DATA ANALYSIS ON INDIRA GANDHI NATIONAL DISABILITY PENSION SCHEME(IGNDPS)

This presentation analyzes data from the Indira Gandhi National Disability Pension Scheme (IGNDPS). The analysis includes data cleaning, visualization, and statistical testing to understand beneficiary distribution, registration coverage, and potential anomalies. The dataset contains information on beneficiaries across various states and districts in India, categorized by social group, Aadhar and mobile registration status, and gender.

Data Overview and Cleaning

The dataset, loaded from "IGNDPS.csv", contains 14,297 entries with 15 columns. The columns include financial year, month, state and district codes, state and district names, scheme code, total beneficiaries, and counts for Scheduled Castes (SC), Scheduled Tribes (ST), General, Other Backward Classes (OBC), Aadhar registrations, and mobile number registrations.

- **Data Types:** The dataset includes integer and object data types.

	fin_year	mnth	lgd_state_code	state_name	lgd_district_code	district_name	scheme_code	total_beneficiaries	total_sc	total_st	total_gen	total_obc	total_aad
0	2024-2025	6	1	JAMMU AND KASHMIR	1	ANANTNAG	IGNDPS	104	0	2	101	1	
1	2024-2025	6	1	JAMMU AND KASHMIR	11	PULWAMA	IGNDPS	100	0	6	86	8	
2	2024-2025	6	1	JAMMU AND KASHMIR	620	KISHTWAR	IGNDPS	107	0	0	107	0	
3	2024-2025	6	10	BIHAR	190	BANKA	IGNDPS	1990	213	61	1500	216	
4	2024-2025	6	10	BIHAR	196	GAYA	IGNDPS	1665	407	31	1142	85	
...	...	...	...		...	...	...	...	...	...	...	...	
14292	2024-2025	2	9	UTTAR PRADESH	170	MIRZAPUR	IGNDPS	1609	492	11	299	807	
14293	2024-2025	2	9	UTTAR PRADESH	180	SHAHJAHANPUR	IGNDPS	949	171	2	305	471	
14294	2024-2025	2	9	UTTAR PRADESH	181	SHRAVASTI	IGNDPS	300	56	2	82	160	
14295	2024-2025	2	9	UTTAR PRADESH	184	SONBHADRA	IGNDPS	1491	284	127	327	753	
14296	2024-2025	2	9	UTTAR PRADESH	186	UNNAO	IGNDPS	2135	683	1	588	863	

	fin_year	mnth	lgd_state_code	state_name	lgd_district_code	district_name	scheme_code	total_beneficiaries	total_sc	total_st	total_gen	total_obc	total_aadhar	t
0	2024-2025	6	1	JAMMU AND KASHMIR	1	ANANTNAG	IGNDPS	104	0	2	101	1	104	
1	2024-2025	6	1	JAMMU AND KASHMIR	11	PULWAMA	IGNDPS	100	0	6	86	8	99	
2	2024-2025	6	1	JAMMU AND KASHMIR	620	KISHTWAR	IGNDPS	107	0	0	107	0	104	
3	2024-2025	6	10	BIHAR	190	BANKA	IGNDPS	1990	213	61	1500	216	1402	
4	2024-2025	6	10	BIHAR	196	GAYA	IGNDPS	1665	407	31	1142	85	1191	

- **Missing Values:** The initial data cleaning process revealed no null values in the dataset.

- **Data Cleaning:** The dataset was cleaned by dropping rows with missing values to ensure data integrity for subsequent analysis.

