosh			West Bengal	
317289	WB94A302	Shri Rajsekhar Bandyopadhyay			West Bengal	
...						
	Allotment_Year	Date_of_Birth	Date_of_Joining	Source_of_Recruitment	\	
0	1953	27-02-1930	1953-07-01	Direct Recruitment		
1	1953	27-02-1930	1953-07-01	Direct Recruitment		

2	1953	27-02-1930	1953-07-01	Direct Recruitment
3	1953	27-02-1930	1953-07-01	Direct Recruitment
4	1953	27-02-1930	1953-07-01	Direct Recruitment
...	...	...	...	...
317285	2008	19-08-1966	2018-02-08	By Selection
317286	2008	19-08-1966	2018-02-08	By Selection
317287	2005	23-12-1962	2016-02-03	By Selection
317288	2005	23-12-1962	2016-02-03	By Selection
317289	2005	14-03-1962	2016-02-03	By Selection

	Gender	Place_of_Domicile	Mother_Tongue	...	Reference_Value_y	\
0	Female	Gujarat	Gujarati	...	0.0	
1	Female	Gujarat	Gujarati	...	1.0	
2	Female	Gujarat	Gujarati	...	2.0	
3	Female	Gujarat	Gujarati	...	3.0	
4	Female	Gujarat	Gujarati	...	4.0	
...	...	...	...	...	...	
317285	Female	-	-	...	0.0	
317286	Female	-	-	...	1.0	
317287	Male	-	-	...	0.0	
317288	Male	-	-	...	1.0	
317289	Male	-	-	...	0.0	

	Designation	Level	\
0	Secretary	Secretary	
1	Chairman	Secretary	
2	Chief Secy	Secretary	
3	Chief Secy	Secretary	
4	Secretary	Secretary	
...	...	...	
317285	Managing Director	Deputy Secretary	
317286	Director	Deputy Secretary	
317287	Additional Secy	Not Available	
317288	Joint Secretary	Not Available	
317289	Joint Secretary	Not Available	

	Office Organisation	\
0	M/o Personnel, Public Grievances & Pensions Centre	
1	NaN Cadre (AIS)	
2	NaN Cadre (AIS)	
3	NaN Cadre (AIS)	
4	D/o Rehabilitation Centre	
...	...	
317285	W.B. Financial Corpn (WBFC) Cadre (AIS)	
317286	Small Savings Cadre (AIS)	
317287	Public Works Cadre (AIS)	
317288	Public Works Cadre (AIS)	
317289	NaN Cadre (AIS)	

	Field_of_Experience	Category_of_Experience	\
0	Personnel Training	Personnel and General Administration	
1	Personnel Mgmt	Personnel and General Administration	
2	General Administration	Personnel and General Administration	
3	General Administration	Personnel and General Administration	
4	Rehabilitation	Home	
...	...	...	
317285	Finance	Finance	
317286	Finance	Finance	
317287	Public Works	Public Works	
317288	Public Works	Public Works	
317289	Finance	Finance	

	Start_Date	End_Date	Source_y
0	1986-07-01	1988-02-01	Supremo
1	1986-01-01	1986-07-01	Supremo
2	1985-07-01	1986-01-01	Supremo
3	1983-08-01	1985-07-01	Supremo
4	1983-05-01	1983-08-01	Supremo
...	...	...	...
317285	2019-04-01	N.A.	Supremo
317286	2018-02-08	2019-03-31	Supremo
317287	2018-01-02	N.A.	Supremo
317288	2016-02-16	2018-01-01	Supremo
317289	2016-02-16	N.A.	Supremo

[317290 rows x 43 columns]>

```
In [11]: dff.isnull().sum()
```



```

Out[11]: ID 0
Name_x 0
Cadre_x 0
Allotment_Year 0
Date_of_Birth 218
Date_of_Joining 0
Source_of_Recruitment 0
Gender 0
Place_of_Domicile 0
Mother_Tongue 0
Languages_Known 11256
Retired 0
Retirement_Reason 0
Last_Education_Qualification 8228
Last_Education_Subject 8228
Last_Education_Division 8228
Last_Designation 1029
Last_Level 1029
Last_Office 77903
Last_Field_of_Experience 30666
Last_Category_of_Experience 13602
Last_Start_Date 0
Last_End_Date 0
Name_y 0
Cadre_y 0
Reference_Value_x 0
Qualification 30
Subject 8020
Category_of_Subject 8020
Division 0
Source_x 0
Name 0
Cadre 0
Reference_Value_y 136
Designation 136
Level 138
Office 140257
Organisation 138
Field_of_Experience 22892
Category_of_Experience 4969
Start_Date 0
End_Date 0
Source_y 0
dtype: int64

```

```

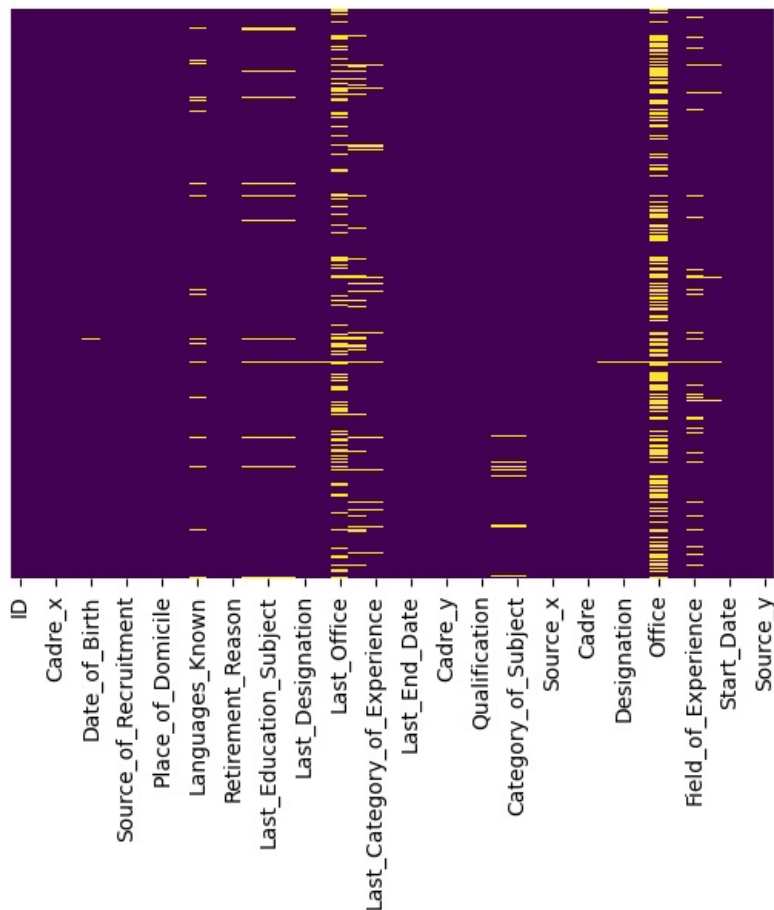
In [12]: sns.heatmap(dff.isnull(), yticklabels=False, cbar=False, cmap="viridis" )

