

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt

# Load final datasets
state_df = pd.read_csv("state_wise_enrolments.csv")
age_df = pd.read_csv("state_age_ratio_clean.csv")
district_df = pd.read_csv("district_anomalies.csv")

# Show sanity check
state_df.head(), age_df.head(), district_df.head()
```

```
Out[1]: (
      state  total_enrolments  records
0  Uttar Pradesh      1018629    110369
1      Bihar      609585    60567
2  Madhya Pradesh      493970    50225
3  West Bengal      375297    76519
4  Maharashtra      369139    77191,

      state  total_child  total_adult \
0      meghalaya      74484      35287
1  dadra and nagar haveli and daman and diu      152      21
2      assam      207320      22877
3      mizoram      5431      495
4      nagaland      14465      1122

      adult_child_ratio
0      0.473753
1      0.138158
2      0.110346
3      0.091143
4      0.077567 ,

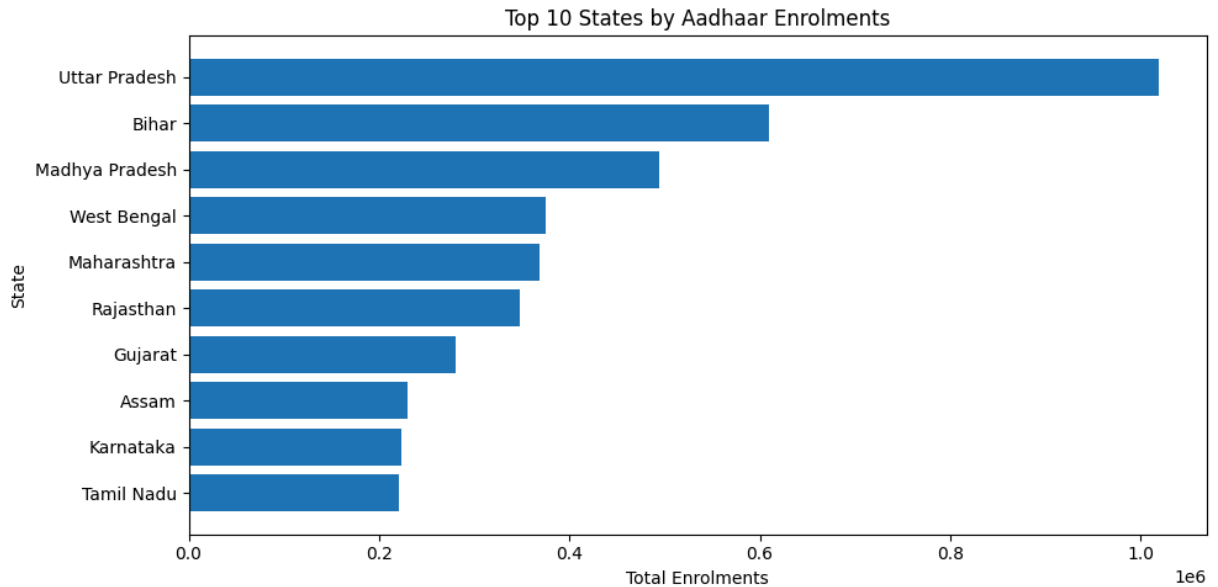
      state      district  total_enrolments  records  anomaly_sc
ore
0  maharashtra      Thane      43688      4236      5.847
145
1      bihar      Sitamarhi      42232      1672      5.625
308
2  uttar pradesh      Bahraich      39338      1585      5.184
376
3  west bengal      Murshidabad      35911      4562      4.662
236
4  west bengal  South 24 Parganas      33540      4559      4.300
989)
```

```
In [2]: # GRAPH 1: Top 10 States by Total Enrolments

top_states = state_df.head(10)

