

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
```

```
files = [
    'api_data_aadhar_biometric_0_500000.csv',
    'api_data_aadhar_biometric_500000_1000000.csv',
    'api_data_aadhar_biometric_1000000_1500000.csv',
    'api_data_aadhar_biometric_1500000_1861108.csv'
]
```

```
df = pd.concat([pd.read_csv(f) for f in files], ignore_index=True)
df['date'] = pd.to_datetime(df['date'], dayfirst=True)
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1861108 entries, 0 to 1861107
Data columns (total 6 columns):
 #   Column      Dtype
---  -
 0   date        datetime64[ns]
 1   state       object
 2   district    object
 3   pincode     int64
 4   bio_age_5_17 int64
 5   bio_age_17_ int64
dtypes: datetime64[ns](1), int64(3), object(2)
memory usage: 85.2+ MB
```

```
df.columns
```

```
Index(['date', 'state', 'district', 'pincode', 'bio_age_5_17', 'bio_age_17_'], dtype='object')
```

```
df.shape
```

```
(1861108, 6)
```

```
total_5_17 = df['bio_age_5_17'].sum()
total_17_plus = df['bio_age_17_'].sum()
print(f"Total Updates: {total_5_17 + total_17_plus:,}")
```

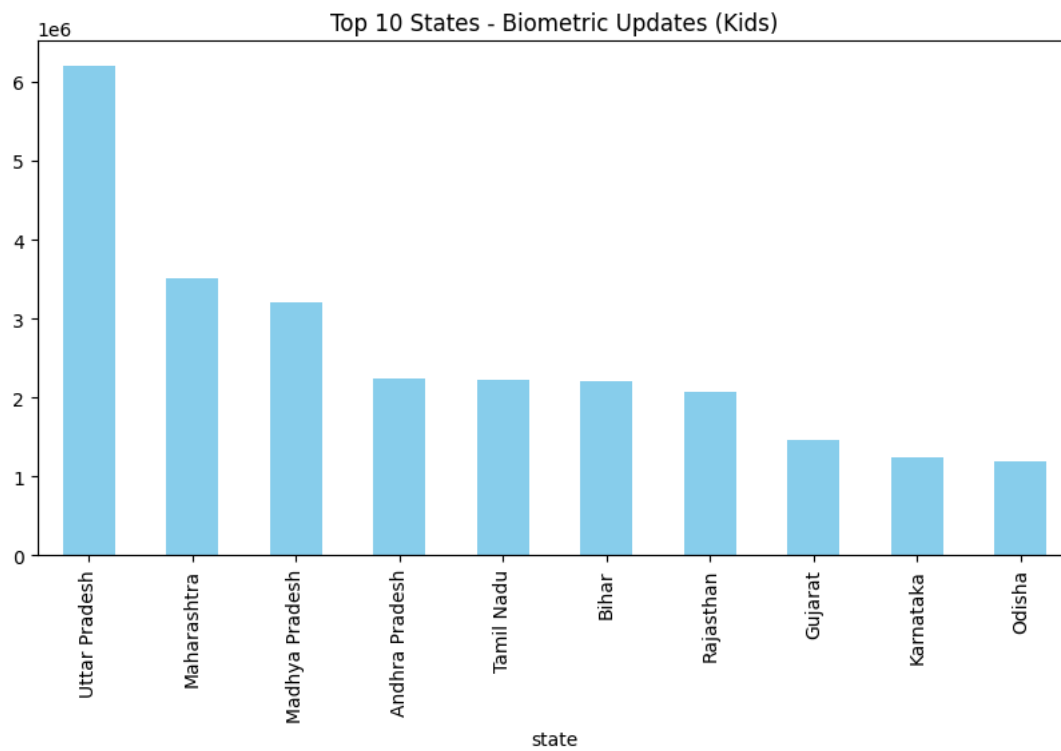
```
Total Updates: 69,763,095
```