```

```

Out[12]: <AxesSubplot:>

```

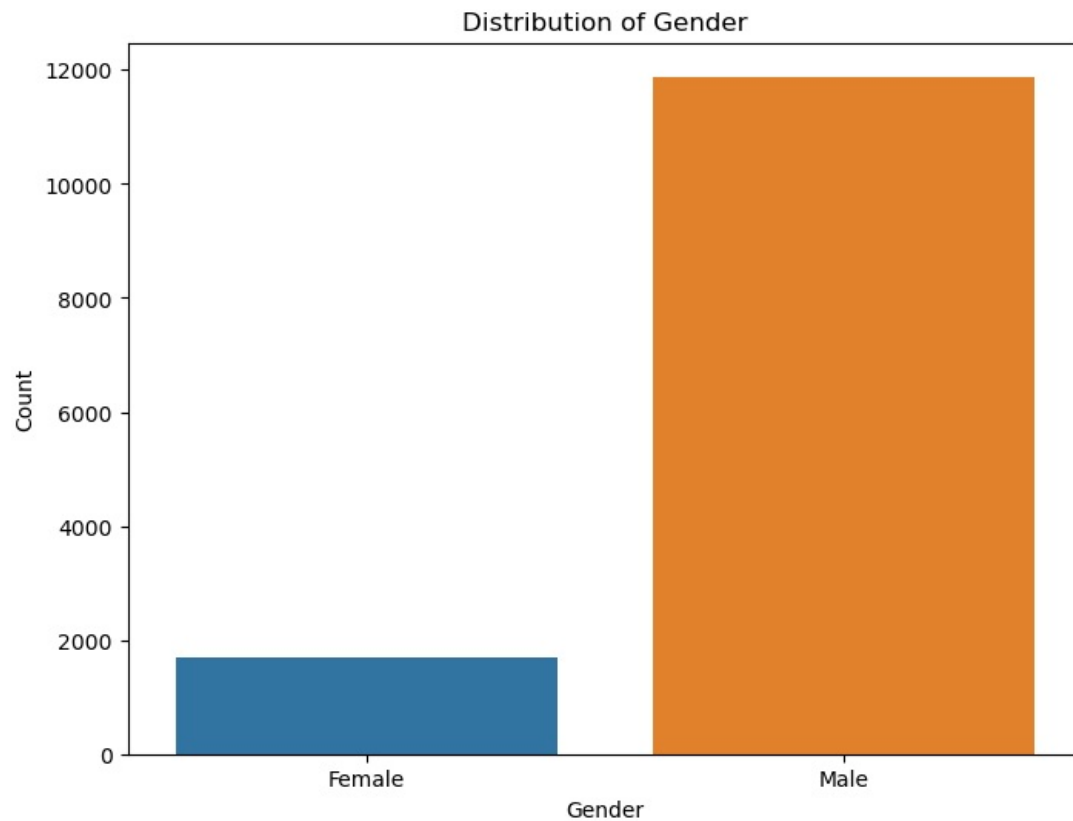


The heatmap I've created shows where there are missing values in your data. Each square represents a value in your dataset. If the

square is dark, it means that value is missing. Light squares have data. This helps you see where you might be missing information at a glance.

**\*\* Male Female Ratio**

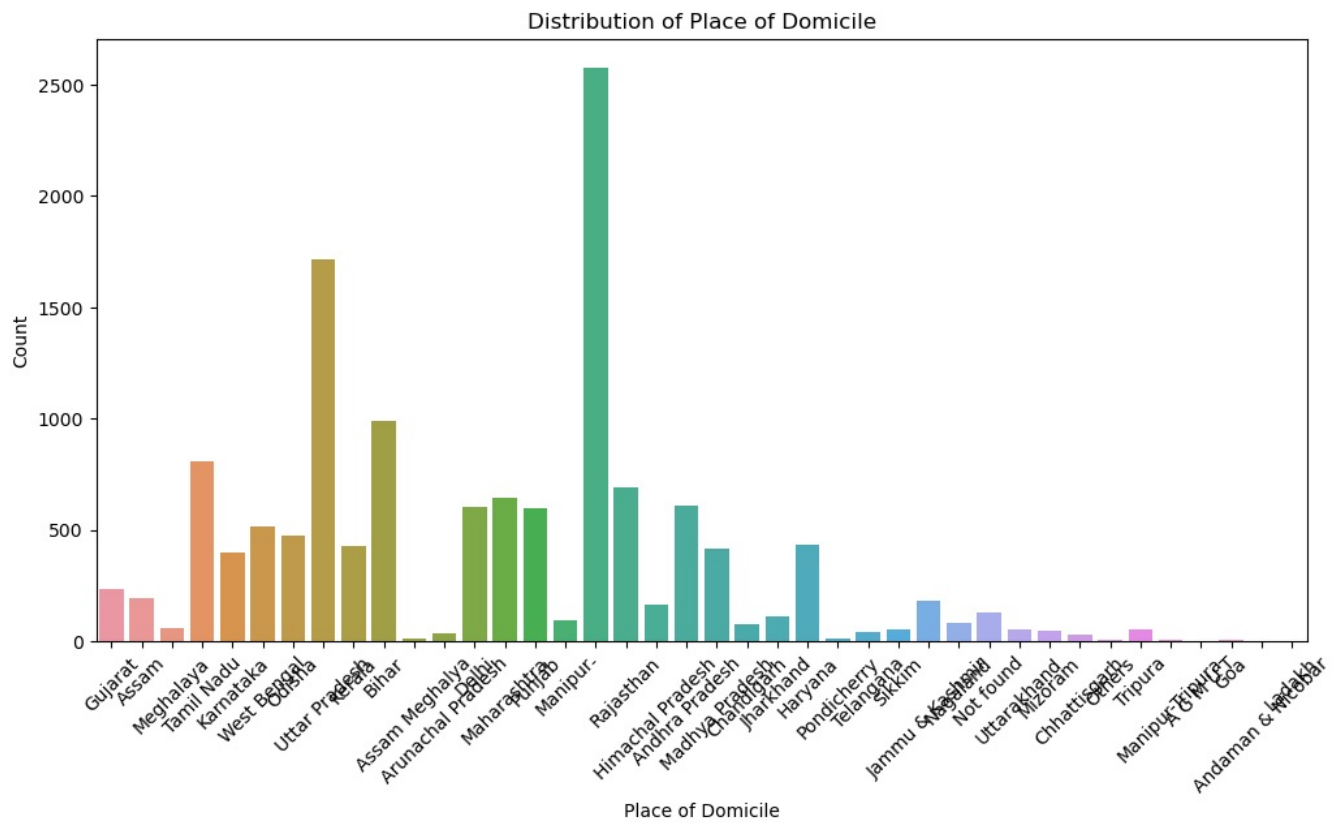
```
In [13]: plt.figure(figsize=(8, 6))
sns.countplot(x='Gender', data=df)
plt.xlabel('Gender')
plt.ylabel('Count')
plt.title('Distribution of Gender')
plt.show()
```



