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

In [3]:
```
data= pd.read_csv(r'D:/prodigy/task1 world population/Metadata_Country_API_SP.POP.T
```

In [4]:
```
data
```

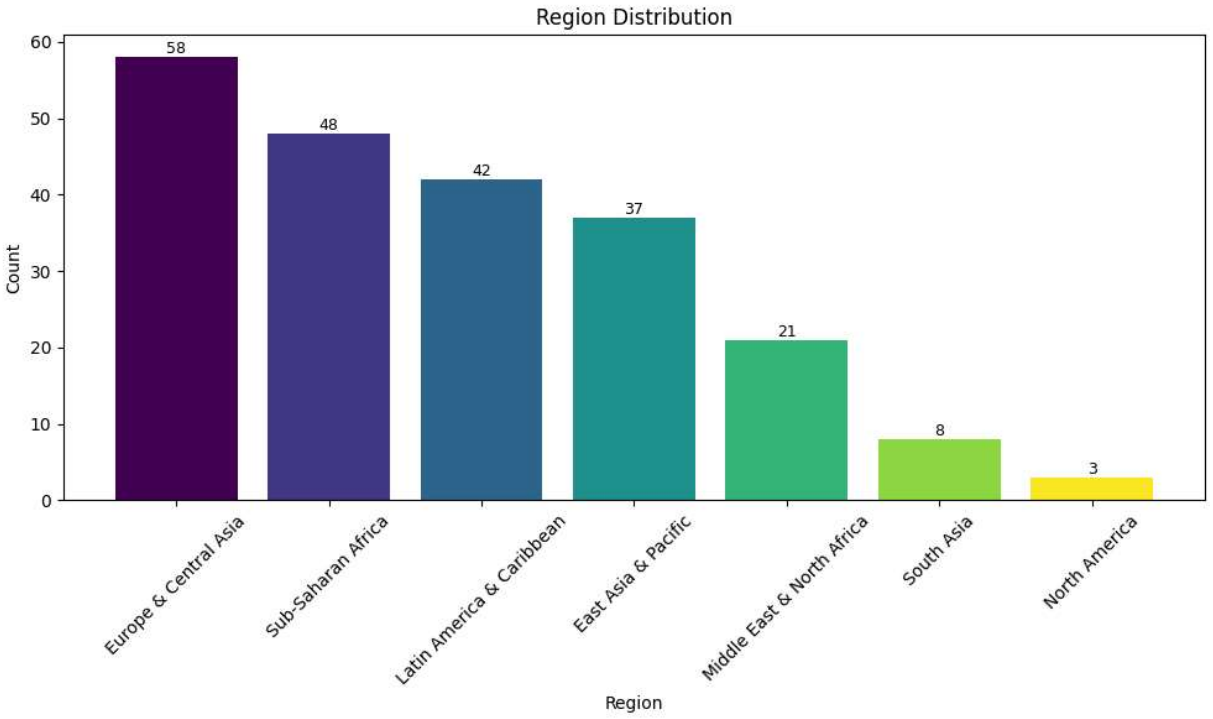
Out[4]:

	Country Code	Region	IncomeGroup	SpecialNotes	TableName	Unnamed: 5
0	ABW	Latin America & Caribbean	High income	NaN	Aruba	NaN
1	AFE	NaN	NaN	26 countries, stretching from the Red Sea in t...	Africa Eastern and Southern	NaN
2	AFG	South Asia	Low income	The reporting period for national accounts dat...	Afghanistan	NaN
3	AFW	NaN	NaN	22 countries, stretching from the westernmost ...	Africa Western and Central	NaN
4	AGO	Sub-Saharan Africa	Lower middle income	The World Bank systematically assesses the app...	Angola	NaN
...	...	...	...	...	...	...
260	XKX	Europe & Central Asia	Upper middle income	NaN	Kosovo	NaN
261	YEM	Middle East & North Africa	Low income	The World Bank systematically assesses the app...	Yemen, Rep.	NaN
262	ZAF	Sub-Saharan Africa	Upper middle income	Fiscal year end: March 31; reporting period fo...	South Africa	NaN
263	ZMB	Sub-Saharan Africa	Lower middle income	National accounts data were rebased to reflect...	Zambia	NaN
264	ZWE	Sub-Saharan Africa	Lower middle income	National Accounts data are reported in Zimbabw...	Zimbabwe	NaN