```
pincode_summary = df.groupby(['pincode', 'state', 'district']).agg({
    'bio_age_5_17': 'sum', 'bio_age_17_': 'sum'
}).reset_index()
pincode_summary['total'] = pincode_summary['bio_age_5_17'] + pincode_summary['bio_age_17_']
```

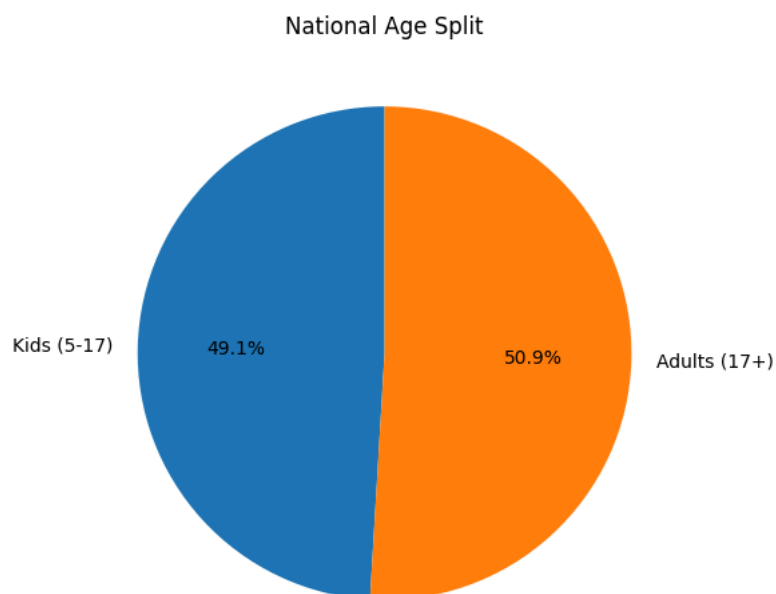
```
scaler = StandardScaler()
X_scaled = scaler.fit_transform(pincode_summary[['bio_age_5_17', 'bio_age_17_']])
kmeans = KMeans(n_clusters=3, random_state=42, n_init=10)
pincode_summary['cluster_id'] = kmeans.fit_predict(X_scaled)
```