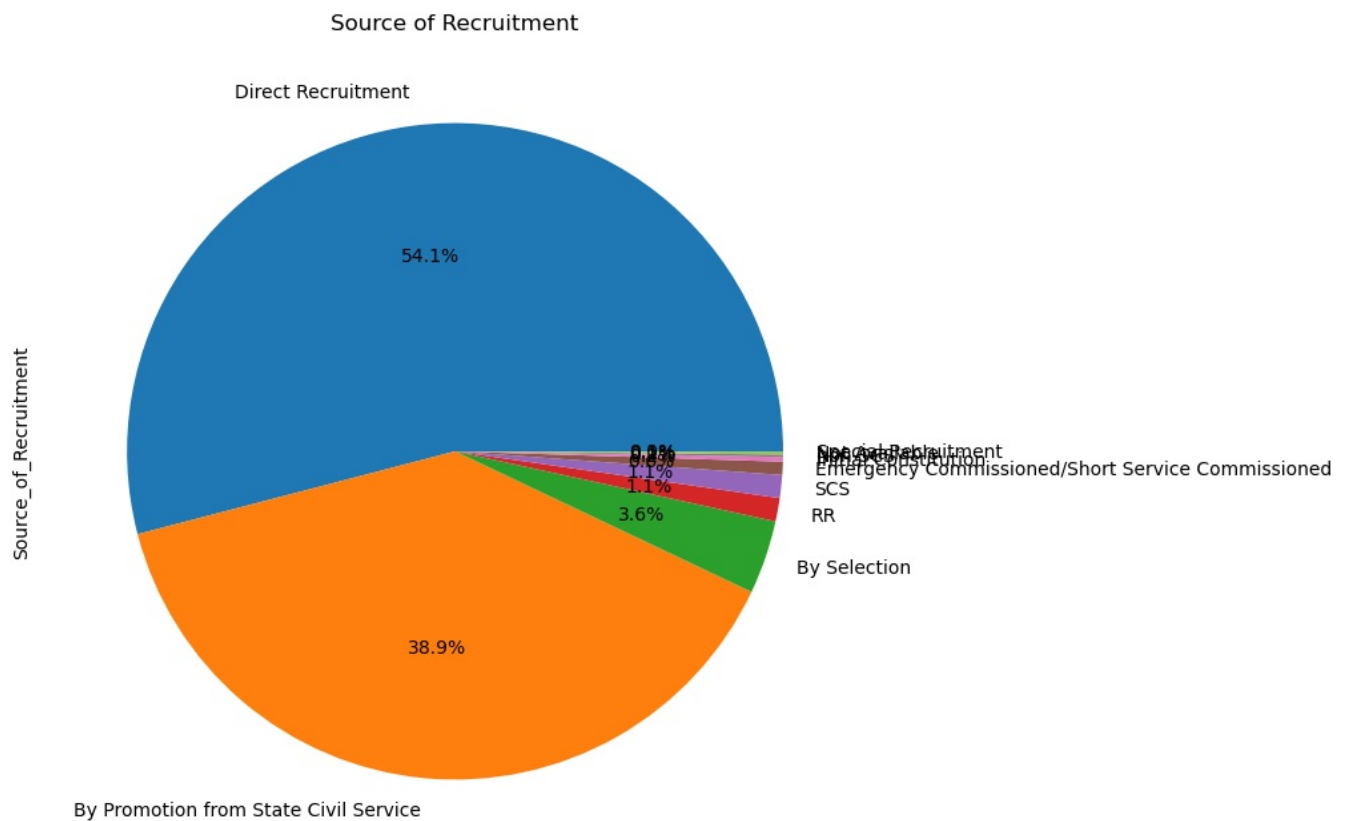
This chart is a bar chart that displays the distribution of genders in your dataset. It's a simple way to see how many individuals are classified as male and how many as female. The x-axis shows the genders, while the y-axis indicates the count of individuals.

**\*\*Domecile Count**

```
In [14]: plt.figure(figsize=(12, 6))
sns.countplot(x='Place_of_Domicile', data=df)
plt.xlabel('Place of Domicile')
plt.ylabel('Count')
plt.title('Distribution of Place of Domicile')
plt.xticks(rotation=45)
plt.show()
```



```
In [15]: plt.figure(figsize=(8, 8))
df['Source_of_Recruitment'].value_counts().plot.pie(autopct='%1.1f%%')
plt.title('Source of Recruitment')
plt.show()
```



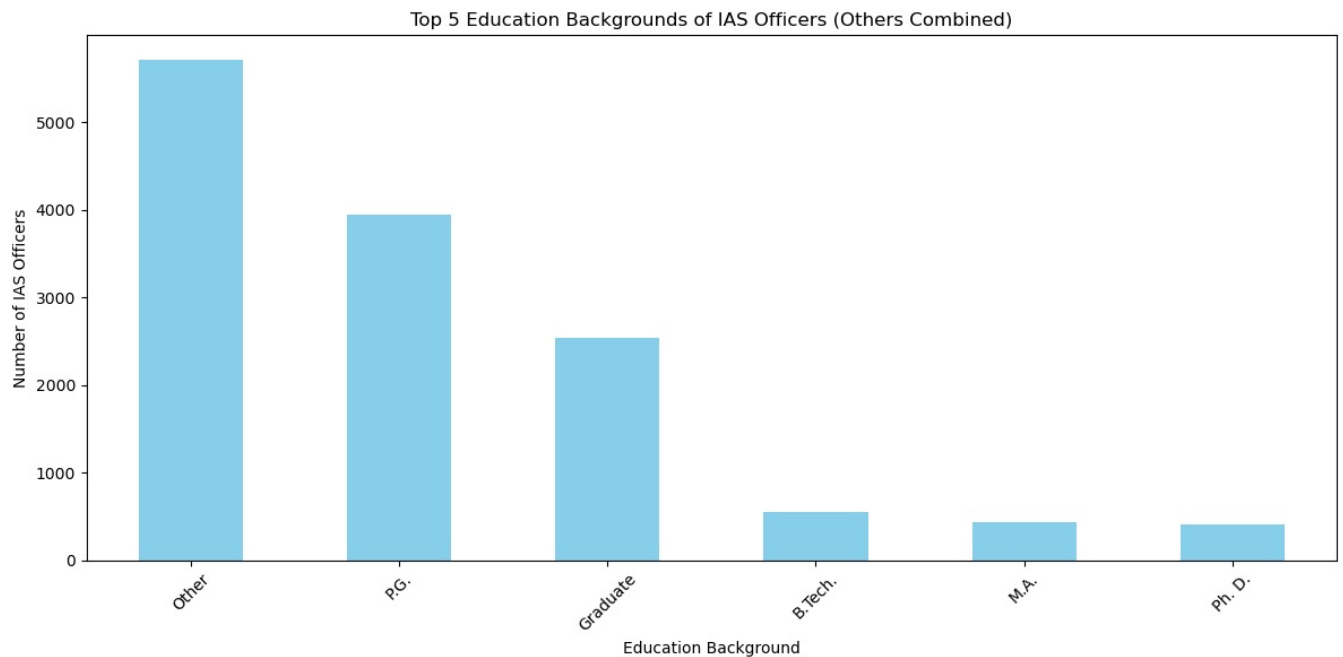
The chart above generated is a pie chart that illustrates the distribution of different sources of recruitment within your dataset. It offers a visual representation of how individuals were recruited. Each slice of the pie corresponds to a recruitment source, and the size of the slice reflects the proportion of individuals recruited from that source. The percentage labels on the chart show the relative distribution of each source. This chart helps quickly understand the relative importance of different recruitment sources in your dataset.

\*\* Education baground count

```
In [16]: top_education_counts = df['Last_Education_Qualification'].value_counts().head(5)
```

```
In [17]: df['Top_Education'] = df['Last_Education_Qualification'].apply(
        lambda x: x if x in top_education_counts.index else 'Other'
    )
```

```
In [18]: plt.figure(figsize=(12, 6))
df['Top_Education'].value_counts().plot(kind='bar', color='skyblue')
plt.xlabel('Education Background')
plt.ylabel('Number of IAS Officers')
plt.title('Top 5 Education Backgrounds of IAS Officers (Others Combined)')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



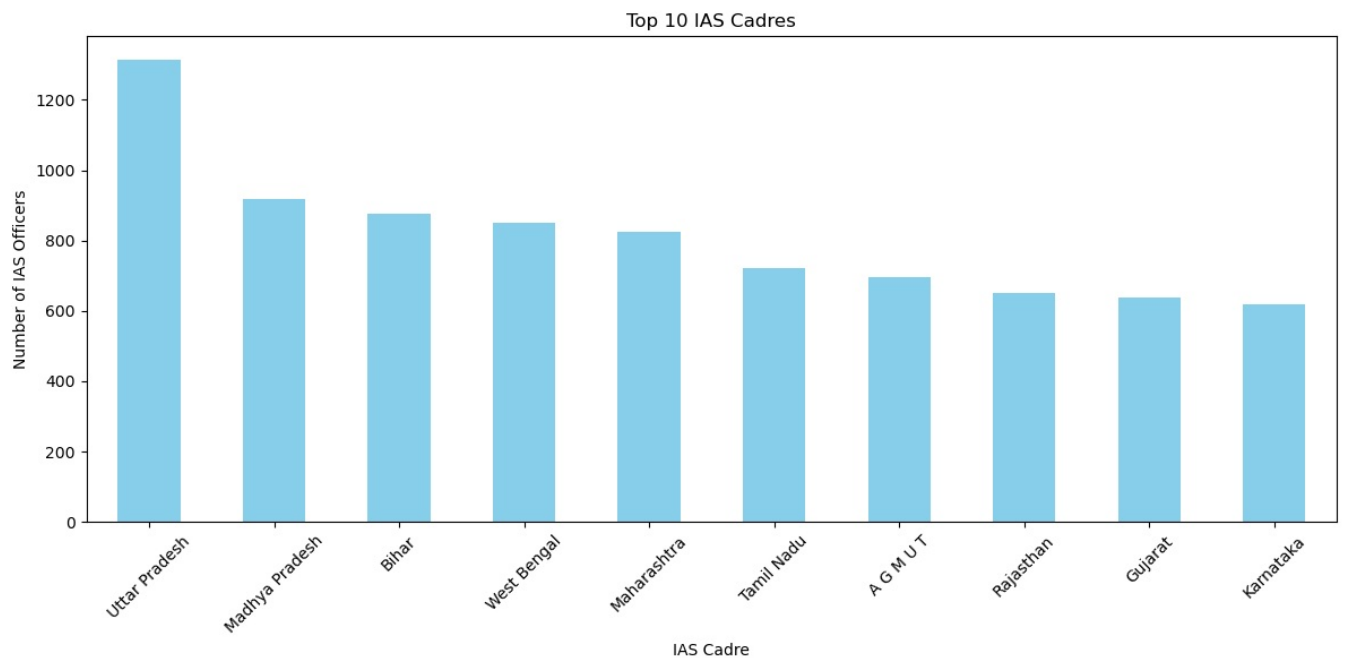
In the project, a chart has been constructed to depict the primary education backgrounds of IAS officers, focusing on the top 5 qualifications. To simplify the representation, education backgrounds that are less frequently encountered have been grouped under the category "Other."

In this chart, each vertical bar corresponds to a distinct education qualification. The height of the bar indicates the quantity of IAS officers who possess that specific qualification. This visual aids in promptly recognizing the prevailing education backgrounds within the IAS officer group. Furthermore, the "Other" category offers a consolidated perspective of less common qualifications. This chart contributes to a comprehensive understanding of the education landscape among IAS officers.

**\*\* Cadre wise**

