In [2]:
```
!pip install pandas numpy matplotlib seaborn plotly scipy scikit-learn
```

```
Requirement already satisfied: pandas in c:\users\sl\anaconda3\lib\site-package
s (2.3.3)
Requirement already satisfied: numpy in c:\users\sl\anaconda3\lib\site-packages
(2.3.5)
Requirement already satisfied: matplotlib in c:\users\sl\anaconda3\lib\site-pac
kages (3.10.6)
Requirement already satisfied: seaborn in c:\users\sl\anaconda3\lib\site-packag
es (0.13.2)
Requirement already satisfied: plotly in c:\users\sl\anaconda3\lib\site-package
s (6.3.0)
Requirement already satisfied: scipy in c:\users\sl\anaconda3\lib\site-packages
(1.16.3)
Requirement already satisfied: scikit-learn in c:\users\sl\anaconda3\lib\site-p
ackages (1.7.2)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\sl\anaconda3\
lib\site-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in c:\users\sl\anaconda3\lib\site-p
ackages (from pandas) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in c:\users\sl\anaconda3\lib\sit
e-packages (from pandas) (2025.2)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\sl\anaconda3\lib\si
te-packages (from matplotlib) (1.3.3)
Requirement already satisfied: cycler>=0.10 in c:\users\sl\anaconda3\lib\site-p
ackages (from matplotlib) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\sl\anaconda3\lib\s
ite-packages (from matplotlib) (4.60.1)
Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\sl\anaconda3\lib\s
ite-packages (from matplotlib) (1.4.9)
Requirement already satisfied: packaging>=20.0 in c:\users\sl\anaconda3\lib\sit
e-packages (from matplotlib) (25.0)
Requirement already satisfied: pillow>=8 in c:\users\sl\anaconda3\lib\site-pack
ages (from matplotlib) (12.0.0)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\sl\anaconda3\lib\si
te-packages (from matplotlib) (3.2.5)
Requirement already satisfied: narwhals>=1.15.1 in c:\users\sl\anaconda3\lib\si
te-packages (from plotly) (2.7.0)
Requirement already satisfied: joblib>=1.2.0 in c:\users\sl\anaconda3\lib\site-
packages (from scikit-learn) (1.5.2)
Requirement already satisfied: threadpoolctl>=3.1.0 in c:\users\sl\anaconda3\li
b\site-packages (from scikit-learn) (3.5.0)
Requirement already satisfied: six>=1.5 in c:\users\sl\anaconda3\lib\site-packa
ges (from python-dateutil>=2.8.2->pandas) (1.17.0)
```

In [3]:
```
import pandas as pd
import numpy as np

import matplotlib.pyplot as plt
import seaborn as sns

import plotly.express as px
import plotly.graph_objects as go
```

```
from scipy import stats
from sklearn.ensemble import IsolationForest
```

In [4]:
```
import os
os.listdir()
```

Out[4]:
```
['.ipynb_checkpoints',
 'Analysis.ipynb',
 'api_data_aadhar_biometric',
 'api_data_aadhar_demographic',
 'api_data_aadhar_enrolment']
```

In [5]:
```
os.listdir('api_data_aadhar_biometric')
```

Out[5]:
```
['api_data_aadhar_biometric_0_500000.csv',
 'api_data_aadhar_biometric_1000000_1500000.csv',
 'api_data_aadhar_biometric_1500000_1861108.csv',
 'api_data_aadhar_biometric_500000_1000000.csv']
```

In [6]:
```
os.listdir('api_data_aadhar_demographic')
```

Out[6]:
```
['api_data_aadhar_demographic_0_500000.csv',
 'api_data_aadhar_demographic_1000000_1500000.csv',
 'api_data_aadhar_demographic_1500000_2000000.csv',
 'api_data_aadhar_demographic_2000000_2071700.csv',
 'api_data_aadhar_demographic_500000_1000000.csv']
```

In [7]:
```
os.listdir('api_data_aadhar_enrolment')
```

Out[7]:
```
['api_data_aadhar_enrolment_0_500000.csv',
 'api_data_aadhar_enrolment_1000000_1006029.csv',
 'api_data_aadhar_enrolment_500000_1000000.csv']
```