265 rows × 6 columns

In [5]:
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

region_counts = data['Region'].value_counts()

colors = plt.cm.viridis(np.linspace(0, 1, len(region_counts.index)))
```

```
plt.figure(figsize=(10, 6))
bars = plt.bar(region_counts.index, region_counts.values, color=colors)

plt.xlabel('Region')
plt.ylabel('Count')
plt.title('Region Distribution')

plt.xticks(rotation=45)

for bar in bars:
    plt.text(bar.get_x() + bar.get_width()/2, bar.get_height(), round(bar.get_heigh
             ha='center', va='bottom', fontsize=9)
plt.tight_layout()
plt.show()
```



In [6]:
```
data.shape
```

Out[6]: (265, 6)

In [7]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 265 entries, 0 to 264
Data columns (total 6 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   Country Code  265 non-null    object 
 1   Region        217 non-null    object 
 2   IncomeGroup   216 non-null    object 
 3   SpecialNotes  127 non-null    object 
 4   TableName     265 non-null    object 
 5   Unnamed: 5    0 non-null      float64
dtypes: float64(1), object(5)
memory usage: 12.6+ KB
```

In [8]:
```
data.describe()
```

Out[8]:

	Unnamed: 5
count	0.0
mean	NaN
std	NaN
min	NaN
25%	NaN
50%	NaN
75%	NaN
max	NaN

In [9]:
```
data.isnull().sum()
```

Out[9]:
```
Country Code      0
Region           48
IncomeGroup      49
SpecialNotes    138
TableName         0
Unnamed: 5      265
dtype: int64
```