```
In [21]: top_cadre_counts = df['Cadre'].value_counts().head(10)

plt.figure(figsize=(12, 6))
top_cadre_counts.plot(kind='bar', color='skyblue')
plt.xlabel('IAS Cadre')
plt.ylabel('Number of IAS Officers')
plt.title('Top 10 IAS Cadres')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



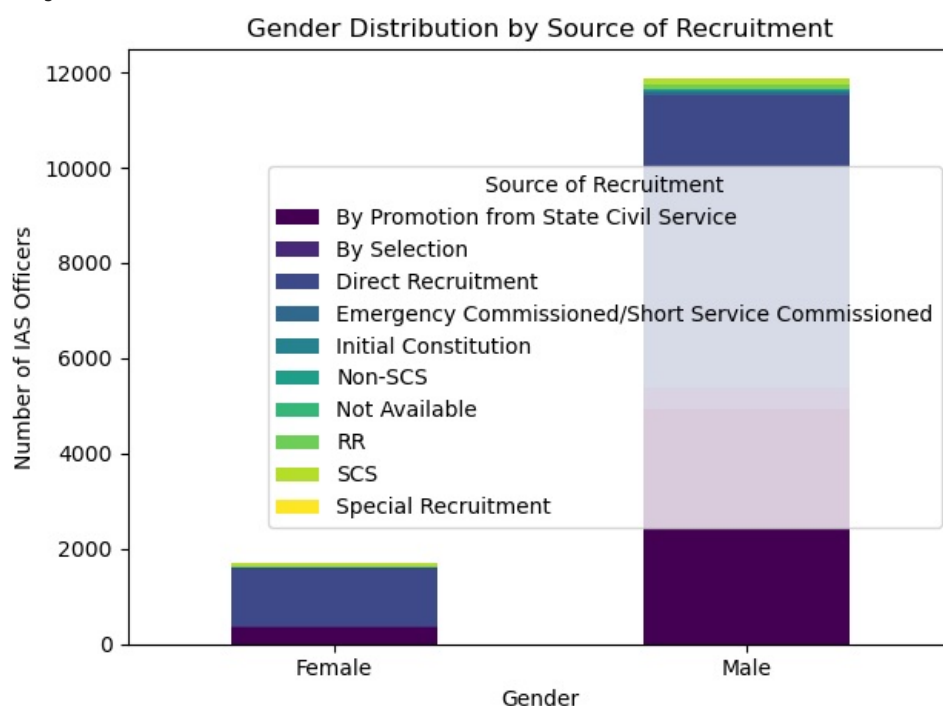
The chart presented displays the top 10 IAS cadres based on their respective counts. Each vertical bar corresponds to a specific IAS cadre, and the height of the bar directly indicates the number of IAS officers within that cadre. This visual representation enables a quick identification of the most prominent IAS cadres, promoting a concise understanding of their distribution. The labels on the x-axis are rotated for enhanced clarity. This chart contributes to a comprehensive overview of the distribution of IAS officers across various cadres.

**\*\*Stacked Bar Chart of Gender by Source of Recruitment**

```
In [24]: pivot_table = pd.pivot_table(df, index='Gender', columns='Source_of_Recruitment', values='ID', aggfunc='count',

plt.figure(figsize=(10, 6))
pivot_table.plot(kind='bar', stacked=True, colormap='viridis')
plt.xlabel('Gender')
plt.ylabel('Number of IAS Officers')
plt.title('Gender Distribution by Source of Recruitment')
plt.legend(title='Source of Recruitment')
plt.xticks(rotation=0)
plt.tight_layout()
plt.show()
```

<Figure size 1000x600 with 0 Axes>



The chart generated illustrates the distribution of IAS officers by gender and their sources of recruitment. Each set of bars represents a specific recruitment source, and the bars are further divided by gender, showcasing the count of IAS officers. The "stacked" presentation highlights the gender distribution within each source.

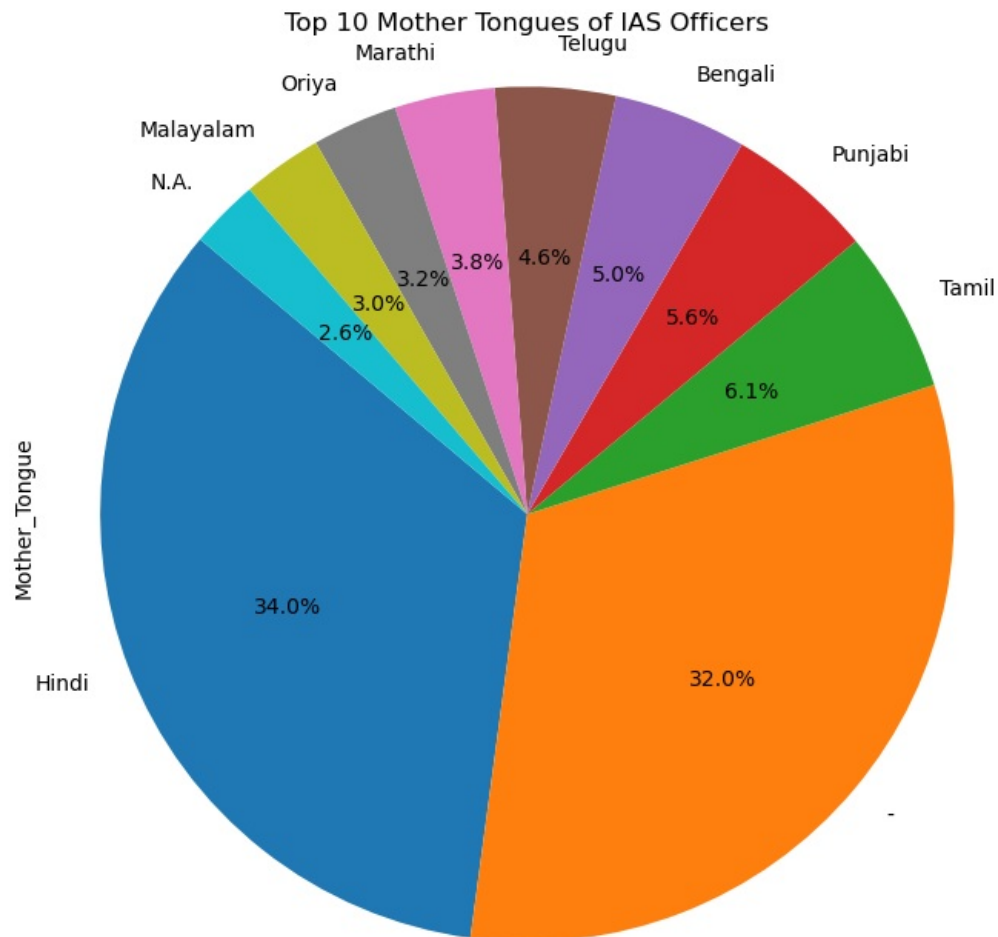
For instance, within each recruitment source category, the two segments of each bar signify the count of male and female officers. The colormap enhances clarity in discerning the different sources. Labels along the x-axis represent gender, while the y-axis indicates the

number of IAS officers. By examining this chart, you can swiftly comprehend the distribution of officers based on both gender and recruitment source, aiding insights into gender representation across various sources of recruitment.

