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

```python
files = [
    '/content/api_data_aadhar_enrolment_0_500000.csv',
    '/content/api_data_aadhar_enrolment_1000000_1006029.csv',
    '/content/api_data_aadhar_enrolment_500000_1000000.csv'
]
df_list = [pd.read_csv(f) for f in files]
df = pd.concat(df_list, ignore_index=True)
```

```python
df['date'] = pd.to_datetime(df['date'], dayfirst=True)
```

```python
df_agg = df.groupby(['date', 'state', 'district', 'pincode']).sum().reset_index()
```

```python
df_agg['total_enrolment'] = df_agg['age_0_5'] + df_agg['age_5_17'] + df_agg['age_18_greater']
```

```python
df_agg.to_csv('consolidated_aadhaar_data.csv', index=False)
```

```python
df_agg['month_year'] = df_agg['date'].dt.to_period('M')
monthly_trends = df_agg.groupby('month_year').agg({
    'age_0_5': 'sum',
    'age_5_17': 'sum',
    'age_18_greater': 'sum',
    'total_enrolment': 'sum'
}).reset_index()
```

```python
monthly_trends['month_year_str'] = monthly_trends['month_year'].astype(str)
```

```python
plt.figure(figsize=(12, 6))
plt.plot(monthly_trends['month_year_str'], monthly_trends['total_enrolment'], marker='o', color='#1f77b4', label='Total Enr
plt.title('National Aadhaar Enrolment Trends (Monthly)', fontsize=14)
plt.xlabel('Timeline', fontsize=12)
plt.ylabel('Volume', fontsize=12)
plt.xticks(rotation=45)
plt.grid(alpha=0.3)
plt.tight_layout()
plt.savefig('national_trends.png')
```
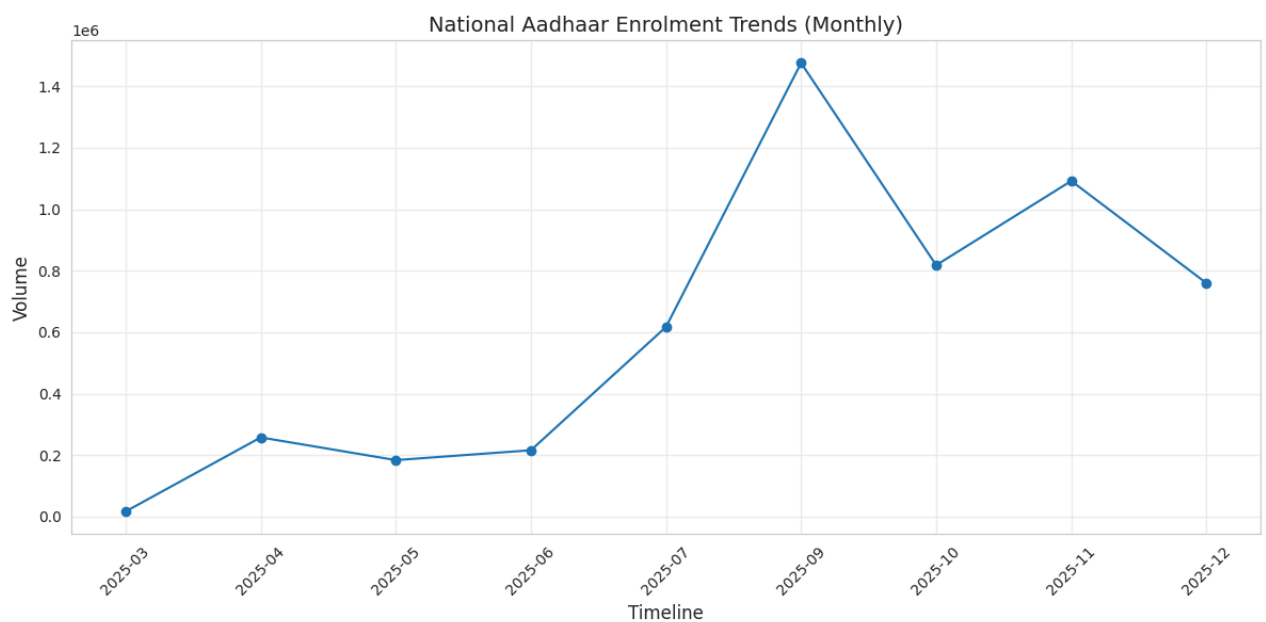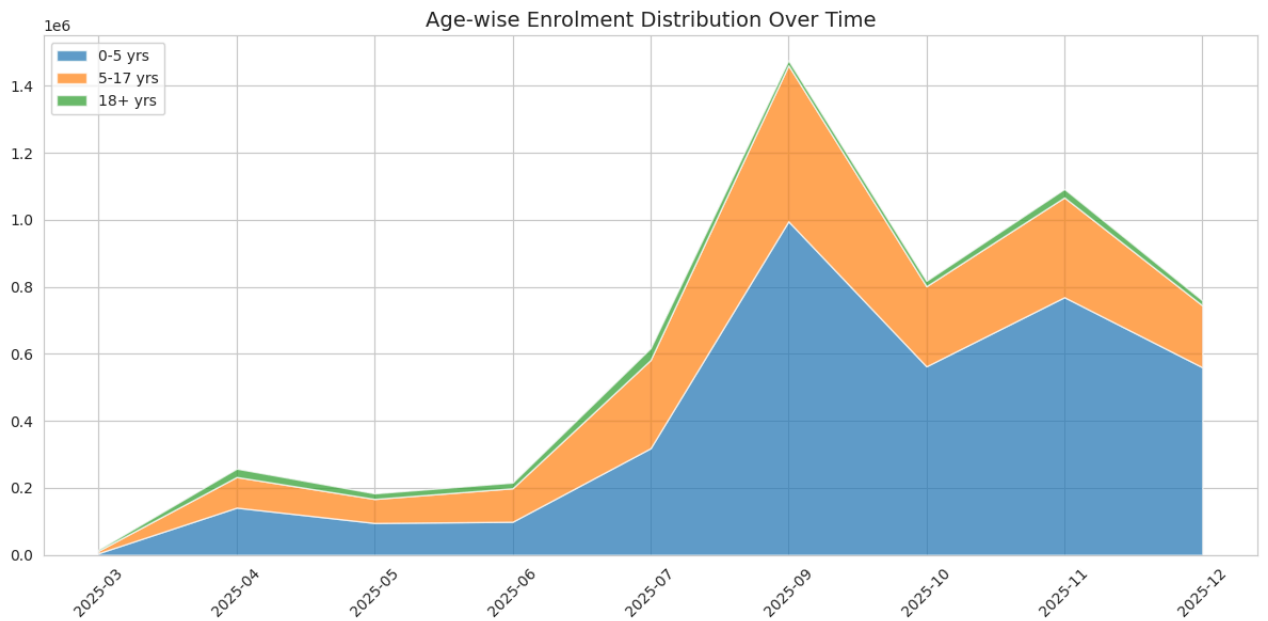


```python
plt.figure(figsize=(12, 6))
plt.stackplot(monthly_trends['month_year_str'],
              monthly_trends['age_0_5'],
```

```
            monthly_trends['age_5_17'],
            monthly_trends['age_18_greater'],
            labels=['0-5 yrs', '5-17 yrs', '18+ yrs'], alpha=0.7)
plt.title('Age-wise Enrolment Distribution Over Time', fontsize=14)
plt.xticks(rotation=45)
plt.legend(loc='upper left')
plt.tight_layout()
plt.savefig('age_wise_trends.png')
```
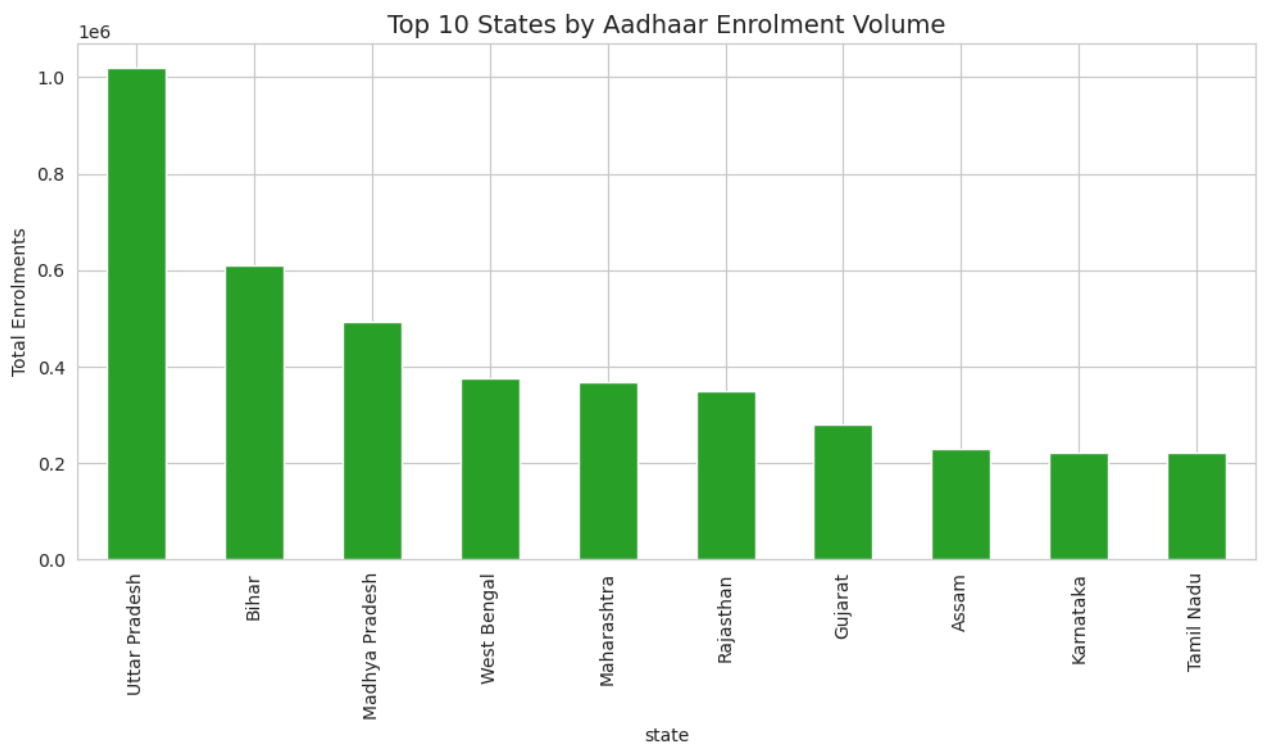


```