Total Beneficiaries per State
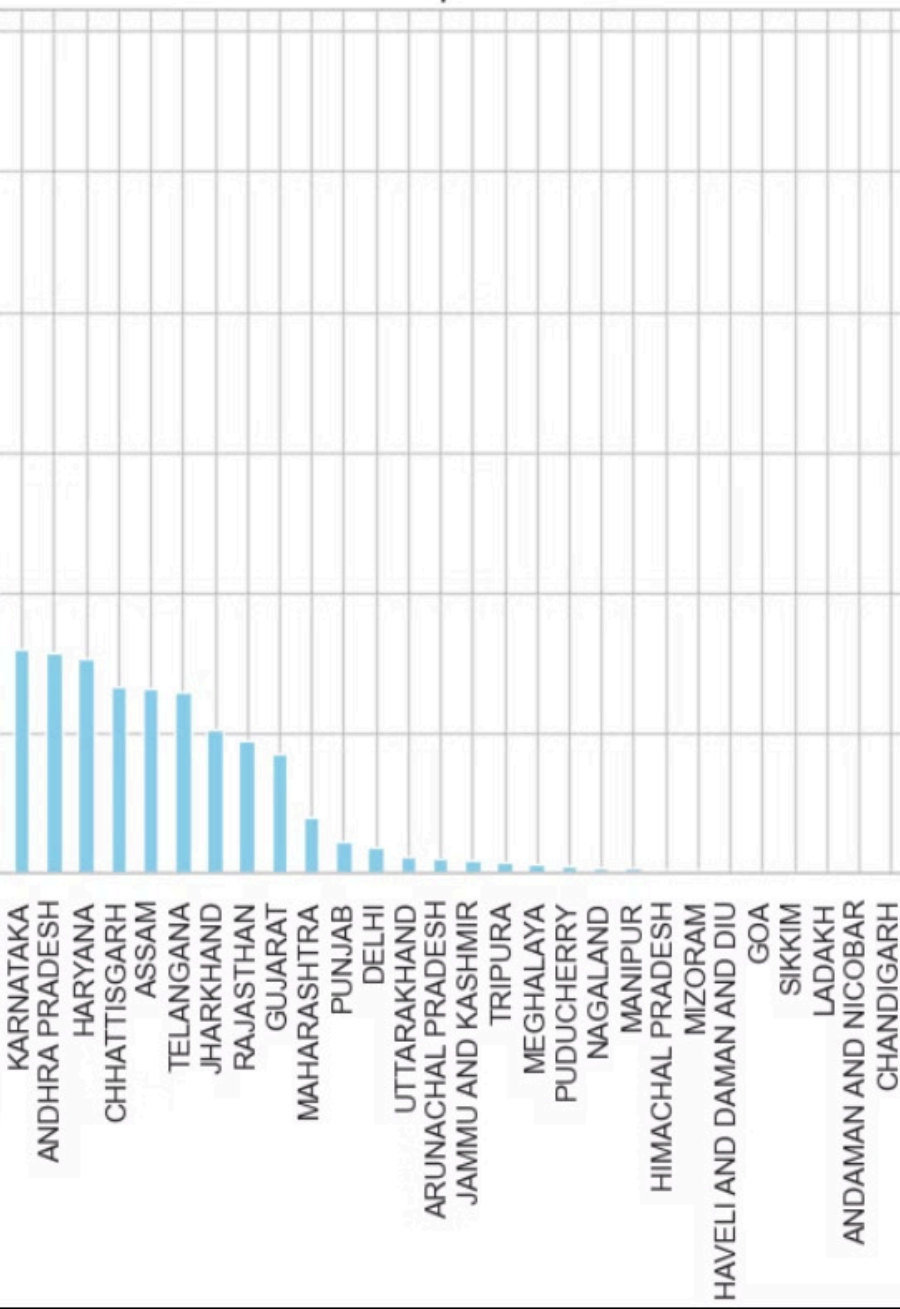
State-wise Beneficiary Analysis

This bar chart illustrates the total number of beneficiaries per state. It provides a clear comparison of beneficiary distribution across different states, highlighting those with the highest and lowest numbers of beneficiaries under the IGNDPS scheme.

```python
plt.figure(figsize=(12, 6)) df.groupby("state_name")
["total_beneficiaries"].sum().sort_values().plot(kind='bar', color='skyblue')
plt.title("Total Beneficiaries per State") plt.xlabel("State") plt.ylabel("Total
Beneficiaries") plt.xticks(rotation=90) plt.show()
```
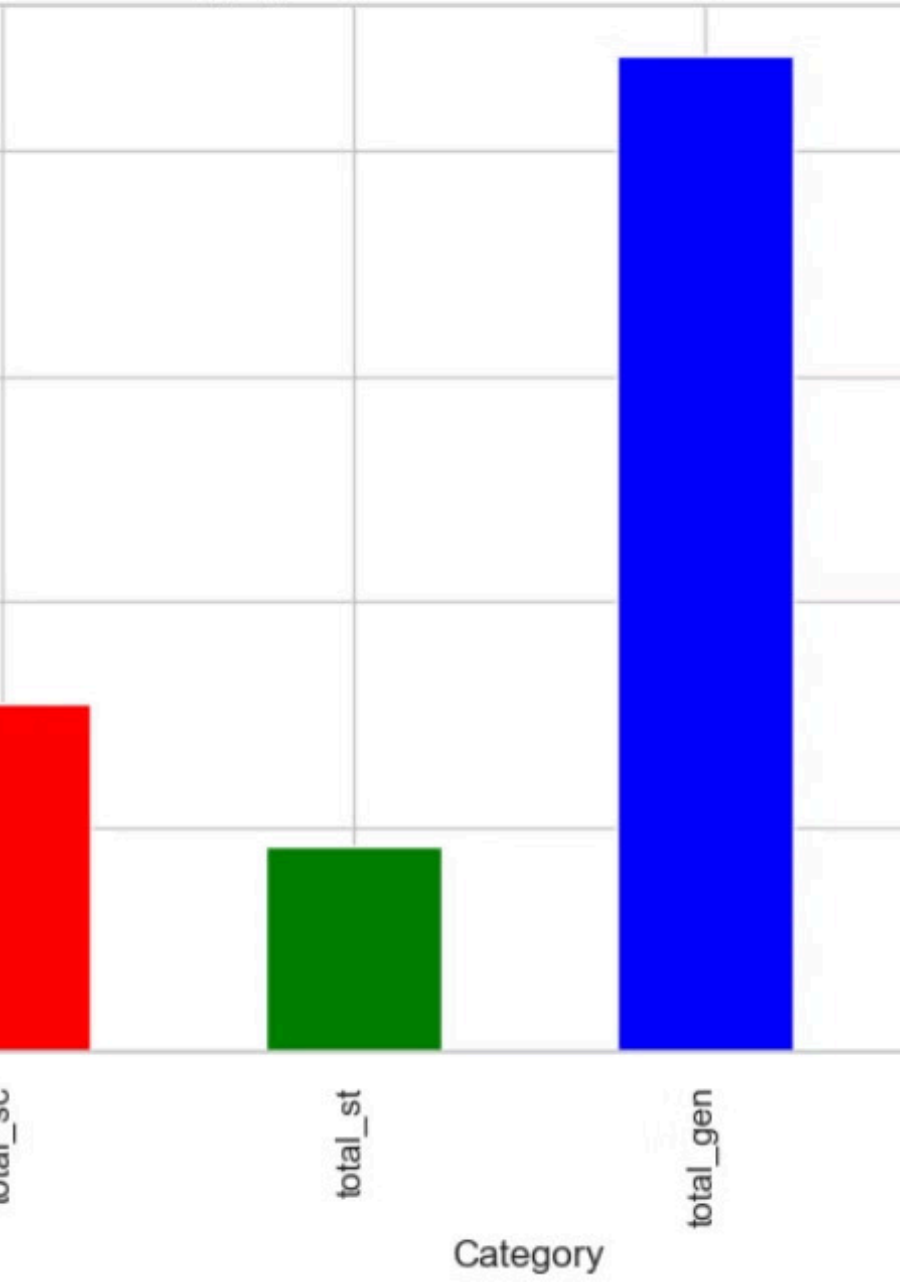
Category-wise Distribution of Beneficiaries

This bar chart shows the distribution of beneficiaries across different social categories: Scheduled Castes (SC), Scheduled Tribes (ST), General, and Other Backward Classes (OBC). It helps in understanding the representation of each category within the IGNDPS scheme.

```python
category_totals = df[["total_sc", "total_st", "total_gen", "total_obc"]].sum()
plt.figure(figsize=(8, 6))
category_totals.plot(kind="bar", color=["red", "green", "blue", "orange"])
plt.title("Category-wise Distribution of Beneficiaries", fontsize=14)
plt.xlabel("Category", fontsize=12)
plt.ylabel("Total Count", fontsize=12)
plt.show()
```
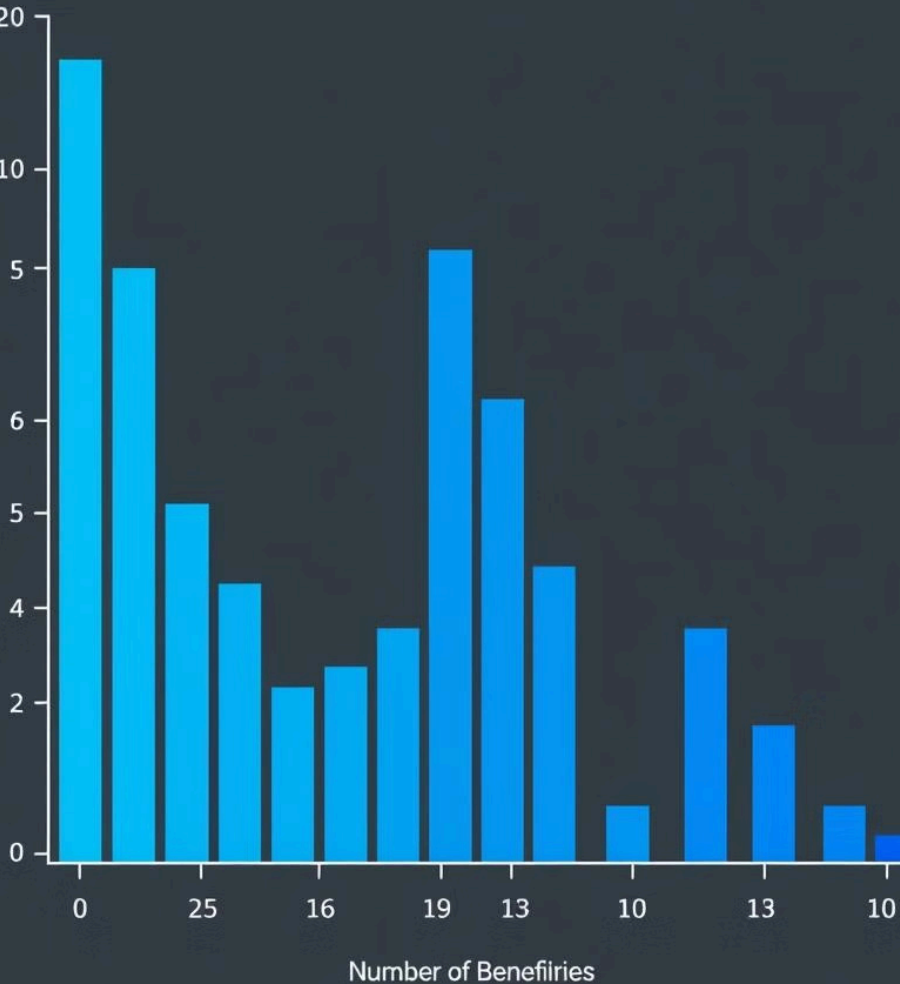
Beneficiary Anomaly Detection

This histogram highlights anomalies where the sum of SC, ST, OBC, and General beneficiaries does not match the total number of beneficiaries. The x-axis represents the difference between the total beneficiaries and the sum of categories, while the y-axis shows the count of such discrepancies.
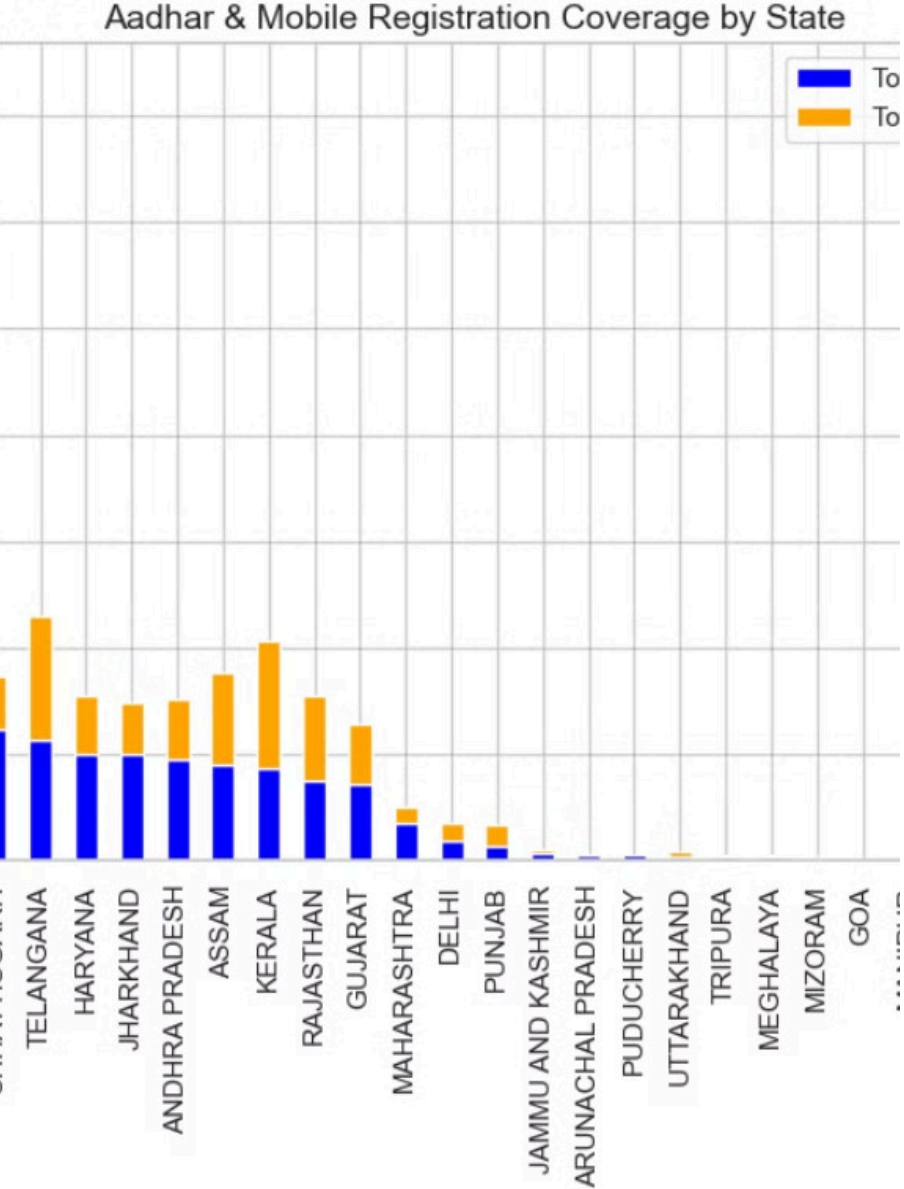
```python
df["total_check"] = df["total_sc"] + df["total_st"] + df["total_gen"] + df["total_obc"] mismatch_beneficiaries = df[df["total_beneficiaries"] != df["total_check"]] plt.figure(figsize=(10, 5)) sns.histplot(mismatch_beneficiaries["total_beneficiaries"] - mismatch_beneficiaries["total_check"], bins=20, kde=True plt.title("Distribution of Beneficiary Anomalies", fontsize=14) plt.xlabel("Difference (Total Beneficiaries - Sum of Categories)", fontsize=12) plt.ylabel("Count", fontsize=12) plt.show()
```

Aadhar and Mobile Registration Coverage

These pie charts illustrate the coverage of Aadhar and mobile number registrations among the beneficiaries. They provide insights into the extent to which beneficiaries are linked to these essential identification and communication tools.

```python
total_aadhar = df["total_aadhar"].sum() total_non_aadhar = df["total_beneficiaries"].sum() - total_aadhar plt.figure(figsize=(6, 6)) plt.pie([total_aadhar, total_non_aadhar], labels=["Aadhar Registered", "Not Registered"], autopct='%1.1f%%', colors plt.title("Aadhar Registration Coverage", fontsize=14) plt.show()
```
```python
total_mobile = df["total_mobileno"].sum() total_non_mobile = df["total_beneficiaries"].sum() - total_mobile plt.figure(figsize=(6, 6)) plt.pie([total_mobile, total_non_mobile], labels=["Mobile Registered", "Not Registered"], autopct='%1.1f%%', colors plt.title("Mobile Registration Coverage", fontsize=14) plt.show()
```

Aadhar & Mobile Registration Coverage by State

Legend: To... (blue), To... (orange)

X-axis labels: TELANGANA, HARYANA, JHARKHAND, ANDHRA PRADESH, ASSAM, KERALA, RAJASTHAN, GUJARAT, MAHARASHTRA, DELHI, PUNJAB, JAMMU AND KASHMIR, ARUNACHAL PRADESH, PUDUCHERRY, UTTARAKHAND, TRIPURA, MEGHALAYA, MIZORAM, GOA
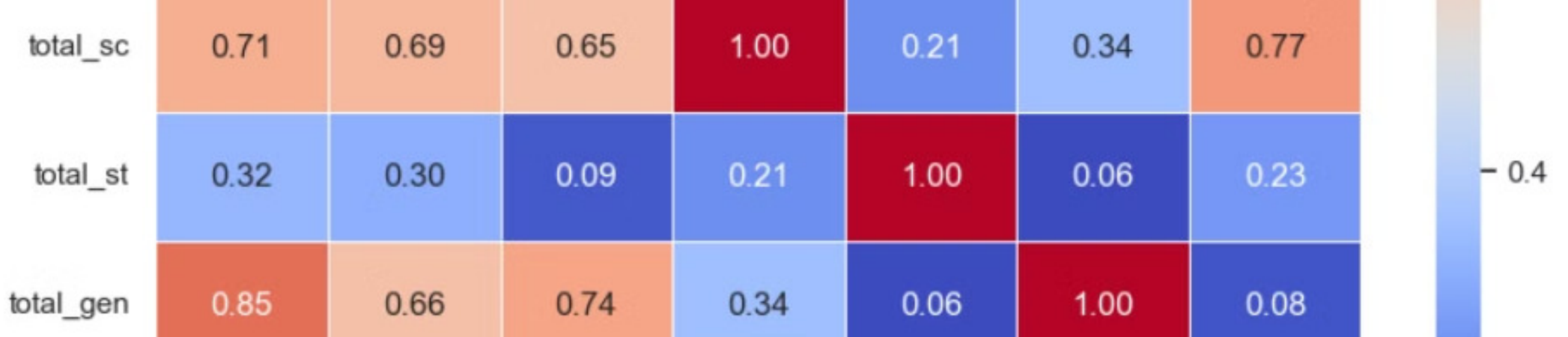
Month-wise Beneficiary Trends

This line chart displays the trend of total beneficiaries over the months. It helps in identifying any seasonal patterns or fluctuations in the number of beneficiaries availing the IGNDPS scheme.

```python
plt.figure(figsize=(10, 5))
df.groupby("mnth")["total_beneficiaries"].sum().plot(marker='o', color='green')
plt.title("Trend of Total Beneficiaries Over Months")
plt.xlabel("Month")
plt.ylabel("Total Beneficiaries")
plt.grid(True)
plt.show()
```

Made with Gamma

total_sc	0.71	0.69	0.65	1.00	0.21	0.34	0.77
total_st	0.32	0.30	0.09	0.21	1.00	0.06	0.23
total_gen	0.85	0.66	0.74	0.34	0.06	1.00	0.08

Correlation Analysis

The correlation matrix and heatmap provide insights into the relationships between different numerical variables in the dataset, such as total beneficiaries, Aadhar registrations, mobile registrations, and beneficiary counts for different social categories.

```python
num_cols = ["total_beneficiaries", "total_aadhar", "total_mobileno", "total_sc", "total_st", "total_gen", "total_obc"]
corr_matrix = df[num_cols].corr()
print("Correlation Matrix:")
print(corr_matrix)
```
```python
plt.figure(figsize=(10, 6))
sns.heatmap(corr_matrix, annot=True, cmap="coolwarm", fmt=".2f", linewidths=0.5)
plt.title("Coefficient Matrix (Correlation Heatmap)", fontsize=14)
plt.show()
```