In [10]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 265 entries, 0 to 264
Data columns (total 6 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   Country Code  265 non-null    object 
 1   Region        217 non-null    object 
 2   IncomeGroup   216 non-null    object 
 3   SpecialNotes  127 non-null    object 
 4   TableName     265 non-null    object 
 5   Unnamed: 5    0 non-null      float64
dtypes: float64(1), object(5)
memory usage: 12.6+ KB
```

In [11]:
```
data['Region'].fillna(data['Region'], inplace=True)
data['IncomeGroup'].fillna(data['IncomeGroup'], inplace=True)
```

In [12]:
```
data['SpecialNotes'] = data['SpecialNotes'].fillna('No Special Notes')
```

In [13]:
```
data = data.rename(columns={'Unnamed: 5': 'Gender'})
data['Gender'] = data['Gender'].fillna('others')
```

In [14]:
```
data
```

Out[14]:

	Country Code	Region	IncomeGroup	SpecialNotes	TableName	Gender
0	ABW	Latin America & Caribbean	High income	No Special Notes	Aruba	others
1	AFE	NaN	NaN	26 countries, stretching from the Red Sea in t...	Africa Eastern and Southern	others
2	AFG	South Asia	Low income	The reporting period for national accounts dat...	Afghanistan	others
3	AFW	NaN	NaN	22 countries, stretching from the westernmost ...	Africa Western and Central	others
4	AGO	Sub-Saharan Africa	Lower middle income	The World Bank systematically assesses the app...	Angola	others
...	...	...	...	...	...	...
260	XKX	Europe & Central Asia	Upper middle income	No Special Notes	Kosovo	others
261	YEM	Middle East & North Africa	Low income	The World Bank systematically assesses the app...	Yemen, Rep.	others
262	ZAF	Sub-Saharan Africa	Upper middle income	Fiscal year end: March 31; reporting period fo...	South Africa	others
263	ZMB	Sub-Saharan Africa	Lower middle income	National accounts data were rebased to reflect...	Zambia	others
264	ZWE	Sub-Saharan Africa	Lower middle income	National Accounts data are reported in Zimbabw...	Zimbabwe	others

265 rows × 6 columns

In [15]:
```
data = data.rename(columns={'Unnamed: 5': 'Gender'})
data
```

Out[15]:

	Country Code	Region	IncomeGroup	SpecialNotes	TableName	Gender
0	ABW	Latin America & Caribbean	High income	No Special Notes	Aruba	others
1	AFE	NaN	NaN	26 countries, stretching from the Red Sea in t...	Africa Eastern and Southern	others