plt.figure(figsize=(10, 5))
plt.barh(top_states['state'], top_states['total_enrolments'])
plt.xlabel("Total Enrolments")
plt.ylabel("State")
plt.title("Top 10 States by Aadhaar Enrolments")
plt.gca().invert_yaxis()
```

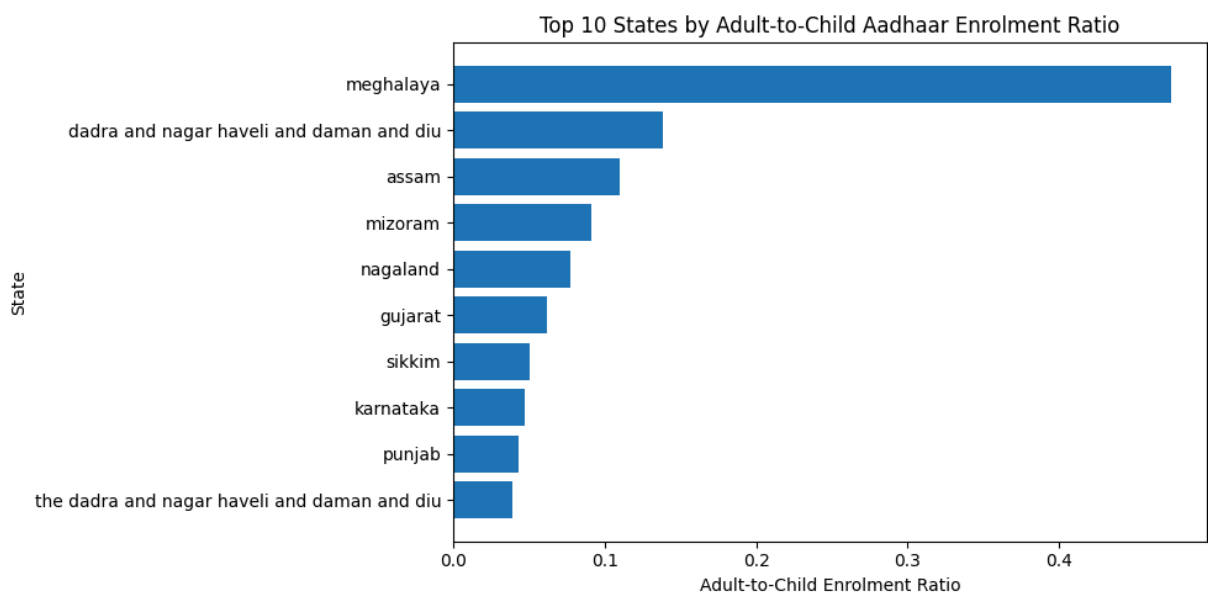
```
plt.tight_layout()
plt.show()
```



In [3]: *# GRAPH 2: Adult vs Child Enrolment Ratio (Top 10 States)*

```
top_ratio = age_df.sort_values(
    by="adult_child_ratio",
    ascending=False
).head(10)

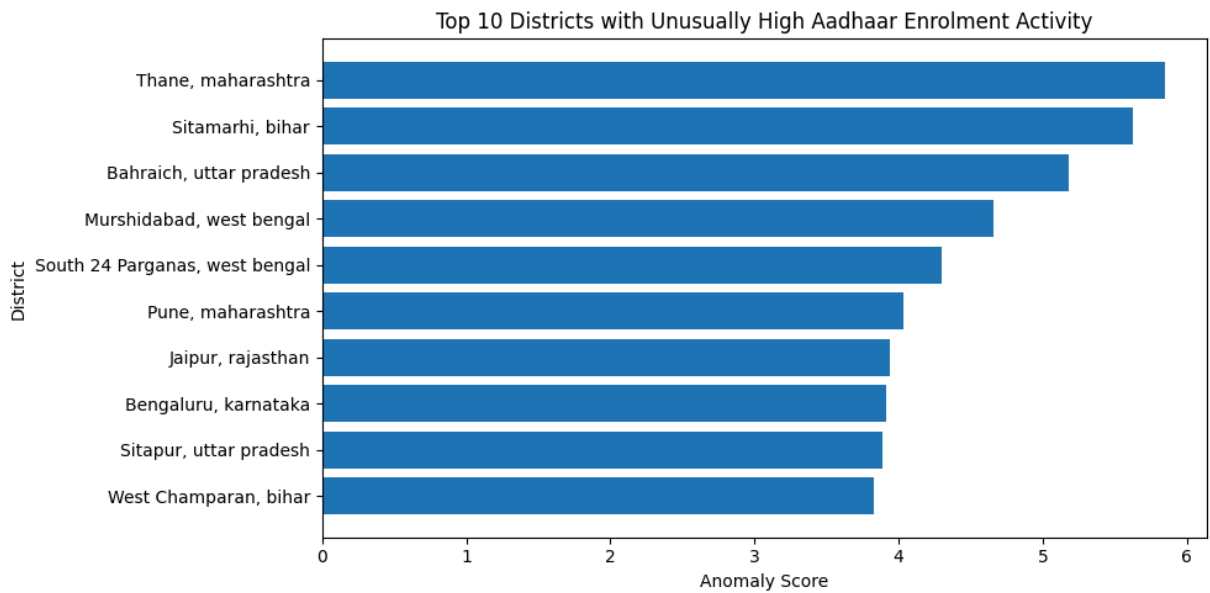
plt.figure(figsize=(10, 5))
plt.barh(top_ratio['state'], top_ratio['adult_child_ratio'])
plt.xlabel("Adult-to-Child Enrolment Ratio")
plt.ylabel("State")
plt.title("Top 10 States by Adult-to-Child Aadhaar Enrolment Ratio")
plt.gca().invert_yaxis()
plt.tight_layout()
plt.show()
```



```
In [4]: # GRAPH 3: Top 10 District-level Enrolment Anomalies

top_districts = district_df.head(10)

plt.figure(figsize=(10, 5))
plt.barh(
    top_districts['district'] + ", " + top_districts['state'],
    top_districts['anomaly_score']
)
plt.xlabel("Anomaly Score")
plt.ylabel("District")
plt.title("Top 10 Districts with Unusually High Aadhaar Enrolment Activity")
plt.gca().invert_yaxis()
plt.tight_layout()
plt.show()
```



In []: