Mall_Customer's Data Analysis

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

In [2]:
```
data = pd.read_csv('Mall_Customers.csv')
```

Data Observation

In [3]:
```
data
```

Out[3]:

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)	City
0	1	Male	19	15	39.0	New York
1	2	Male	21	15	81.0	Seattle
2	3	Female	20	16	6.0	Los Angeles.
3	4	Female	23	16	77.0	Chicago.
4	5	Female	31	17	40.0	Houston.
...	...	...	...	...	...	...
1175	1176	Female	47	88	73.0	Chicago.
1176	1177	Male	48	88	10.0	Houston.
1177	1178	Male	49	88	72.0	Phoenix.
1178	1179	Male	50	93	5.0	Philadelphia.
1179	1180	Male	51	93	93.0	San Antonio.

1180 rows × 6 columns

In [4]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1180 entries, 0 to 1179
Data columns (total 6 columns):
 #   Column                  Non-Null Count  Dtype  
---  ------                  --------------  -----  
 0   CustomerID              1180 non-null   int64  
 1   Gender                  1180 non-null   object 
 2   Age                     1180 non-null   int64  
 3   Annual Income (k$)      1180 non-null   int64  
 4   Spending Score (1-100)  942 non-null    float64
 5   City                    1180 non-null   object 
dtypes: float64(1), int64(3), object(2)
memory usage: 55.4+ KB
```

In [5]:
```
data.describe()
```

Out[5]:

	CustomerID	Age	Annual Income (k$)	Spending Score (1-100)
count	1180.000000	1180.000000	1180.000000	942.000000
mean	590.500000	37.718644	59.727119	50.152866
std	340.780966	12.703662	25.574620	25.255422
min	1.000000	18.000000	15.000000	1.000000
25%	295.750000	28.000000	40.000000	35.000000
50%	590.500000	36.000000	61.000000	50.000000
75%	885.250000	47.000000	77.000000	72.000000
max	1180.000000	70.000000	137.000000	99.000000

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

Out[6]:
```
(1180, 6)
```

In [7]:
```
data.dtypes
```

Out[7]:
```
CustomerID                  int64
Gender                     object
Age                         int64
Annual Income (k$)          int64
Spending Score (1-100)    float64
City                       object
dtype: object
```

In [8]:
```
data.columns
```

Out[8]:
```
Index(['CustomerID', 'Gender', 'Age', 'Annual Income (k$)',
       'Spending Score (1-100)', 'City'],
      dtype='object')
```

In [197]:
```
data.sample(5)
```

Out[197]:

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)	City
118	119	Female	51	67	43.0	Houston.
479	480	Female	41	54	13.0	San Antonio.
596	597	Female	35	126	46.0	Seattle
413	414	Female	51	20	43.0	Chicago.
25	26	Male	29	28	82.0	San Diego

In [198]:
```
df = data.copy()
```

Data Cleansing

In [5]:
```
df
```

Out[5]:

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)	City
0	1	Male	19	15	39.0	New York
1	2	Male	21	15	81.0	Seattle
2	3	Female	20	16	6.0	Los Angeles.
3	4	Female	23	16	77.0	Chicago.
4	5	Female	31	17	40.0	Houston.
...	...	...	...	...	...	...
1175	1176	Female	47	88	73.0	Chicago.
1176	1177	Male	48	88	10.0	Houston.
1177	1178	Male	49	88	72.0	Phoenix.
1178	1179	Male	50	93	5.0	Philadelphia.
1179	1180	Male	51	93	93.0	San Antonio.

1180 rows × 6 columns

In [6]:
```
data.isnull().sum()
```

Out[6]:
```
CustomerID                0
Gender                    0
Age                       0
Annual Income (k$)        0
Spending Score (1-100)  238
City                      0
dtype: int64
```

Above Data have of Spending Score have 238 Null Values

In [7]:
```
df.fillna(df['Spending Score (1-100)'].median() ,inplace=True)
```

In [8]:
```
df.isnull().sum()
```

Out[8]:
```
CustomerID                0
Gender                    0
Age                       0
Annual Income (k$)        0
Spending Score (1-100)    0
City                      0
dtype: int64
```

In [9]:
```
df['Spending Score (1-100)'].median()
```

Out[9]:
```
50.0
```

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1180 entries, 0 to 1179
Data columns (total 6 columns):
 #   Column                  Non-Null Count  Dtype  
---  ------                  --------------  -----  
 0   CustomerID              1180 non-null   int64  
 1   Gender                  1180 non-null   object 
 2   Age                     1180 non-null   int64  
 3   Annual Income (k$)      1180 non-null   int64  
 4   Spending Score (1-100)  1180 non-null   float64
 5   City                    1180 non-null   object 
dtypes: float64(1), int64(3), object(2)
memory usage: 55.4+ KB
```

In [11]:
```
df.describe()
```

Out[11]:

	CustomerID	Age	Annual Income (k$)	Spending Score (1-100)
count	1180.000000	1180.000000	1180.000000	1180.000000
mean	590.500000	37.718644	59.727119	50.122034
std	340.780966	12.703662	25.574620	22.562860
min	1.000000	18.000000	15.000000	1.000000
25%	295.750000	28.000000	40.000000	41.000000
50%	590.500000	36.000000	61.000000	50.000000
75%	885.250000	47.000000	77.000000	60.000000
max	1180.000000	70.000000	137.000000	99.000000

Creating an Age Group Table

In [12]:
```
df['Age'].max()
```

Out[12]:
```
70
```

In [13]:
```
bins = np.arange(17,74,4)
```

In [14]:
```
bins
```

Out[14]:
```
array([17, 21, 25, 29, 33, 37, 41, 45, 49, 53, 57, 61, 65, 69, 73])
```

In [15]:
```
df['Age_Group']=pd.cut(df['Age'],bins)
```

In [16]:
```
df.head()
```

Out[16]:

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)	City	Age_Group
0	1	Male	19	15	39.0	New York	(17, 21]
1	2	Male	21	15	81.0	Seattle	(17, 21]
2	3	Female	20	16	6.0	Los Angeles.	(17, 21]
3	4	Female	23	16	77.0	Chicago.	(21, 25]
4	5	Female	31	17	40.0	Houston.	(29, 33]

Creating an Categories of Segment

In [17]:
```
lst = []
m = ''
g = ''
t = ''
ct = ''
sp = ''

for i in df['Spending Score (1-100)']:
    if i > 90:
        sp = 'Spendthrift'
        lst.append(sp)
    elif i > 75:
        ct = 'Centroid'
        lst.append(ct)
    elif i > 50:
        t = 'Target'
        lst.append(t)

    elif i > 35:
        g = 'General'
        lst.append(g)
    elif i > 20:
        c = 'Careful'
        lst.append(c)
    elif i >= 1:
        m = 'Miser'
        lst.append(m)


print(lst)
#lst.append(i)
```

```
['General', 'Centroid', 'Miser', 'Centroid', 'General', 'Centroid', 'Miser',
'Spendthrift', 'Miser', 'Target', 'Miser', 'Spendthrift', 'Miser', 'Centroi
d', 'Miser', 'Centroid', 'Careful', 'Target', 'Careful', 'Spendthrift', 'Care
ful', 'Target', 'Miser', 'Target', 'Miser', 'Centroid', 'Careful', 'Target',
'Careful', 'Centroid', 'Miser', 'Target', 'Miser', 'Spendthrift', 'Miser', 'C
entroid', 'Miser', 'Target', 'Careful', 'Target', 'Careful', 'Spendthrift',
'General', 'Target', 'Careful', 'Target', 'Target', 'General', 'General', 'Ge
neral', 'Target', 'Target', 'Target', 'Target', 'General', 'General', 'Genera
l', 'General', 'Target', 'General', 'Target', 'Target', 'Target', 'Target',
'Target', 'Target', 'General', 'General', 'Target', 'General', 'Target', 'Gen
eral', 'General', 'Target', 'General', 'Target', 'Target', 'General', 'Targe
t', 'General', 'Target', 'Target', 'General', 'General', 'Target', 'General',
'Target', 'Target', 'Target', 'General', 'Target', 'General', 'General', 'Gen
eral', 'General', 'Target', 'General', 'General', 'General', 'General', 'Gene
ral', 'General', 'Target', 'Target', 'Target', 'General', 'General', 'Genera
l', 'General', 'General', 'Target', 'Target', 'General', 'General', 'Genera
l', 'General', 'General', 'Target', 'General', 'Target', 'Target', 'General',
'Target', 'Spendthrift', 'Careful', 'Centroid', 'Careful', 'Spendthrift', 'Mi
ser', 'Target', 'Miser', 'Target', 'Careful', 'Target', 'Miser', 'Centroid',
'Miser', 'Target', 'Miser', 'Target', 'Miser', 'Spendthrift', 'General', 'Cen
troid', 'Miser', 'Spendthrift', 'General', 'Target', 'Careful', 'Centroid',
'Miser', 'Centroid', 'Miser', 'Centroid', 'Miser', 'Centroid', 'Miser', 'Cent
roid', 'Miser', 'Target', 'Careful', 'Centroid', 'Miser', 'Spendthrift', 'Car
eful', 'Target', 'Miser', 'Spendthrift', 'Careful', 'Target', 'Miser', 'Targe
```

```
t', 'Miser', 'Spendthrift', 'Miser', 'Centroid', 'Miser', 'Target', 'Miser',
'Centroid', 'Careful', 'Centroid', 'Miser', 'Centroid', 'General', 'Spendthri
ft', 'Careful', 'Target', 'Miser', 'Centroid', 'Careful', 'Target', 'Miser',
'Spendthrift', 'Miser', 'Centroid', 'Careful', 'Target', 'Miser', 'Centroid',
'General', 'General', 'General', 'General', 'General', 'General', 'General',
'General', 'General', 'General', 'General', 'General', 'General', 'General',
'General', 'General', 'General', 'General', 'General', 'General', 'General',
'General', 'General', 'General', 'General', 'General', 'General', 'General',
'General', 'General', 'General', 'General', 'General', 'General', 'General',
'General', 'General', 'General', 'General', 'General', 'General', 'General',
'General', 'General', 'General', 'General', 'General', 'General', 'General',
'General', 'General', 'General', 'General', 'General', 'General', 'General',
'General', 'General', 'General', 'General', 'General', 'General', 'General',
'General', 'General', 'General', 'General', 'General', 'General', 'General',
'General', 'General', 'General', 'General', 'General', 'General', 'General',
'General', 'General', 'General', 'General', 'General', 'General', 'General',
'General', 'General', 'General', 'General', 'General', 'General', 'General',
'General', 'General', 'General', 'General', 'General', 'General', 'General',
'General', 'General', 'General', 'General', 'General', 'General', 'General',
'General', 'Centroid', 'Miser', 'Centroid', 'General', 'Centroid', 'Miser',
'Spendthrift', 'Miser', 'Target', 'Miser', 'Spendthrift', 'Miser', 'Centroi
d', 'Miser', 'Centroid', 'Careful', 'Target', 'Careful', 'Spendthrift', 'Care
ful', 'Target', 'Miser', 'Target', 'Miser', 'Centroid', 'Careful', 'Target',
'Careful', 'Centroid', 'Miser', 'Target', 'Miser', 'Spendthrift', 'Miser', 'C
entroid', 'Miser', 'Target', 'Careful', 'Target', 'Careful', 'Spendthrift',
'General', 'Target', 'Careful', 'Target', 'Target', 'General', 'General', 'Ge
neral', 'Target', 'Target', 'Target', 'Target', 'General', 'General', 'Genera
l', 'General', 'Target', 'General', 'Target', 'Target', 'Target', 'Target',
'Target', 'Target', 'General', 'General', 'Target', 'General', 'Target', 'Gen
eral', 'General', 'Target', 'General', 'Target', 'Target', 'General', 'Targe
t', 'General', 'Target', 'Target', 'General', 'General', 'Target', 'General',
'Target', 'Target', 'Target', 'General', 'Target', 'General', 'General', 'Gen
eral', 'General', 'Target', 'General', 'General', 'General', 'General', 'Gene
ral', 'General', 'Target', 'Target', 'Target', 'General', 'General', 'Genera
l', 'General', 'General', 'Target', 'Target', 'General', 'General', 'Genera
l', 'General', 'General', 'Target', 'General', 'Target', 'Target', 'General',
'Target', 'Spendthrift', 'Careful', 'Centroid', 'Careful', 'Spendthrift', 'Mi
ser', 'Target', 'Miser', 'Target', 'Careful', 'Target', 'Miser', 'Centroid',
'Miser', 'Target', 'Miser', 'Target', 'Miser', 'Spendthrift', 'General', 'Cen
troid', 'Miser', 'Spendthrift', 'General', 'Target', 'Careful', 'Centroid',
'Miser', 'Centroid', 'Miser', 'Centroid', 'Miser', 'Centroid', 'Miser', 'Cent
roid', 'Miser', 'Target', 'Careful', 'Centroid', 'Miser', 'Spendthrift', 'Car
eful', 'Target', 'Miser', 'Spendthrift', 'Careful', 'Target', 'Miser', 'Targe
t', 'Miser', 'Spendthrift', 'Miser', 'Centroid', 'Miser', 'Target', 'Miser',
'Centroid', 'Careful', 'Centroid', 'Miser', 'Centroid', 'General', 'Spendthri
ft', 'Careful', 'Target', 'Miser', 'Centroid', 'Careful', 'Target', 'Miser',
'Spendthrift', 'Miser', 'Centroid', 'Careful', 'Target', 'Miser', 'Centroid',
'General', 'General', 'General', 'Centroid', 'Miser', 'Centroid', 'General',
'Centroid', 'Miser', 'Spendthrift', 'Miser', 'Target', 'Miser', 'Spendthrif
t', 'Miser', 'Centroid', 'Miser', 'Centroid', 'Careful', 'Target', 'Careful',
'Spendthrift', 'Careful', 'Target', 'Miser', 'Target', 'Miser', 'Centroid',
'Careful', 'Target', 'Careful', 'Centroid', 'Miser', 'Target', 'Miser', 'Spen
dthrift', 'Miser', 'Centroid', 'Miser', 'Target', 'Careful', 'Target', 'Caref
ul', 'Spendthrift', 'General', 'Target', 'Careful', 'Target', 'Target', 'Gene
ral', 'General', 'General', 'Target', 'Target', 'Target', 'Target', 'Genera
l', 'General', 'General', 'General', 'Target', 'General', 'Target', 'Target',
'Target', 'Target', 'Target', 'Target', 'General', 'General', 'Target', 'Gene
```

```
ral', 'Target', 'General', 'General', 'Target', 'General', 'Target', 'Targe
t', 'General', 'Target', 'General', 'Target', 'Target', 'General', 'General',
'Target', 'General', 'Target', 'Target', 'Target', 'General', 'Target', 'Gene
ral', 'General', 'General', 'General', 'Target', 'General', 'General', 'Gener
al', 'General', 'General', 'General', 'Target', 'Target', 'Target', 'Genera
l', 'General', 'General', 'General', 'General', 'Target', 'Target', 'Genera
l', 'General', 'General', 'General', 'General', 'Target', 'General', 'Targe
t', 'Target', 'General', 'Target', 'Spendthrift', 'Careful', 'Centroid', 'Car
eful', 'Spendthrift', 'Miser', 'Target', 'Miser', 'Target', 'Careful', 'Targe
t', 'Miser', 'Centroid', 'Miser', 'Target', 'Miser', 'Target', 'Miser', 'Spen
dthrift', 'General', 'Centroid', 'Miser', 'Spendthrift', 'General', 'Target',
'Careful', 'Centroid', 'Miser', 'Centroid', 'Miser', 'Centroid', 'Miser', 'Ce
ntroid', 'Miser', 'Centroid', 'Miser', 'Target', 'Careful', 'Centroid', 'Mise
r', 'Spendthrift', 'Careful', 'Target', 'Miser', 'Spendthrift', 'Careful', 'T
arget', 'Miser', 'Target', 'Miser', 'Spendthrift', 'Miser', 'Centroid', 'Mise
r', 'Target', 'Miser', 'Centroid', 'Careful', 'Centroid', 'Miser', 'Centroi
d', 'General', 'Spendthrift', 'Careful', 'Target', 'Miser', 'Centroid', 'Care
ful', 'Target', 'Miser', 'Spendthrift', 'Miser', 'Centroid', 'Careful', 'Targ
et', 'Miser', 'Centroid', 'General', 'Centroid', 'Miser', 'Centroid', 'Genera
l', 'Centroid', 'Miser', 'Spendthrift', 'Miser', 'Target', 'Miser', 'Spendthr
ift', 'Miser', 'Centroid', 'Miser', 'Centroid', 'Careful', 'Target', 'Carefu
l', 'Spendthrift', 'Careful', 'Target', 'Miser', 'Target', 'Miser', 'Centroi
d', 'Careful', 'Target', 'Careful', 'Centroid', 'Miser', 'Target', 'Miser',
'Spendthrift', 'Miser', 'Centroid', 'Miser', 'Target', 'Careful', 'Target',
'Careful', 'Spendthrift', 'General', 'Target', 'Careful', 'Target', 'Target',
'General', 'General', 'General', 'Target', 'Target', 'Target', 'Target', 'Gen
eral', 'General', 'General', 'General', 'Target', 'General', 'Target', 'Targe
t', 'Target', 'Target', 'Target', 'Target', 'General', 'General', 'Target',
'General', 'Target', 'General', 'General', 'Target', 'General', 'Target', 'Ta
rget', 'General', 'Target', 'General', 'Target', 'Target', 'General', 'Genera
l', 'Target', 'General', 'Target', 'Target', 'Target', 'General', 'Target',
'General', 'General', 'General', 'General', 'Target', 'General', 'General',
'General', 'General', 'General', 'General', 'Target', 'Target', 'Target', 'Ge
neral', 'General', 'General', 'General', 'General', 'Target', 'Target', 'Gene
ral', 'General', 'General', 'General', 'General', 'Target', 'General', 'Targe
t', 'Target', 'General', 'Target', 'Spendthrift', 'Careful', 'Centroid', 'Car
eful', 'Spendthrift', 'Miser', 'Target', 'Miser', 'Target', 'Careful', 'Targe
t', 'Miser', 'Centroid', 'Miser', 'Target', 'Miser', 'Target', 'Miser', 'Spen
dthrift', 'General', 'Centroid', 'Miser', 'Spendthrift', 'General', 'Target',
'Careful', 'Centroid', 'Miser', 'Centroid', 'Miser', 'Centroid', 'Miser', 'Ce
ntroid', 'Miser', 'Centroid', 'Miser', 'Target', 'Careful', 'Centroid', 'Mise
r', 'Spendthrift', 'Careful', 'Target', 'Miser', 'Spendthrift', 'Careful', 'T
arget', 'Miser', 'Target', 'Miser', 'Spendthrift', 'Miser', 'Centroid', 'Mise
r', 'Target', 'Miser', 'Centroid', 'Careful', 'Centroid', 'Miser', 'Centroi
d', 'General', 'Spendthrift', 'Careful', 'Target', 'Miser', 'Centroid', 'Care
ful', 'Target', 'Miser', 'Spendthrift', 'Miser', 'Centroid', 'Careful', 'Targ
et', 'Miser', 'Centroid', 'General', 'General', 'General', 'General', 'Genera
l', 'General', 'General', 'General', 'General', 'General', 'General', 'Genera
l', 'General', 'General', 'General', 'General', 'General', 'General', 'Genera
l', 'General', 'General', 'General', 'General', 'General', 'General', 'Genera
l', 'General', 'General', 'General', 'General', 'General', 'General', 'Genera
l', 'General', 'General', 'General', 'General', 'General', 'General', 'Genera
l', 'General', 'General', 'General', 'General', 'General', 'General', 'Genera
l', 'General', 'General', 'General', 'General', 'General', 'General', 'Genera
l', 'General', 'General', 'General', 'General', 'General', 'General', 'Genera
l', 'General', 'General', 'General', 'General', 'General', 'General', 'Genera
l', 'General', 'General', 'General', 'General', 'General', 'General', 'Genera
```

```
l', 'General', 'General', 'General', 'General', 'General', 'General', 'Genera
l', 'General', 'General', 'General', 'General', 'General', 'General', 'Genera
l', 'General', 'General', 'General', 'General', 'General', 'General', 'Genera
l', 'General', 'General', 'General', 'General', 'General', 'General', 'Genera
l', 'General', 'General', 'General', 'General', 'General', 'General', 'Genera
l', 'General', 'General', 'General', 'General', 'General', 'General', 'Genera
l', 'General', 'General', 'General', 'General', 'General', 'General', 'Genera
l', 'General', 'General', 'General', 'General', 'General', 'General', 'Genera
l', 'General', 'Centroid', 'Miser', 'Centroid', 'General', 'Centroid', 'Mise
r', 'Spendthrift', 'Miser', 'Target', 'Miser', 'Spendthrift', 'Miser', 'Centr
oid', 'Miser', 'Centroid', 'Careful', 'Target', 'Careful', 'Spendthrift', 'Ca
reful', 'Target', 'Miser', 'Target', 'Miser', 'Centroid', 'Careful', 'Targe
t', 'Careful', 'Centroid', 'Miser', 'Target', 'Miser', 'Spendthrift', 'Mise
r', 'Centroid', 'Miser', 'Target', 'Careful', 'Target', 'Careful', 'Spendthri
ft', 'General', 'Target', 'Careful', 'Target', 'Target', 'General', 'Genera
l', 'General', 'Target', 'Target', 'Target', 'Target', 'General', 'General',
'General', 'General', 'Target', 'General', 'Target', 'Target', 'Target', 'Tar
get', 'Target', 'Target', 'General', 'General', 'Target', 'General', 'Targe
t', 'General', 'General', 'Target', 'General', 'Target', 'Target', 'General',
'Target', 'General', 'Target', 'Target', 'General', 'General', 'Target', 'Gen
eral', 'Target', 'Target', 'Target', 'General', 'Target', 'General', 'Genera
l', 'General', 'General', 'Target', 'General', 'General', 'General', 'Genera
l', 'General', 'General', 'Target', 'Target', 'Target', 'General', 'General',
'General', 'General', 'General', 'Target', 'Target', 'General', 'General', 'G
eneral', 'General', 'General', 'Target', 'General', 'Target', 'Target', 'Gene
ral', 'Target', 'Spendthrift', 'Careful', 'Centroid', 'Careful', 'Spendthrif
t', 'Miser', 'Target', 'Miser', 'Target', 'Careful', 'Target', 'Miser', 'Cent
roid', 'Miser', 'Target', 'Miser', 'Target', 'Miser', 'Spendthrift']
```

In [18]:
```
lst
```

```
 'Target',
 'Target',
 'Target',
 'General',
 'General',
 'General',
 'General',
 'Target',
 'General',
 'Target',
 'Target',
 'Target',

 'Target',
 'Target',
 'Target',
 'General',
 'General',
 'Target',
 'General',
 'Target'
```

In [159]:
```
segment = []
```

In [160]:
```
segment = pd.DataFrame(lst)
```

In [161]:
```
segment['Segment'] = pd.DataFrame(lst)
```

In [162]:
```
segment
```

Out[162]:

	0	Segment
0	General	General
1	Centroid	Centroid
2	Miser	Miser
3	Centroid	Centroid
4	General	General
...	...	...
1175	Target	Target
1176	Miser	Miser
1177	Target	Target
1178	Miser	Miser
1179	Spendthrift	Spendthrift

1180 rows × 2 columns

In [163]:
```
segment.drop(columns=0,axis=1)
```

Out[163]:

	Segment
0	General
1	Centroid
2	Miser
3	Centroid
4	General
...	...
1175	Target
1176	Miser
1177	Target
1178	Miser
1179	Spendthrift

1180 rows × 1 columns

In [164]:
```
df1 = df.join(segment)
```

In [165]:
```
df1.drop(columns=0,axis=1,inplace=True)
```

In [166]:
```
df1['City'].value_counts()
```

Out[166]:
```
Los Angeles.     135
Chicago.         135
Houston.         135
Phoenix.         135
Philadelphia.    135
San Antonio.     135
San Diego        134
Washington       100
New York          68
Seattle           68
Name: City, dtype: int64
```

Removing '.' from City columns

In [167]:
```
df1['City'] = df1['City'].replace(['Los Angeles.','Chicago.','Houston.','Phoenix.
```

In [169]:
```
df1['City'].value_counts()
```

Out[169]:
```
Los Angeles     135
Chicago         135
Houston         135
Phoenix         135
Philadelphia    135
San Antonio     135
San Diego       134
Washington      100
New York         68
Seattle          68
Name: City, dtype: int64
```

In [29]:
```
df1.head(7)
```

Out[29]:

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)	City	Age_Group	Segment
0	1	Male	19	15	39.0	New York	(17, 21]	General
1	2	Male	21	15	81.0	Seattle	(17, 21]	Centroid
2	3	Female	20	16	6.0	Los Angeles	(17, 21]	Miser
3	4	Female	23	16	77.0	Chicago	(21, 25]	Centroid
4	5	Female	31	17	40.0	Houston	(29, 33]	General
5	6	Female	22	17	76.0	Phoenix	(21, 25]	Centroid
6	7	Female	35	18	6.0	Philadelphia	(33, 37]	Miser

In [30]:
```
df1.isnull().sum()
```

Out[30]:
```
CustomerID              0
Gender                  0
Age                     0
Annual Income (k$)      0
Spending Score (1-100)  0
City                    0
Age_Group               0
Segment                 0
dtype: int64
```

Craeting Generation Column on the bases of Age_Group

In [31]:
```python
gen = []
y = ''
ad = ''
ee= ''
old = ''
for i in df['Age']:
    if i > 17 and i <= 25 :
        y = 'Young'
        gen.append(y)
    elif i > 25 and i <= 35  :
        ad = 'AdultHood'
        gen.append(ad)
    elif i > 35 and i <= 50:
        ee = 'Early Elder'
        gen.append(ee)
    elif i > 50 and i <= 70 :
        old = 'Elder/Old'
        gen.append(old)
    else :
        pass
print (gen)
```

```
r', 'AdultHood', 'AdultHood', 'AdultHood', 'Early Elder', 'Early Elder', 'Ear
ly Elder', 'Early Elder', 'Early Elder', 'AdultHood', 'Early Elder', 'AdultHo
od', 'AdultHood', 'AdultHood', 'Elder/Old', 'AdultHood', 'Young', 'AdultHoo
d', 'Early Elder', 'Early Elder', 'Early Elder', 'AdultHood', 'Early Elder',
'AdultHood', 'Early Elder', 'AdultHood', 'Early Elder', 'Early Elder', 'Elde
r/Old', 'AdultHood', 'Elder/Old', 'AdultHood', 'Elder/Old', 'AdultHood', 'Ear
ly Elder', 'AdultHood', 'Early Elder', 'AdultHood', 'Early Elder', 'AdultHoo
d', 'Elder/Old', 'AdultHood', 'Early Elder', 'Early Elder', 'AdultHood', 'Adu
ltHood', 'AdultHood', 'Early Elder', 'Early Elder', 'AdultHood', 'Early Elde

r', 'AdultHood', 'AdultHood', 'AdultHood', 'Young', 'Young', 'Young', 'Youn
g', 'Young', 'Young', 'AdultHood', 'AdultHood', 'AdultHood', 'AdultHood', 'Ad
ultHood', 'AdultHood', 'AdultHood', 'AdultHood', 'AdultHood', 'AdultHood', 'E
arly Elder', 'Early Elder', 'Early Elder', 'Early Elder', 'Early Elder', 'Ear
ly Elder', 'Early Elder', 'Early Elder', 'Early Elder', 'Early Elder', 'Early
 Elder', 'Early Elder', 'Early Elder', 'Early Elder', 'Early Elder', 'Young',
'Young', 'Young', 'Young', 'Young', 'Young', 'AdultHood', 'AdultHood', 'Adult
Hood', 'AdultHood', 'AdultHood', 'AdultHood', 'AdultHood', 'AdultHood', 'Adul
tHood', 'AdultHood', 'Early Elder', 'Early Elder', 'Early Elder', 'Early Elde
r', 'Early Elder', 'Early Elder', 'Early Elder', 'Early Elder', 'Early Elde
```

In [32]:
```python
df1['Generation'] = pd.DataFrame(gen)
```

In [33]:
```
df1.isnull().sum()
```

Out[33]:
```
CustomerID                0
Gender                    0
Age                       0
Annual Income (k$)        0
Spending Score (1-100)    0
City                      0
Age_Group                 0
Segment                   0
Generation                0
dtype: int64
```

In [34]:
```
df1.sample(5)
```

Out[34]:

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)	City	Age_Group	Segment	Generation
363	364	Female	19	81	51.0	Houston	(17, 21]	Target	Young
369	370	Male	59	87	51.0	Los Angeles	(57, 61]	Target	Elder/Old
1005	1006	Female	27	17	50.0	San Diego	(25, 29]	General	AdultHood
1112	1113	Female	47	64	47.0	Los Angeles	(45, 49]	General	Early Elder
352	353	Female	69	78	47.0	Los Angeles	(65, 69]	General	Elder/Old

Creating an column of Annual Income_Group

In [35]:
```
bins1 = np.arange(1,101,11)
```

In [36]:
```
bins1
```

Out[36]:
```
array([  1,  12,  23,  34,  45,  56,  67,  78,  89, 100])
```

In [37]:
```
df1['Annual Income_Group'] = pd.cut(df1['Annual Income (k$)'],bins1)
```

In [38]:
```
df1
```

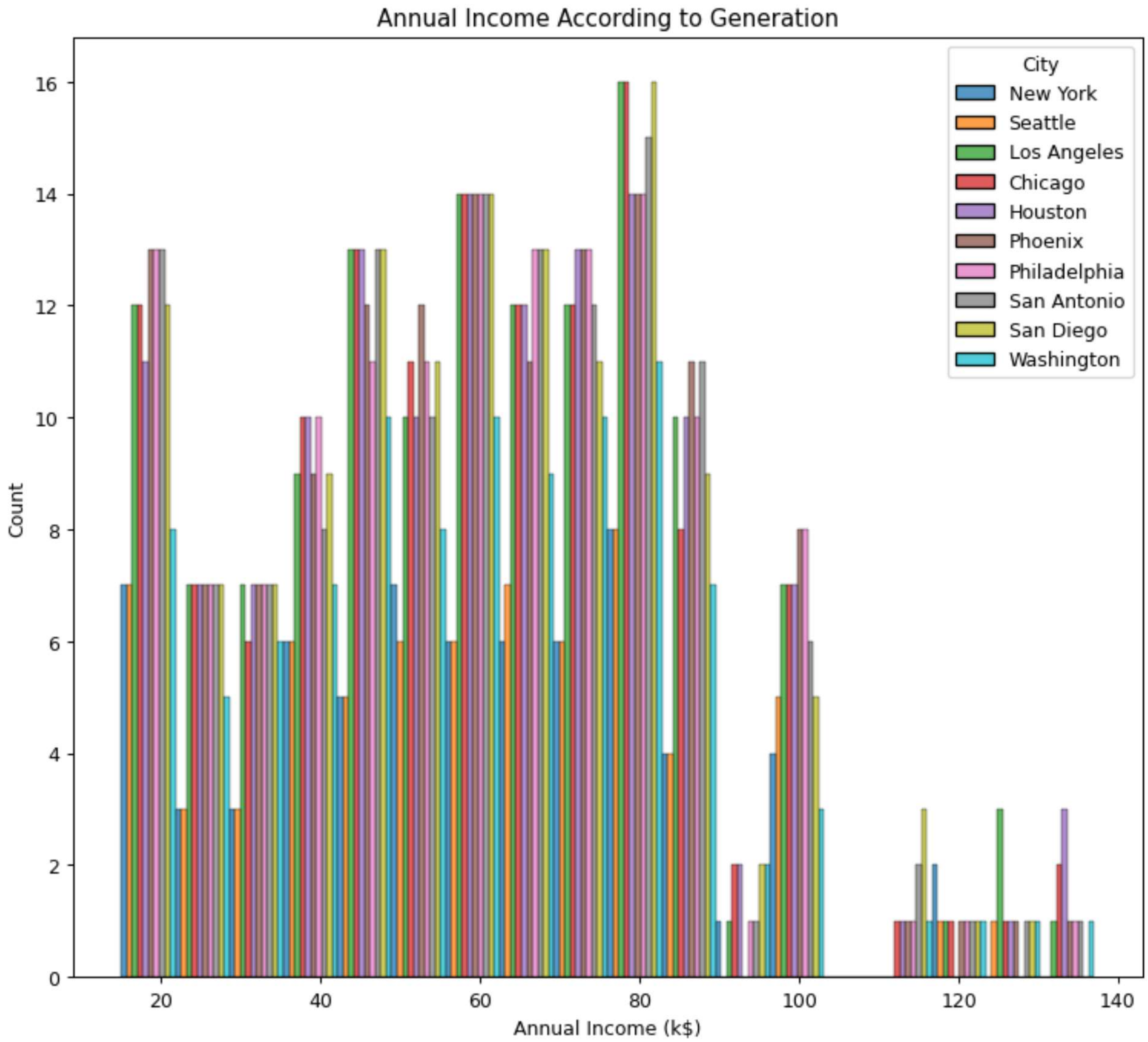
Out[38]:

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)	City	Age_Group	Segment	Generatio
0	1	Male	19	15	39.0	New York	(17, 21]	General	Your
1	2	Male	21	15	81.0	Seattle	(17, 21]	Centroid	Your
2	3	Female	20	16	6.0	Los Angeles	(17, 21]	Miser	Your
3	4	Female	23	16	77.0	Chicago	(21, 25]	Centroid	Your
4	5	Female	31	17	40.0	Houston	(29, 33]	General	AdultHoc
...	...	...	...	...	...	...	...	...	
1175	1176	Female	47	88	73.0	Chicago	(45, 49]	Target	Early Eld
1176	1177	Male	48	88	10.0	Houston	(45, 49]	Miser	Early Eld
1177	1178	Male	49	88	72.0	Phoenix	(45, 49]	Target	Early Eld
1178	1179	Male	50	93	5.0	Philadelphia	(49, 53]	Miser	Early Eld
1179	1180	Male	51	93	93.0	San Antonio	(49, 53]	Spendthrift	Elder/O

1180 rows × 10 columns

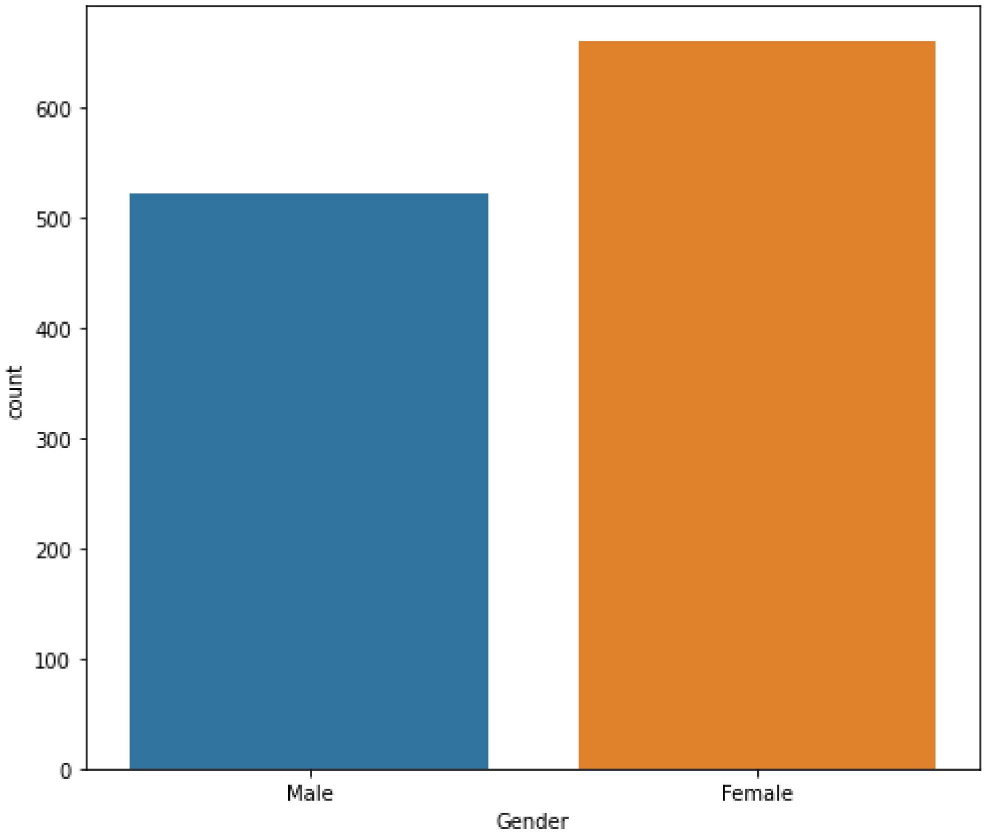
Data Visualization

In [170]:
```
fig = plt.figure(figsize=(10,9),dpi=90)
sns.histplot(data=df1,x='Annual Income (k$)',hue='City',multiple='dodge',palette=
plt.show()
```



The Above graph shows that 15 to 100 dolars approximation are having all city's

In [44]:
```
fig = plt.figure(figsize=(8,7))
sns.countplot(data=df1,x='Gender')
plt.show()
```

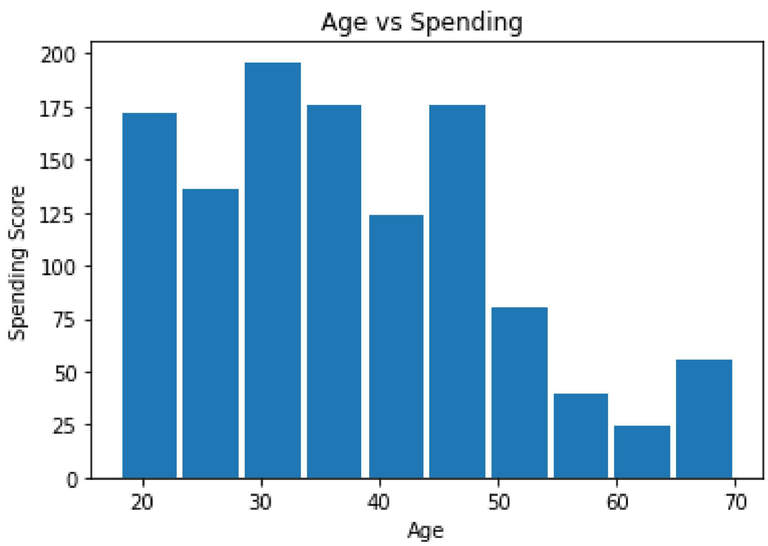


In above graph shows that Male count are above 500 and Female count are above 600 in all city's

In [181]:
```
df1['Gender'].value_counts()
```

Out[181]:
```
Female    659
Male      521
Name: Gender, dtype: int64
```

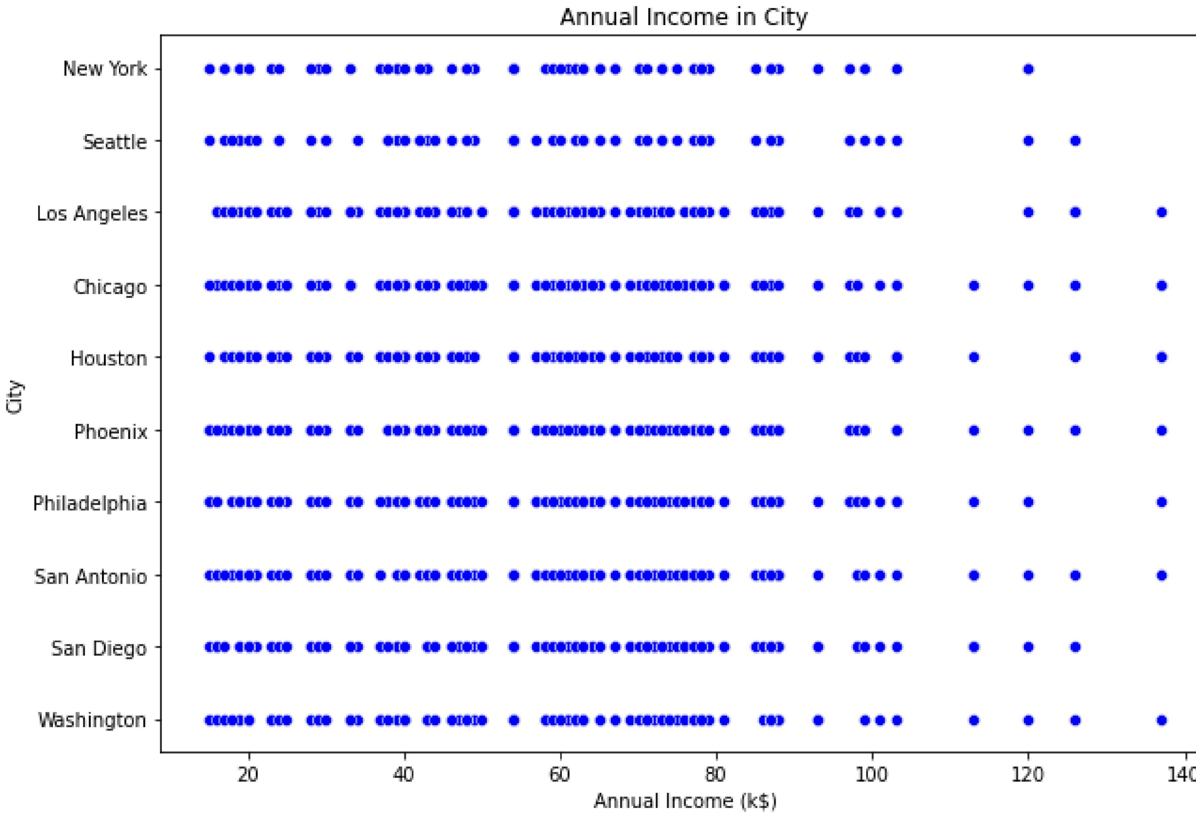
In [45]:
```
fig=plt.figure()
plt.hist(df1['Age'],rwidth=0.9)
plt.xlabel('Age')
plt.ylabel('Spending Score')
plt.title('Age vs Spending ')

plt.show()
```



In the Above graph shows that Age of 30 people are spending more in all City's

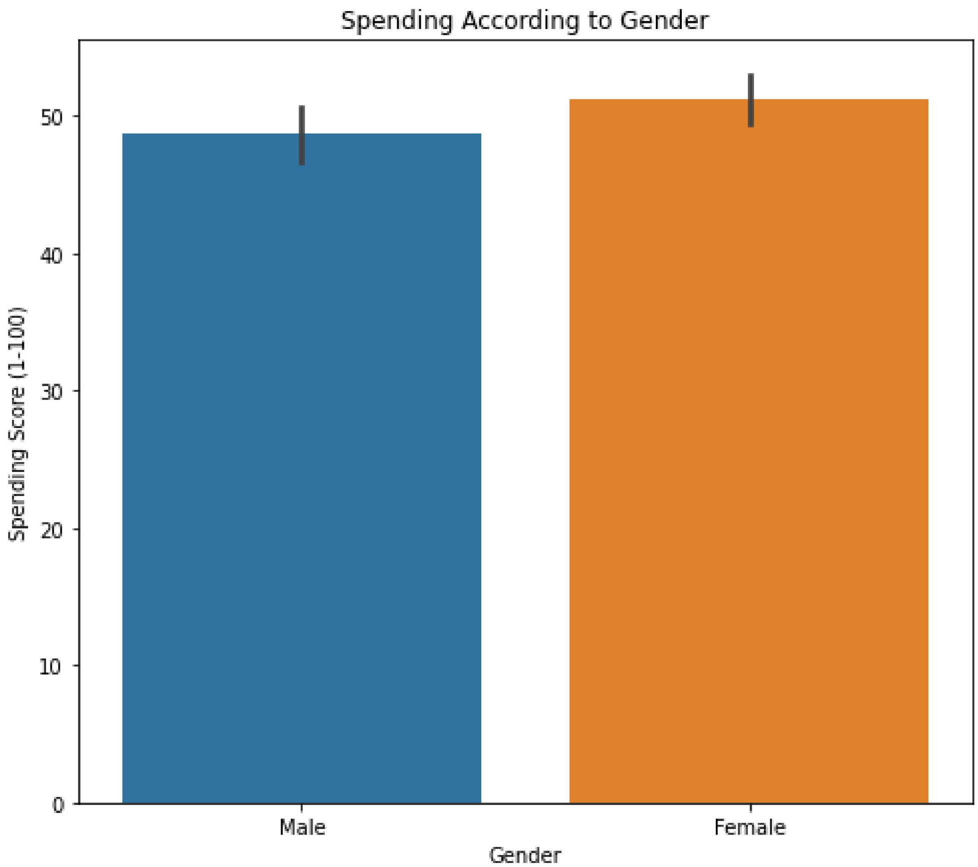
In [46]:

```
fig = plt.figure(figsize=(10,7))
sns.scatterplot(data=df1,x =df1['Annual Income (k$)'],y = df1['City'],color='blue
plt.show()
```



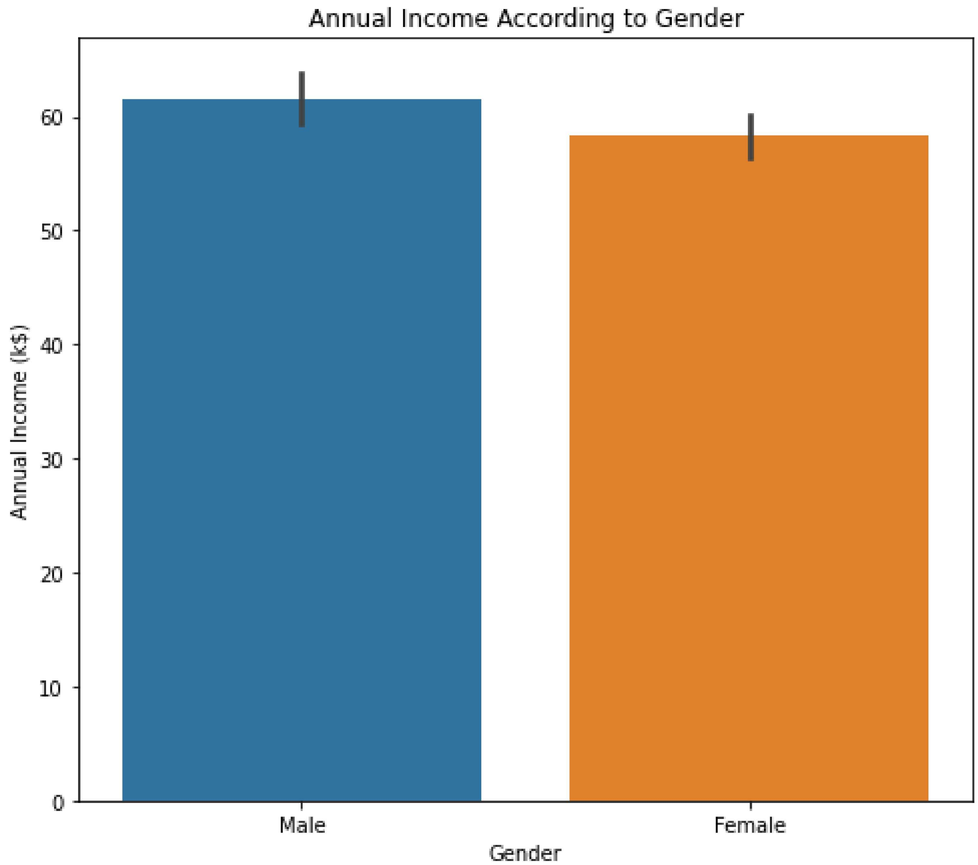
In Above graph shows that AnnualIncome on the bases of city's

In [47]:
```
fig = plt.figure(figsize=(8,7))
sns.barplot(x=df1['Gender'] ,y=df1['Spending Score (1-100)'] ,data=df1).set(title
plt.show()
```



In the Above graph shows that Female are spending more than Male in all City's

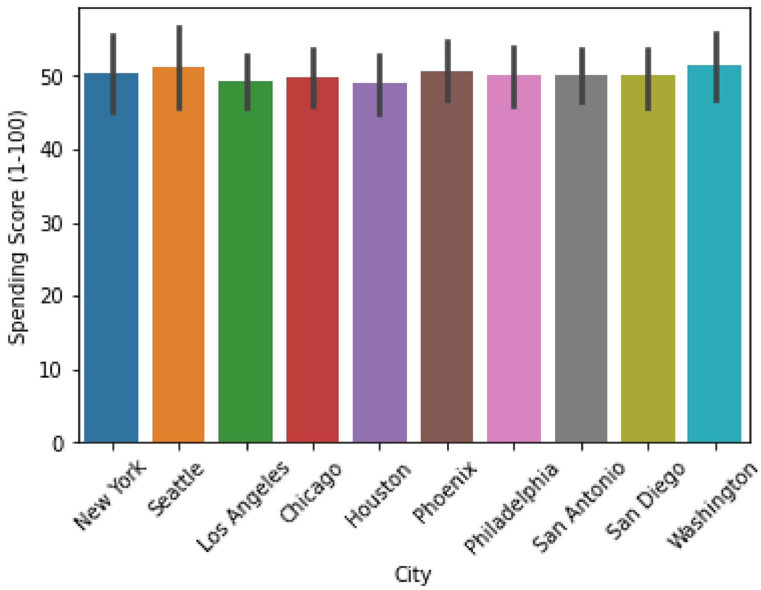
In [150]:
```
fig = plt.figure(figsize=(8,7))
sns.barplot(x=df1['Gender'] ,y=df1['Annual Income (k$)'] ,data=df1).set(title='Ar
plt.show()
```



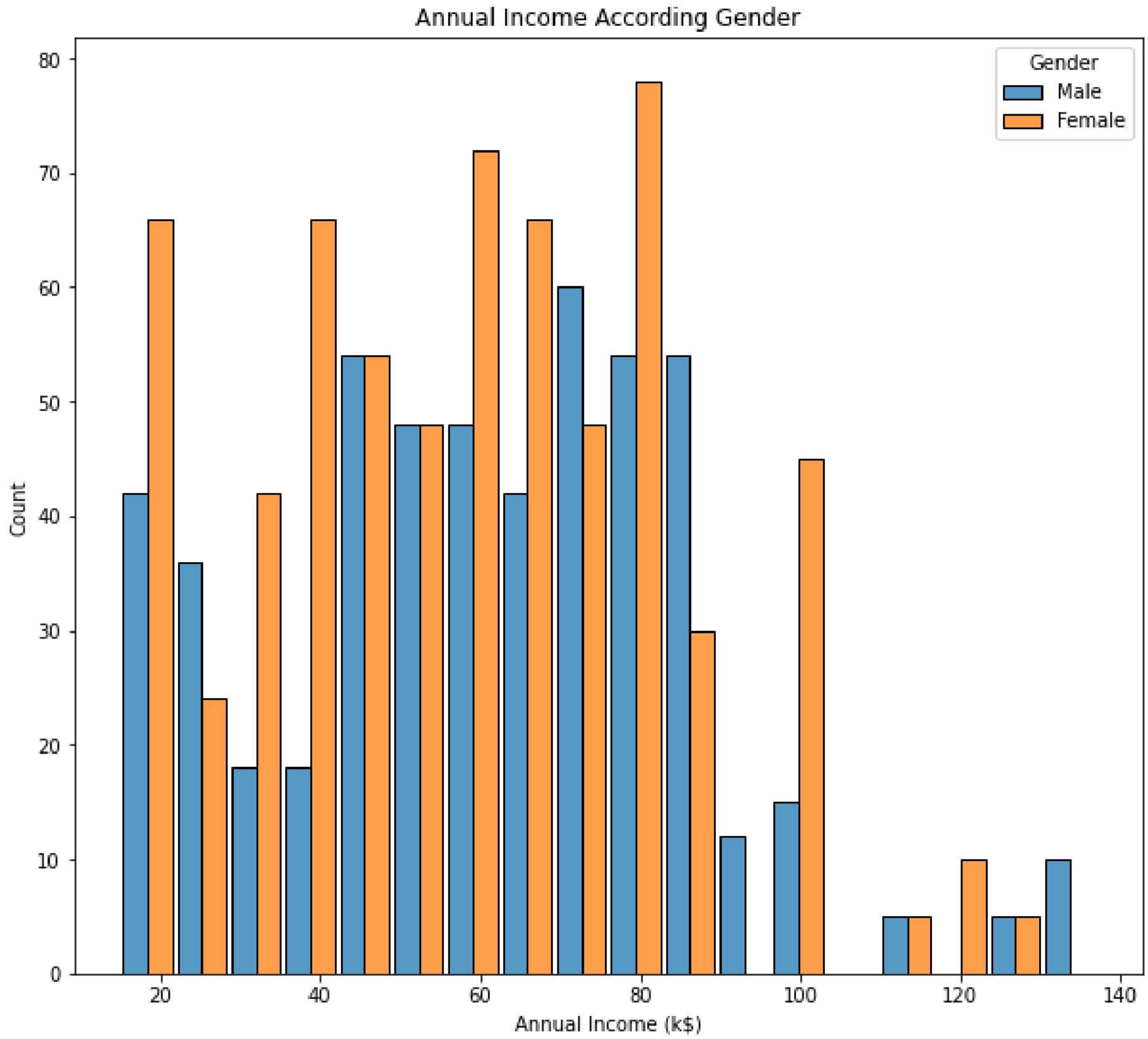
In the Above graph shows that Male AnnualIncome are more then Female in all City's

Graph of Spending Score in City's

In [149]:
```
sns.barplot(x=df1['City'] ,y=df1['Spending Score (1-100)'] ,data=df1)
plt.xticks(rotation = 45)
plt.show()
```