```
zone_map = {
    '1': 'North (DL, HR, PB)', '2': 'North-2 (UP, UK)', '3': 'West (RJ, GJ)',
    '4': 'West/Central (MH, MP)', '5': 'South (AP, KA)', '6': 'South-2 (TN, KL)',
    '7': 'East (WB, NE)', '8': 'East-2 (BR, JH)', '9': 'Army'
}
pincode_summary['Zone'] = pincode_summary['pincode'].astype(str).str[0].map(zone_map)
```

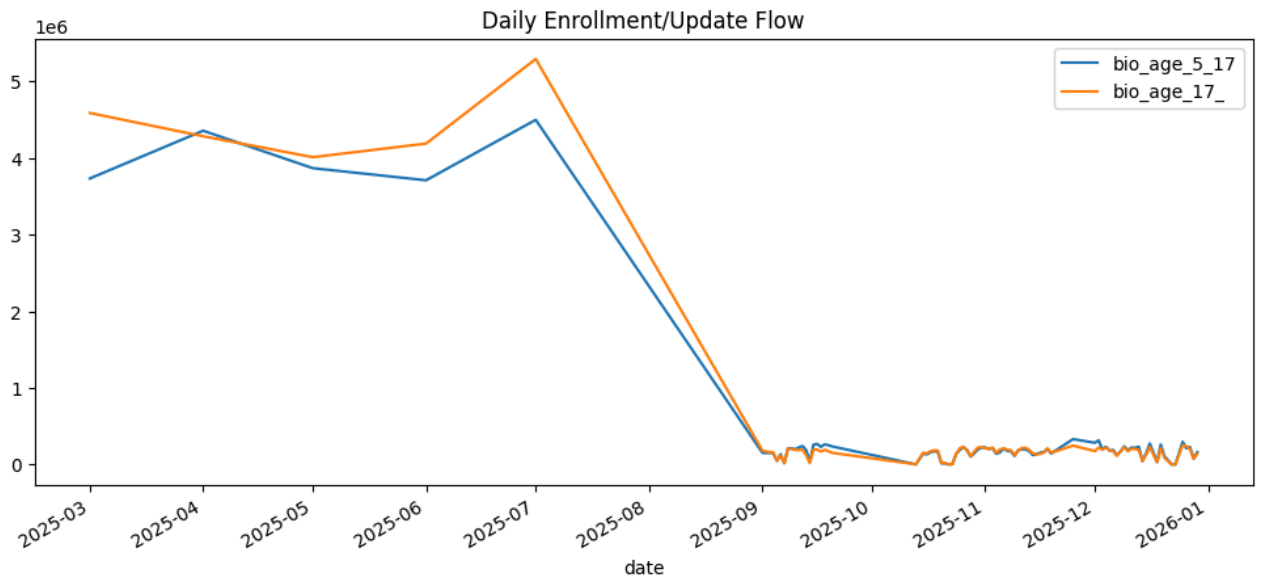
```
plt.figure(figsize=(10, 5))
state_top = df.groupby('state')['bio_age_5_17'].sum().nlargest(10) # Example: Kids updates
state_top.plot(kind='bar', color='skyblue')
plt.title('Top 10 States - Biometric Updates (Kids)')
plt.savefig('1_state_analysis.png')
```



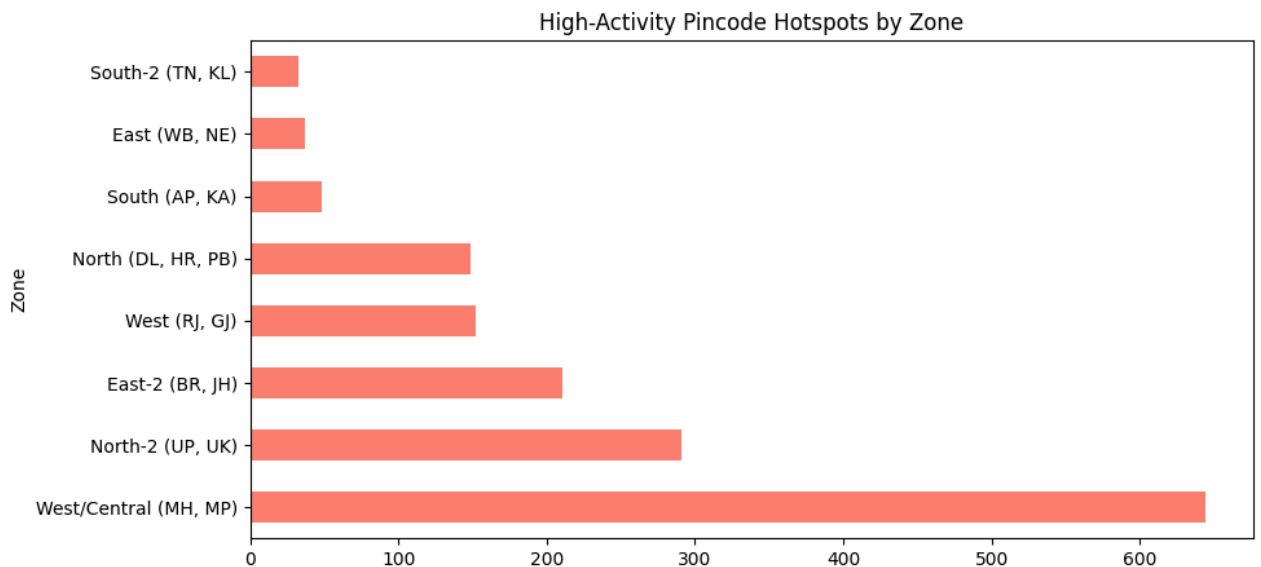
```
plt.figure(figsize=(6, 6))
plt.pie([total_5_17, total_17_plus], labels=['Kids (5-17)', 'Adults (17+)'], autopct='%1.1f%%', startangle=90)
plt.title('National Age Split')
plt.savefig('2_age_distribution.png')
```



```
daily = df.groupby('date')[['bio_age_5_17', 'bio_age_17_']].sum()
daily.plot(figsize=(12, 5))
plt.title('Daily Update Flow')
plt.savefig('3_time_trend.png')
```



```
high_activity = pincode_summary[pincode_summary['total'] > pincode_summary['total'].quantile(0.95)]
plt.figure(figsize=(10, 5))
high_activity['Zone'].value_counts().plot(kind='barh', color='salmon')
plt.title('High-Activity Pincode Hotspots by Zone')
plt.savefig('4_zone_hotspots.png')
```



```
pincode_summary.to_csv('final_biometric_report.csv', index=False)
```