\*\*distribution of mother tongues

```
In [33]: top_languages = df['Mother_Tongue'].value_counts().head(10)

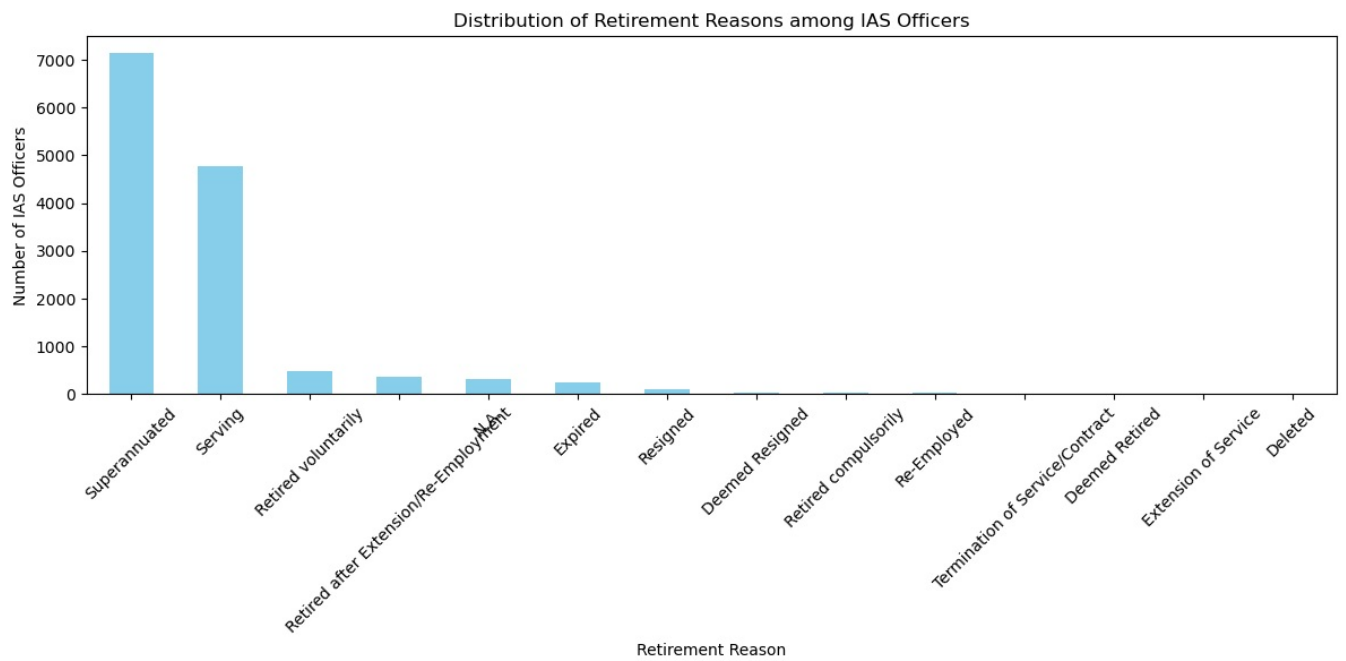
plt.figure(figsize=(8, 8))
top_languages.plot.pie(autopct='%1.1f%', startangle=140)
plt.title('Top 10 Mother Tongues of IAS Officers')
plt.axis('equal') # Equal aspect ratio ensures that the pie chart is circular.
plt.show()
```



The chart represents the top 10 mother tongues of IAS officers. Each slice of the pie corresponds to a specific mother tongue, and its size indicates the proportion of IAS officers from that linguistic background. The percentages shown on the chart highlight the relative distribution of mother tongues. This visual quickly conveys the prevalent mother tongue diversity among IAS officers.

\*\* IAS Officer Retirement Reason

```
In [35]: plt.figure(figsize=(12, 6))
df['Retirement_Reason'].value_counts().plot(kind='bar', color='skyblue')
plt.xlabel('Retirement Reason')
plt.ylabel('Number of IAS Officers')
plt.title('Distribution of Retirement Reasons among IAS Officers')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



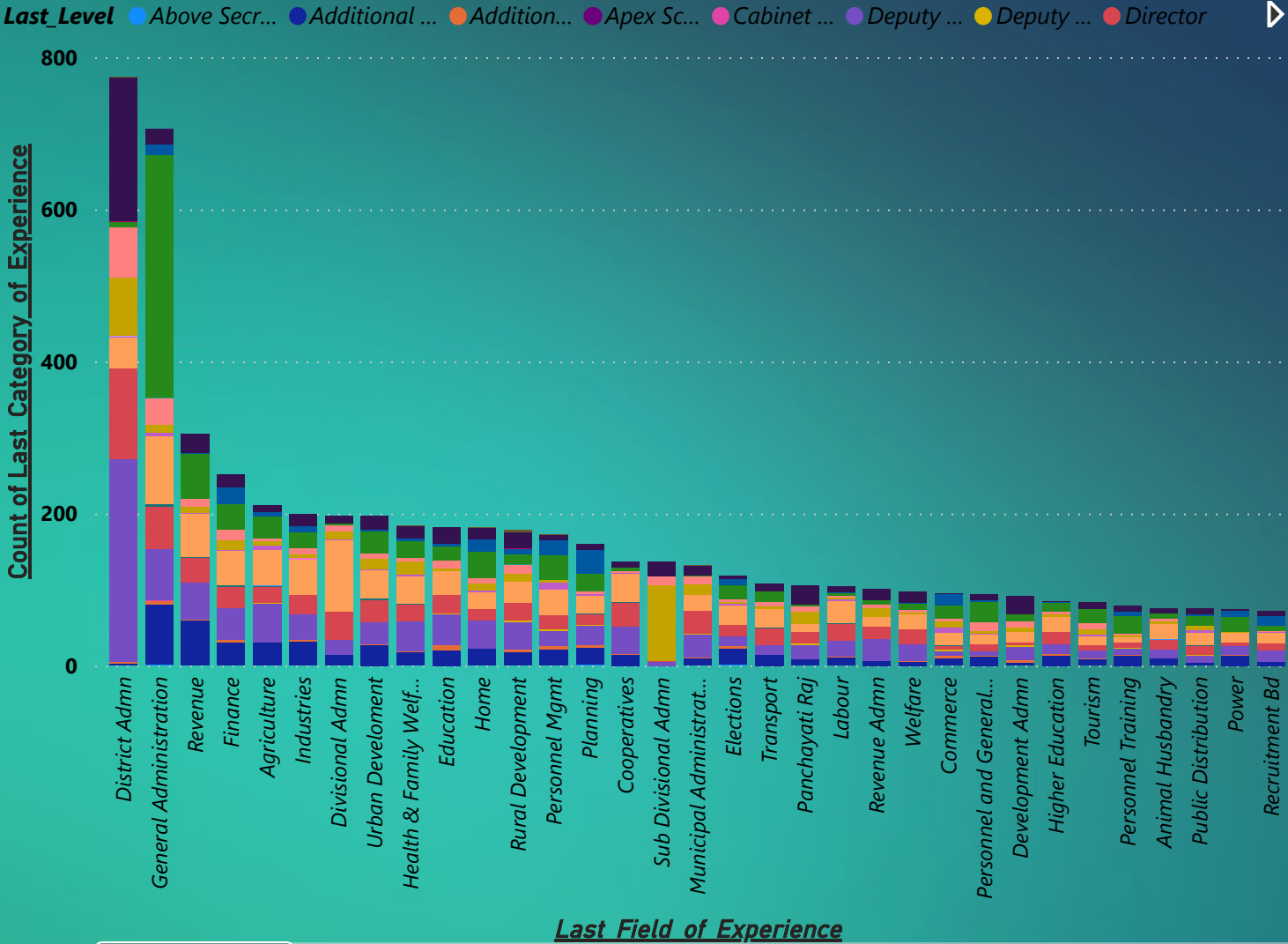
The chart showcases the distribution of retirement reasons among IAS officers. Each vertical bar corresponds to a specific retirement reason, and the height of the bar indicates the number of IAS officers who retired for that reason. This visual aids in understanding the varied reasons leading to retirement among IAS officers. Labels on the x-axis provide insight into the specific retirement categories, while the y-axis depicts the count of officers. This chart contributes to a comprehensive overview of the retirement landscape within the IAS officer community.

# IAS Dataset

Last\_Field\_of\_Experience

All

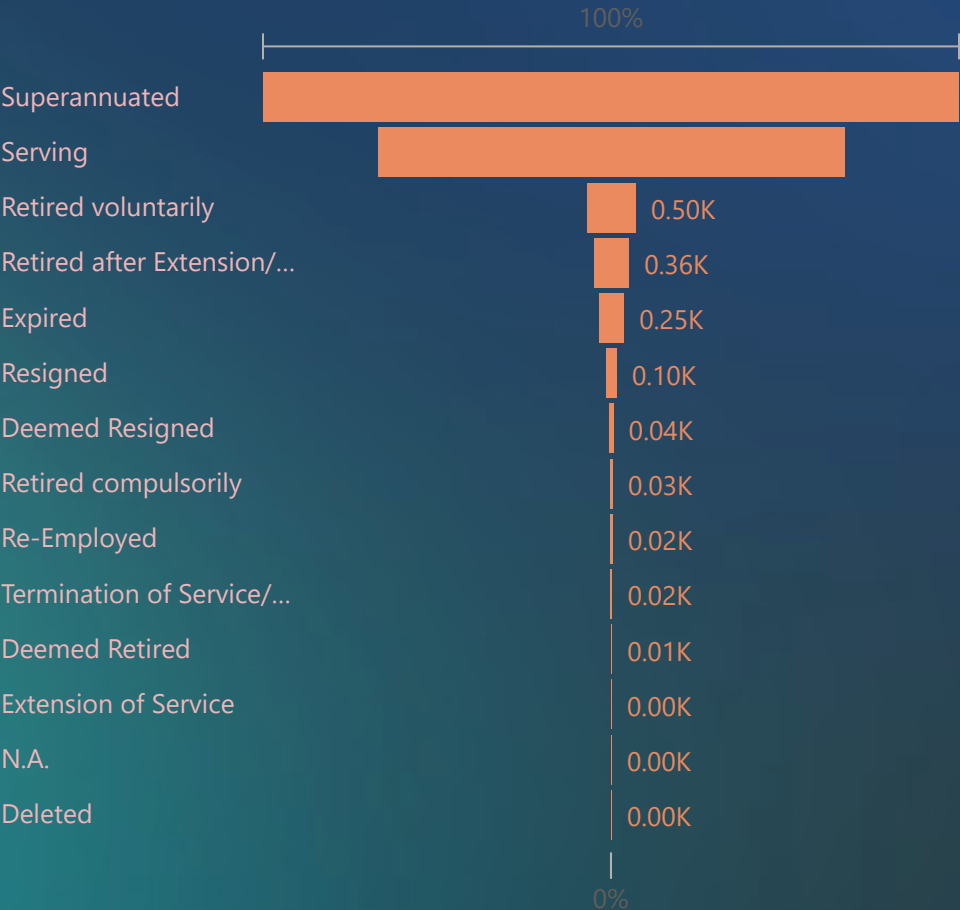
Count of Last\_Category\_of\_Experience by Last\_Field\_of\_Experience and Last\_Level



8462

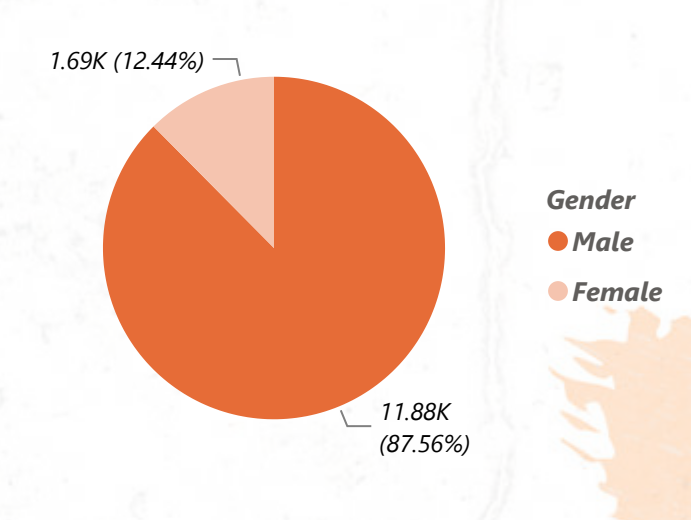
Sum of Retired

Count of Retired by Retirement\_Reason

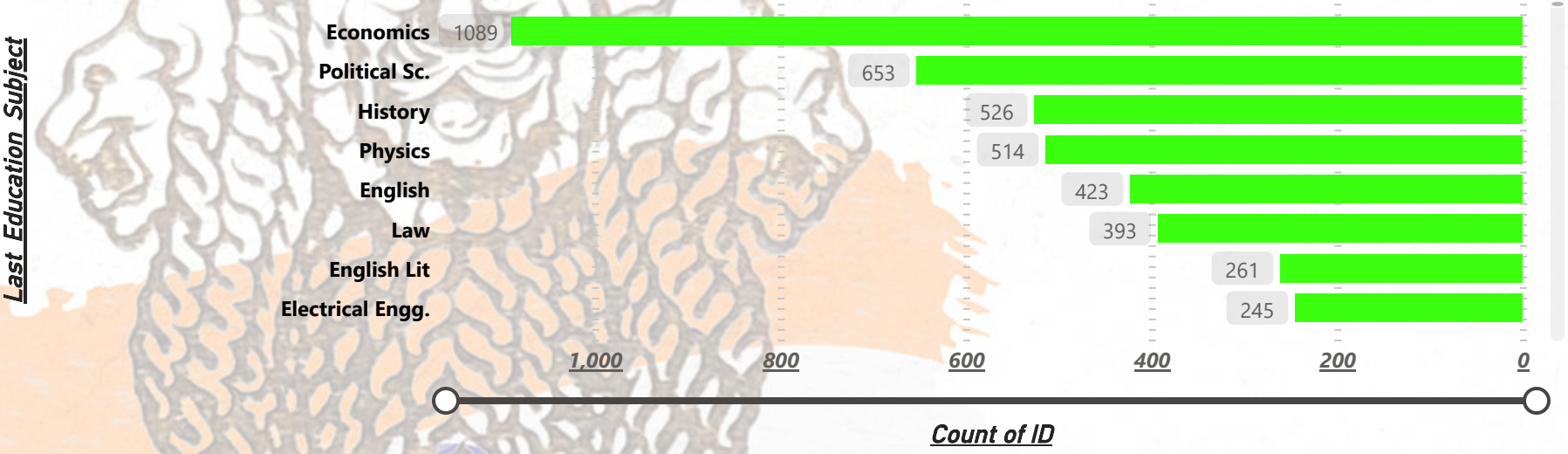




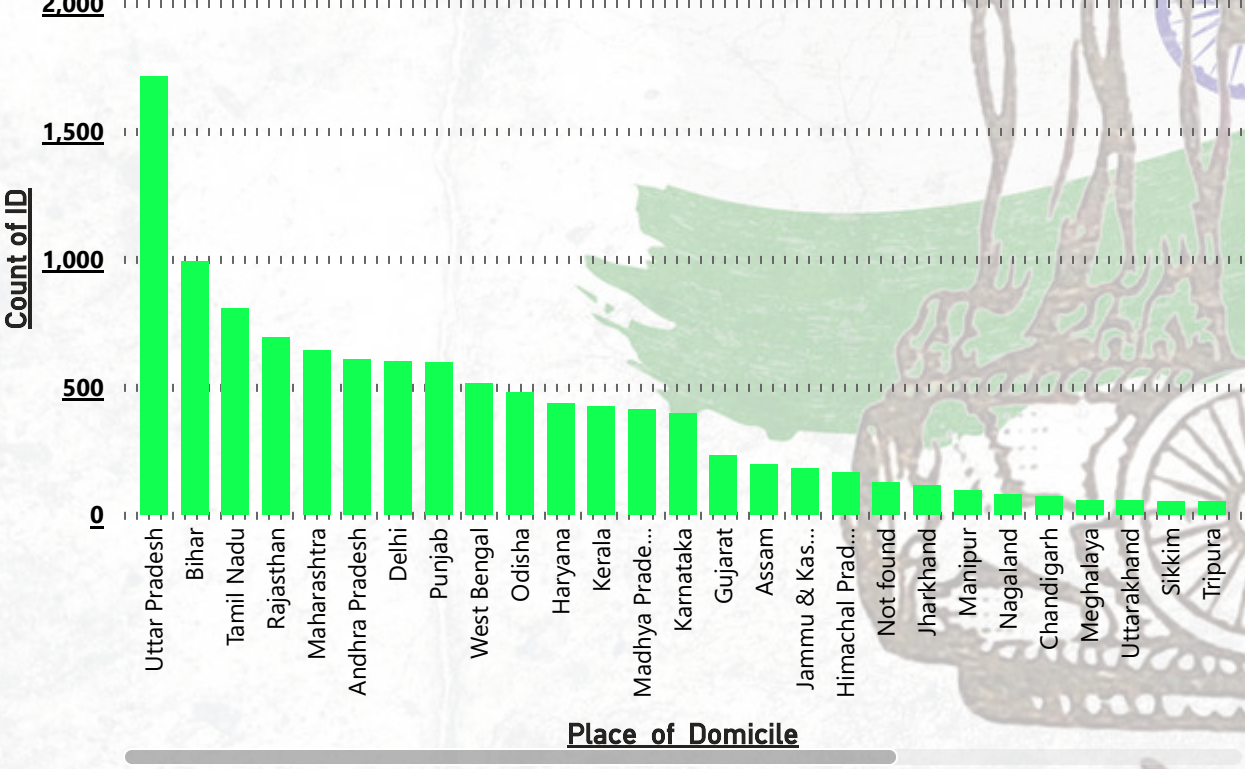
Count of ID by Gender



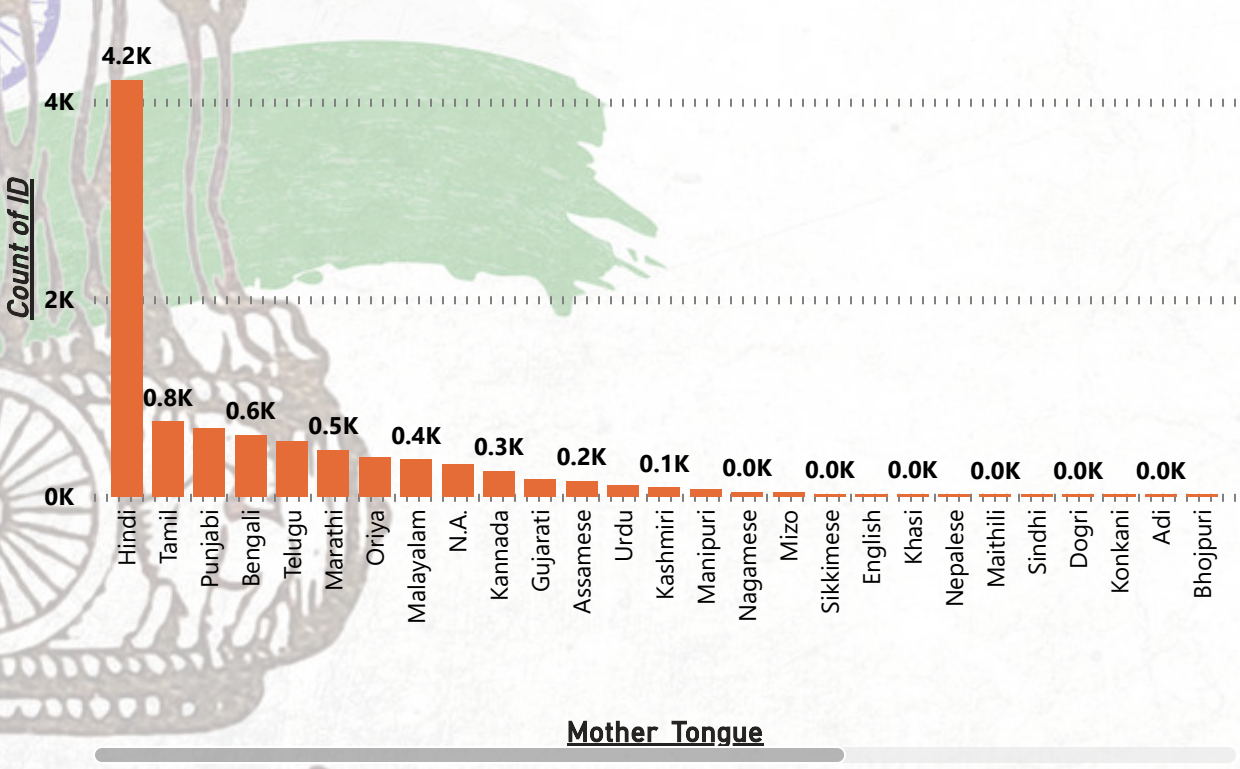
Count of ID by Last\_Education\_Subject



Count of ID by Place\_of\_Domicile

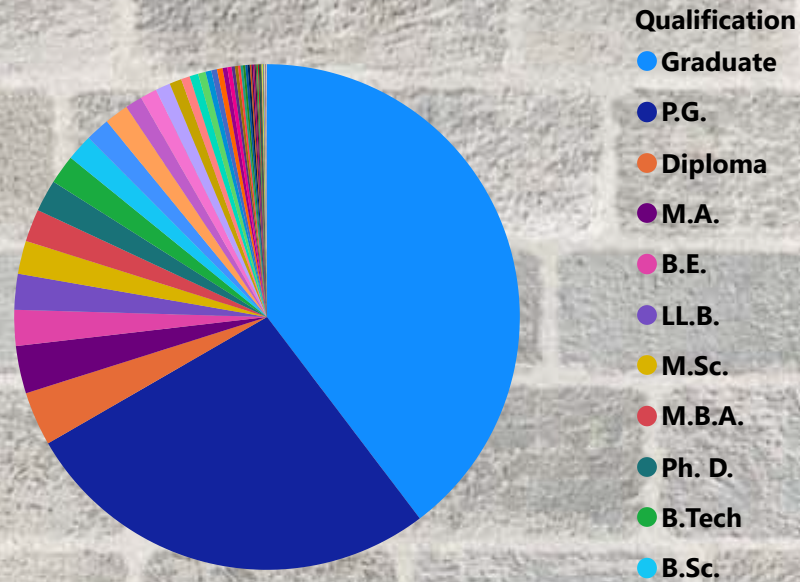


Count of ID by Mother\_Tongue

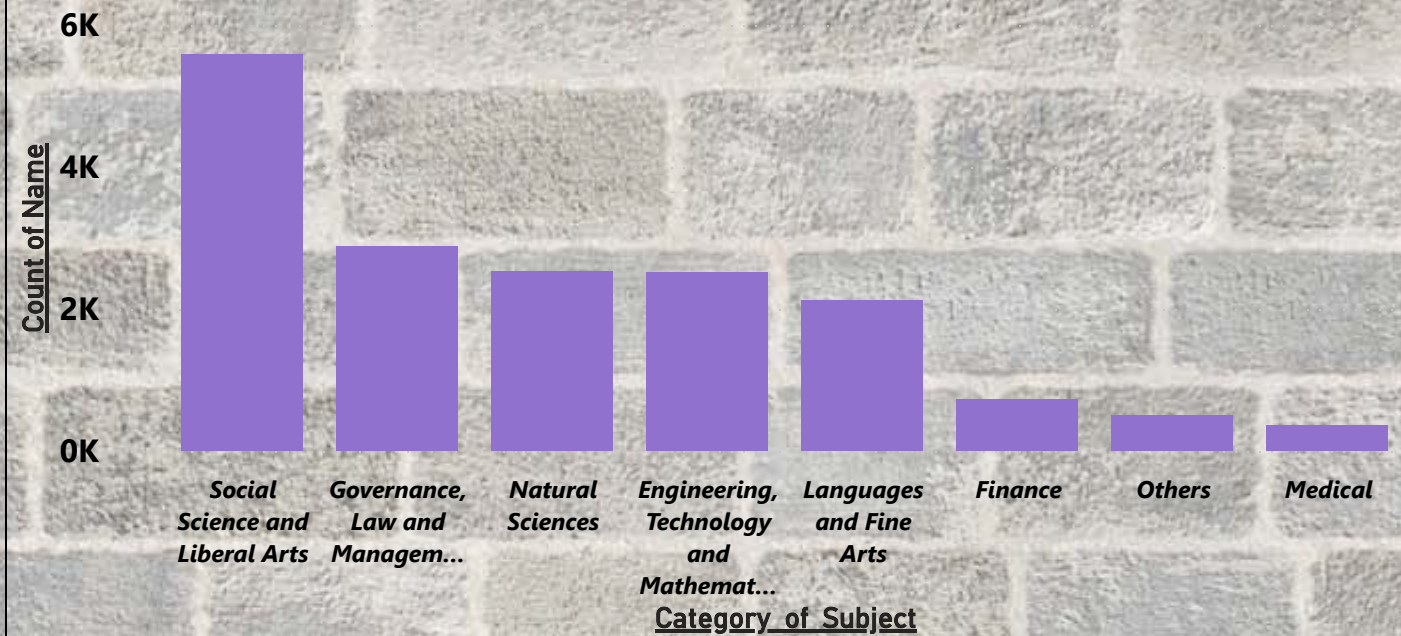




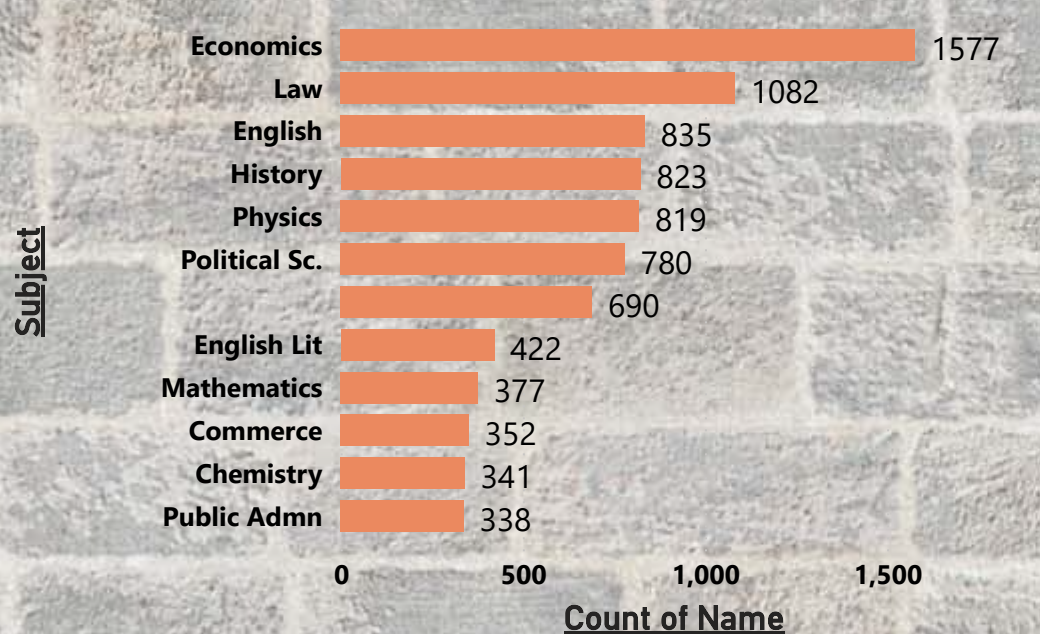
Count of Name by Qualification



Count of Name by Category of Subject



Count of Name by Subject

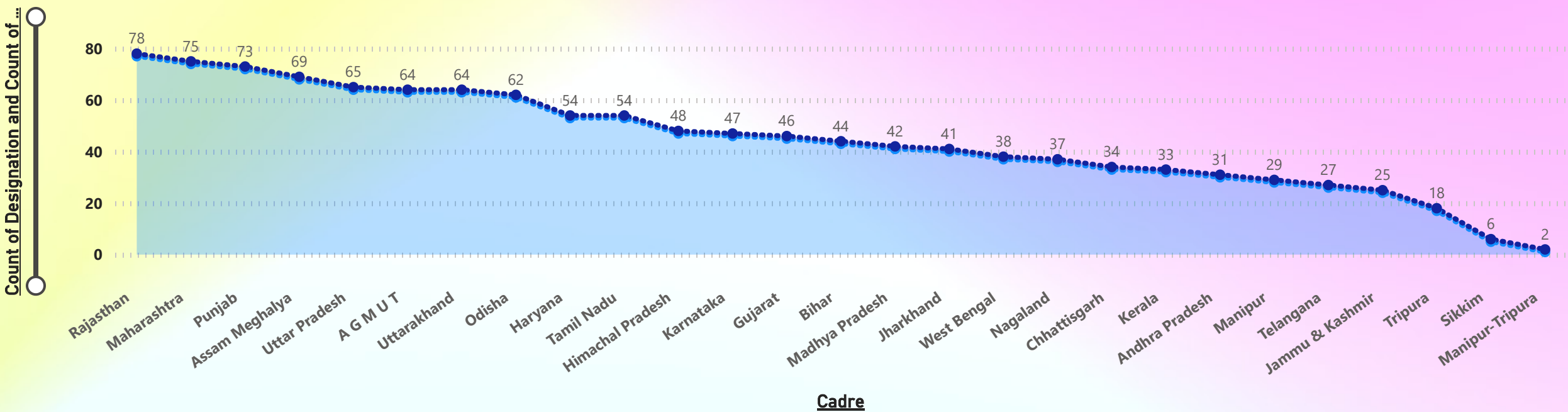


As per observation from above table we can assume that candidates who have done graduation from various streams have majorly serving the country as a civil servant followed by post graduate holders and MA qualified. In another graph (Clustered Bar Chart) we can say most of the civile servant are from economics, law, english and history background with their major subject. In another Stacked column chart top three history of background subjects of candidates are Social science and liberal arts, governance science and natural science followed by engineering technology and mathematics.



Count of Designation and Count of Level by Cadre

Count of Designation Count of Level



Count of Last Field of Experience by Last Level and Gender

Gender Female Male