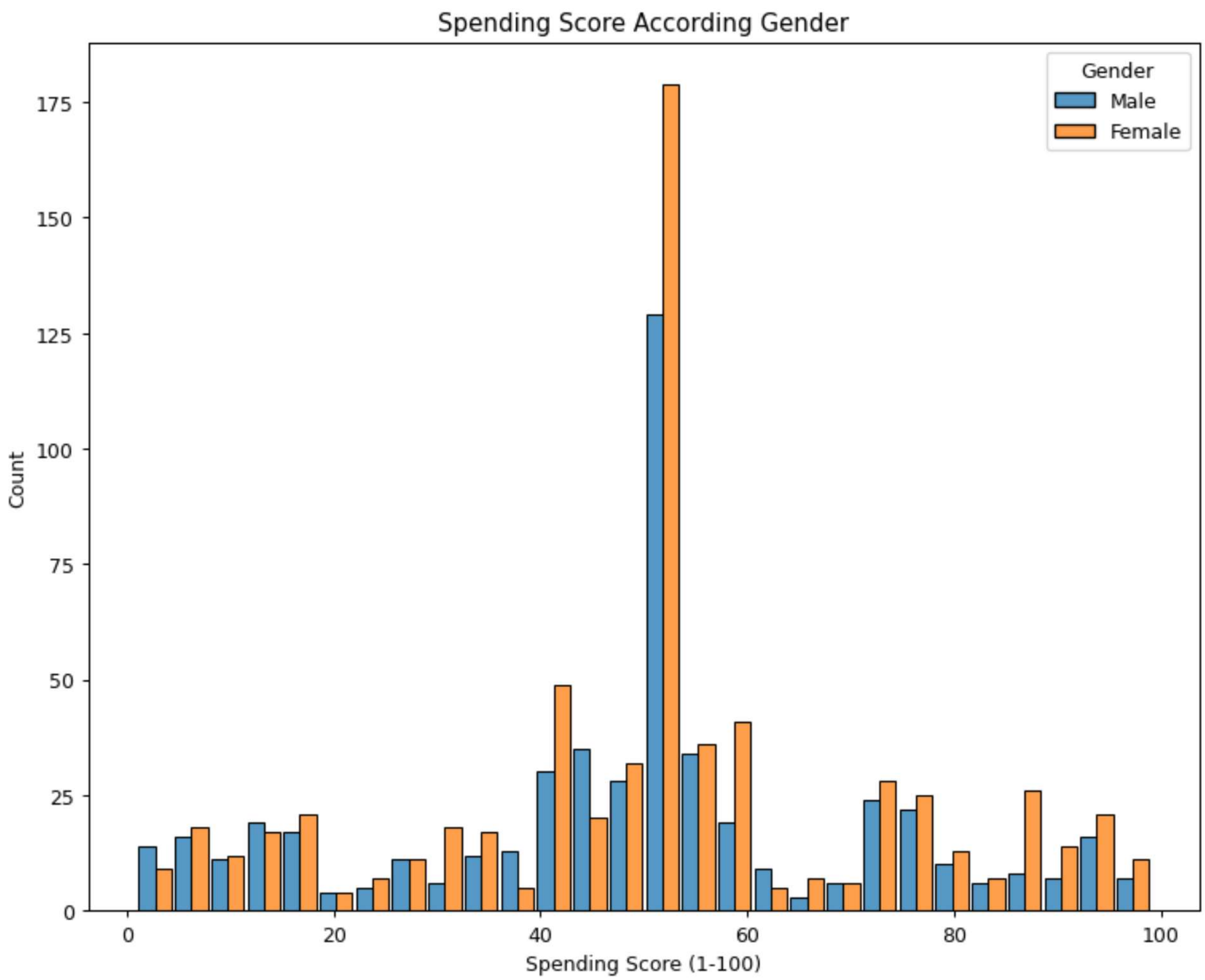


In [89]:
```
fig = plt.figure(figsize=(10,9))
sns.histplot(data=df,x=df['Annual Income (k$)'] ,hue=df['Gender'] ,color='b',mult
plt.show()
```



In Above graph shows that AnnualIncome of 80 $ Dollars are Higest in Female in all City's

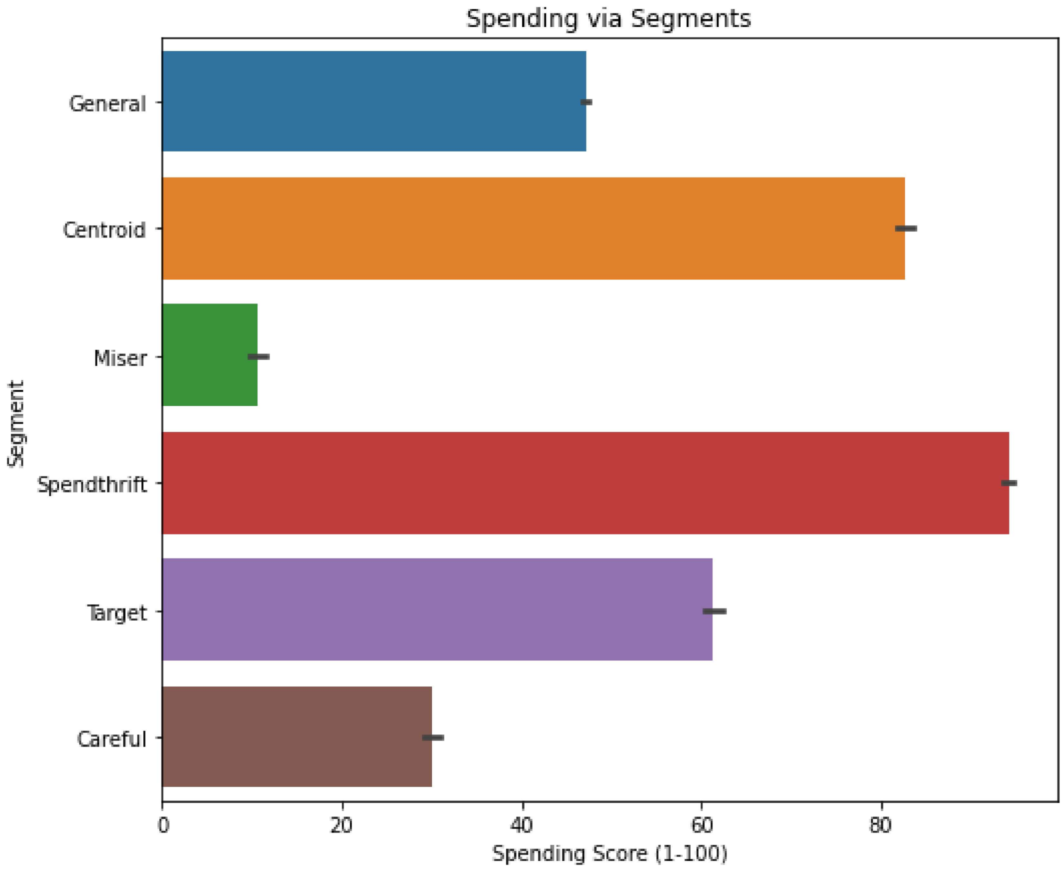
In [148]:
```
fig = plt.figure(figsize=(10,8),dpi=90)
sns.histplot(data=df,x=df1['Spending Score (1-100)'] ,hue=df1['Gender'] ,color='b
plt.show()
```



In Above Graph shows that spending score approximately 50 are Female in all City's

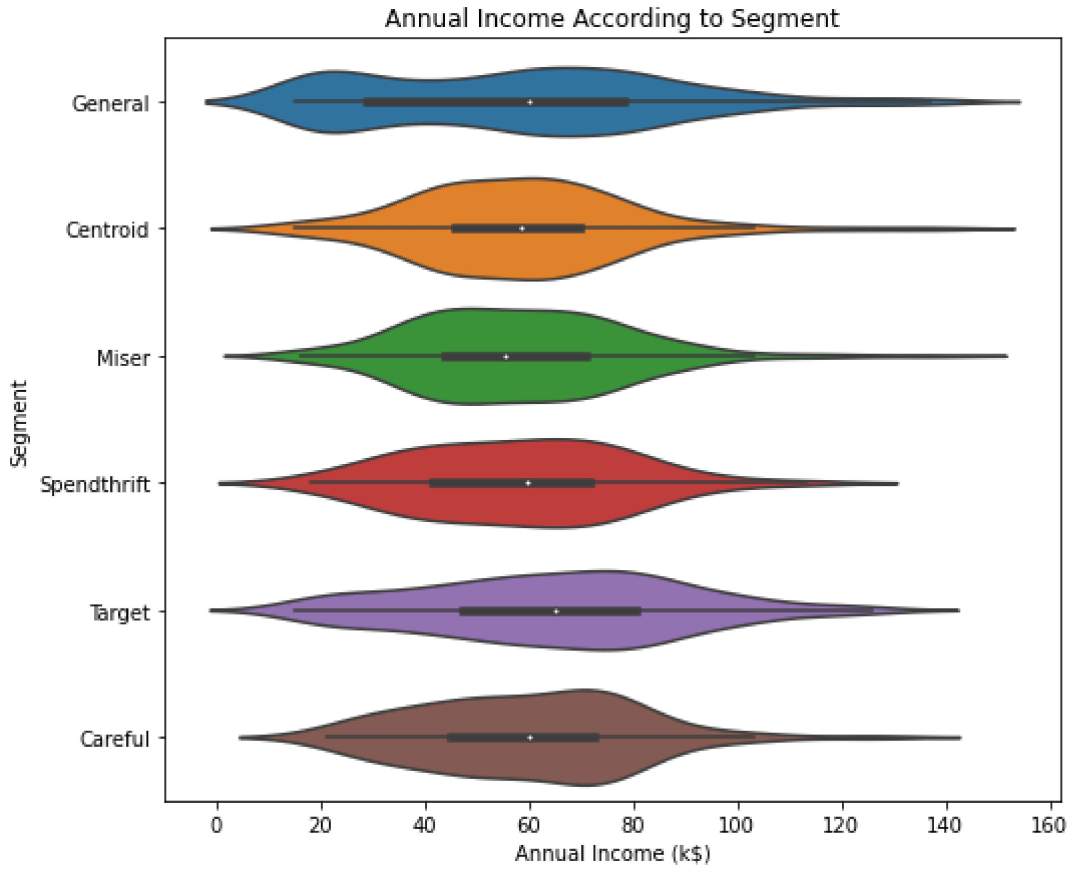
In [213]:

```
fig = plt.figure(figsize=(8,7))
sns.barplot(data=df1,x = df1['Spending Score (1-100)'],y= df1['Segment']).set(tit
plt.show()
```



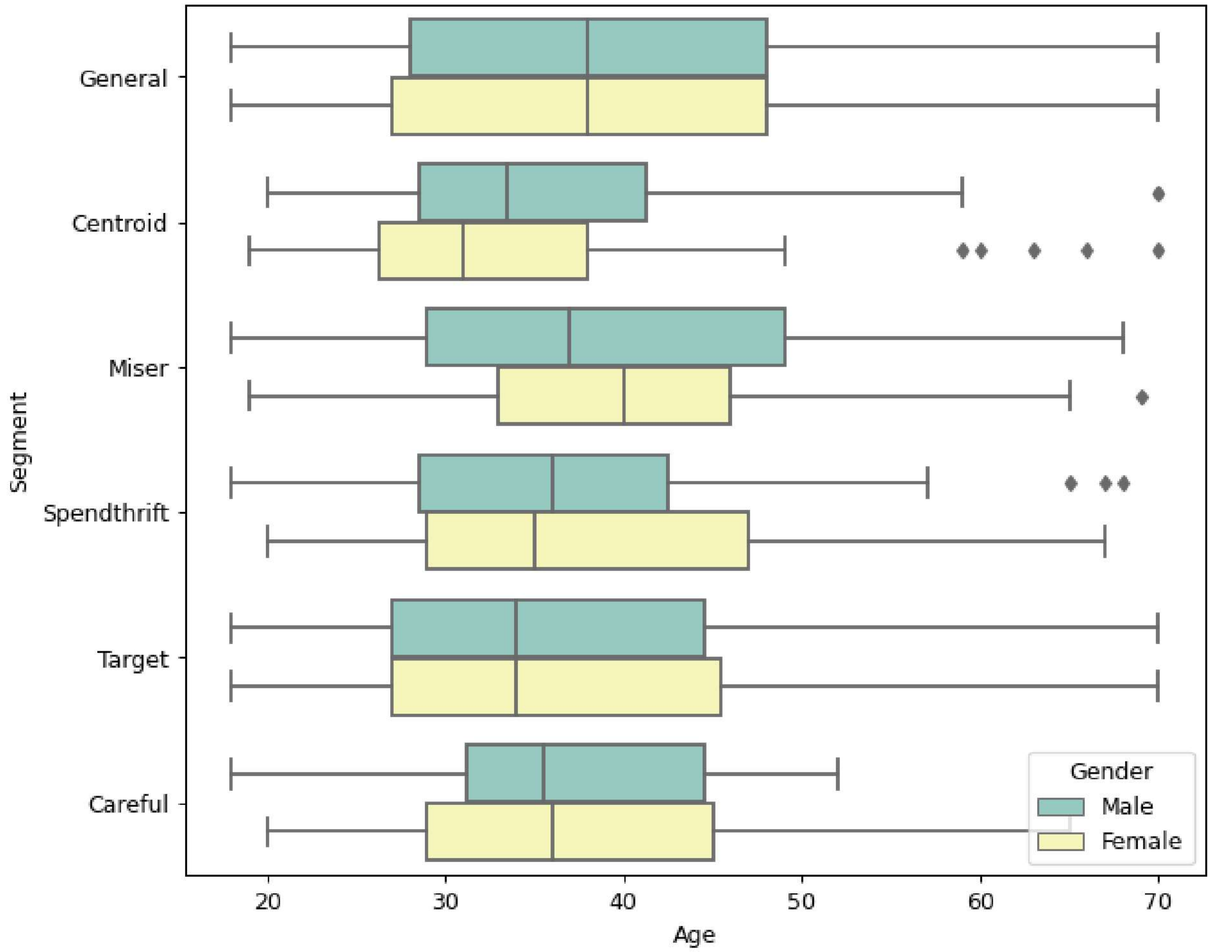
In Above graph shows that most of people in Spending score are spendthrift in all city's

In [214]:
```
fig = plt.figure(figsize=(8,7))
sns.violinplot(data=df1 ,x=df1['Annual Income (k$)'],y=df1['Segment']).set(title=
plt.show()
```



In the Above graph shows that most of people in AnnualIncome are Centroid in all City's

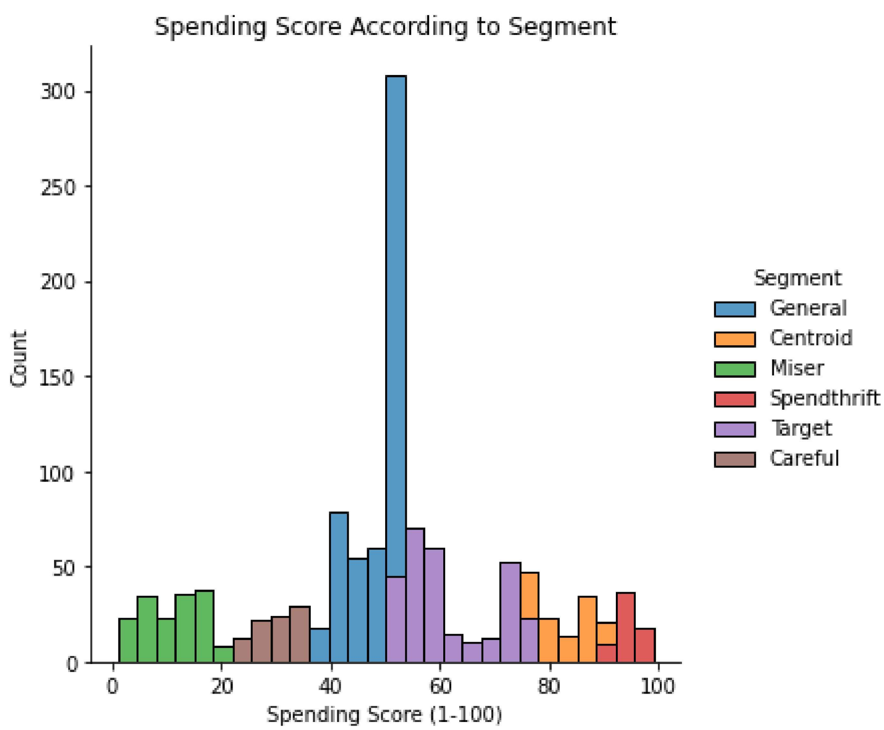
In [215]:
```
fig = plt.figure(figsize=(8,7),dpi=90)
sns.boxplot(data=df1,x=df1['Age'],y=df1['Segment'],hue=df1['Gender'],palette="Set
plt.show()
```



In the Above graph shows that Age in segment most of people are general and target in all City's

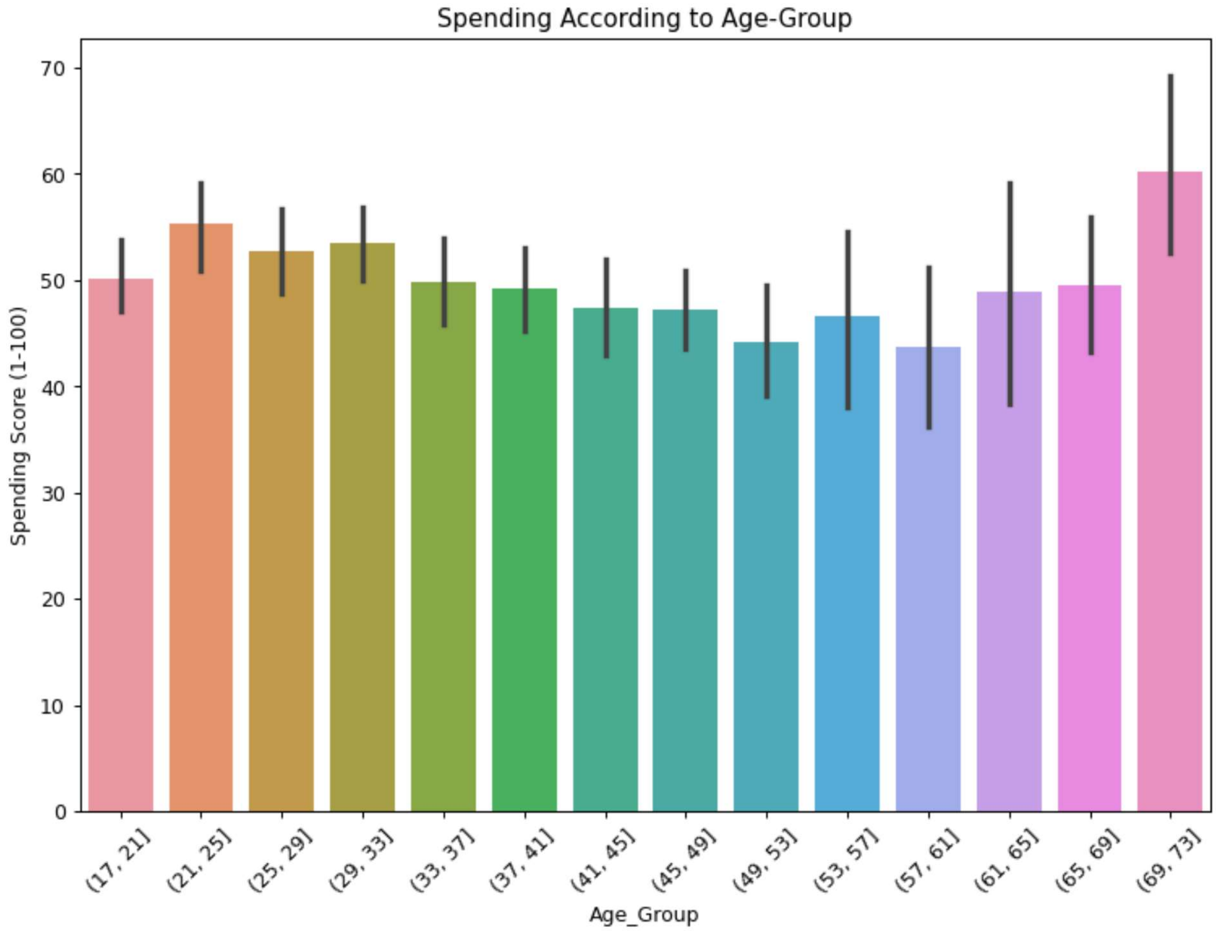
```
In [94]: fig = plt.figure(figsize=(10,7))
sns.displot(x=df1['Spending Score (1-100)'],hue=df1['Segment'],multiple='stack').
plt.show()
```

```
<Figure size 720x504 with 0 Axes>
```



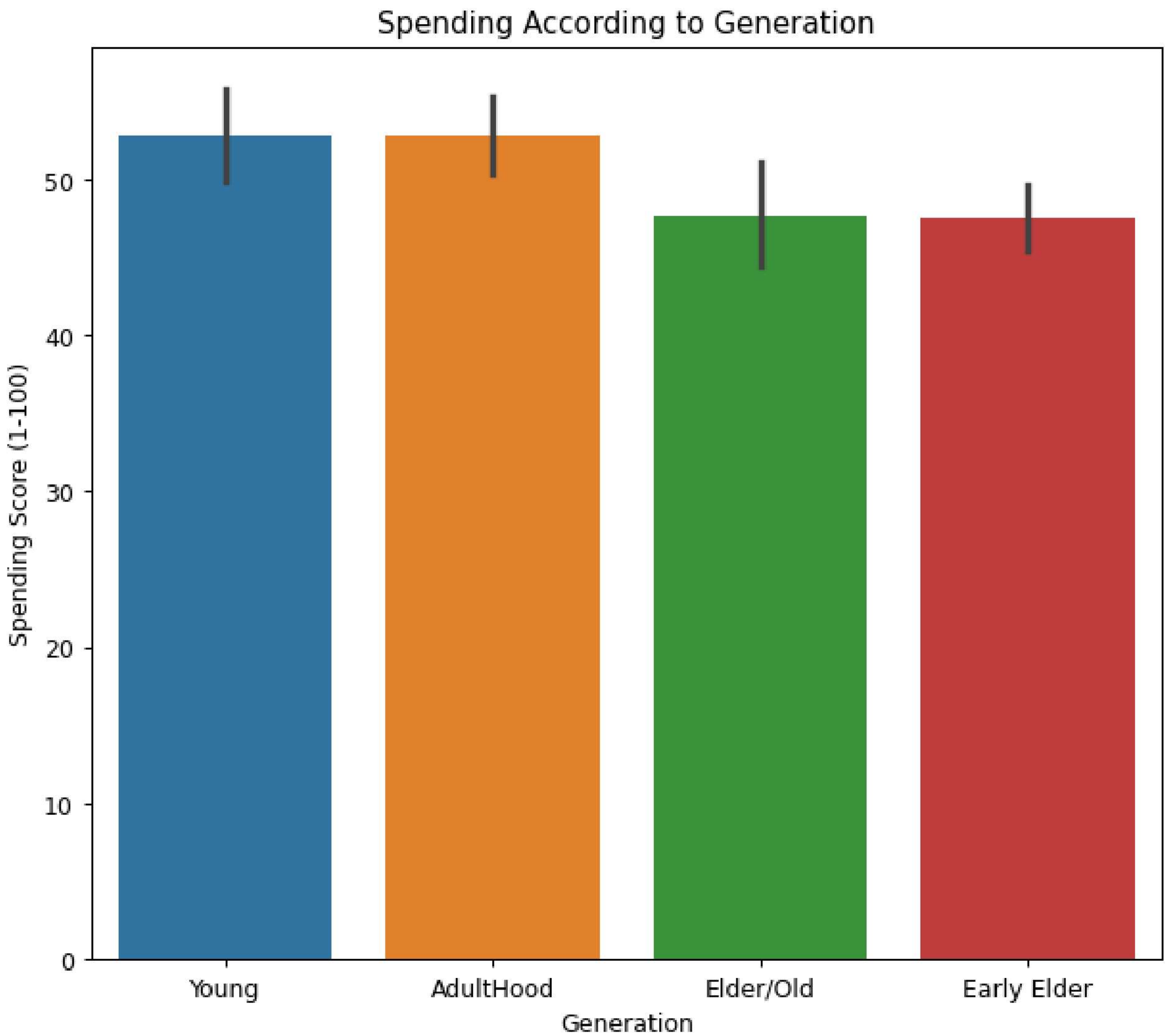
In the Above graph shows that Spending score most of people are general in all city's

In [46]:
```
fig = plt.figure(figsize=(10,7),dpi=90)
rs = sns.barplot(data=df1,x=df1['Age_Group'],y=df1['Spending Score (1-100)']).set
plt.xticks(rotation = 45)
plt.show()
```



In the Above graph shows Spending Score on the bases of Age Group most of people are 69-73 Ages

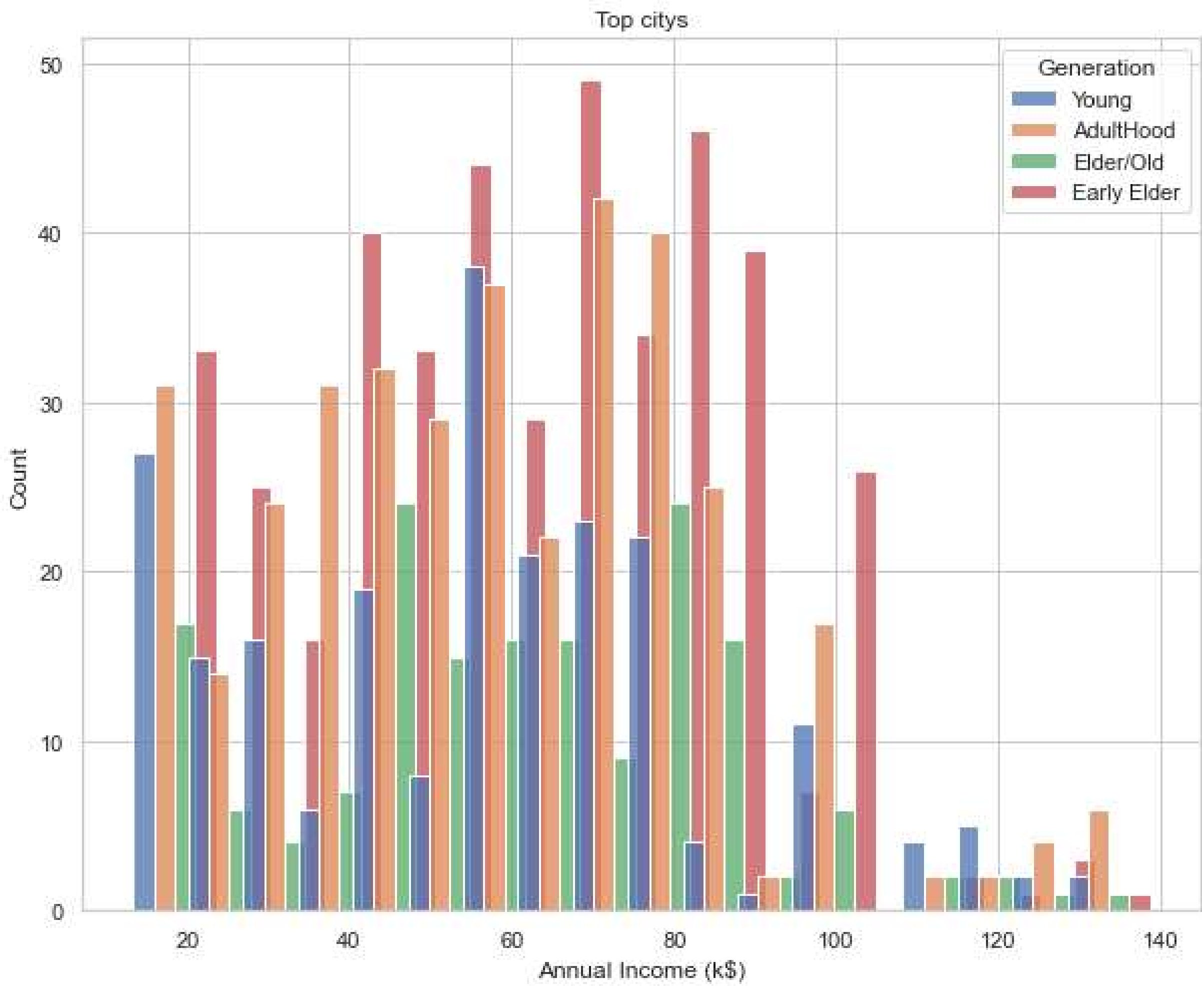
In [73]:
```
fig = plt.figure(figsize=(8,7),dpi=90)
rs = sns.barplot(data=df1,x=df1['Generation'],y=df1['Spending Score (1-100)']).se
plt.show()
```



In above graph shows that spending score on bases of Generation are Young and Adulthood

In [306]:

```
fig = plt.figure(figsize=(10,8))
ab=sns.histplot(data=df1,x='Annual Income (k$)',hue='Generation',multiple='dodge'
plt.show()
```



In Above graph shows that Annual Income on the bases of Generation are EarlyElder in all citys

Below DataFrame shows that city wise segregation

In [183]:

```
city = df1[(df1['City']=='Los Angeles') | (df1['City']=='Chicago') | (df1['City']
```

In [184]:
```
df1['City'].value_counts()
```

Out[184]:
```
Los Angeles     135
Chicago         135
Houston         135
Phoenix         135
Philadelphia    135
San Antonio     135
San Diego       134
Washington      100
New York         68
Seattle          68
Name: City, dtype: int64
```

In [185]:
```
city_male = df1[df1['Gender']=='Male']
```

In [186]:
```
city_male.value_counts().sum()
```

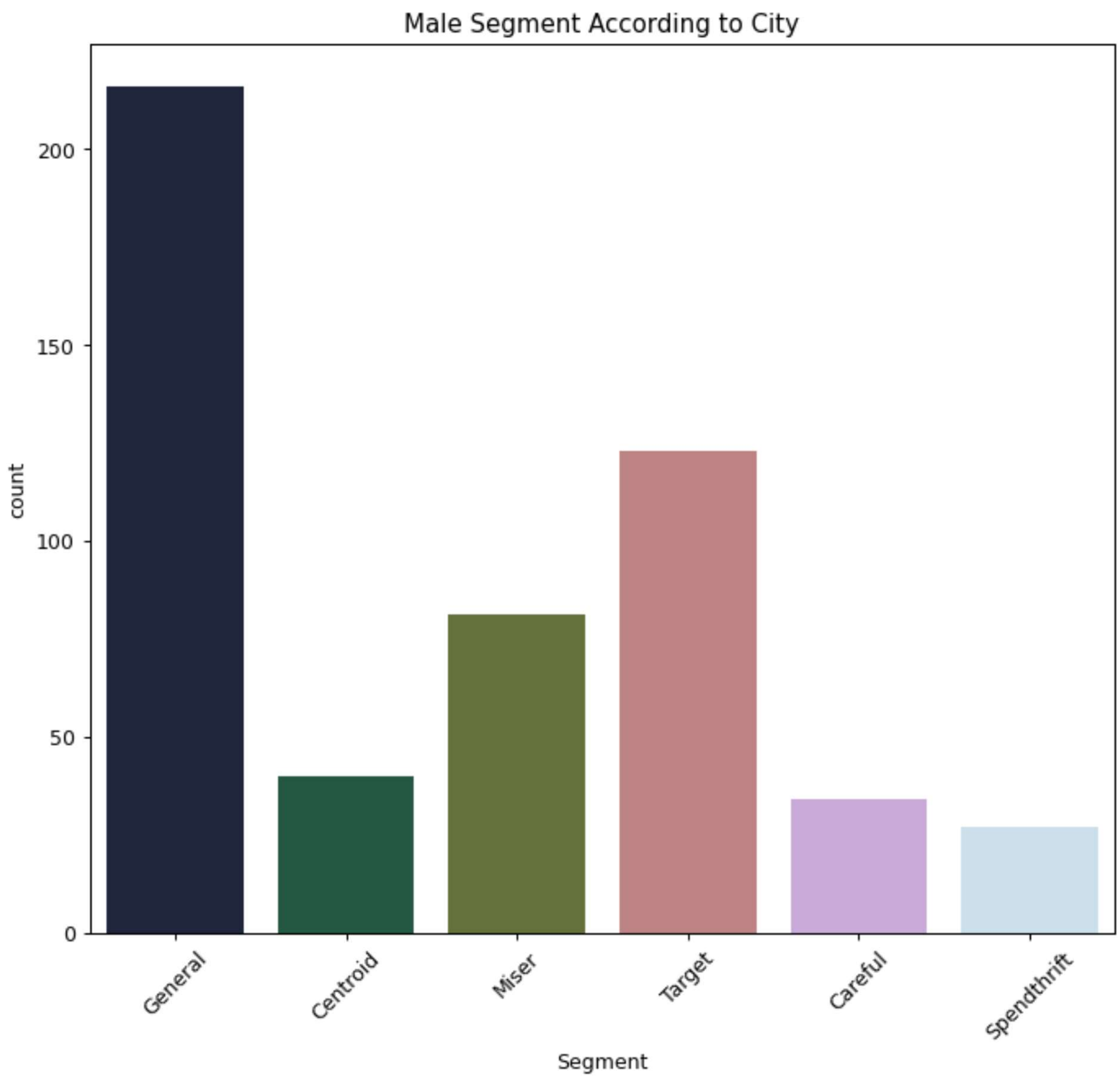
Out[186]:
```
521
```

In [187]:
```
city_female = df1[df['Gender']=='Female']
```

In [188]:
```
city_female.value_counts().sum()
```

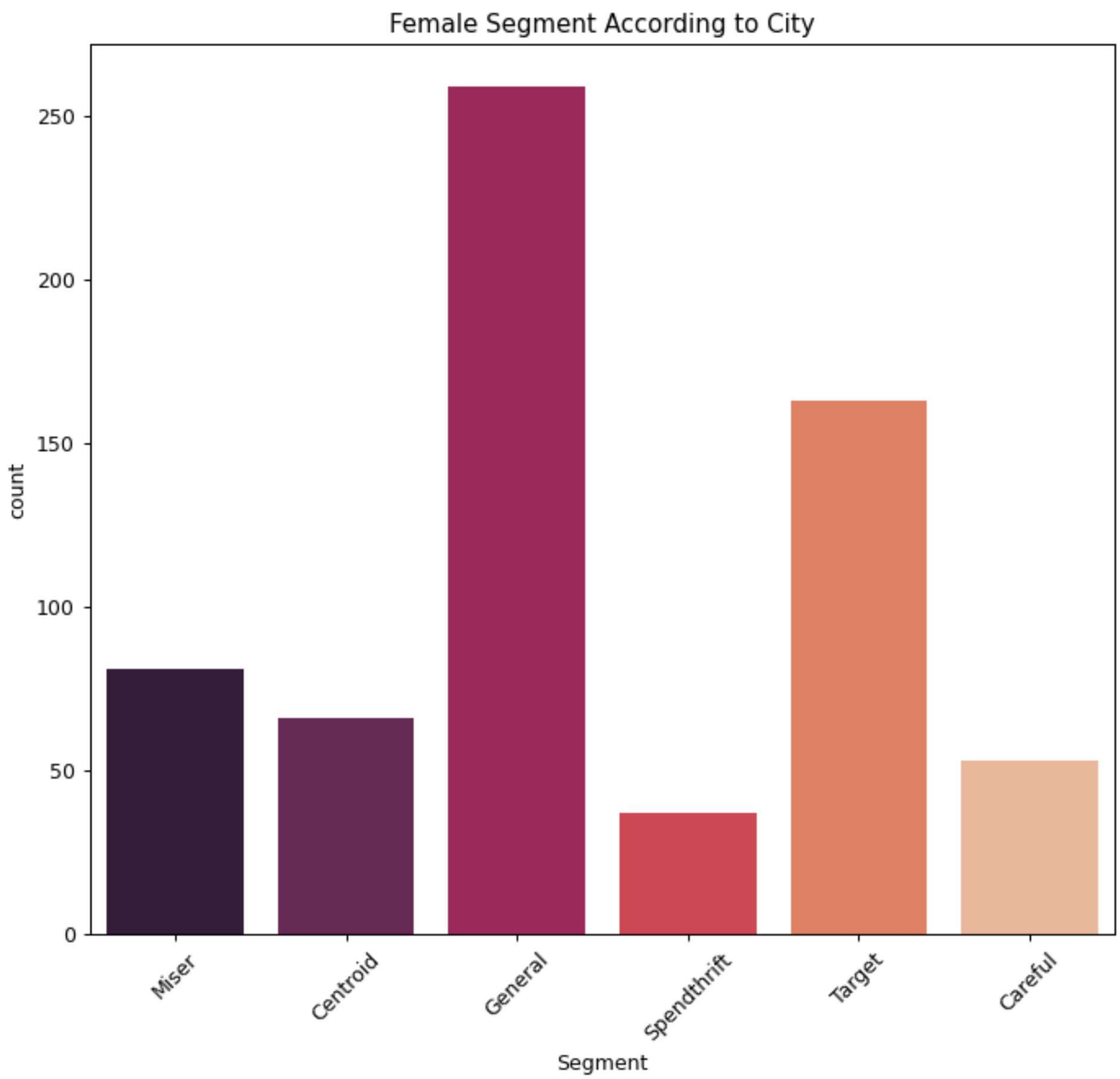
Out[188]:
```
659
```

In [189]:
```
fig = plt.figure(figsize=(9,8),dpi=90)
sns.countplot(data=city_male,x='Segment',dodge=True,palette="cubehelix").set(titl
plt.xticks(rotation=45)
plt.show()
```



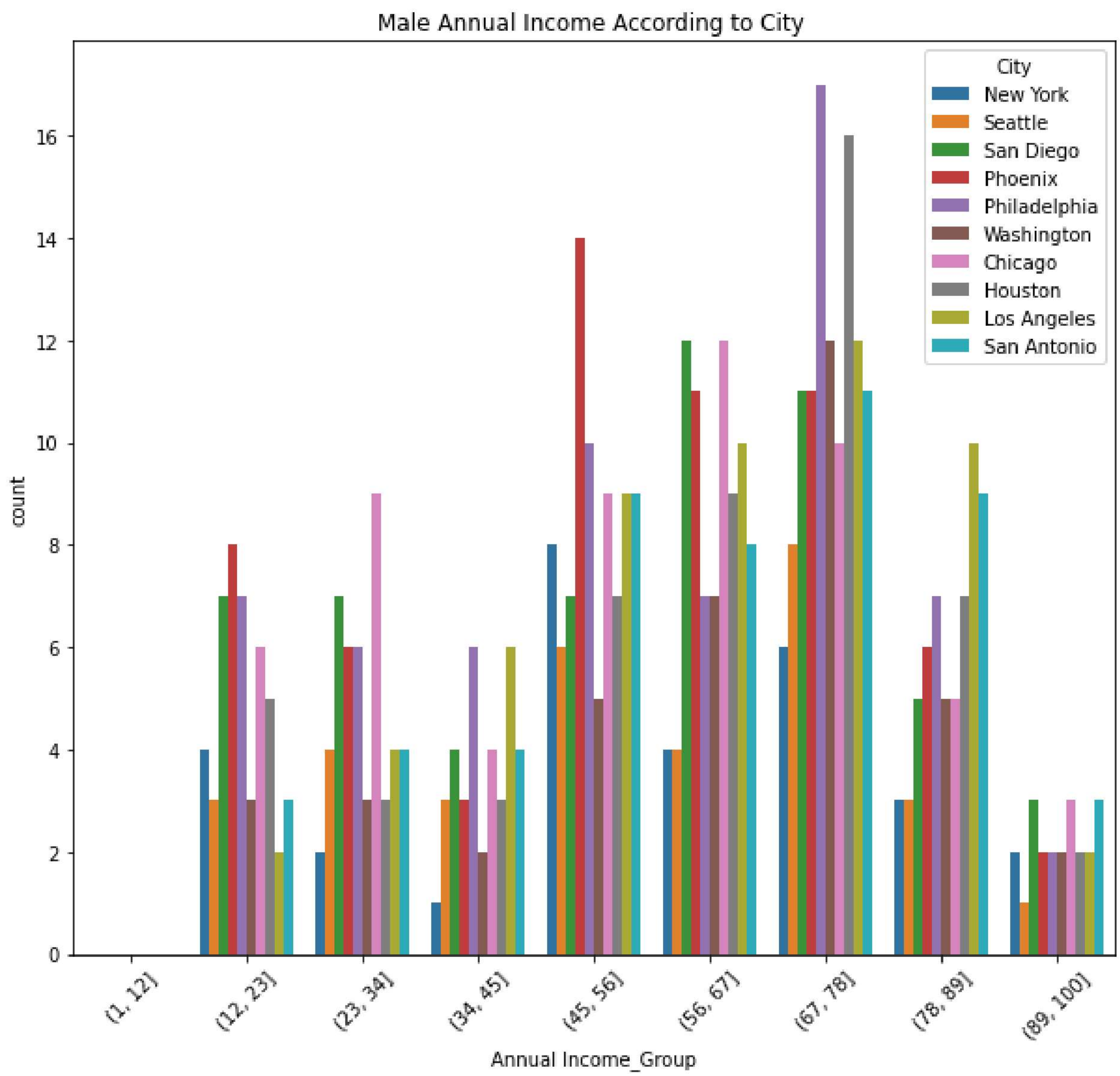
In the Above shows that City wise Segment in Male more people are General in all city's

In [131]:
```
fig = plt.figure(figsize=(9,8),dpi=90)
sns.countplot(data=city_female,x='Segment',dodge=True,palette="rocket").set(title
plt.xticks(rotation=45)
plt.show()
```



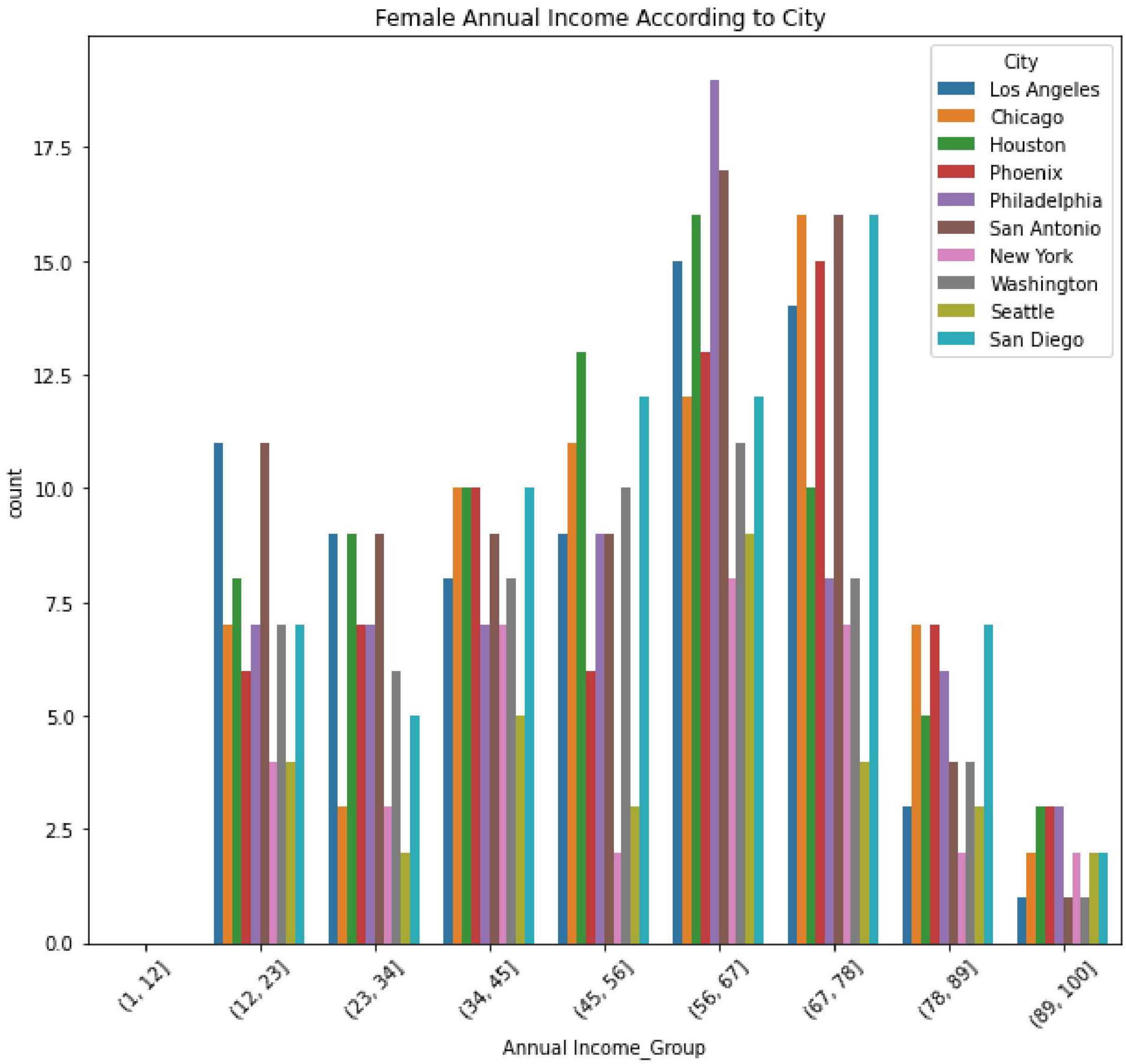
In the Above shows that City wise Segment in Female more people are General in all city's

In [142]:
```
fig = plt.figure(figsize=(10,9))
sns.countplot(data=city_male,x='Annual Income_Group',palette='tab10',hue='City').
plt.xticks(rotation=45)
plt.show()
```



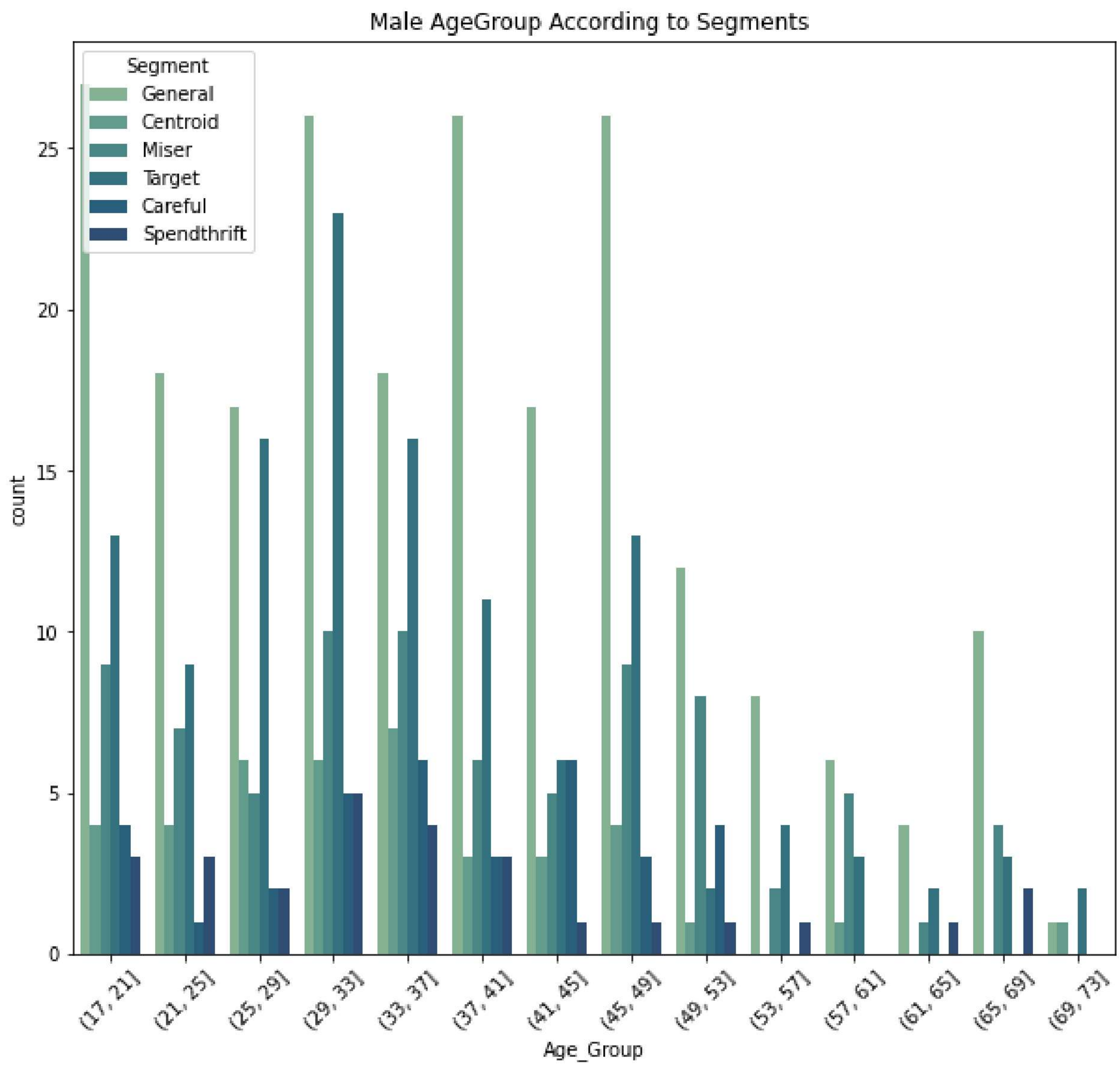
In the Above graph shows that AnnualIncome_Group in city most of people are 67-78 dollars are more philadelphia in all city's

In [143]:
```
fig = plt.figure(figsize=(10,9))
sns.countplot(data=city_female,x='Annual Income_Group',palette='tab10',hue='City'
plt.xticks(rotation=45)
plt.show()
```



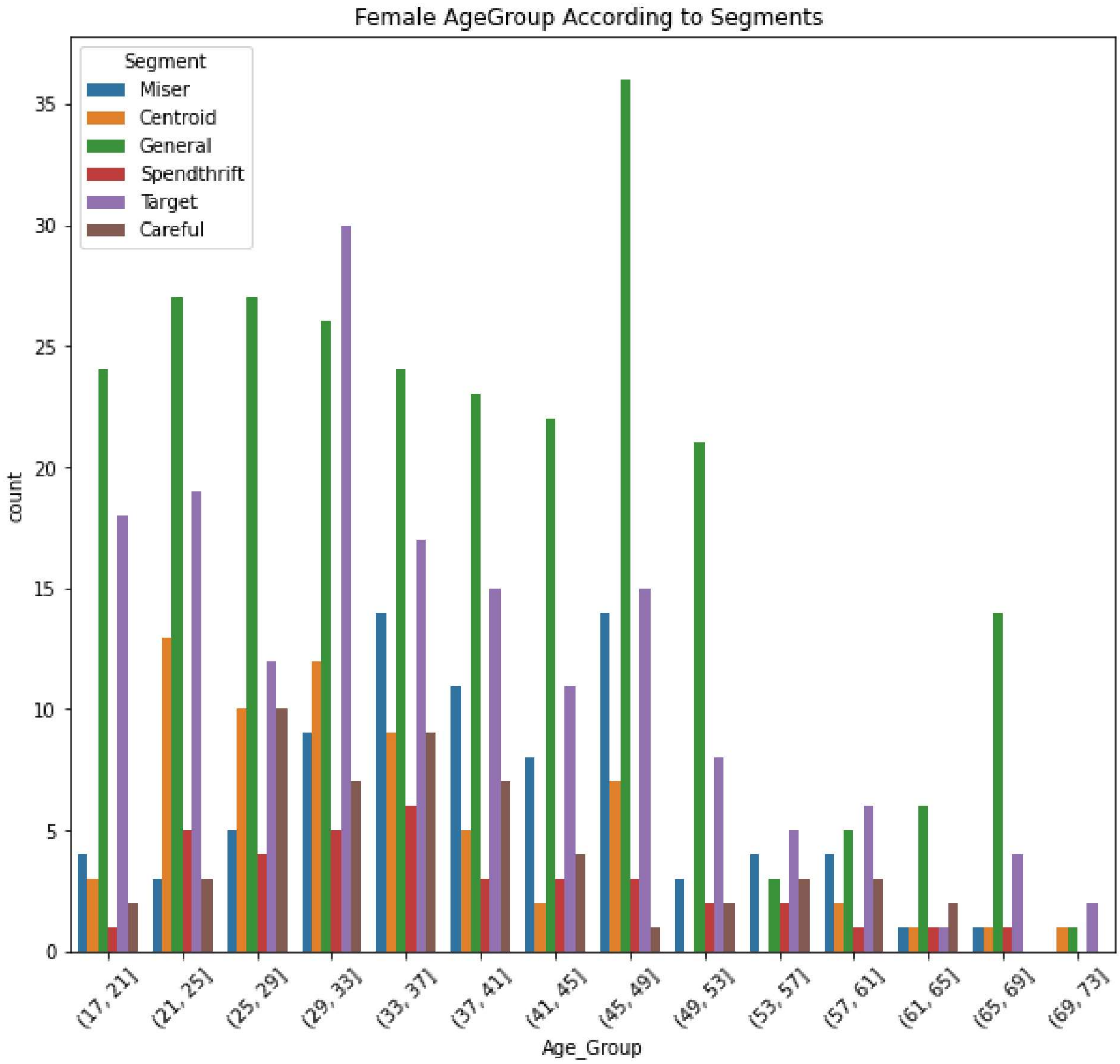
In the Above graph shows that AnnualIncome_Group in city most of people are 56-67 dollars are more in philadelphia in all city's

In [146]:
```
fig = plt.figure(figsize=(10,9))
sns.countplot(data=city_male,x='Age_Group',palette='crest',hue='Segment').set(tit
plt.xticks(rotation=45)
plt.show()
```



In the Above graph shows that Male AgeGroup in segment are General 17-21,29-33,37-41 and 45-49 AgeGroup in all Citys

```
In [190]: fig = plt.figure(figsize=(10,9))
          sns.countplot(data=city_female,x='Age_Group',palette='tab10',hue='Segment').set(t
          plt.xticks(rotation=45)
          plt.show()
```



In the Above graph shows that Female AgeGroup in Segment are General&Targeted 17-21,21-25,25-29,33-37,37-41,41-45,45-49,49-53 and 65-69 in all city's

Separating each City and Analysis of it

In [152]:
```
df1['City'].value_counts()
```

Out[152]:
```
Los Angeles     135
Chicago         135
Houston         135
Phoenix         135
Philadelphia    135
San Antonio     135
San Diego       134
Washington      100
New York         68
Seattle          68
Name: City, dtype: int64
```

In [177]:
```
city_los = df1[df1['City'] == 'Los Angeles']
```

In [182]:
```
city_los['City'].value_counts()
```

Out[182]:
```
Los Angeles    135
Name: City, dtype: int64
```

In [173]:
```
city_chicago = df1[df1['City'] == 'Chicago']
```

In [183]:
```
city_chicago['City'].value_counts()
```

Out[183]:
```
Chicago    135
Name: City, dtype: int64
```

In [184]:
```
city_hou = df1[df1['City'] == 'Houston']
```

In [185]:
```
city_hou['City'].value_counts()
```

Out[185]:
```
Houston    135
Name: City, dtype: int64
```

In [186]:
```
city_pho = df1[df1['City'] == 'Phoenix']
```

In [187]:
```
city_pho['City'].value_counts()
```

Out[187]:
```
Phoenix    135
Name: City, dtype: int64
```

In [188]:
```
city_phi = df1[df1['City'] == 'Philadelphia']
```

In [189]:
```
city_phi['City'].value_counts()
```

Out[189]:
```
Philadelphia    135
Name: City, dtype: int64
```

In [196]:
```
city_san = df1[df1['City']=='San Antonio']
```

In [197]:
```
city_san['City'].value_counts()
```

Out[197]:
```
San Antonio    135
Name: City, dtype: int64
```

In [202]:
```
city_san_d = df1[df1['City'] == 'San Diego']
```

In [204]:
```
city_san_d['City'] .value_counts()
```

Out[204]:
```
San Diego    134
Name: City, dtype: int64
```

In [205]:
```
city_wash = df1[df1['City'] == 'Washington']
```

In [206]:
```
city_wash['City'].value_counts()
```

Out[206]:
```
Washington    100
Name: City, dtype: int64
```

In [207]:
```
city_new = df1[df1['City'] == 'New York']
```

In [208]:
```
city_new['City'].value_counts()
```

Out[208]:
```
New York    68
Name: City, dtype: int64
```

In [209]:
```
city_sea = df1[df1['City'] == 'Seattle']
```

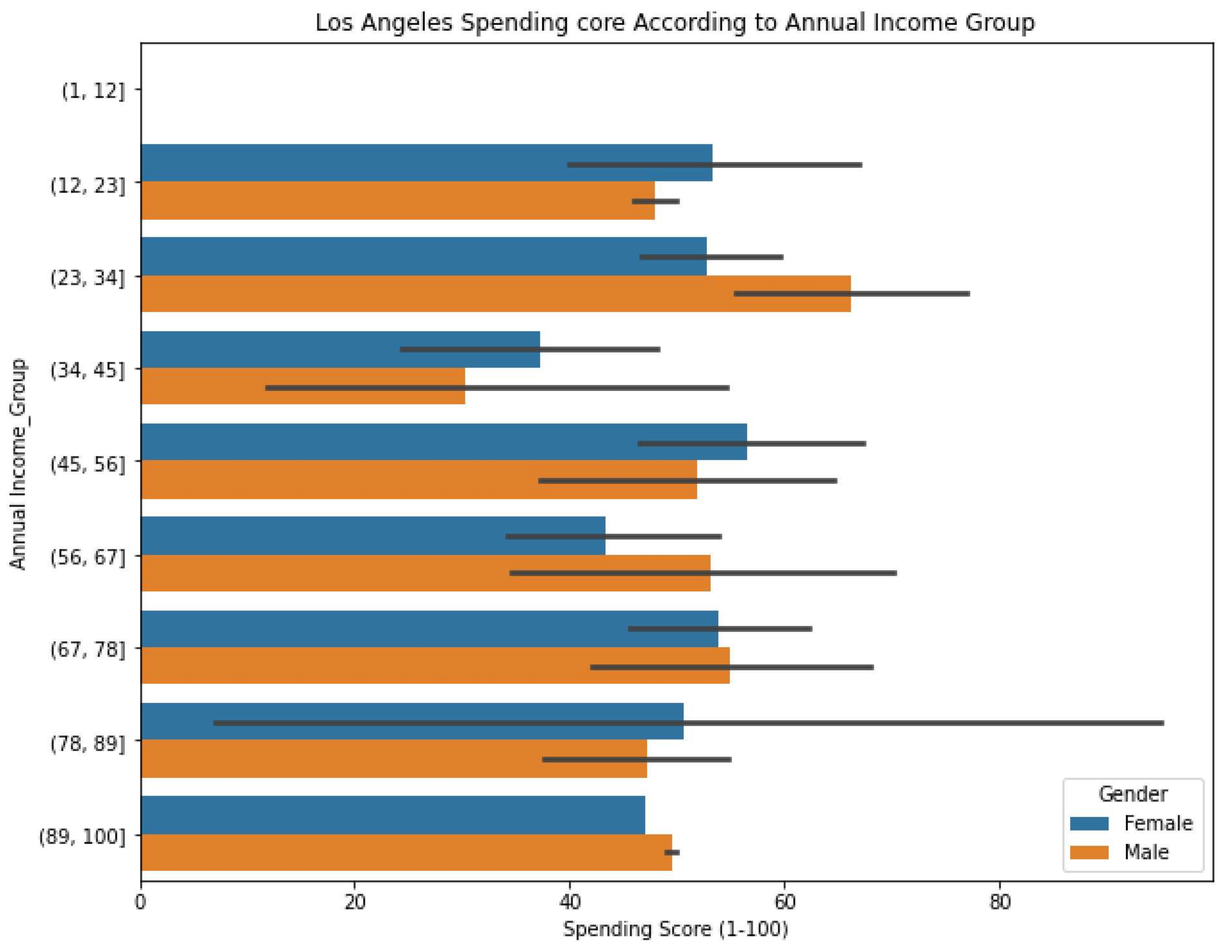
In [210]:
```
city_sea['City'].value_counts()
```

Out[210]:
```
Seattle    68
Name: City, dtype: int64
```

Analysis of Individual City

Los Angeles

In [244]:
```
fig = plt.figure(figsize=(10,8))
sns.barplot(data=city_los,x='Spending Score (1-100)',y='Annual Income_Group' ,hue
plt.show()
```



-- In the Above Graph we can see that

1)Annul Income of Male under 12-13 dollar there spending score is less than Female

2)Annual Income of Male under 23-24 dollar there spending score is more than Femals

3)Annual Income of Male under 34-45 dollar there spending score is less than Female

4)Annual Income of Male under 45-56 dollar there spending score is less than Female

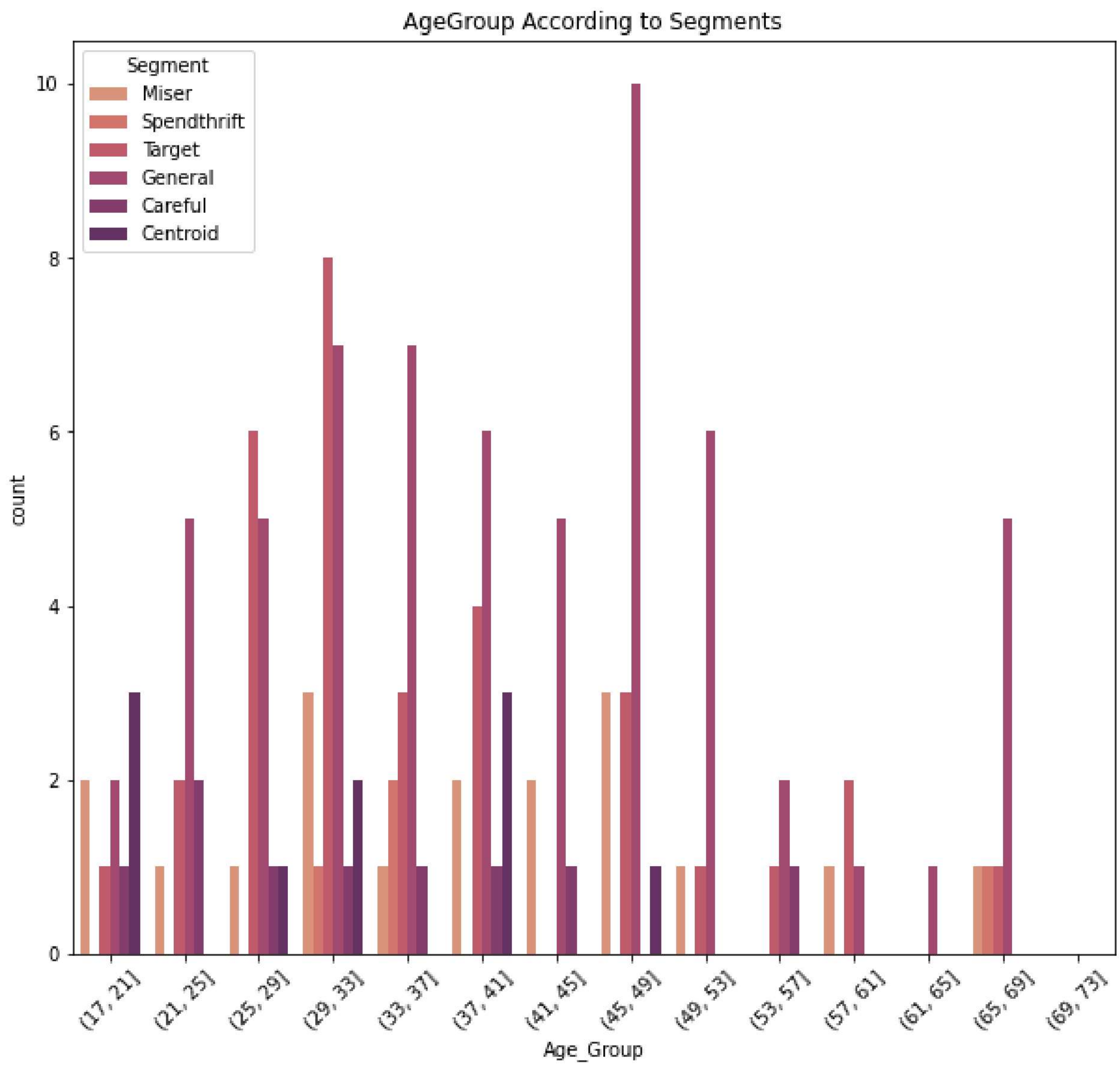
5)Annual Income of Male under 56-67 dollar there spendind score is more than Female

6)Annual Income of Male under 67-78 dollar there spendind score is more than Female

7)Annual Income of Male under 78-89 dollar there spendind score is less than Female

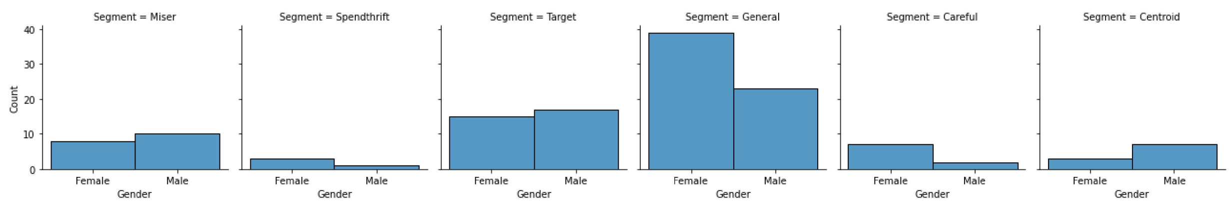
8)Annual Income of Male under 89-100 dollar there spendind score is more than Female

In [260]:
```
fig = plt.figure(figsize=(10,9))
sns.countplot(data=city_los,x='Age_Group', palette='flare',hue='Segment').set(tit
plt.xticks(rotation=45)
plt.show()
```



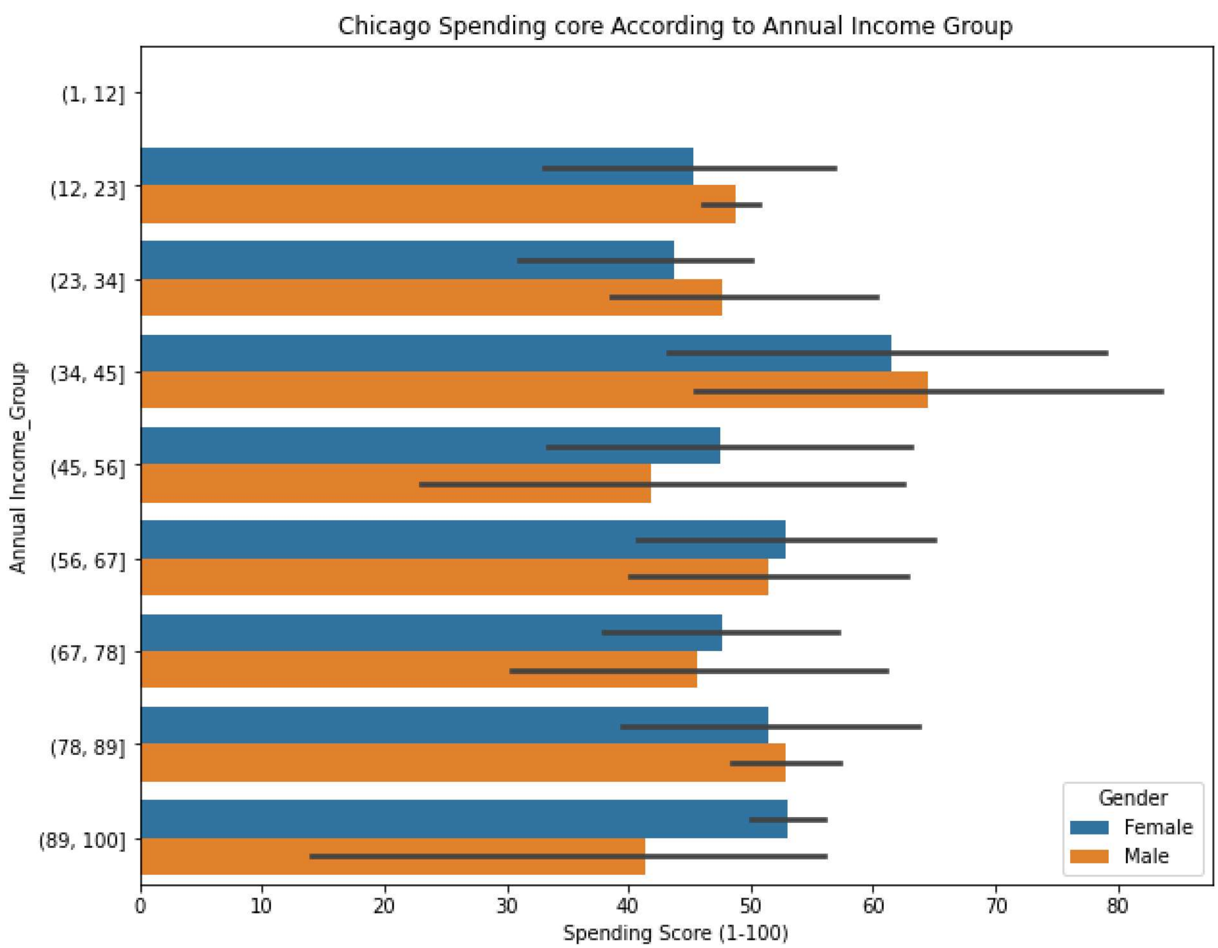
In [261]:
```
fig = plt.figure(figsize=(10,8))
g = sns.FacetGrid(city_los, col="Segment")
g.map(sns.histplot, "Gender")
plt.show()
```

```
<Figure size 720x576 with 0 Axes>
```



Chicago

In [262]:
```
fig = plt.figure(figsize=(10,8))
sns.barplot(data=city_chicago,x='Spending Score (1-100)',y='Annual Income_Group'
plt.show()
```



-- -- In the Above Graph we can see that

1)Annul Income of Male under 12-13 dollar there spending score is more than Female

2)Annual Income of Male under 23-24 dollar there spending score is more than Femals

3)Annual Income of Male under 34-45 dollar there spending score is more than Female

4)Annual Income of Male under 45-56 dollar there spending score is less than Female

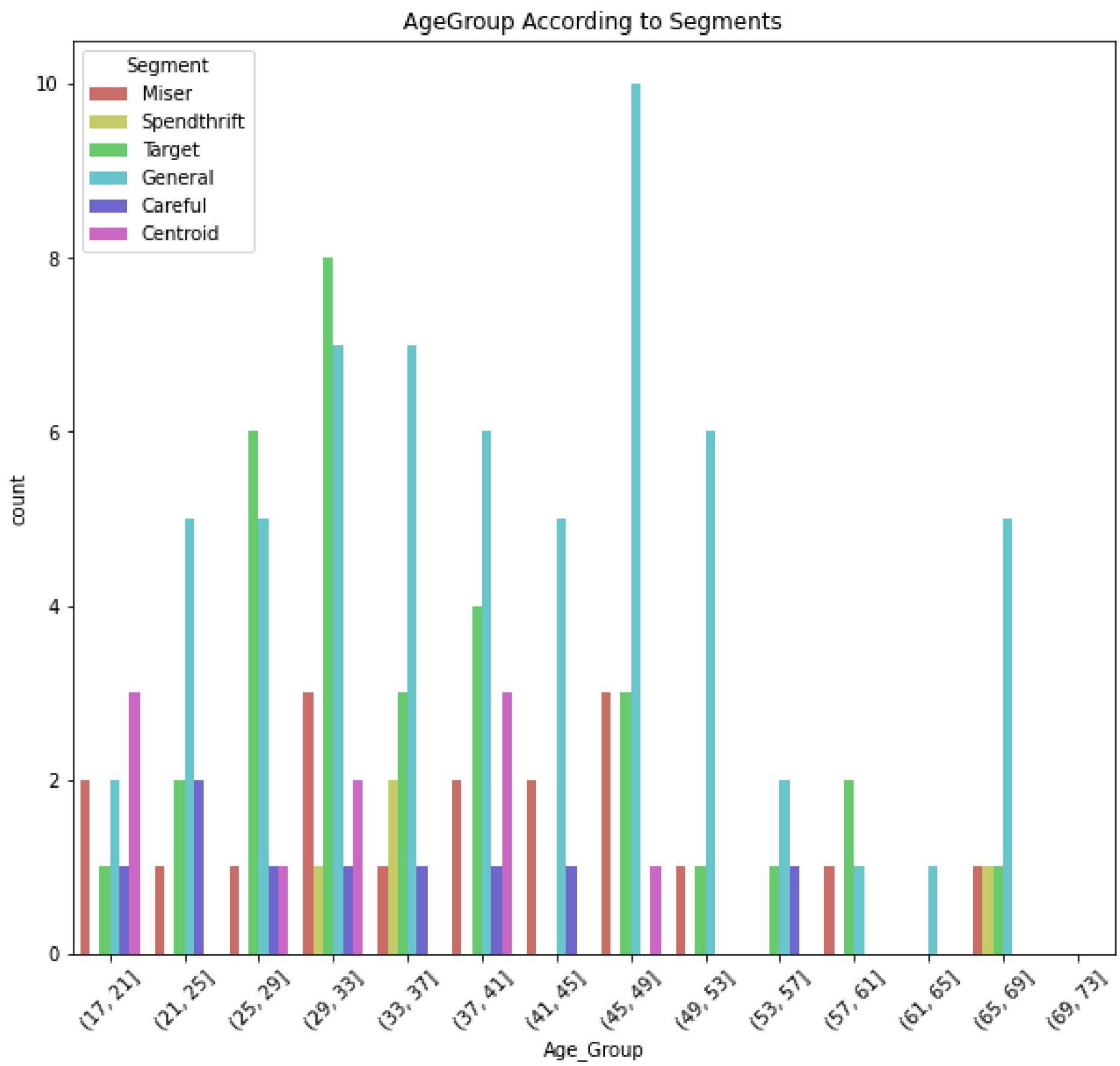
5)Annual Income of Male under 56-67 dollar there spendind score is less than Female

6)Annual Income of Male under 67-78 dollar there spendind score is less than Female

7)Annual Income of Male under 78-89 dollar there spendind score is more than Female

8)Annual Income of Male under 89-100 dollar there spendind score is less than Female

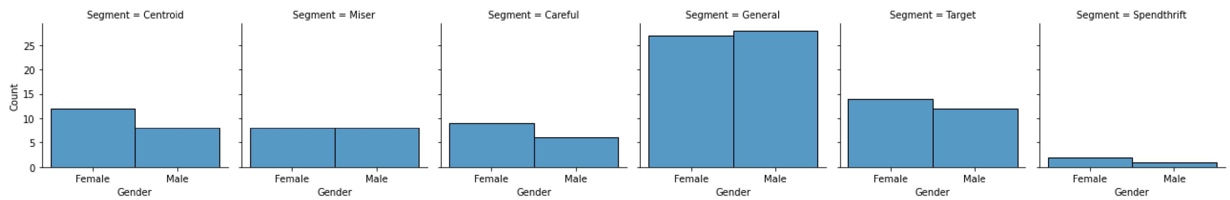
In [263]:
```
fig = plt.figure(figsize=(10,9))
sns.countplot(data=city_los,x='Age_Group', palette='hls',hue='Segment').set(title
plt.xticks(rotation=45)
plt.show()
```



-- In Chicago the Most of People are General Spending

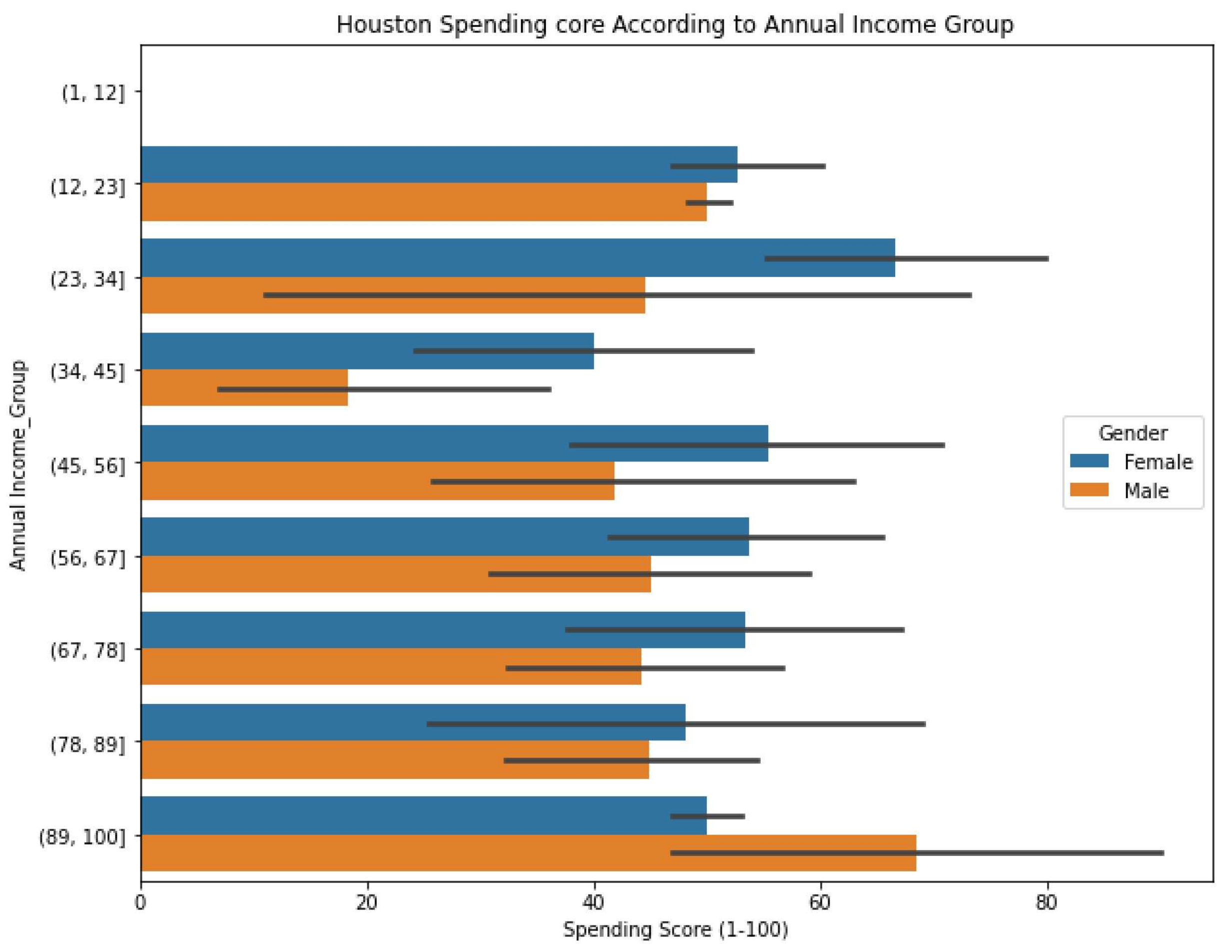
In [264]:
```python
fig = plt.figure(figsize=(10,8))
g = sns.FacetGrid(city_chicago, col="Segment")
g.map(sns.histplot, "Gender")
plt.show()
```

<Figure size 720x576 with 0 Axes>



Houston

In [265]:
```python
fig = plt.figure(figsize=(10,8))
sns.barplot(data=city_hou,x='Spending Score (1-100)',y='Annual Income_Group' ,hue
plt.show()
```



-- In the Above Graph we can see that

1)Annul Income of Male under 12-13 dollar there spending score is less than Female

2)Annual Income of Male under 23-24 dollar there spending score is less than Femals

3)Annual Income of Male under 34-45 dollar there spending score is less than Female

4)Annual Income of Male under 45-56 dollar there spending score is less than Female

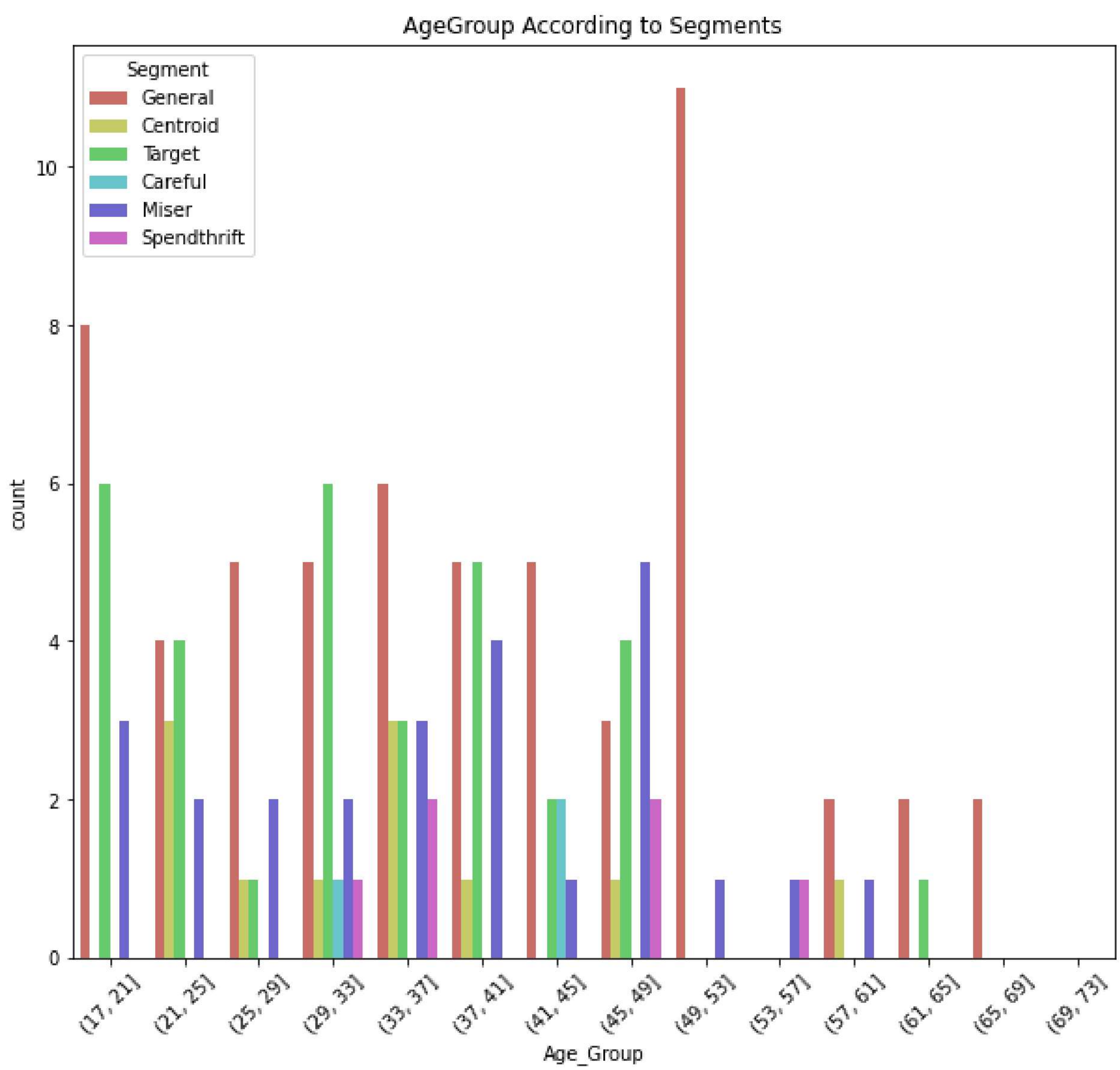
5)Annual Income of Male under 56-67 dollar there spendind score is less than Female

6)Annual Income of Male under 67-78 dollar there spendind score is less than Female

7)Annual Income of Male under 78-89 dollar there spendind score is less than Female

8)Annual Income of Male under 89-100 dollar there spendind score is more than Female

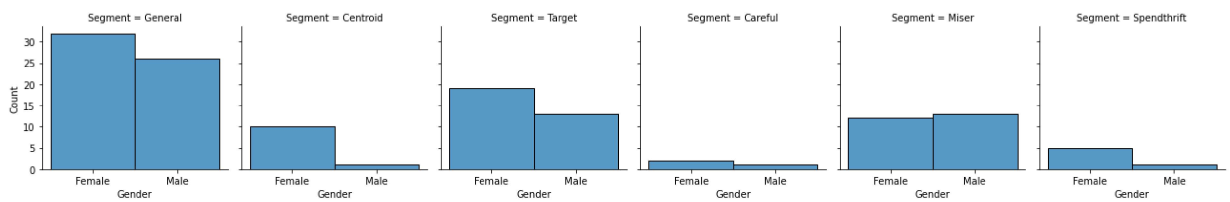
In [266]:
```
fig = plt.figure(figsize=(10,9))
sns.countplot(data=city_hou,x='Age_Group', palette='hls',hue='Segment').set(title
plt.xticks(rotation=45)
plt.show()
```



In Houston Most Of People Are Targeted,Miser & General Spending

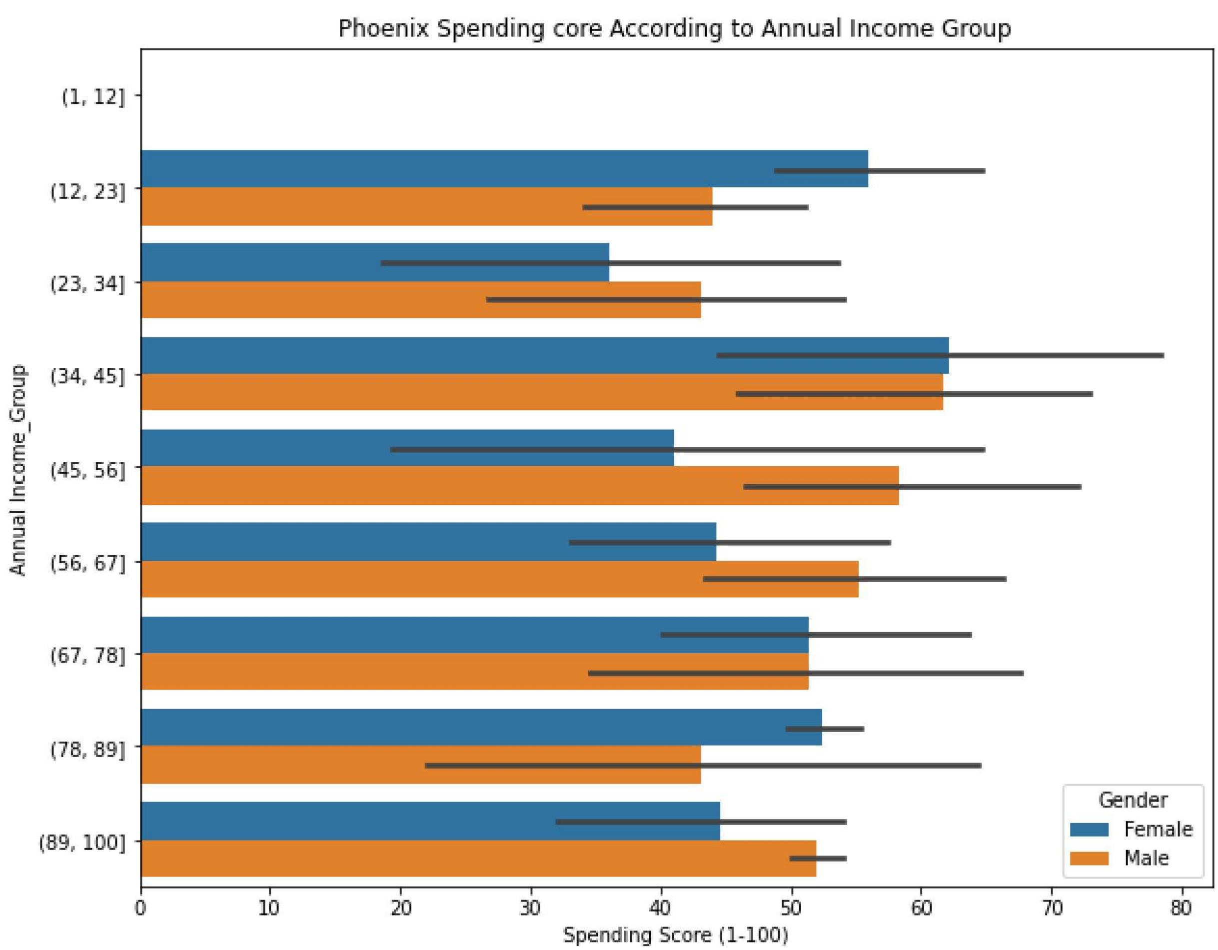
In [268]:
```
fig = plt.figure(figsize=(10,8))
g = sns.FacetGrid(city_hou, col="Segment")
g.map(sns.histplot, "Gender")
plt.show()
```

```
<Figure size 720x576 with 0 Axes>
```



Phoenix

In [269]:
```
fig = plt.figure(figsize=(10,8))
sns.barplot(data=city_pho,x='Spending Score (1-100)',y='Annual Income_Group' ,hue
plt.show()
```



-- In the Above Graph we can see that

1)Annul Income of Male under 12-13 dollar there spending score is less than Female

2)Annual Income of Male under 23-24 dollar there spending score is more than Femals

3)Annual Income of Male under 34-45 dollar there spending score is some what equal than Female

4)Annual Income of Male under 45-56 dollar there spending score is more than Female

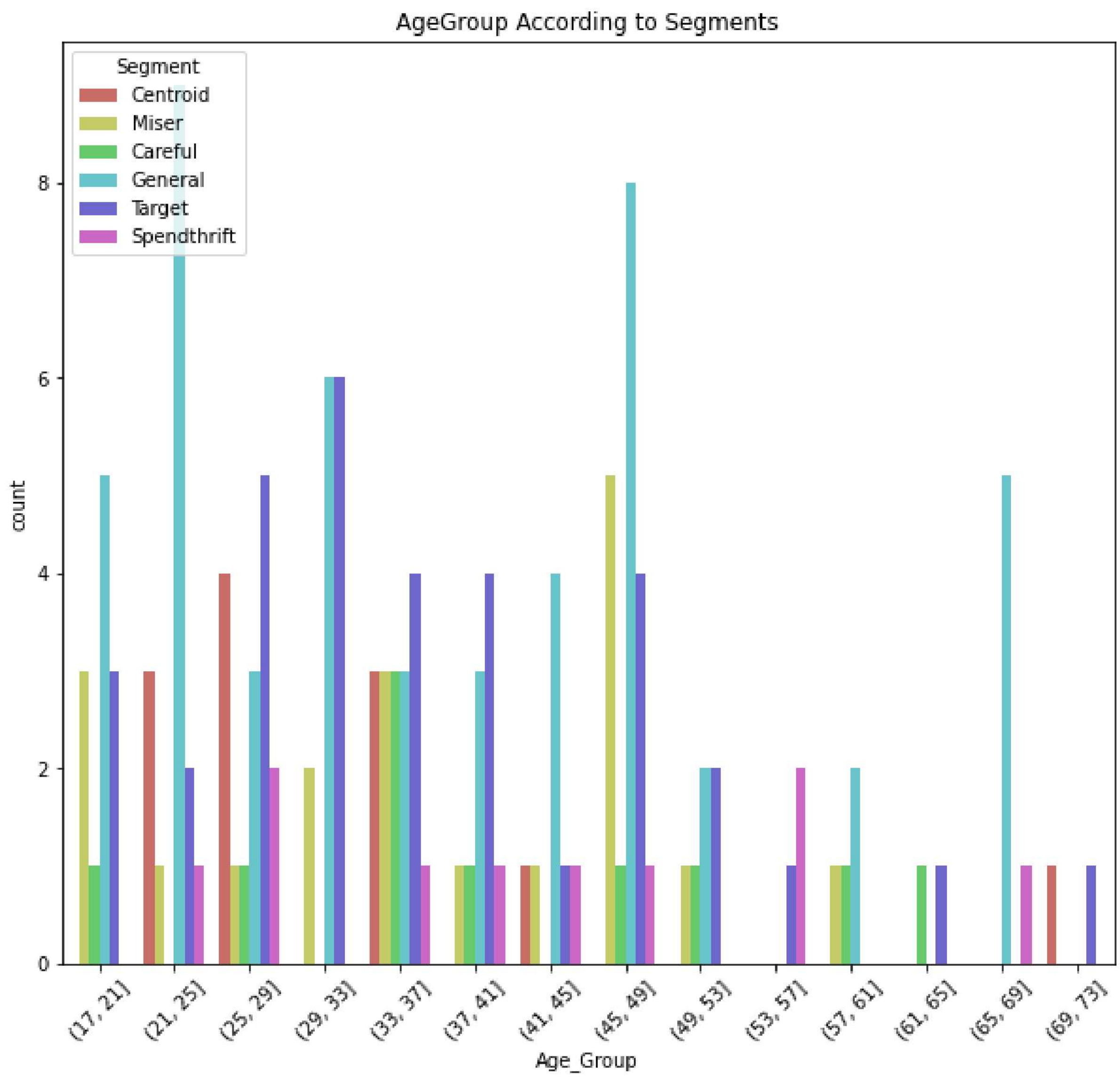
5)Annual Income of Male under 56-67 dollar there spendind score is more than Female

6)Annual Income of Male under 67-78 dollar there spendind score is equal than Female

7)Annual Income of Male under 78-89 dollar there spendind score is less than Female

8)Annual Income of Male under 89-100 dollar there spendind score is more than Female

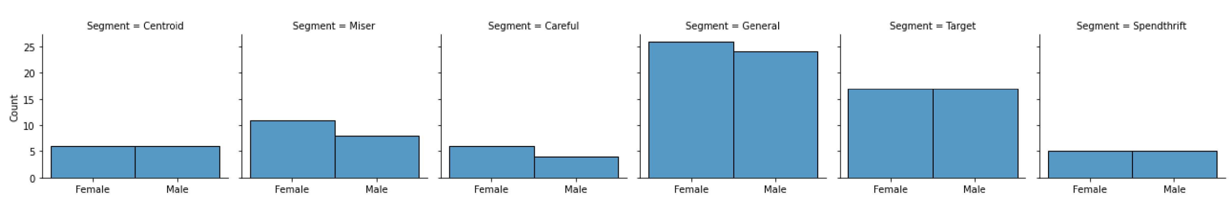
In [270]:
```
fig = plt.figure(figsize=(10,9))
sns.countplot(data=city_pho,x='Age_Group', palette='hls',hue='Segment').set(title
plt.xticks(rotation=45)
plt.show()
```



In Phoenix Most of people are Targeted,Miser & General Spending

In [317]:
```
fig = plt.figure(figsize=(10,8))
g = sns.FacetGrid(city_pho, col="Segment")
g.map(sns.histplot, "Gender")
plt.show()
```

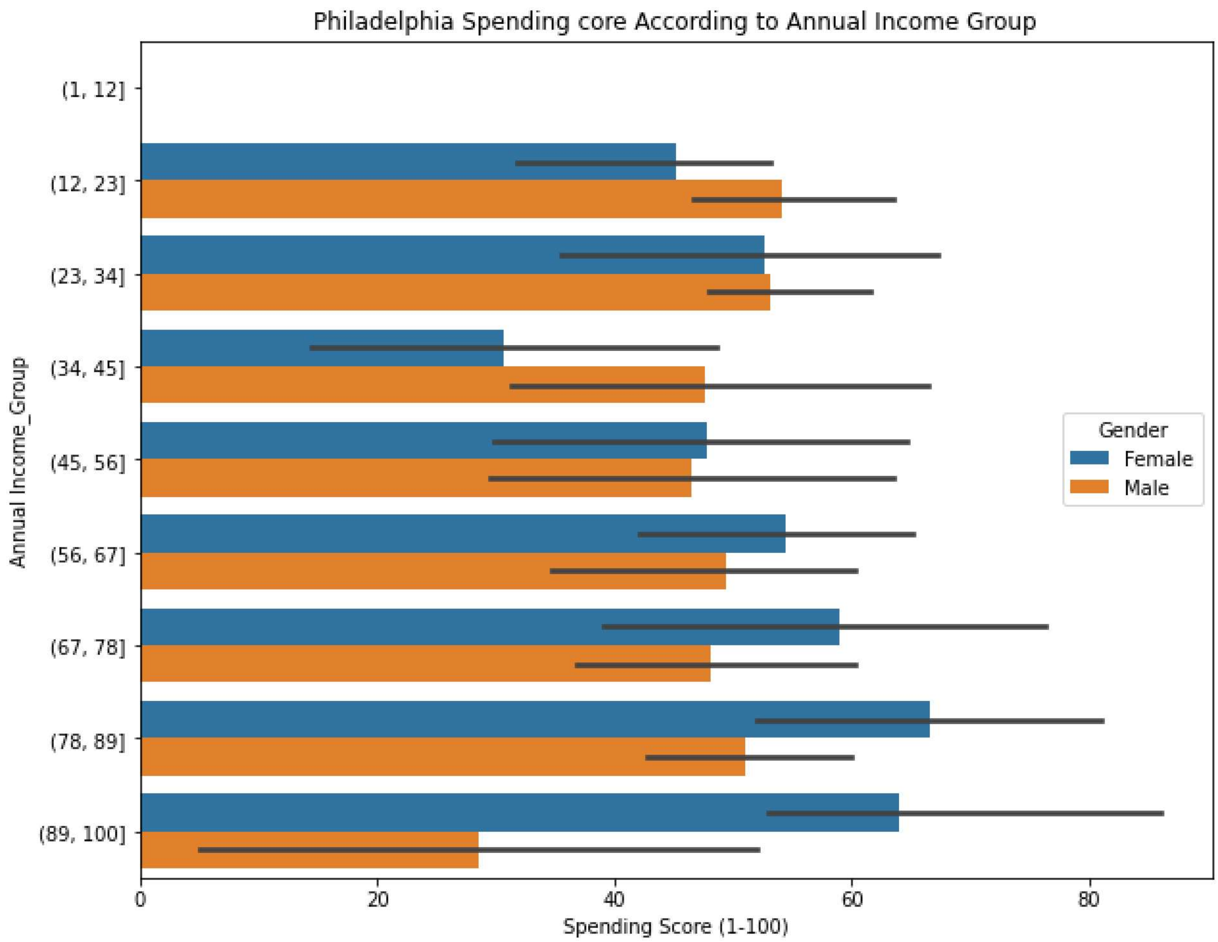
```
<Figure size 720x576 with 0 Axes>
```



Philadelphia

In [315]:

```
fig = plt.figure(figsize=(10,8))
sns.barplot(data=city_phi,x='Spending Score (1-100)',y='Annual Income_Group' ,hue
plt.show()
```



-- In the Above Graph we can see that

1)Annul Income of Male under 12-13 dollar there spending score is more than Female

2)Annual Income of Male under 23-24 dollar there spending score is more than Femals

3)Annual Income of Male under 34-45 dollar there spending score is more than Female

4)Annual Income of Male under 45-56 dollar there spending score is less than Female

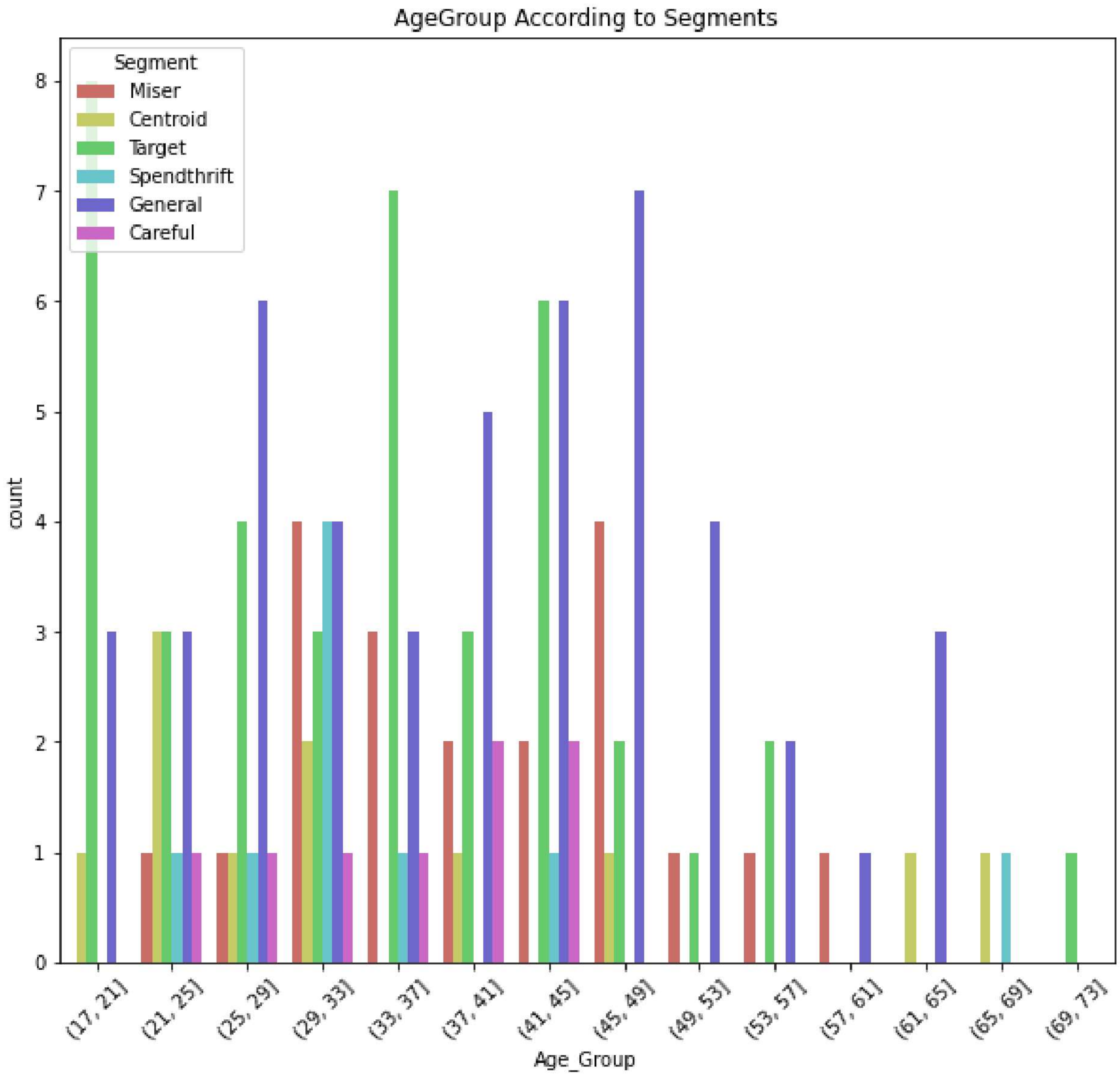
5)Annual Income of Male under 56-67 dollar there spendind score is less than Female

6)Annual Income of Male under 67-78 dollar there spendind score is less than Female

7)Annual Income of Male under 78-89 dollar there spendind score is less than Female

8)Annual Income of Male under 89-100 dollar there spendind score is less than Female

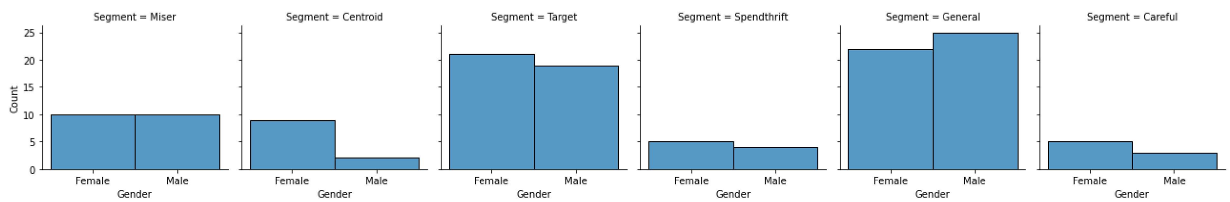
In [316]:
```
fig = plt.figure(figsize=(10,9))
sns.countplot(data=city_phi,x='Age_Group', palette='hls',hue='Segment').set(title
plt.xticks(rotation=45)
plt.show()
```



In Philadelphia Most of People is General,Miser & Target Spending

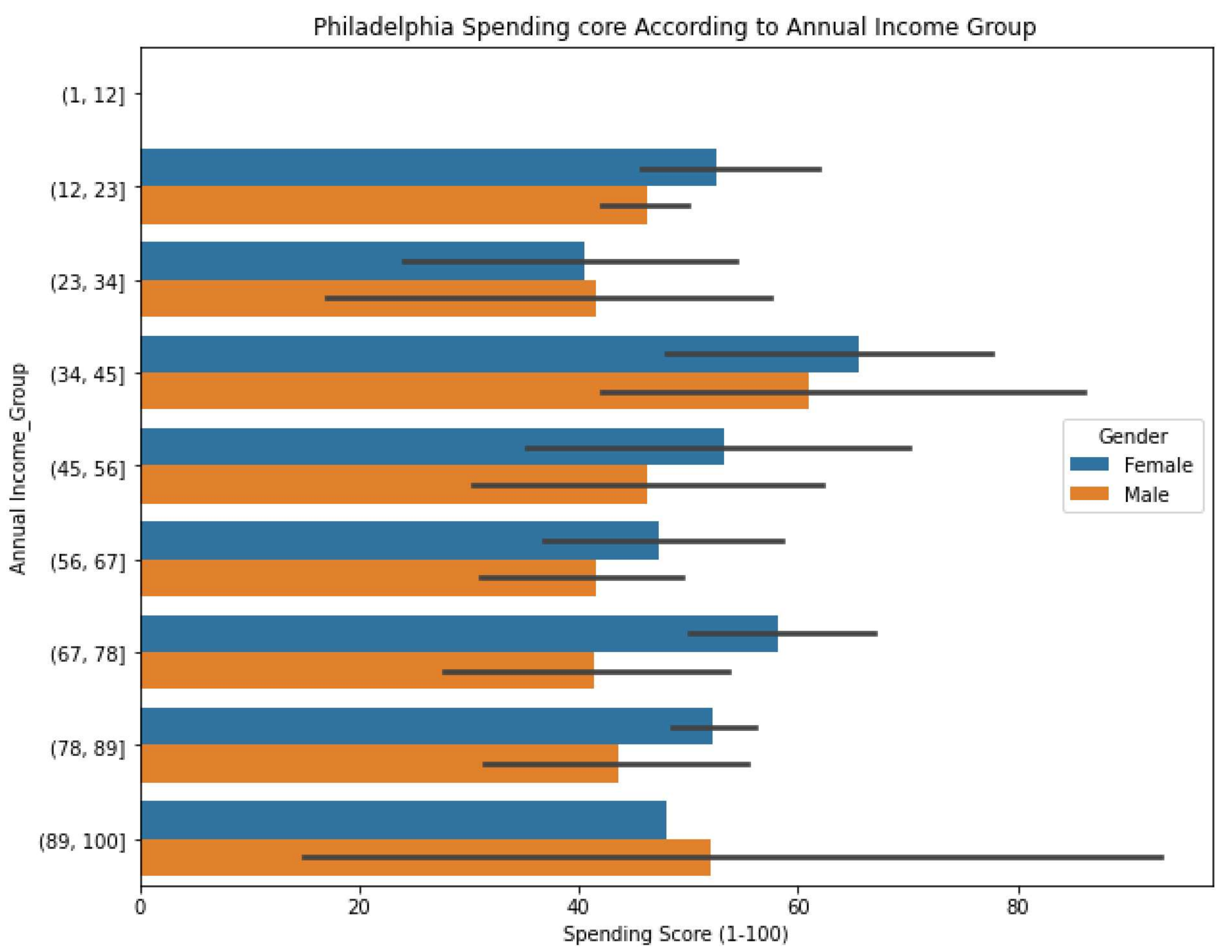
```
In [318]: fig = plt.figure(figsize=(10,8))
          g = sns.FacetGrid(city_phi, col="Segment")
          g.map(sns.histplot, "Gender")
          plt.show()
```

```
<Figure size 720x576 with 0 Axes>
```



San Antonio

```
In [319]: fig = plt.figure(figsize=(10,8))
          sns.barplot(data=city_san,x='Spending Score (1-100)',y='Annual Income_Group' ,hue
          plt.show()
```



-- In the Above Graph we can see that

1)Annul Income of Male under 12-13 dollar there spending score is less than Female

2)Annual Income of Male under 23-24 dollar there spending score is more than Femals

3)Annual Income of Male under 34-45 dollar there spending score is less than Female

4)Annual Income of Male under 45-56 dollar there spending score is less than Female

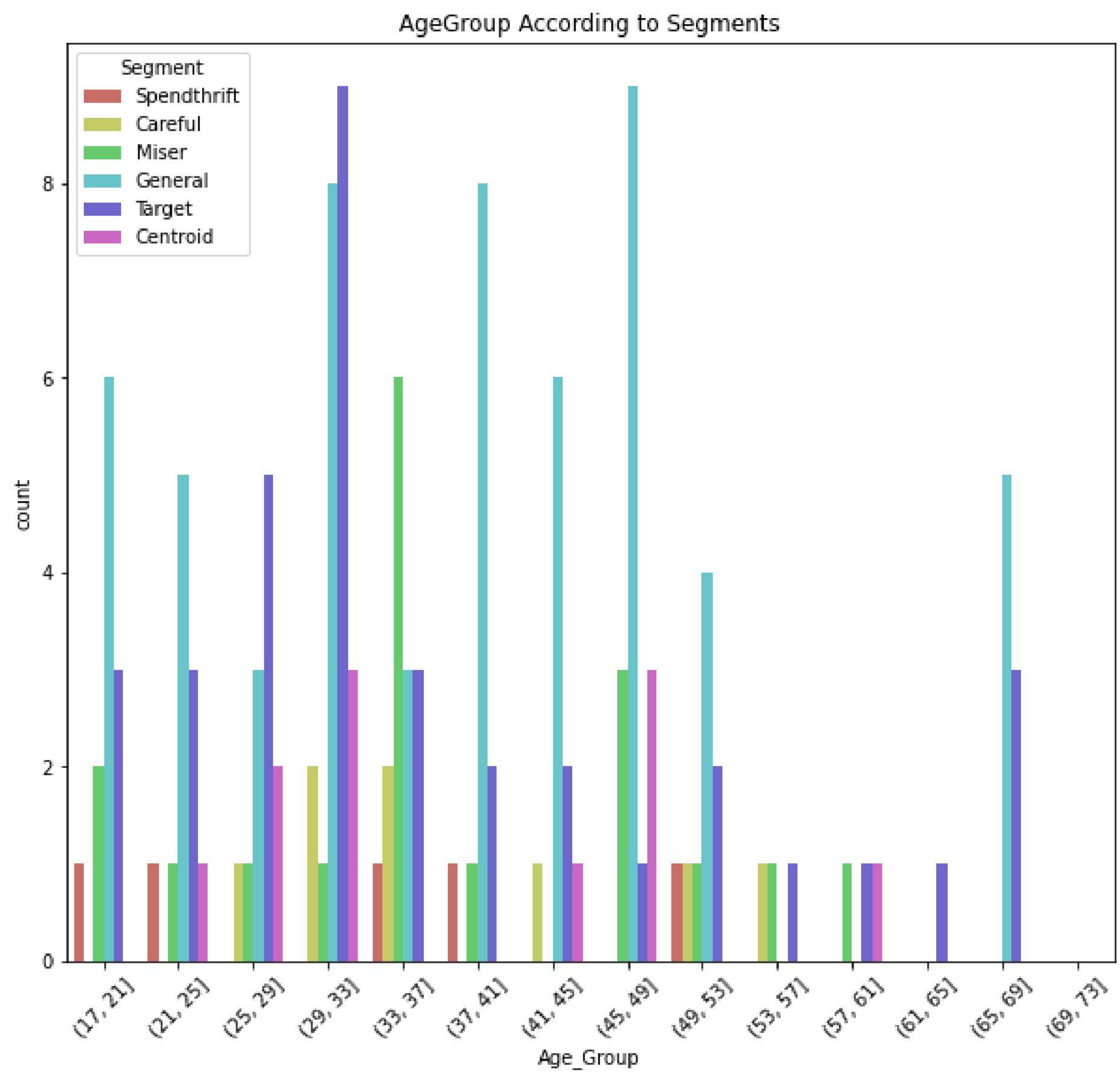
5)Annual Income of Male under 56-67 dollar there spendind score is less than Female

6)Annual Income of Male under 67-78 dollar there spendind score is less than Female

7)Annual Income of Male under 78-89 dollar there spendind score is less than Female

8)Annual Income of Male under 89-100 dollar there spendind score is more than Female

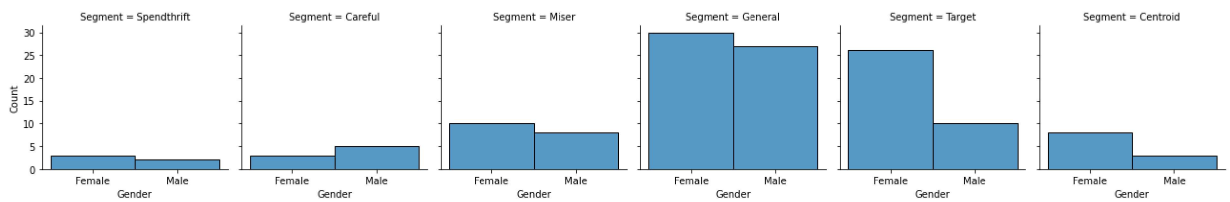
In [320]:
```
fig = plt.figure(figsize=(10,9))
sns.countplot(data=city_san,x='Age_Group', palette='hls',hue='Segment').set(title
plt.xticks(rotation=45)
plt.show()
```



In San Antonio most of people are Targeted and General Spending

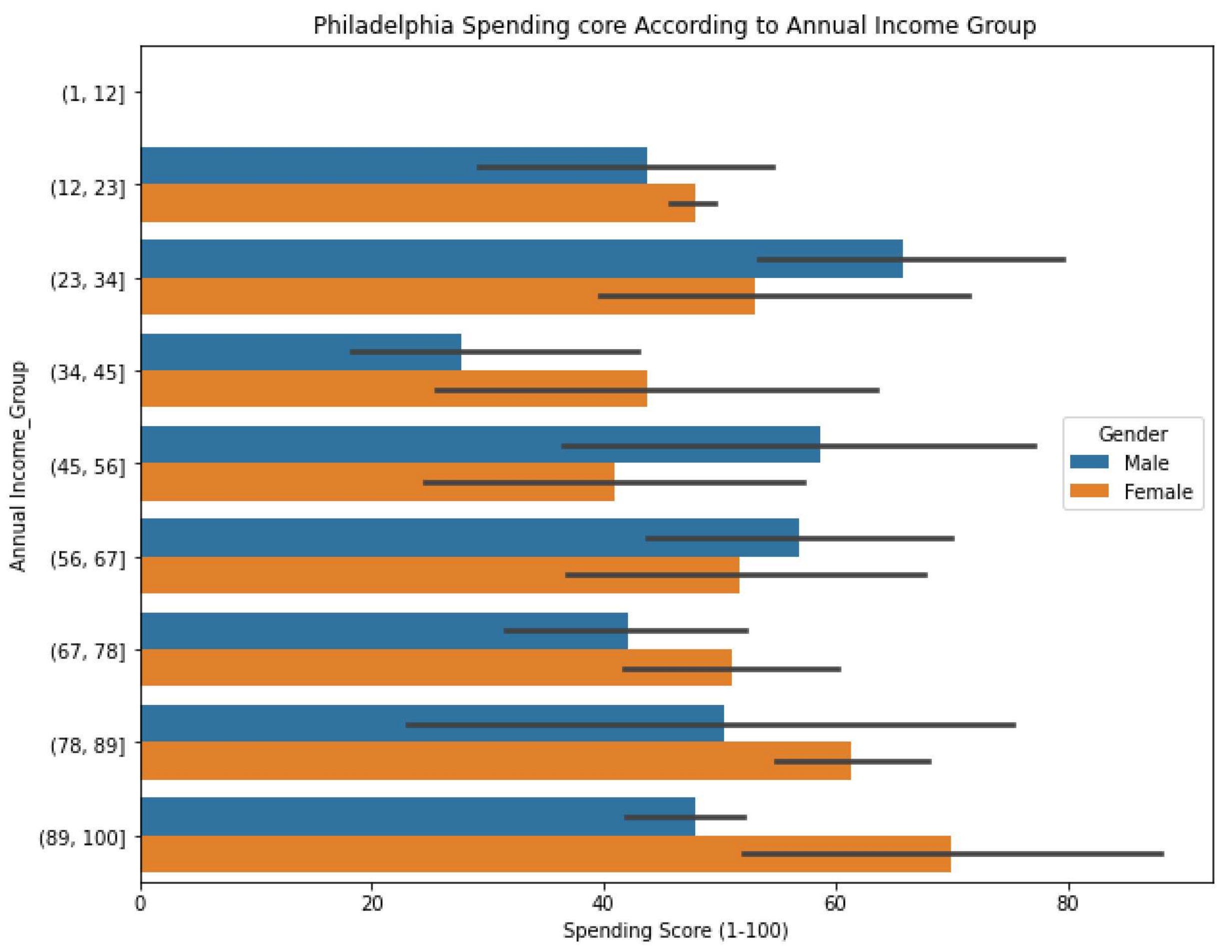
In [321]:
```
fig = plt.figure(figsize=(10,8))
g = sns.FacetGrid(city_san, col="Segment")
g.map(sns.histplot, "Gender")
plt.show()
```

```
<Figure size 720x576 with 0 Axes>
```



San Diego

In [322]:
```
fig = plt.figure(figsize=(10,8))
sns.barplot(data=city_san_d,x='Spending Score (1-100)',y='Annual Income_Group' ,h
plt.show()
```



-- In the Above Graph we can see that

1)Annul Income of Male under 12-13 dollar there spending score is more than Female

2)Annual Income of Male under 23-24 dollar there spending score is less than Femals

3)Annual Income of Male under 34-45 dollar there spending score is more than Female

4)Annual Income of Male under 45-56 dollar there spending score is less than Female

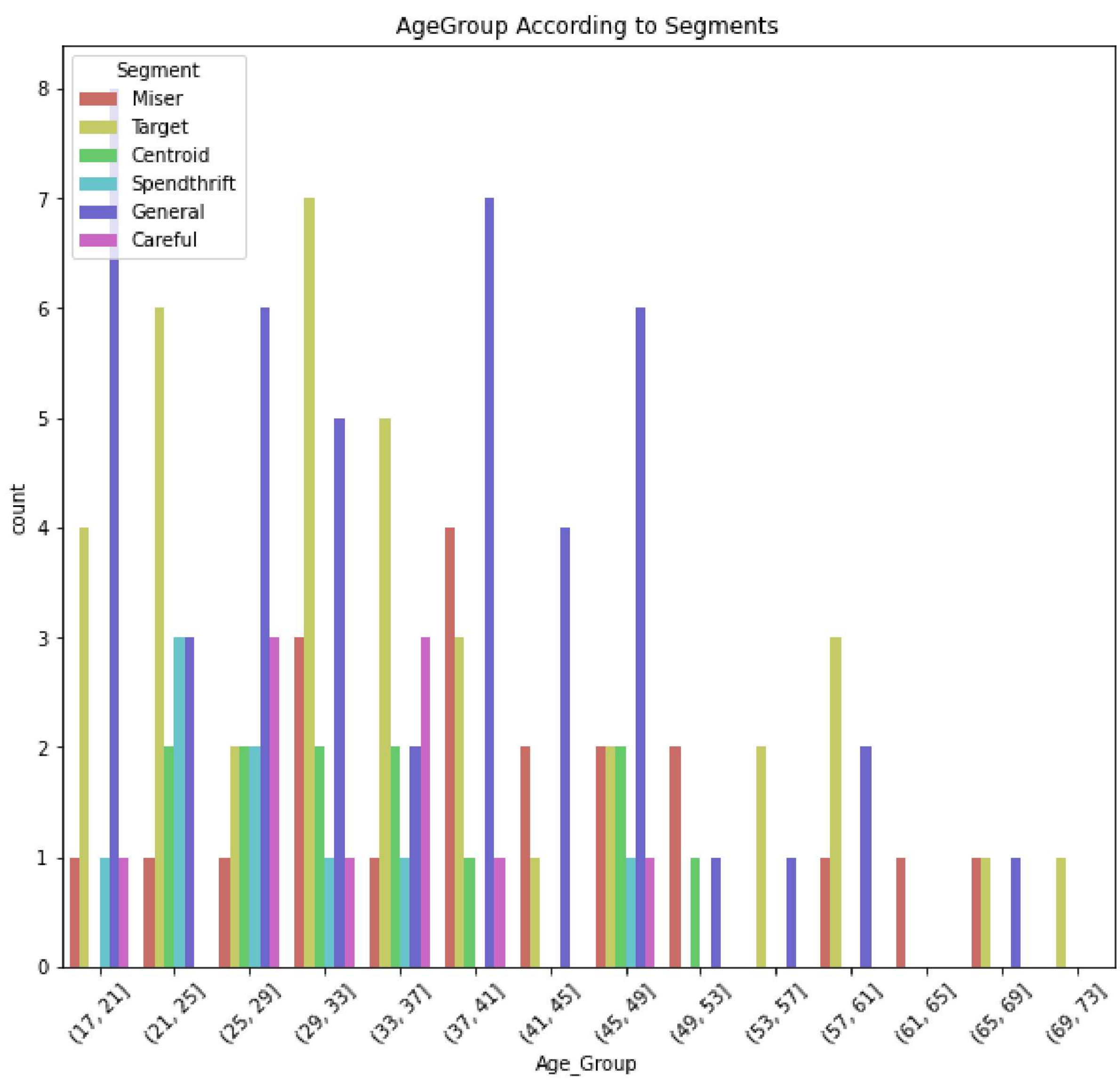
5)Annual Income of Male under 56-67 dollar there spendind score is less than Female

6)Annual Income of Male under 67-78 dollar there spendind score is more than Female

7)Annual Income of Male under 78-89 dollar there spendind score is more than Female

8)Annual Income of Male under 89-100 dollar there spendind score is more than Female

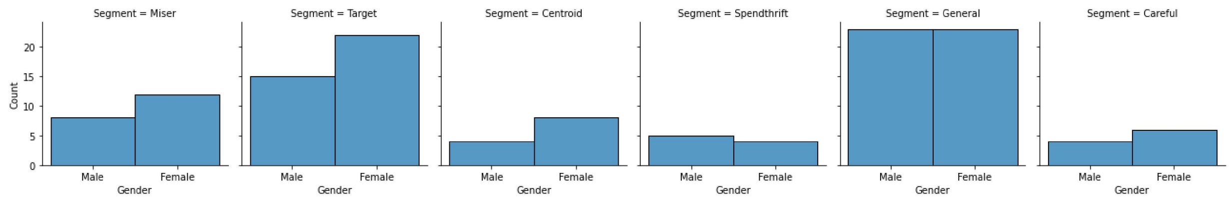
In [323]:
```
fig = plt.figure(figsize=(10,9))
sns.countplot(data=city_san_d,x='Age_Group', palette='hls',hue='Segment').set(tit
plt.xticks(rotation=45)
plt.show()
```



In San Diego people are General,Tageted and Miser Spending

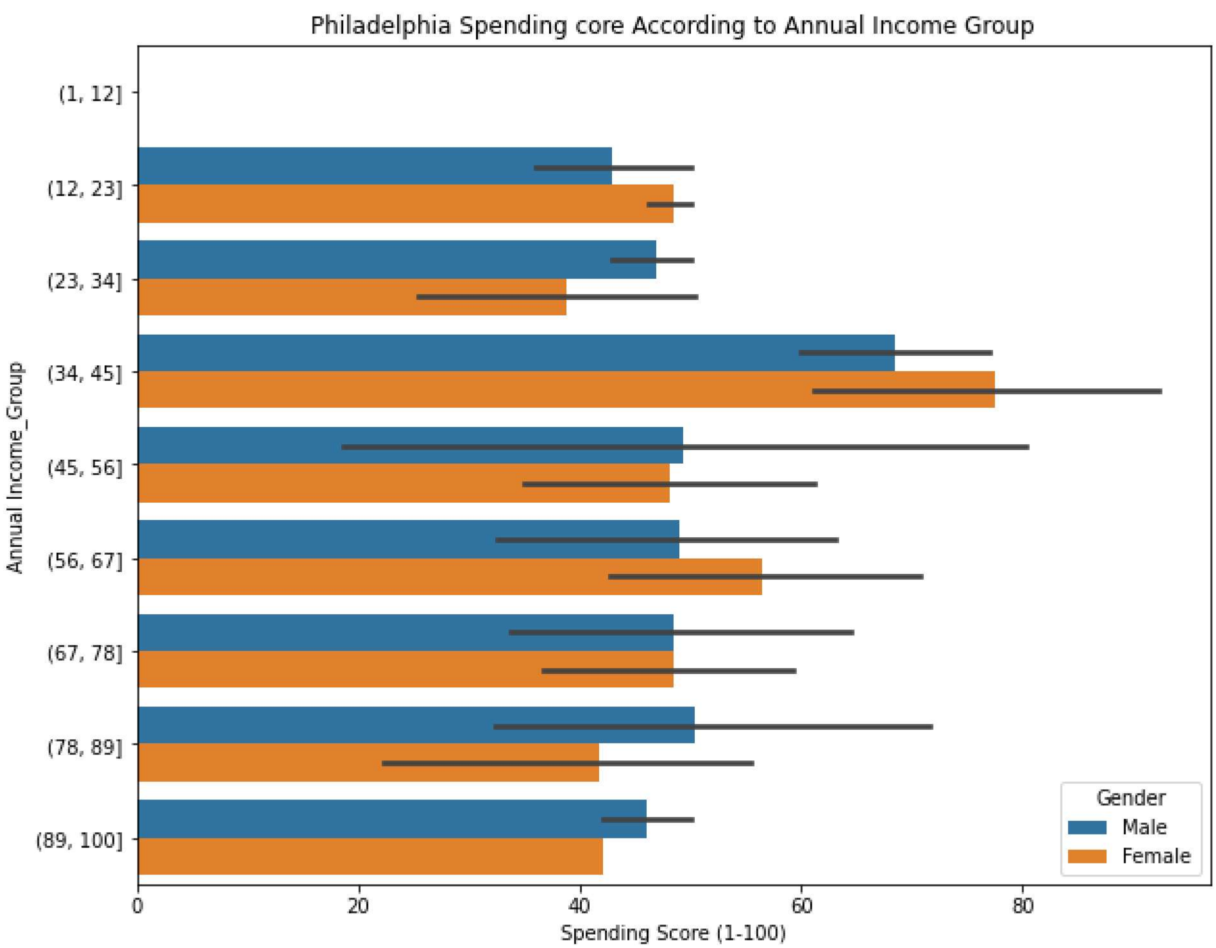
In [325]:
```
fig = plt.figure(figsize=(10,8))
g = sns.FacetGrid(city_san_d, col="Segment")
g.map(sns.histplot, "Gender")
plt.show()
```

```
<Figure size 720x576 with 0 Axes>
```



Washington

In [326]:
```
fig = plt.figure(figsize=(10,8))
sns.barplot(data=city_wash,x='Spending Score (1-100)',y='Annual Income_Group' ,hu
plt.show()
```



-- In the Above Graph we can see that

1)Annul Income of Male under 12-13 dollar there spending score is more than Female

2)Annual Income of Male under 23-24 dollar there spending score is less than Femals

3)Annual Income of Male under 34-45 dollar there spending score is more than Female

4)Annual Income of Male under 45-56 dollar there spending score is less than Female

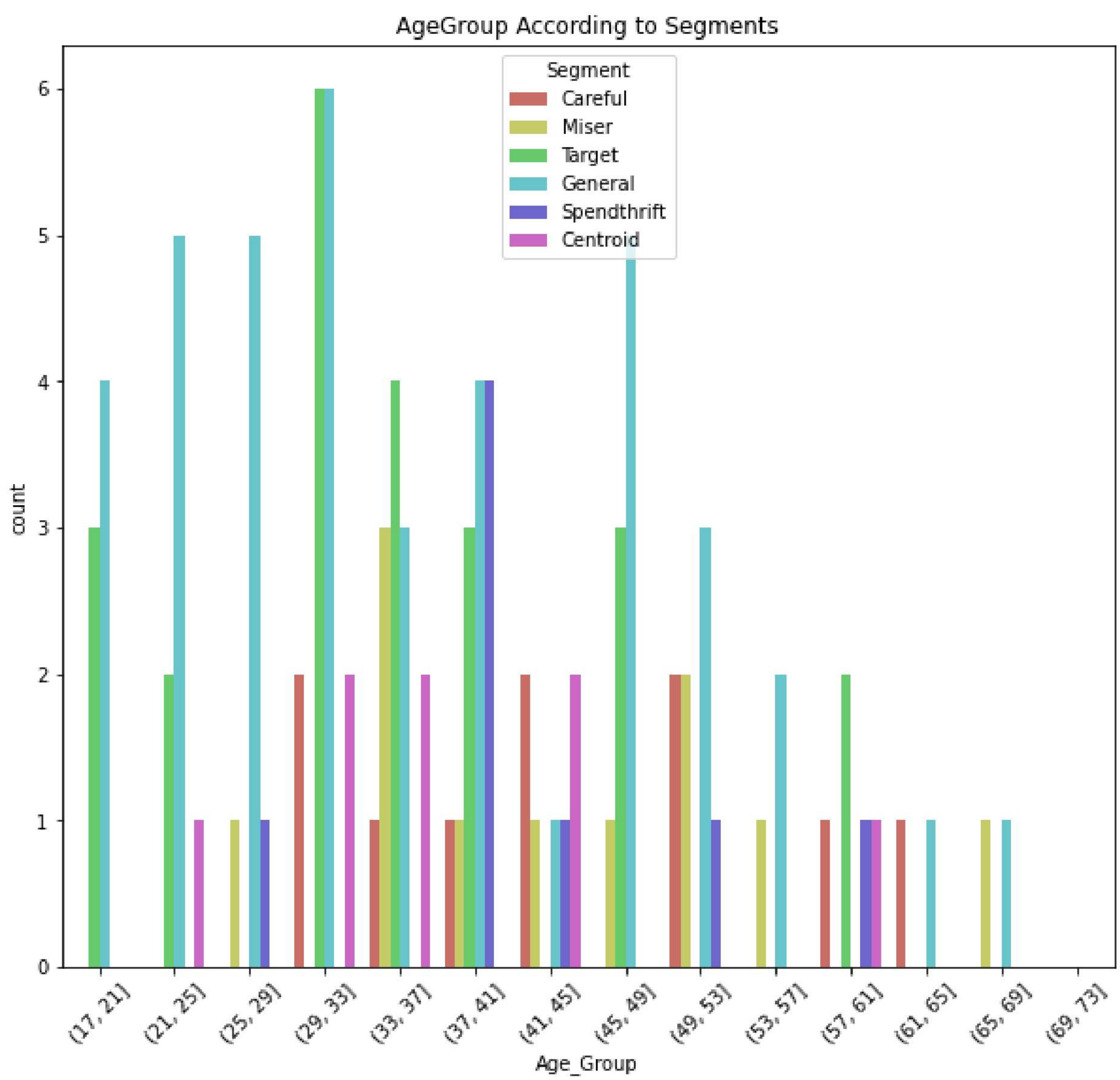
5)Annual Income of Male under 56-67 dollar there spendind score is more than Female

6)Annual Income of Male under 67-78 dollar there spendind score is equal than Female

7)Annual Income of Male under 78-89 dollar there spendind score is less than Female

8)Annual Income of Male under 89-100 dollar there spendind score is less than Female

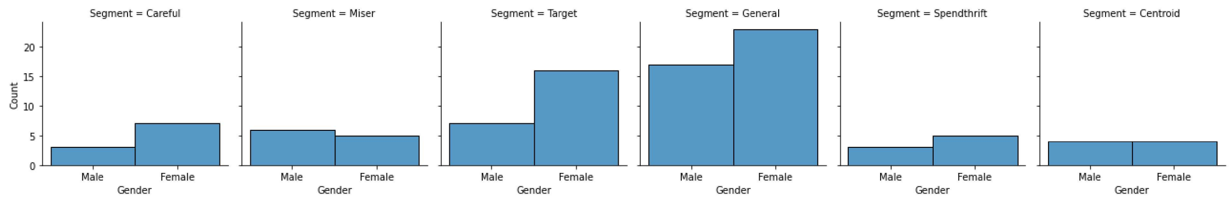
In [327]:
```
fig = plt.figure(figsize=(10,9))
sns.countplot(data=city_wash,x='Age_Group', palette='hls',hue='Segment').set(titl
plt.xticks(rotation=45)
plt.show()
```



In Washington people are General,Targeted and Careful Spending

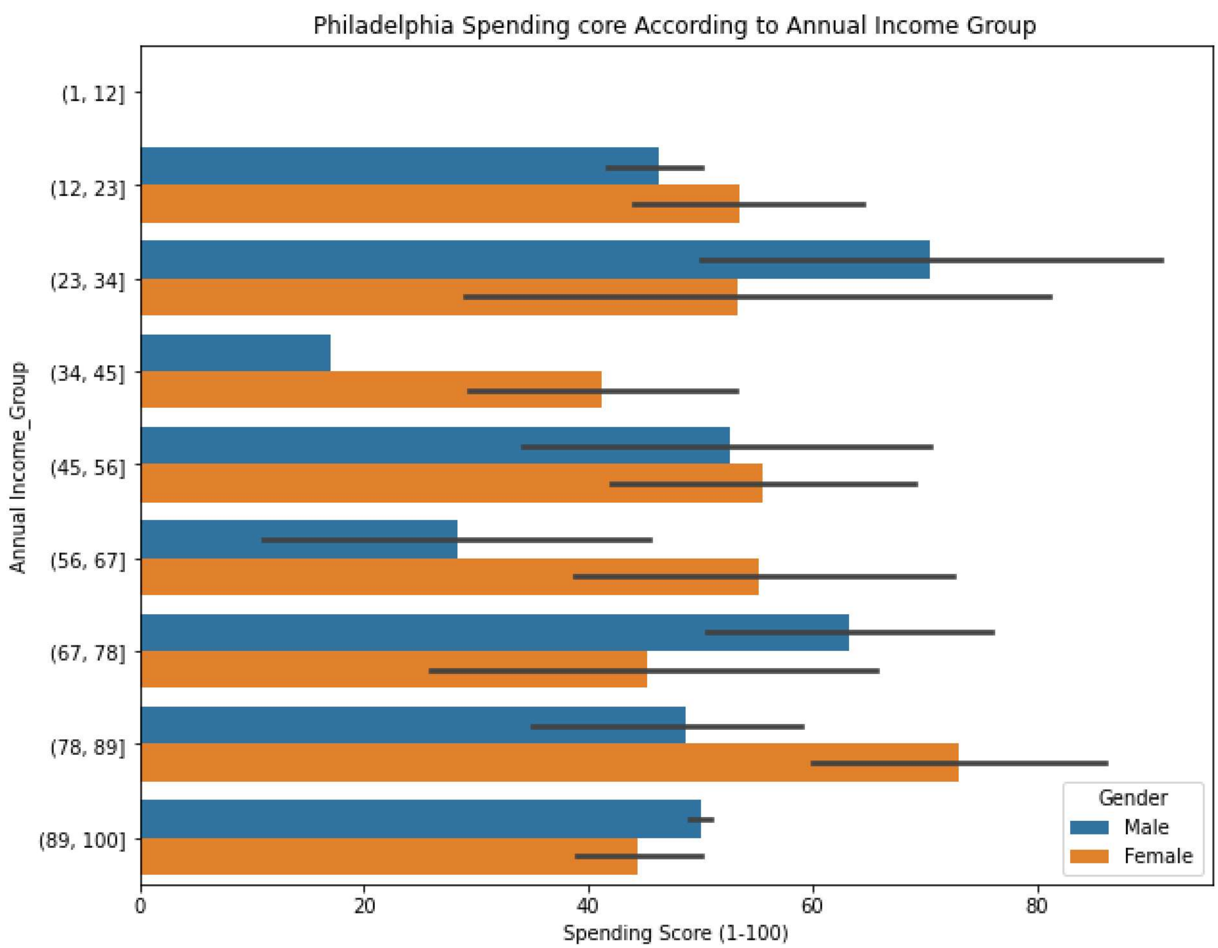
In [328]:
```
fig = plt.figure(figsize=(10,8))
g = sns.FacetGrid(city_wash, col="Segment")
g.map(sns.histplot, "Gender")
plt.show()
```

```
<Figure size 720x576 with 0 Axes>
```



New York

In [329]:
```
fig = plt.figure(figsize=(10,8))
sns.barplot(data=city_new,x='Spending Score (1-100)',y='Annual Income_Group' ,hue
plt.show()
```



-- In the Above Graph we can see that

1)Annul Income of Male under 12-13 dollar there spending score is more than Female

2)Annual Income of Male under 23-24 dollar there spending score is less than Femals

3)Annual Income of Male under 34-45 dollar there spending score is more than Female

4)Annual Income of Male under 45-56 dollar there spending score is more than Female

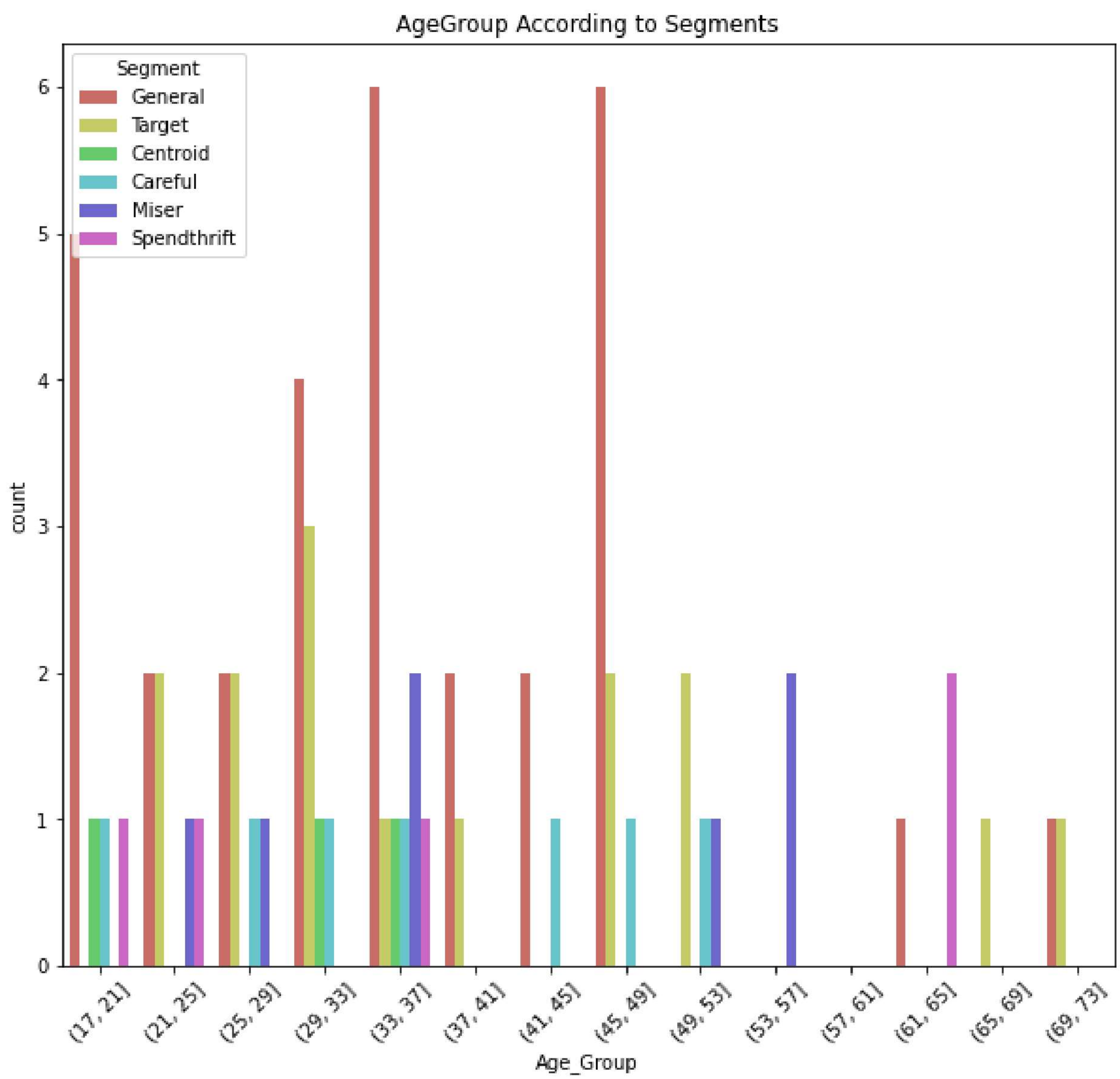
5)Annual Income of Male under 56-67 dollar there spendind score is more than Female

6)Annual Income of Male under 67-78 dollar there spendind score is less than Female

7)Annual Income of Male under 78-89 dollar there spendind score is more than Female

8)Annual Income of Male under 89-100 dollar there spendind score is less than Female

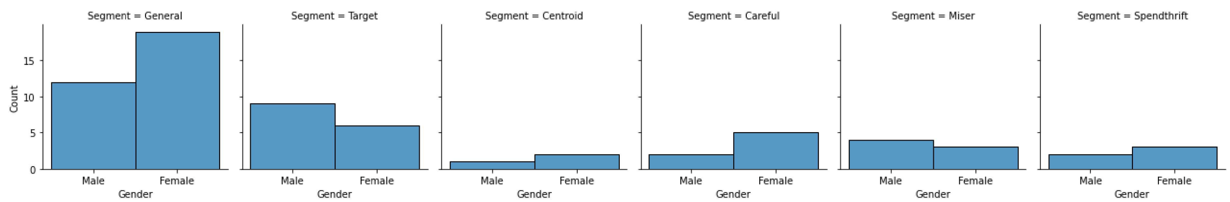
In [330]:
```
fig = plt.figure(figsize=(10,9))
sns.countplot(data=city_new,x='Age_Group', palette='hls',hue='Segment').set(title
plt.xticks(rotation=45)
plt.show()
```



In New York people are General,Targrted and careful Spending

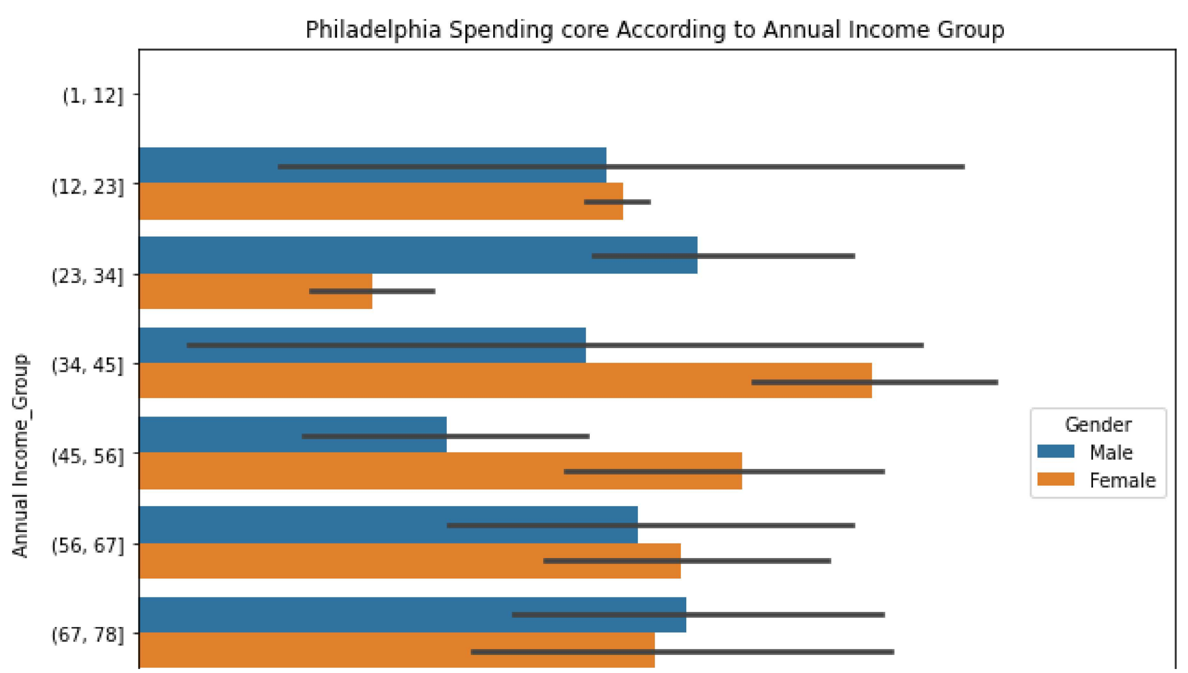
In [331]:
```
fig = plt.figure(figsize=(10,8))
g = sns.FacetGrid(city_new, col="Segment")
g.map(sns.histplot, "Gender")
plt.show()
```

```
<Figure size 720x576 with 0 Axes>
```



Seattle

In [332]:
```
fig = plt.figure(figsize=(10,8))
sns.barplot(data=city_sea,x='Spending Score (1-100)',y='Annual Income_Group' ,hue
plt.show()
```



-- In the Above Graph we can see that

1)Annul Income of Male under 12-13 dollar there spending score is more than Female

2)Annual Income of Male under 23-24 dollar there spending score is less than Femals

3)Annual Income of Male under 34-45 dollar there spending score is more than Female

4)Annual Income of Male under 45-56 dollar there spending score is more than Female

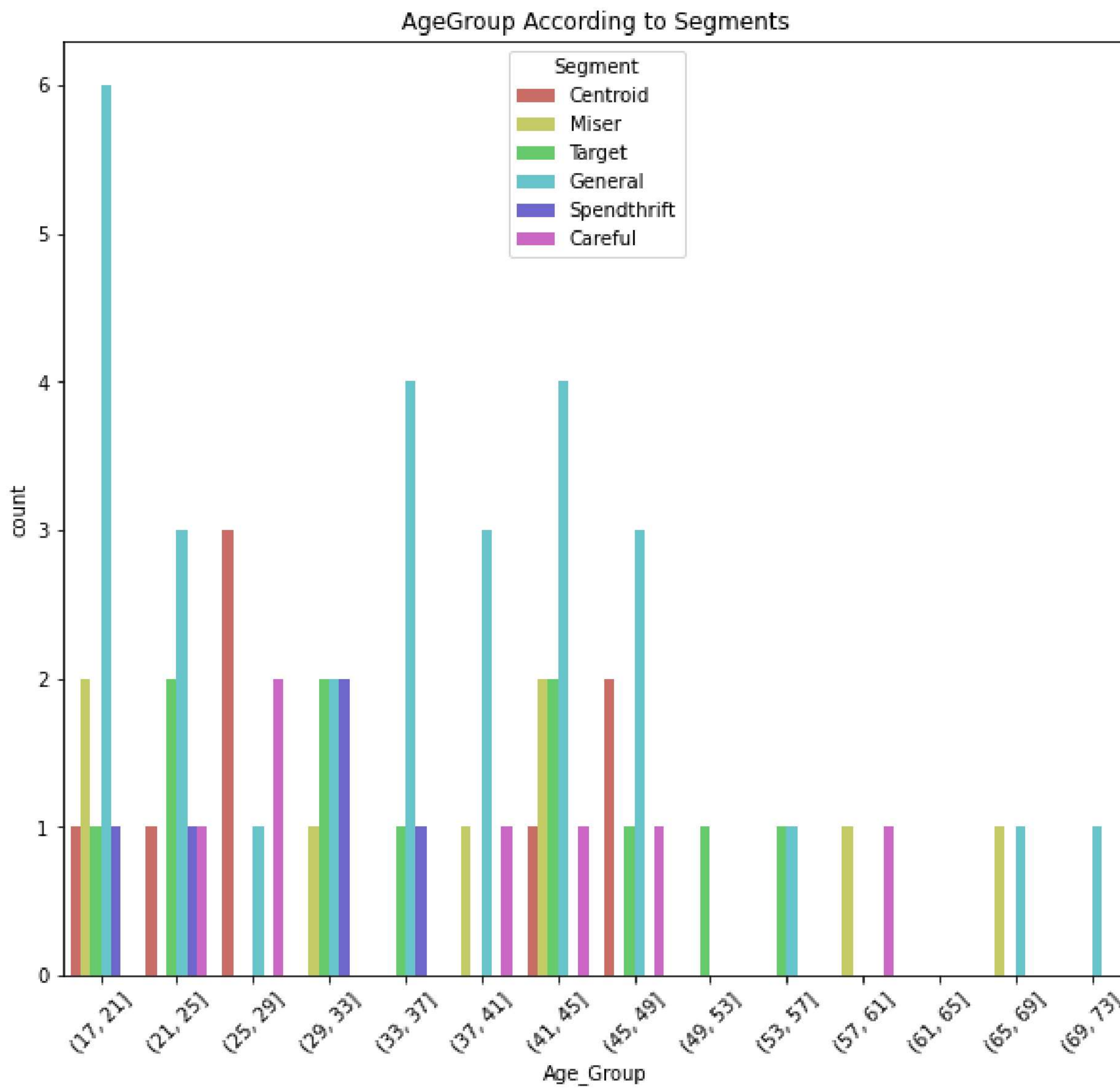
5)Annual Income of Male under 56-67 dollar there spendind score is more than Female

6)Annual Income of Male under 67-78 dollar there spendind score is less than Female

7)Annual Income of Male under 78-89 dollar there spendind score is more than Female

8)Annual Income of Male under 89-100 dollar there spendind score is less than Female

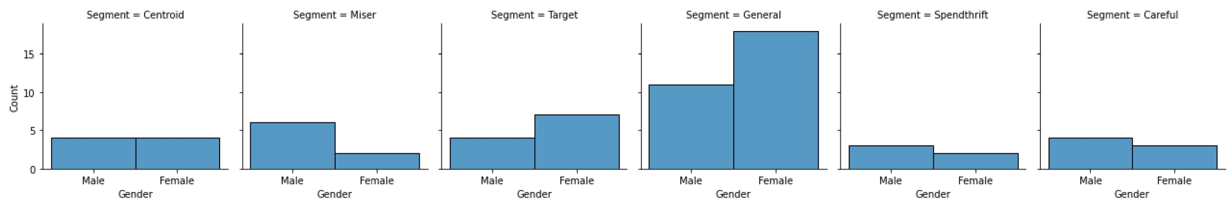
In [333]:
```
fig = plt.figure(figsize=(10,9))
sns.countplot(data=city_sea,x='Age_Group', palette='hls',hue='Segment').set(title
plt.xticks(rotation=45)
plt.show()
```



In Seattle people are General,Miser and Centroid Spending

In [334]:
```
fig = plt.figure(figsize=(10,8))
g = sns.FacetGrid(city_sea, col="Segment")
g.map(sns.histplot, "Gender")
plt.show()
```

<Figure size 720x576 with 0 Axes>



In [ ]:

In [265]:
```
df1
```

Out[265]:

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)	City	Age_Group	Segment	Generati
0	1	Male	19	15	39.0	New York	(17, 21]	General	You
1	2	Male	21	15	81.0	Seattle	(17, 21]	Centroid	You
2	3	Female	20	16	6.0	Los Angeles.	(17, 21]	Miser	You
3	4	Female	23	16	77.0	Chicago.	(21, 25]	Centroid	You
4	5	Female	31	17	40.0	Houston.	(29, 33]	General	AdultHo
...	...	...	...	...	...	...	...	...	
1175	1176	Female	47	88	73.0	Chicago.	(45, 49]	Target	Early Eld
1176	1177	Male	48	88	10.0	Houston.	(45, 49]	Miser	Early Eld
1177	1178	Male	49	88	72.0	Phoenix.	(45, 49]	Target	Early Eld
1178	1179	Male	50	93	5.0	Philadelphia.	(49, 53]	Miser	Early Eld
1179	1180	Male	51	93	93.0	San Antonio.	(49, 53]	Spendthrift	Elder/C

1180 rows × 10 columns

Machine Learning Algorithm

In [47]:

In [206]:
```
from sklearn.model_selection import train_test_split
```

In [207]:
```
x = df1[['Annual Income (k$)','Spending Score (1-100)']]
```

In [208]:
```
y =df1['Segment']
```

In [209]:
```
x
```

Out[209]:

	Annual Income (k$)	Spending Score (1-100)
0	15	39.0
1	15	81.0
2	16	6.0
3	16	77.0
4	17	40.0
...	...	...
1175	88	73.0
1176	88	10.0
1177	88	72.0
1178	93	5.0
1179	93	93.0

1180 rows × 2 columns

In [210]:
```
y
```

Out[210]:
```
0           General
1          Centroid
2             Miser
3          Centroid
4           General
           ...
1175         Target
1176          Miser
1177         Target
1178          Miser
1179    Spendthrift
Name: Segment, Length: 1180, dtype: object
```

In [211]:
```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3,random_state=0)
```

In [212]:
```
x_test
```

Out[212]:

	Annual Income (k$)	Spending Score (1-100)
1099	61	55.0
351	78	55.0
704	62	74.0
434	33	75.0
513	64	6.0
...	...	...
561	79	45.0
54	43	45.0
1005	17	50.0
526	71	98.0
1062	47	14.0

354 rows × 2 columns

In [213]:
```
x_train
```

Out[213]:

	Annual Income (k$)	Spending Score (1-100)
634	33	95.0
1006	18	50.0
686	57	90.0
492	60	68.0
1010	19	50.0
...	...	...
1033	33	50.0
763	81	50.0
835	33	11.0
559	78	54.0
684	54	69.0

826 rows × 2 columns

In [214]:
```
y_test
```

Out[214]:
```
1099          Target
351           Target
704           Target
434           Target
513            Miser
            ...
561          General
54           General
1005         General
526      Spendthrift
1062           Miser
Name: Segment, Length: 354, dtype: object
```

In [215]:
```
y_train
```

Out[215]:
```
634      Spendthrift
1006         General
686         Centroid
492           Target
1010         General
            ...
1033         General
763          General
835            Miser
559           Target
684           Target
Name: Segment, Length: 826, dtype: object
```

In [216]:
```
from sklearn.neighbors import KNeighborsClassifier
```

In [128]:
```
knn = KNeighborsClassifier(n_neighbors=1)
```

In [131]:
```
 knn.fit(x_train,y_train)
```

Out[131]:
```
KNeighborsClassifier(n_neighbors=30)
```

In [145]:
```
knn.score( x_test,y_test)
```

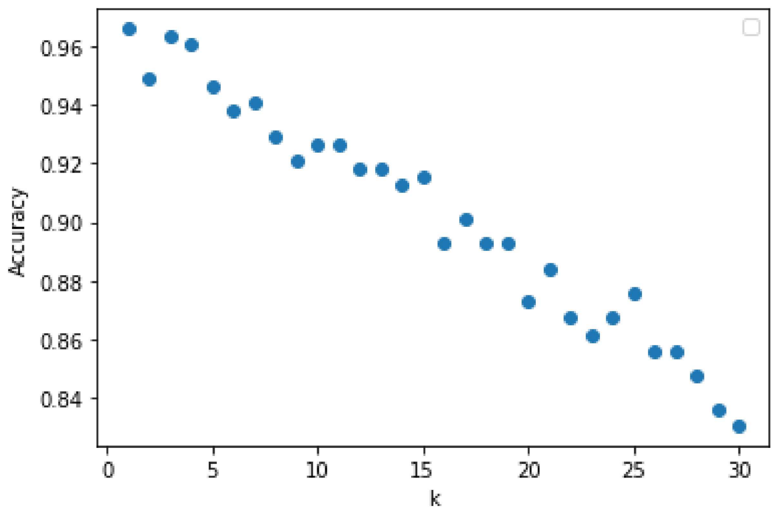
Out[145]:
```
0.8305084745762712
```

- Above Score 83% prediction of Segments on the basis of Annual Income and Spending Score

In [155]:
```
import matplotlib.pyplot as plt
```

In [156]:
```
k_range = range(1,31)
score = []
for k in k_range:
    knn = KNeighborsClassifier(n_neighbors=k)
    knn.fit(x_train,y_train)
    score.append(knn.score(x_test,y_test))
plt.xlabel("k")
plt.ylabel("Accuracy")
plt.legend()
plt.scatter(k_range,score)
plt.show()
```

```
No artists with labels found to put in legend.  Note that artists whose label s
tart with an underscore are ignored when legend() is called with no argument.
```



In [ ]:

In [217]:
```
from sklearn.ensemble import RandomForestClassifier
```

In [219]:
```
rfc = RandomForestClassifier(random_state=0)
```

In [220]:
```
rfc.fit(x_test,y_test)
```

Out[220]:
```
RandomForestClassifier(random_state=0)
```

In [221]:
```
y_pred = rfc.predict(x_test)
```

In [222]:
```
y_pred
```

Out[222]:
```
array(['Target', 'Target', 'Target', 'Target', 'Miser', 'Careful',
       'Target', 'General', 'General', 'Target', 'Centroid', 'General',
       'Miser', 'Target', 'Target', 'General', 'Centroid', 'Miser',
       'Spendthrift', 'Target', 'Careful', 'General', 'Target', 'Target',
       'General', 'General', 'General', 'Miser', 'Careful', 'Spendthrift',
       'Miser', 'Careful', 'Miser', 'General', 'General', 'Centroid',
       'Target', 'General', 'General', 'Careful', 'Miser', 'General',
       'Target', 'Spendthrift', 'General', 'General', 'Careful',
       'Centroid', 'Target', 'General', 'General', 'Target', 'Target',
       'General', 'General', 'Careful', 'Spendthrift', 'General',
       'General', 'General', 'Target', 'Centroid', 'Target', 'Careful',
       'Centroid', 'Miser', 'Target', 'General', 'General', 'Spendthrift',
       'General', 'Spendthrift', 'Spendthrift', 'Target', 'Target',
       'General', 'Careful', 'Centroid', 'General', 'General', 'General',
       'Target', 'Miser', 'General', 'General', 'Careful', 'Careful',
       'Miser', 'Target', 'General', 'Spendthrift', 'General', 'Target',
       'Miser', 'General', 'Target', 'Careful', 'General', 'Miser',
       'General', 'General', 'Target', 'General', 'Target', 'General',
       'Miser', 'General', 'Target', 'General', 'General', 'Target',
       'General', 'Target', 'General', 'General', 'Target', 'Careful',
       'Target', 'General', 'Careful', 'Miser', 'Miser', 'General',
       'Target', 'Target', 'General', 'Target', 'Miser', 'General',
       'General', 'Miser', 'Spendthrift', 'General', 'General', 'Target',
       'Target', 'General', 'General', 'Target', 'General', 'Careful',
       'Miser', 'Miser', 'General', 'Target', 'Careful', 'Target',
       'Target', 'Target', 'General', 'General', 'Centroid', 'Miser',
       'Miser', 'Centroid', 'Target', 'Careful', 'Centroid', 'General',
       'Target', 'Miser', 'Centroid', 'Miser', 'General', 'Target',
       'Centroid', 'Target', 'General', 'General', 'Target', 'Miser',
       'General', 'General', 'Target', 'Centroid', 'General', 'Centroid',
       'Miser', 'General', 'General', 'Careful', 'General', 'General',
       'General', 'Target', 'Target', 'Centroid', 'Target', 'Target',
       'Centroid', 'General', 'General', 'General', 'Target', 'General',
       'General', 'Miser', 'General', 'General', 'General', 'Miser',
       'General', 'General', 'General', 'Careful', 'Miser', 'Spendthrift',
       'Target', 'Miser', 'General', 'Miser', 'Target', 'Target',
       'Target', 'General', 'General', 'Target', 'Target', 'Target',
       'Target', 'Centroid', 'General', 'General', 'Target', 'Centroid',
       'Target', 'Target', 'Target', 'General', 'General', 'Centroid',
       'Target', 'Miser', 'Target', 'General', 'Target', 'Miser',
       'Target', 'Target', 'General', 'General', 'General', 'Target',
       'General', 'Target', 'Target', 'Careful', 'General', 'General',
       'General', 'Miser', 'General', 'Careful', 'Careful', 'Centroid',
       'General', 'General', 'General', 'General', 'Target', 'Miser',
       'General', 'Centroid', 'Careful', 'Target', 'General', 'Miser',
       'General', 'Miser', 'Target', 'Miser', 'General', 'Centroid',
       'General', 'General', 'Target', 'Miser', 'Target', 'Careful',
       'General', 'General', 'General', 'Miser', 'Centroid',
       'Spendthrift', 'Miser', 'General', 'Miser', 'General', 'Miser',
       'General', 'General', 'Target', 'General', 'Target', 'General',
       'Target', 'Centroid', 'Miser', 'General', 'General', 'General',
       'General', 'Target', 'Target', 'General', 'Centroid', 'General',
       'General', 'Target', 'General', 'General', 'Spendthrift', 'Target',
       'General', 'Careful', 'Careful', 'Target', 'Miser', 'Careful',
```

```
 'Miser', 'General', 'General', 'Centroid', 'Spendthrift',
 'General', 'Target', 'Careful', 'General', 'General',
 'Spendthrift', 'Centroid', 'Miser', 'Target', 'General', 'Careful',
 'General', 'General', 'General', 'Target', 'General', 'General',
 'General', 'Miser', 'Target', 'General', 'Miser', 'General',
 'Target', 'General', 'General', 'General', 'Spendthrift', 'Miser'],
dtype=object)
```

In [223]:
```
len(y_pred)
```

Out[223]:
```
354
```

In [224]:
```
from sklearn.metrics import accuracy_score
```

In [226]:
```
print ('Model  Score is',accuracy_score(y_test,y_pred))
```

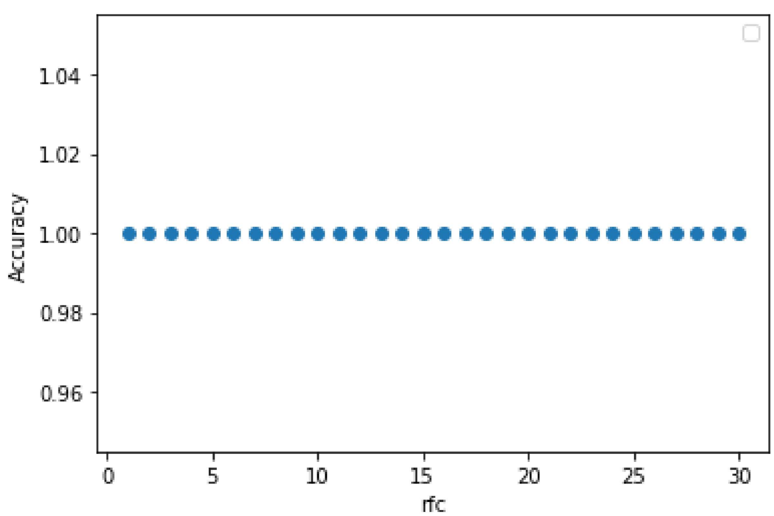
```
Model  Score is 1.0
```

In [227]:
```
print ('Model Score using impurity',accuracy_score(y_test,y_pred))
```

```
Model Score using impurity 1.0
```

In [157]:
```
k_range = range(1,31)
score = []
for k in k_range:
    rfc = RandomForestClassifier(random_state=k)
    rfc.fit(x_train,y_train)
    score.append(rfc.score(x_test,y_test))
plt.xlabel("rfc")
plt.ylabel("Accuracy")
plt.legend()
plt.scatter(k_range,score)
plt.show()
```

```
No artists with labels found to put in legend.  Note that artists whose label s
tart with an underscore are ignored when legend() is called with no argument.
```



Conclusion :-

- This Project is Mall Customer Reviews Data
- In this project done cleansing , analysis and Visualization of the Data
- Apart from this we have done prediction of Segment on the basis of Annual Income and Spending Score Of the Mall Customer Data
- So, we used KNN to understand customer data. KNN is a good clustering algorithm. Almost all the clusters have similar density. It is also fast and efficient in terms of computational cost.