state_data = df_agg.groupby('state')['total_enrolment'].sum().sort_values(ascending=False).head(10)

plt.figure(figsize=(10, 6))
state_data.plot(kind='bar', color='#2ca02c')
plt.title('Top 10 States by Aadhaar Enrolment Volume', fontsize=14)
plt.ylabel('Total Enrolments')
plt.tight_layout()
plt.savefig('top_states.png')
```



```
pincode_metrics = df_agg.groupby('pincode').agg({
    'total_enrolment': 'sum',
```

```
        'age_0_5': 'sum',
        'age_5_17': 'sum',
        'age_18_greater': 'sum'
}).reset_index()
```

```
pincode_metrics['child_ratio'] = (pincode_metrics['age_0_5'] + pincode_metrics['age_5_17']) / pincode_metrics['total_enrolm
pincode_metrics['adult_ratio'] = pincode_metrics['age_18_greater'] / pincode_metrics['total_enrolment']
```

```
features = ['total_enrolment', 'child_ratio', 'adult_ratio']
cluster_data = pincode_metrics[features].fillna(0)
```

```
scaler = StandardScaler()
scaled_features = scaler.fit_transform(cluster_data)
```

```
kmeans = KMeans(n_clusters=3, random_state=42, n_init=10)
pincode_metrics['cluster'] = kmeans.fit_predict(scaled_features)
```

```
plt.figure(figsize=(10, 7))
sns.scatterplot(data=pincode_metrics, x='total_enrolment', y='child_ratio', hue='cluster', palette='viridis', alpha=0.5)
plt.title('Pincode Segmentation: Enrolment Volume vs Child Ratio', fontsize=14)
plt.xlabel('Total Enrolment per PIN')
plt.ylabel('Ratio of Child Enrolments (0-17)')
plt.grid(True, linestyle='--', alpha=0.6)
plt.tight_layout()
plt.savefig('pincode_clustering.png')
```
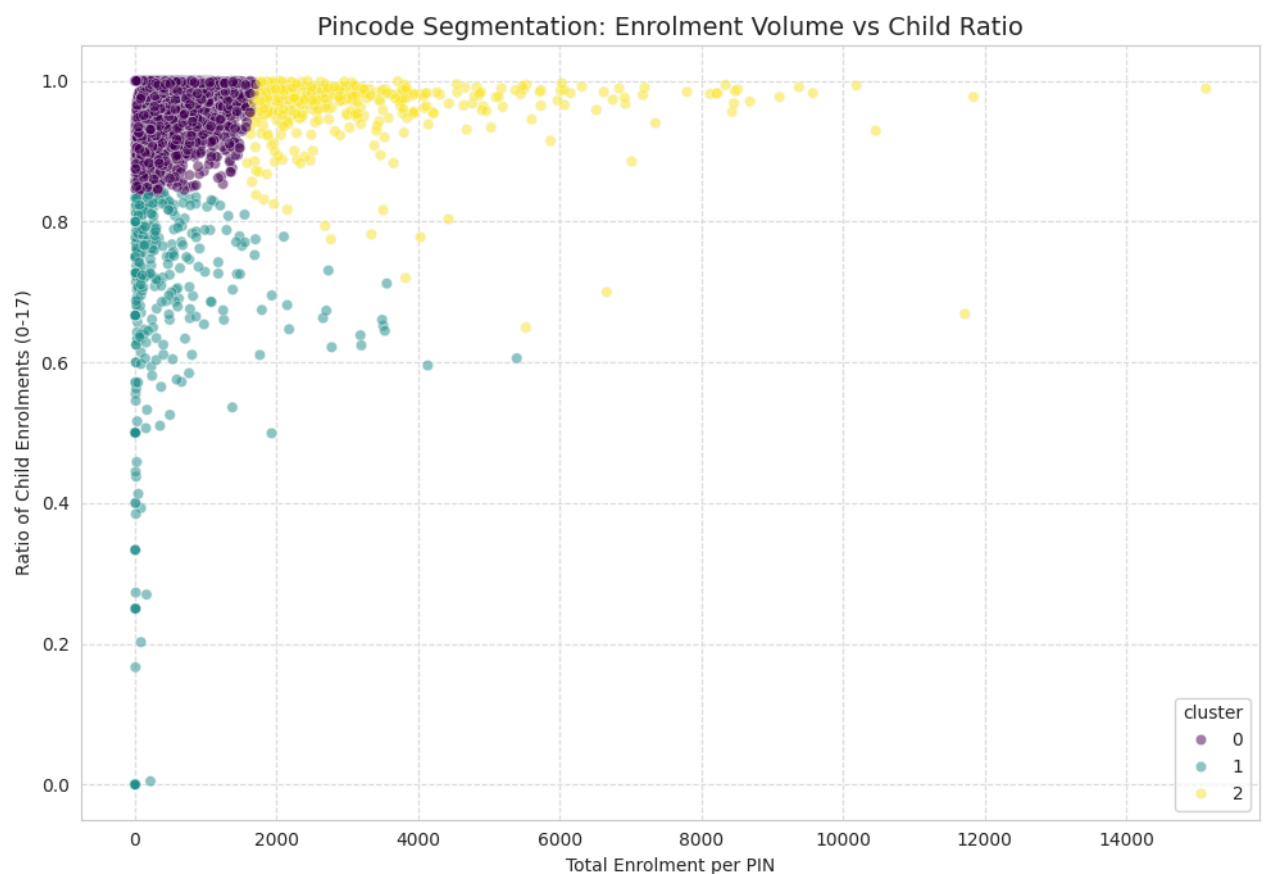


```
cluster_summary = pincode_metrics.groupby('cluster').agg({
    'pincode': 'count',
    'total_enrolment': 'mean',
    'child_ratio': 'mean'
}).rename(columns={'pincode': 'Pin_Code_Count'})

print("\n--- Cluster Profile Summary ---")
print(cluster_summary)
```

```
--- Cluster Profile Summary ---
        Pin_Code_Count  total_enrolment  child_ratio
cluster
0                18529       210.215284     0.984868
1                  493       317.969574     0.698351
```

2	441	3138.013605	0.963582