In [8]:
```
import pandas as pd

files = os.listdir('api_data_aadhar_biometric')

biometric = pd.concat(
    [pd.read_csv(f'api_data_aadhar_biometric/{f}') for f in files],
    ignore_index=True
)

biometric.head()
```

Out[8]:

	date	state	district	pincode	bio_age_5_17	bio_age_17_
0	01-03-2025	Haryana	Mahendragarh	123029	280	577
1	01-03-2025	Bihar	Madhepura	852121	144	369
2	01-03-2025	Jammu and Kashmir	Punch	185101	643	1091
3	01-03-2025	Bihar	Bhojpur	802158	256	980
4	01-03-2025	Tamil Nadu	Madurai	625514	271	815

In [9]:

```
files = os.listdir('api_data_aadhar_enrolment')

enrolment = pd.concat(
    [pd.read_csv(f'api_data_aadhar_enrolment/{f}') for f in files],
    ignore_index=True
)

enrolment.head()
```

Out[9]:

	date	state	district	pincode	age_0_5	age_5_17	age_18_greater
0	02-03-2025	Meghalaya	East Khasi Hills	793121	11	61	37
1	09-03-2025	Karnataka	Bengaluru Urban	560043	14	33	39
2	09-03-2025	Uttar Pradesh	Kanpur Nagar	208001	29	82	12
3	09-03-2025	Uttar Pradesh	Aligarh	202133	62	29	15
4	09-03-2025	Karnataka	Bengaluru Urban	560016	14	16	21

In [10]:

```
files = os.listdir('api_data_aadhar_demographic')

demographic = pd.concat(
    [pd.read_csv(f'api_data_aadhar_demographic/{f}') for f in files],
    ignore_index=True
)

demographic.head()
```

Out[10]:

	date	state	district	pincode	demo_age_5_17	demo_age_17_
0	01-03-2025	Uttar Pradesh	Gorakhpur	273213	49	529
1	01-03-2025	Andhra Pradesh	Chittoor	517132	22	375
2	01-03-2025	Gujarat	Rajkot	360006	65	765
3	01-03-2025	Andhra Pradesh	Srikakulam	532484	24	314
4	01-03-2025	Rajasthan	Udaipur	313801	45	785

In [11]:
```
print("Enrolment:", enrolment.shape)
print("Demographic:", demographic.shape)
print("Biometric:", biometric.shape)
```

```
Enrolment: (1006029, 7)
Demographic: (2071700, 6)
Biometric: (1861108, 6)
```

In [12]:
```
enrolment.columns = enrolment.columns.str.strip().str.lower()
demographic.columns = demographic.columns.str.strip().str.lower()
biometric.columns = biometric.columns.str.strip().str.lower()
```

In [13]:
```
enrolment['date'] = pd.to_datetime(enrolment['date'], errors='coerce')
demographic['date'] = pd.to_datetime(demographic['date'], errors='coerce')
biometric['date'] = pd.to_datetime(biometric['date'], errors='coerce')
```

In [14]:
```
print(enrolment.isnull().sum())
print(demographic.isnull().sum())
print(biometric.isnull().sum())
```

```
date              682238
state                  0
district               0
pincode                0
age_0_5                0
age_5_17               0
age_18_greater         0
dtype: int64
date             1187968
state                  0
district               0
pincode                0
demo_age_5_17          0
demo_age_17_           0
dtype: int64
date             944100
state                 0
district              0
pincode               0
bio_age_5_17          0
bio_age_17_           0
dtype: int64
```

In [15]:
```
enrolment['date'] = pd.to_datetime(enrolment['date'], errors='coerce')
demographic['date'] = pd.to_datetime(demographic['date'], errors='coerce')
biometric['date'] = pd.to_datetime(biometric['date'], errors='coerce')
```

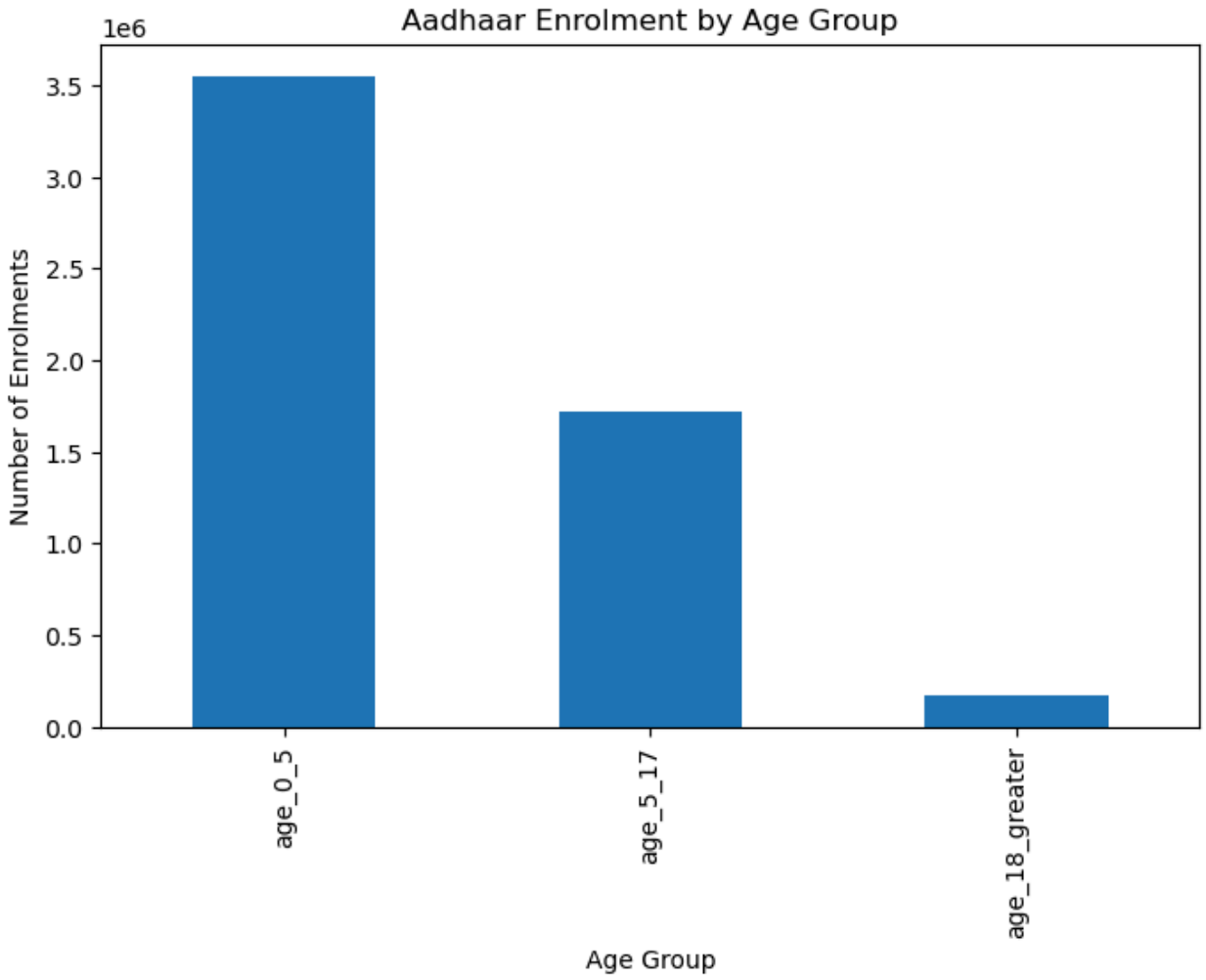
In [16]:
```
enrolment_full = enrolment.copy()
demographic_full = demographic.copy()
biometric_full = biometric.copy()
```

In [17]:
```
enrolment_date = enrolment.dropna(subset=['date'])
demographic_date = demographic.dropna(subset=['date'])
biometric_date = biometric.dropna(subset=['date'])
```

In [18]:
```
print(enrolment_date.shape)
print(demographic_date.shape)
print(biometric_date.shape)
```

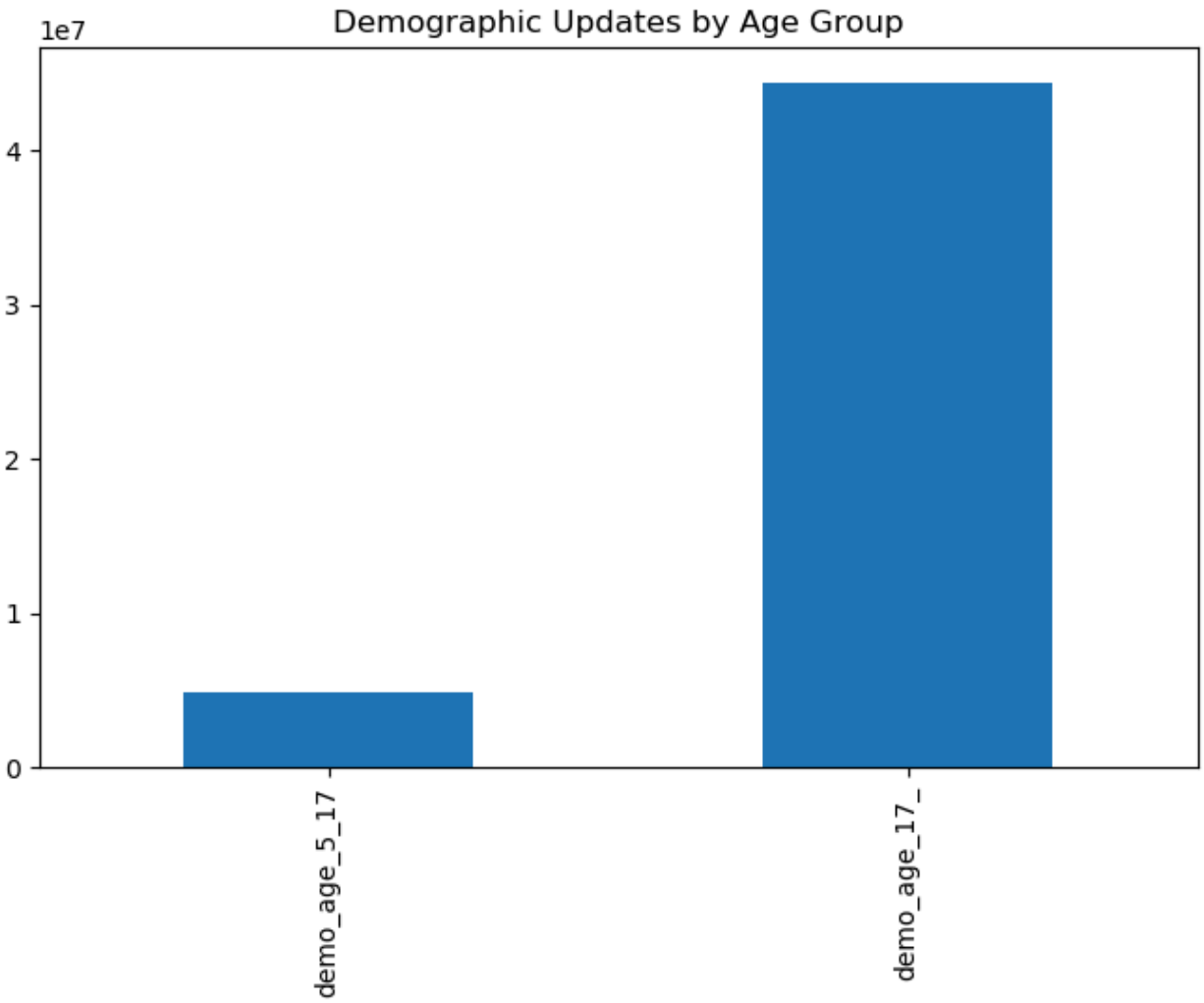
```
(323791, 7)
(883732, 6)
(917008, 6)
```

In [19]:
```
enrolment_full[['age_0_5','age_5_17','age_18_greater']].sum().plot(
    kind='bar',
    title='Aadhaar Enrolment by Age Group',
    figsize=(8,5)
)
plt.xlabel('Age Group')
plt.ylabel('Number of Enrolments')
plt.show()
```



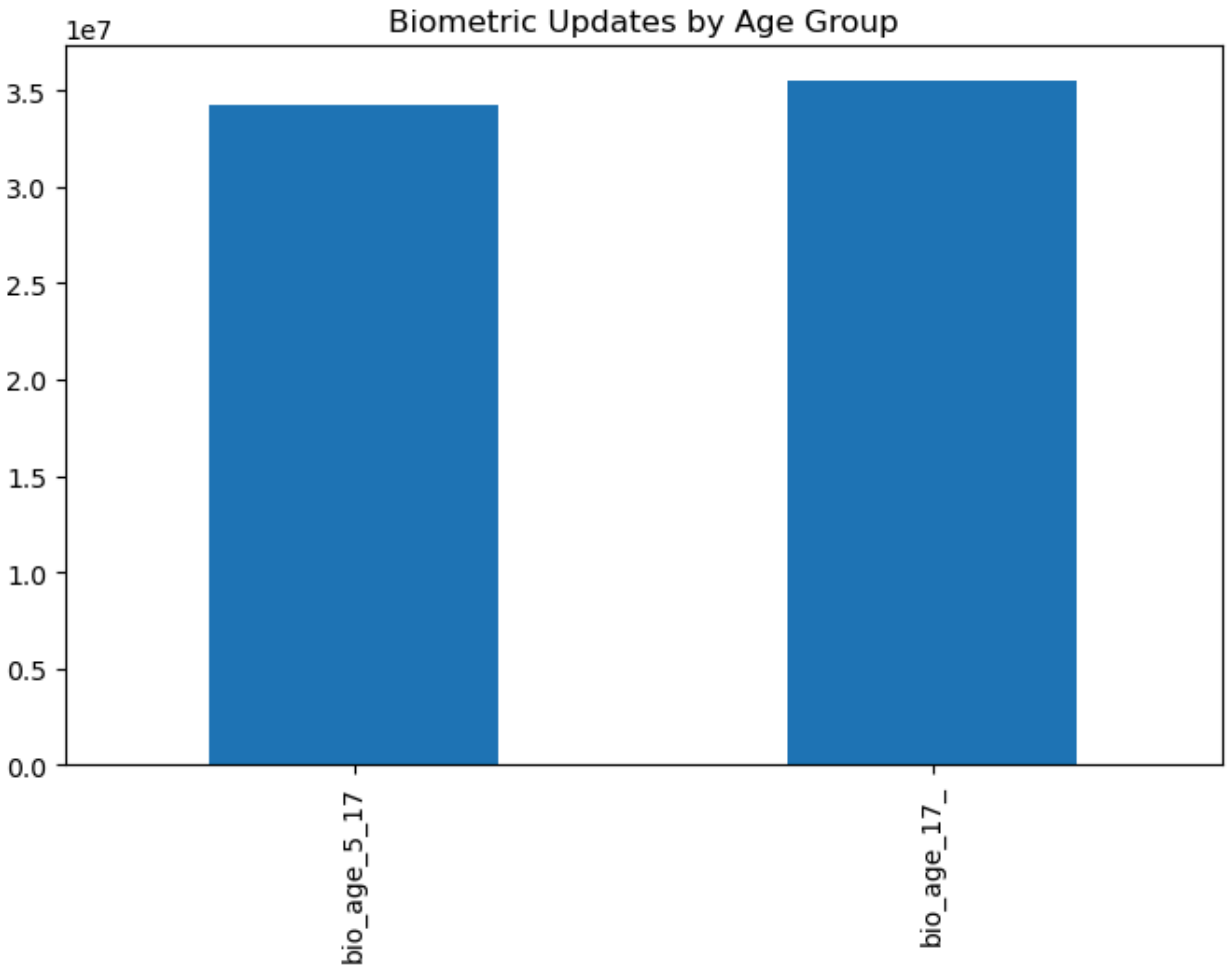
In [20]:
```
demo_cols = ['demo_age_5_17','demo_age_17_']

demographic_full[demo_cols].sum().plot(
    kind='bar',
    title='Demographic Updates by Age Group',
    figsize=(8,5)
)
plt.show()
```



In [21]:
```
bio_cols = ['bio_age_5_17','bio_age_17_']

biometric_full[bio_cols].sum().plot(
    kind='bar',
    title='Biometric Updates by Age Group',
    figsize=(8,5)
)
plt.show()
```



In [22]:
```
enrolment_full.groupby('state')[['age_0_5','age_5_17','age_18_greater']].sum()
    .sum(axis=1) \
    .sort_values(ascending=False) \
    .head(10)
```

Out[22]:
```
state
Uttar Pradesh     1018629
Bihar              609585
Madhya Pradesh     493970
West Bengal        375297
Maharashtra        369139
Rajasthan          348458
Gujarat            280549
Assam              230197
Karnataka          223235
Tamil Nadu         220789
dtype: int64
```

In [23]:
```
demo_state = demographic_full.groupby('state')[demo_cols].sum().sum(axis=1)
bio_state = biometric_full.groupby('state')[bio_cols].sum().sum(axis=1)

comparison = pd.DataFrame({
    'demographic_updates': demo_state,
```

```
    'biometric_updates': bio_state
})

comparison.head()
```

Out[23]:

state	demographic_updates	biometric_updates
100000	2.0	NaN
Andaman & Nicobar Islands	1059.0	2384.0
Andaman and Nicobar Islands	6187.0	18314.0
Andhra Pradesh	2295505.0	3714592.0
Arunachal Pradesh	36443.0	72394.0

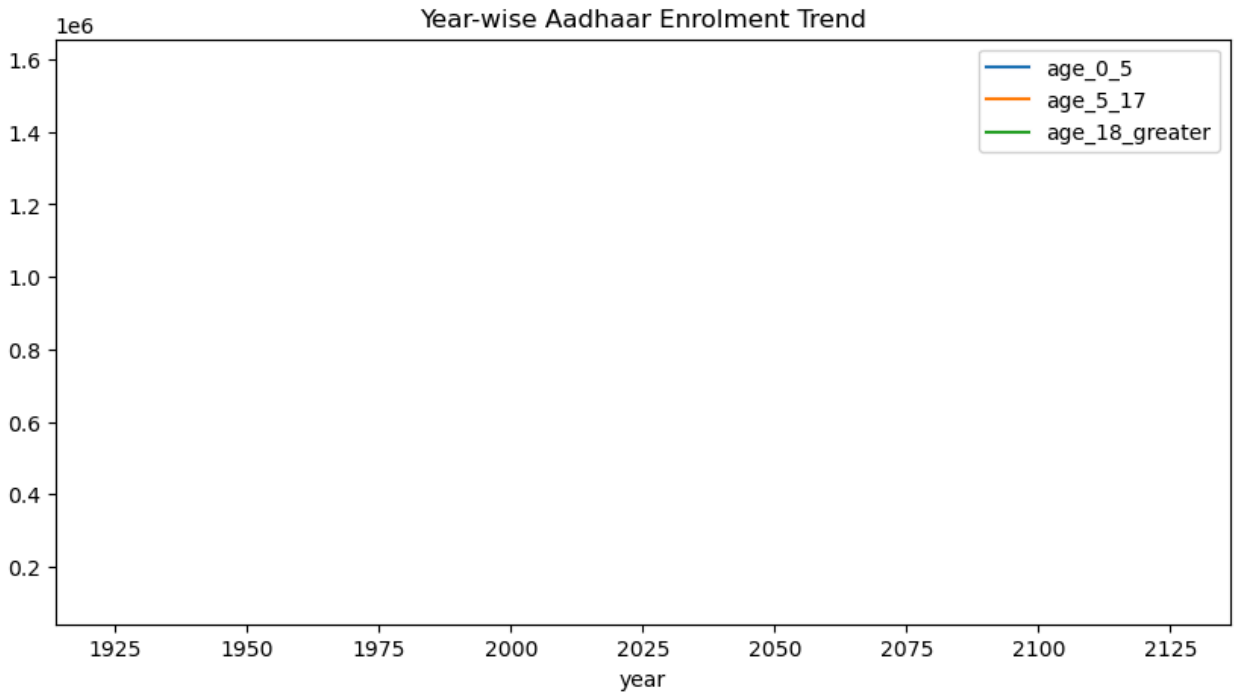
In [24]:

```
enrolment_date['year'] = enrolment_date['date'].dt.year

enrolment_date.groupby('year')[['age_0_5','age_5_17','age_18_greater']].sum().
    title='Year-wise Aadhaar Enrolment Trend',
    figsize=(10,5)
)
plt.show()
```

```
C:\Users\sl\AppData\Local\Temp\ipykernel_15868\2309746322.py:1: SettingWithCopy
Warning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/sta
ble/user_guide/indexing.html#returning-a-view-versus-a-copy
  enrolment_date['year'] = enrolment_date['date'].dt.year
```



Assistant

The selected code performs data aggregation and visualization of Aadhaar enrollment trends by year. Here's a step-by-step explanation:

1. `enrolment_date['year'] = enrolment_date['date'].dt.year` - This line creates a new column called 'year' in the 'enrolment_date' DataFrame by extracting the year component from the 'date' column using the pandas datetime accessor `.dt.year`.

2. `enrolment_date.groupby('year')[['age_0_5','age_5_17','age_18_g` - This line:

    - Groups the data by the newly created 'year' column
    - Selects three age group columns ('age_0_5', 'age_5_17', 'age_18_greater')
    - Calculates the sum of enrollments for each age group per year
    - Creates a plot of the aggregated data

3. `title='Year-wise Aadhaar Enrolment Trend'` - Sets the title of the plot

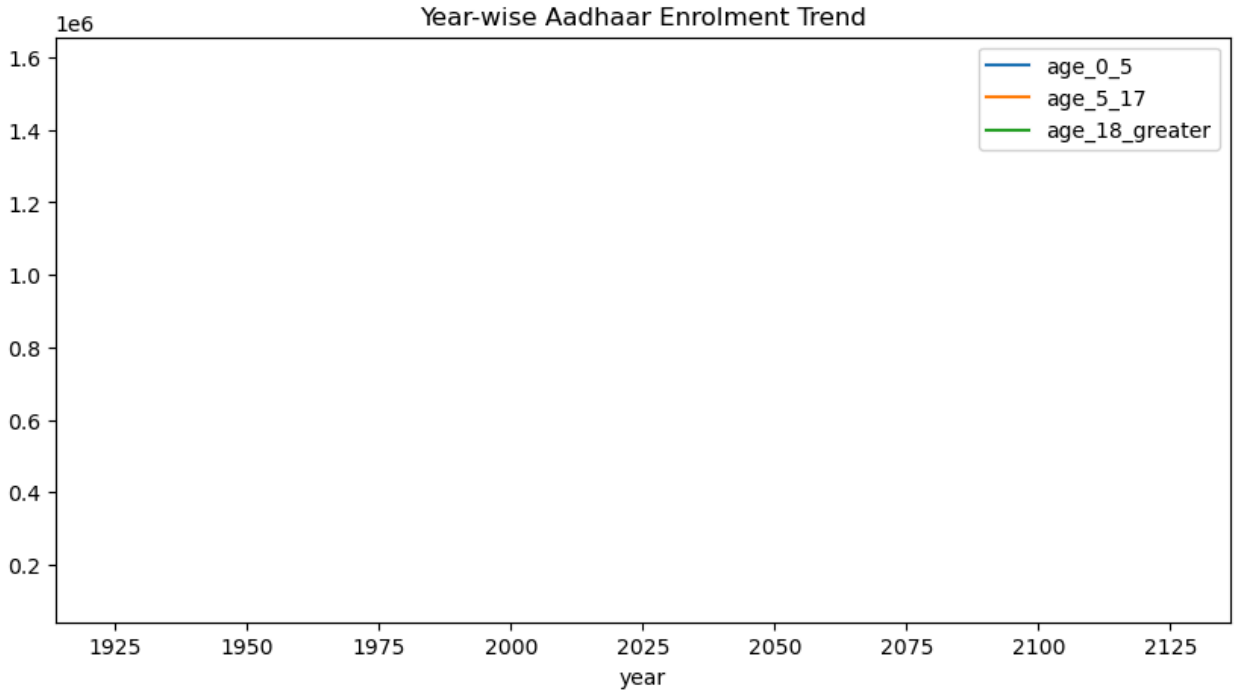
4. `figsize=(10,5)` - Sets the figure size to 10 inches wide by 5 inches tall

5. `plt.show()` - Displays the plot

This code visualizes how Aadhaar enrollments across different age groups have changed over the years.

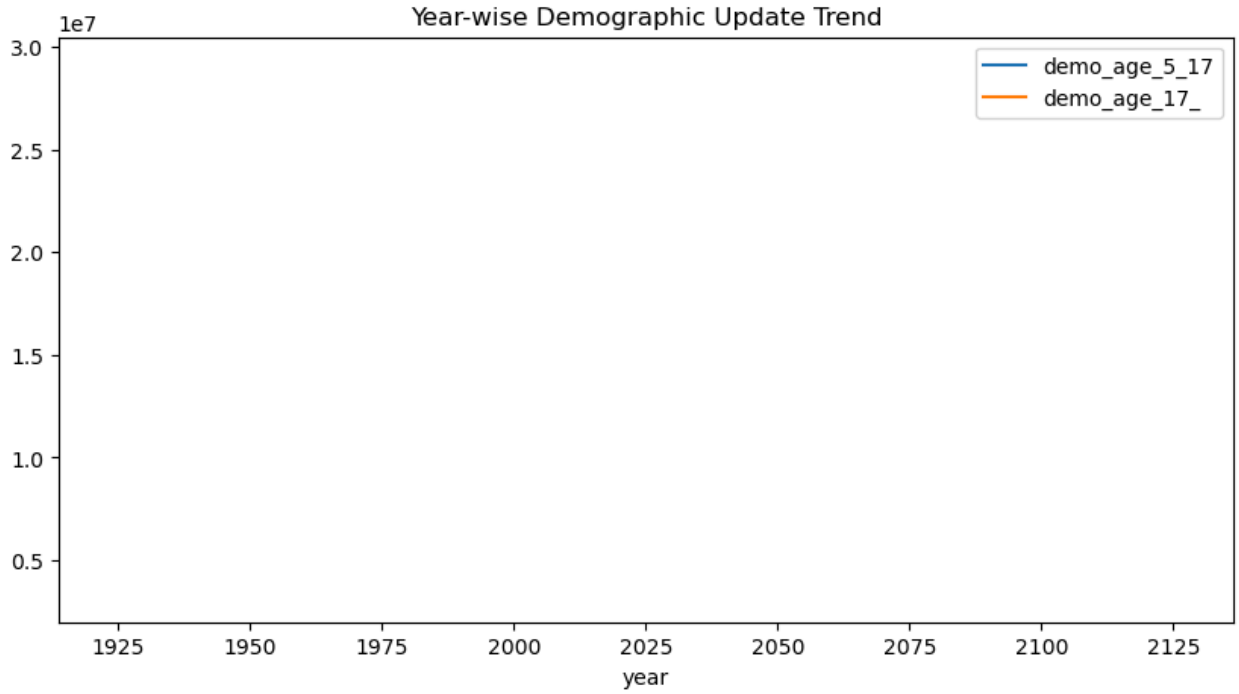
In [25]:
```
enrolment_date = enrolment_date.copy()
enrolment_date['year'] = enrolment_date['date'].dt.year
```

In [26]:
```
enrolment_date.groupby('year')[['age_0_5','age_5_17','age_18_greater']].sum().
    title='Year-wise Aadhaar Enrolment Trend',
    figsize=(10,5)
)
plt.show()
```



In [27]:
```
demo_date = demographic_date.copy()
demo_date['year'] = demo_date['date'].dt.year

demo_date.groupby('year')[['demo_age_5_17','demo_age_17_']].sum().plot(
    title='Year-wise Demographic Update Trend',
    figsize=(10,5)
)
plt.show()
```


Year-wise Demographic Update Trend
1e7
3.0
2.5
2.0
1.5
1.0
0.5
demo_age_5_17
demo_age_17_
1925
1950
1975
2000
2025
2050
2075
2100
2125
year

In [ ]: