Black Friday

a)Load the 'Black Friday' dataset into one of the data structures (NumPy or Pandas).

b)Display header rows and description of the loaded dataset.

c) Remove unnecessary features (E.g. drop unwanted columns) from the dataset such as 'User_ID', 'Product_ID ' 'Stay_In_Current_City_Years' .

d) Manipulate data by replacing empty column values in 'City_Category' with a default value for the city.

e) Perform the following visualizations on the loaded dataset:

i) Tally of the Number of Male & Female persons who bought 'Product_Category_1' and 'Product_Category_2'.

ii) Total Number of Male & Female persons belonging to each city category

```
In [15]: import seaborn as sns
         import matplotlib.pyplot as plt
         import pandas as pd

         df = pd.read_csv('blackfri.csv')
```

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

Out[16]:

er_ID	Product_ID	Gender	Age	Occupation	City_Category	Stay_In_Current_City_Years	Marital_Status	Baseball_Caps	Wine_Tumblers	Pet_Raincoats	Purchase
0001	P00069042	F	0-17	10	A	2	0	3	NaN	NaN	8370
0001	P00248942	F	0-17	10	A	2	0	1	6.0	14.0	15200
0001	P00087842	F	0-17	10	A	2	0	12	NaN	NaN	1422
0001	P00085442	F	0-17	10	NaN	2	0	12	14.0	5.0	1057
0002	P00285442	M	55+	16	C	4+	0	8	NaN	NaN	7969

```
In [17]: df.describe()
```

Out[17]:

	User_ID	Occupation	Marital_Status	Baseball_Caps	Wine_Tumblers	Pet_Raincoats	Purchase
count	2.100000e+01	21.000000	21.000000	21.000000	15.000000	12.000000	21.000000
mean	1.000004e+06	12.047619	0.523810	4.428571	9.133333	11.416667	9912.809524
std	2.226464e+00	5.444963	0.511766	3.668398	5.111146	5.017394	5785.918774
min	1.000001e+06	1.000000	0.000000	1.000000	2.000000	4.000000	1057.000000
25%	1.000003e+06	9.000000	0.000000	1.000000	5.000000	5.000000	5378.000000
50%	1.000005e+06	10.000000	1.000000	4.000000	8.000000	14.000000	8584.000000
75%	1.000006e+06	16.000000	1.000000	8.000000	14.000000	14.250000	15227.000000
max	1.000008e+06	20.000000	1.000000	12.000000	16.000000	17.000000	19614.000000

In [18]:
```
df.drop(['User_ID','Product_ID','Stay_In_Current_City_Years'], axis=1, inplace=True)
df.head()
```

Out[18]:

	Gender	Age	Occupation	City_Category	Marital_Status	Baseball_Caps	Wine_Tumblers	Pet_Raincoats	Purchase
0	F	0-17	10	A	0	3	NaN	NaN	8370
1	F	0-17	10	A	0	1	6.0	14.0	15200
2	F	0-17	10	A	0	12	NaN	NaN	1422
3	F	0-17	10	NaN	0	12	14.0	5.0	1057
4	M	55+	16	C	0	8	NaN	NaN	7969

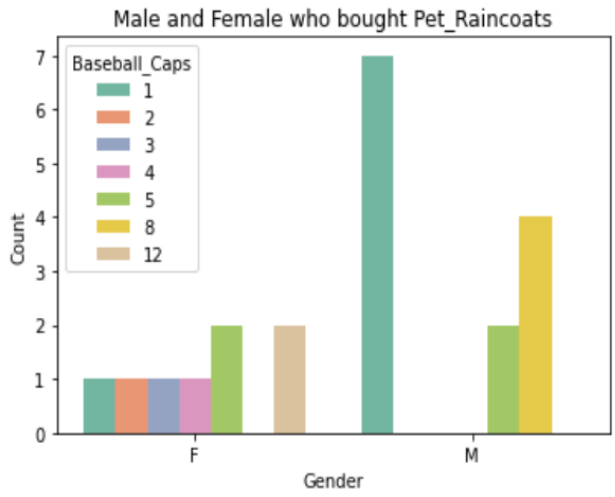
In [19]:
```
print("Filling empty values")
df['City_Category'] = df['City_Category'].fillna("x")
df.head()
```

```
Filling empty values
```

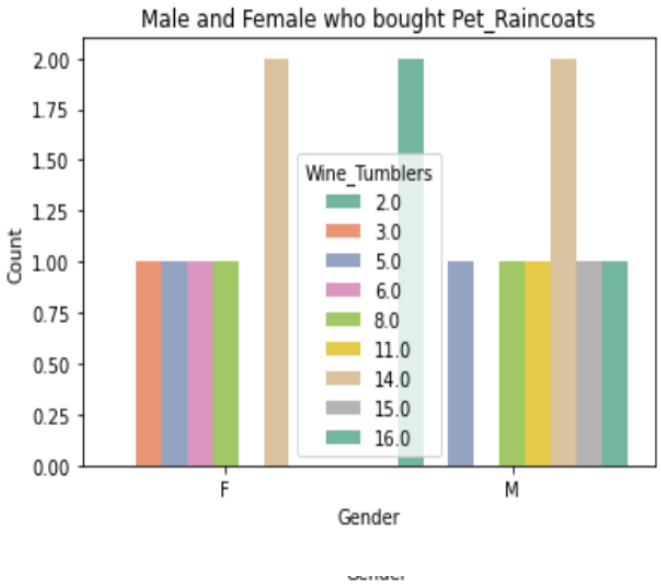
Out[19]:

	Gender	Age	Occupation	City_Category	Marital_Status	Baseball_Caps	Wine_Tumblers	Pet_Raincoats	Purchase
0	F	0-17	10	A	0	3	NaN	NaN	8370
1	F	0-17	10	A	0	1	6.0	14.0	15200
2	F	0-17	10	A	0	12	NaN	NaN	1422
3	F	0-17	10	x	0	12	14.0	5.0	1057
4	M	55+	16	C	0	8	NaN	NaN	7969

In [22]:
```
ax = sns.countplot(data=df,x='Gender',hue='Baseball_Caps',palette='Set2')
ax.set(title='Male and Female who bought Pet_Raincoats',xlabel='Gender',ylabel='Count')
plt.show()
```



In [23]:
```
ax = sns.countplot(data=df,x='Gender',hue='Wine_Tumblers',palette='Set2')
ax.set(title='Male and Female who bought Pet_Raincoats',xlabel='Gender',ylabel='Count')
plt.show()
```



In [24]:
```
ax = sns.countplot(data=df,x='Gender',hue='City_Category',palette='Set1')
ax.set(title='Male and Female belonging to each city',xlabel='Gender',ylabel='Count')
plt.show()
```