2	AFG	South Asia	Low income	The reporting period for national accounts dat...	Afghanistan	others
3	AFW	NaN	NaN	22 countries, stretching from the westernmost ...	Africa Western and Central	others
4	AGO	Sub-Saharan Africa	Lower middle income	The World Bank systematically assesses the app...	Angola	others
...	...	...	...	...	...	...
260	XKX	Europe & Central Asia	Upper middle income	No Special Notes	Kosovo	others
261	YEM	Middle East & North Africa	Low income	The World Bank systematically assesses the app...	Yemen, Rep.	others
262	ZAF	Sub-Saharan Africa	Upper middle income	Fiscal year end: March 31; reporting period fo...	South Africa	others
263	ZMB	Sub-Saharan Africa	Lower middle income	National accounts data were rebased to reflect...	Zambia	others
264	ZWE	Sub-Saharan Africa	Lower middle income	National Accounts data are reported in Zimbabw...	Zimbabwe	others

265 rows × 6 columns

In [16]:
```
data
```

Out[16]:

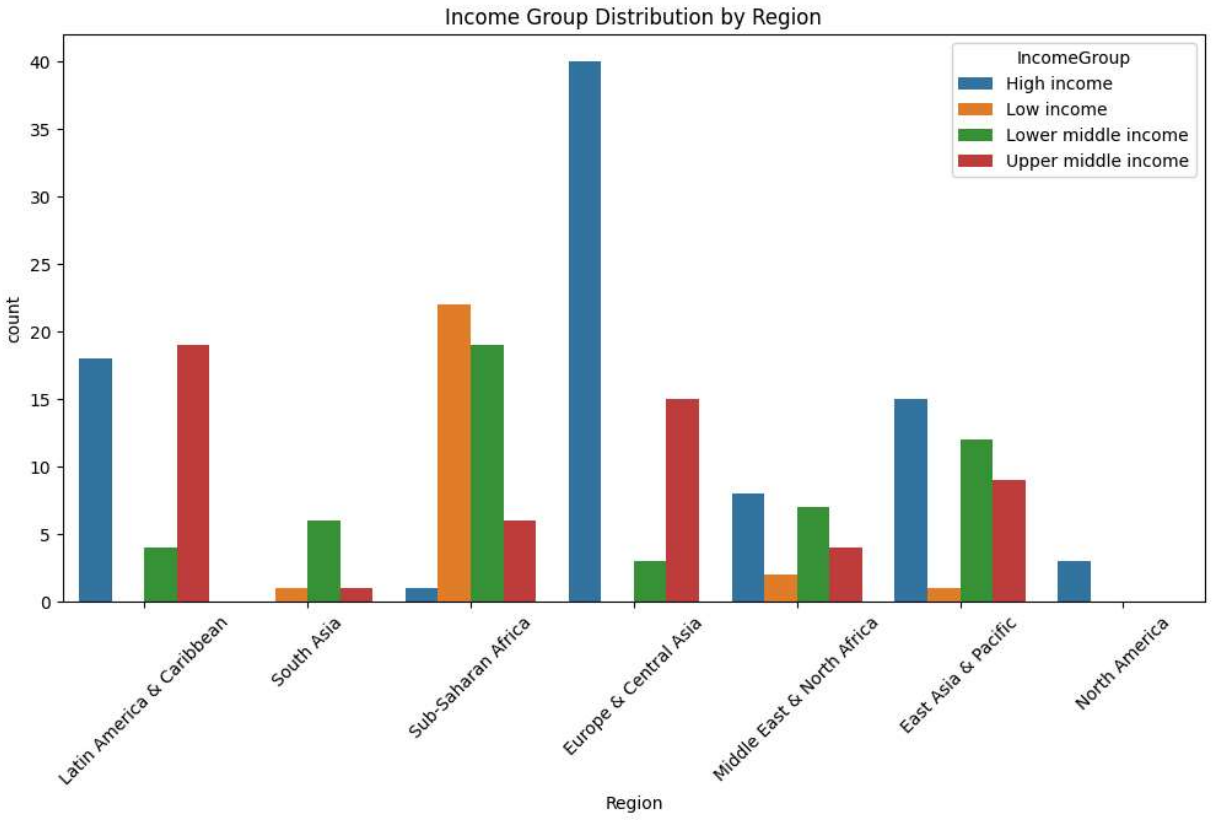
	Country Code	Region	IncomeGroup	SpecialNotes	TableName	Gender
0	ABW	Latin America & Caribbean	High income	No Special Notes	Aruba	others
1	AFE	NaN	NaN	26 countries, stretching from the Red Sea in t...	Africa Eastern and Southern	others
2	AFG	South Asia	Low income	The reporting period for national accounts dat...	Afghanistan	others
3	AFW	NaN	NaN	22 countries, stretching from the westernmost ...	Africa Western and Central	others
4	AGO	Sub-Saharan Africa	Lower middle income	The World Bank systematically assesses the app...	Angola	others
...	...	...	...	...	...	...
260	XKX	Europe & Central Asia	Upper middle income	No Special Notes	Kosovo	others
261	YEM	Middle East & North Africa	Low income	The World Bank systematically assesses the app...	Yemen, Rep.	others
262	ZAF	Sub-Saharan Africa	Upper middle income	Fiscal year end: March 31; reporting period fo...	South Africa	others
263	ZMB	Sub-Saharan Africa	Lower middle income	National accounts data were rebased to reflect...	Zambia	others
264	ZWE	Sub-Saharan Africa	Lower middle income	National Accounts data are reported in Zimbabw...	Zimbabwe	others

265 rows × 6 columns

In [17]:
```
data.isnull().sum()
```

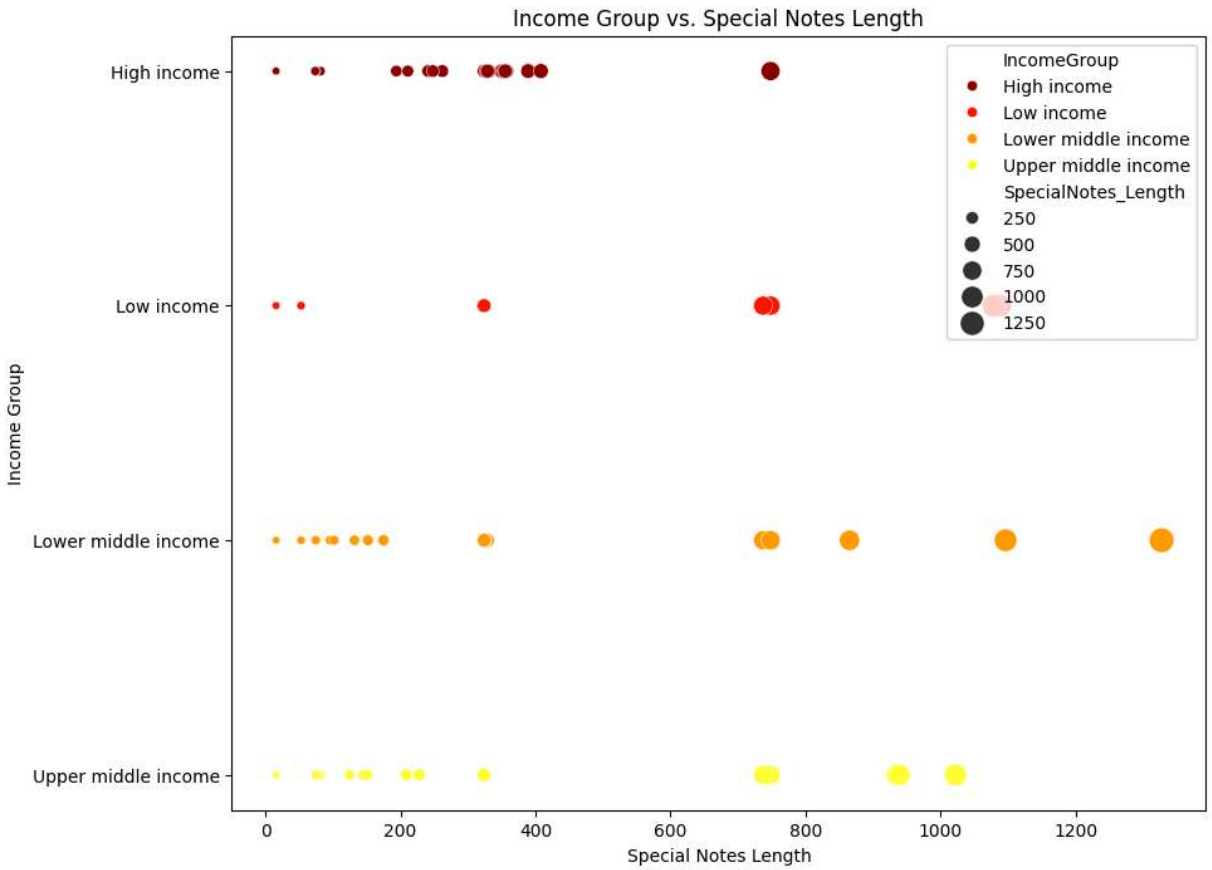
Out[17]:
```
Country Code     0
Region          48
IncomeGroup     49
SpecialNotes     0
TableName        0
Gender           0
dtype: int64
```

In [18]:
```
import seaborn as sns
import matplotlib.pyplot as plt

# Count plot for Income Group by Region
plt.figure(figsize=(12, 6))
sns.countplot(data=data, x='Region', hue='IncomeGroup')
plt.title('Income Group Distribution by Region')
plt.xticks(rotation=45)
plt.show()
```



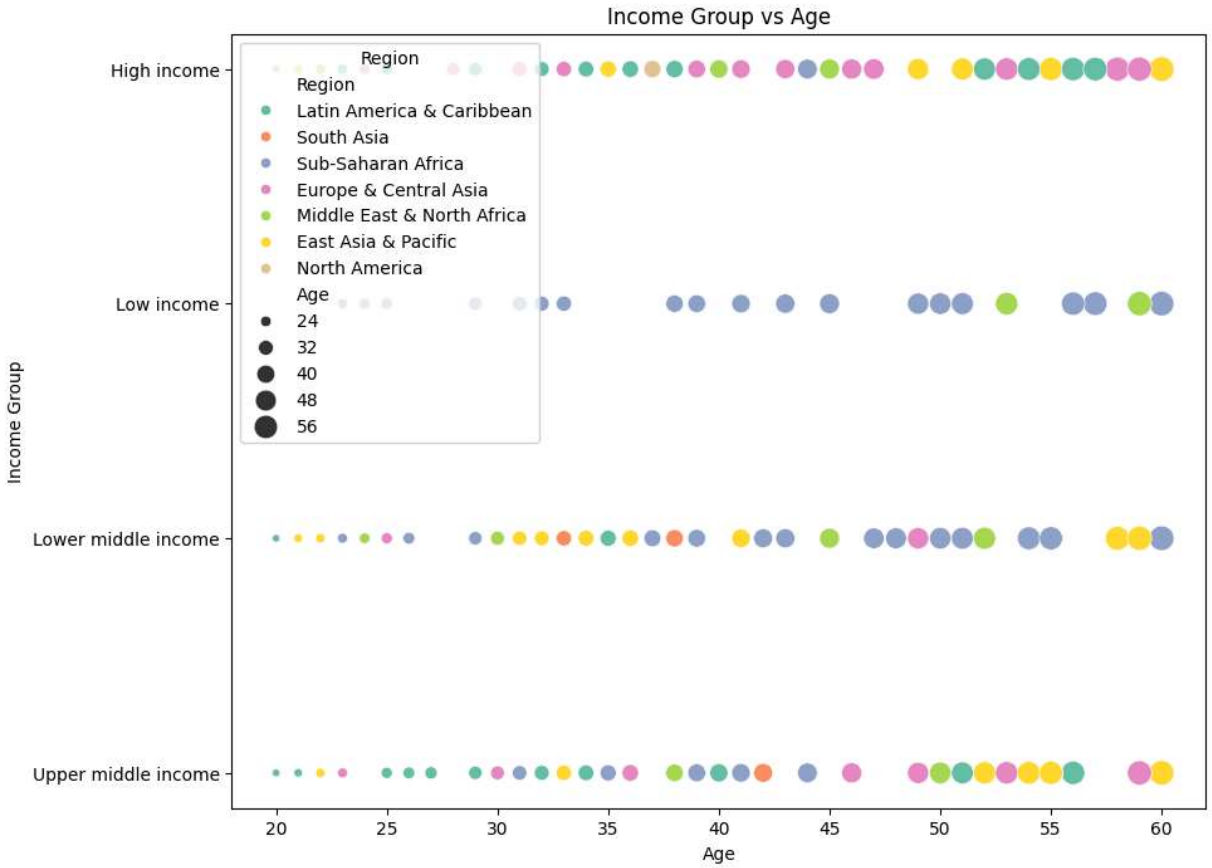
In [19]:
```
from wordcloud import WordCloud

text = ' '.join(data['SpecialNotes'].dropna())

wordcloud = WordCloud(width=800, height=400, background_color='white').generate(tex

plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.title('Word Cloud of Special Notes')
plt.show()
```

In [20]:
```
data['SpecialNotes_Length'] = data['SpecialNotes'].str.len()

# Plotting the scatter plot
plt.figure(figsize=(10, 8))
sns.scatterplot(x='SpecialNotes_Length', y='IncomeGroup', hue='IncomeGroup', data=d
plt.title('Income Group vs. Special Notes Length')
plt.xlabel('Special Notes Length')
plt.ylabel('Income Group')
plt.show()
```

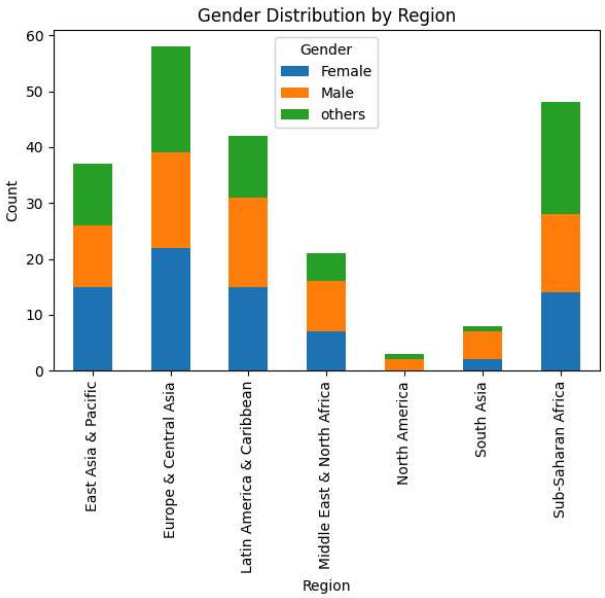


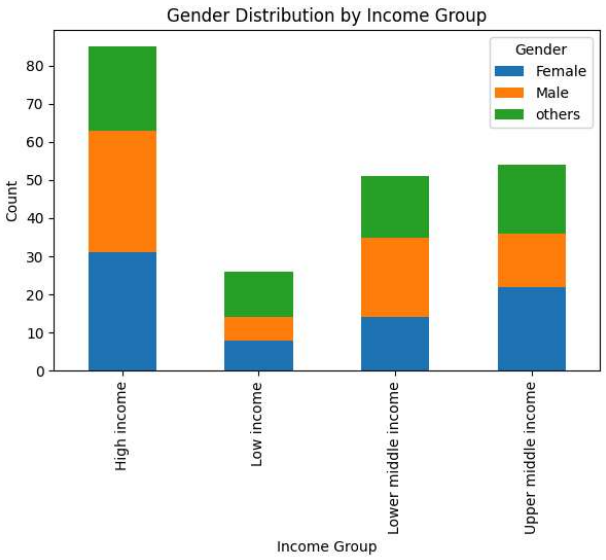
In [21]:
```
import numpy as np
```

In [22]:
```
data = pd.DataFrame(data)

np.random.seed(0)  # For reproducibility
ages = np.random.randint(20, 61, size=len(data))

data['Age'] = ages
plt.figure(figsize=(10, 8))
sns.scatterplot(x='Age', y='IncomeGroup', hue='Region', size='Age', sizes=(20, 200)
plt.title('Income Group vs Age')
plt.xlabel('Age')
plt.ylabel('Income Group')
plt.legend(title='Region', loc='upper left')
plt.show()
```



In [23]:
```
genders = ['Male', 'Female', 'others']
data['Gender'] = np.random.choice(genders, size=len(data))

# Plotting bar chart for Gender vs Region
plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
region_gender_counts = data.groupby(['Region', 'Gender']).size().unstack(fill_value
region_gender_counts.plot(kind='bar', stacked=True, ax=plt.gca())
plt.title('Gender Distribution by Region')
plt.xlabel('Region')
plt.ylabel('Count')

# Plotting bar chart for Gender vs IncomeGroup
plt.subplot(1, 2, 2)
income_gender_counts = data.groupby(['IncomeGroup', 'Gender']).size().unstack(fill_
income_gender_counts.plot(kind='bar', stacked=True, ax=plt.gca())
plt.title('Gender Distribution by Income Group')
plt.xlabel('Income Group')
plt.ylabel('Count')

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


Gender Distribution by Region
Gender
Female
Male
others
Count
0
10
20
30
40
50
60
East Asia & Pacific
Europe & Central Asia
Latin America & Caribbean
Middle East & North Africa
North America
South Asia
Sub-Saharan Africa
Region


Gender Distribution by Income Group
Gender
Female
Male
others
Count
0
10
20
30
40
50
60
70
80
High income
Low income
Lower middle income
Upper middle income
Income Group

In [ ]: