

```
In [1]: import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import plotly.express as px
import warnings
warnings.filterwarnings("ignore")
#To import the necessary libraries
```

```
In [2]: data=pd.read_csv("Europe Hotel Booking Satisfaction Score.csv")
#to read the data from the csv file
```

```
In [3]: data.head()
#To display the first five record of the data
```

Out[3]:

	id	Gender	Age	purpose_of_travel	Type of Travel	Type Of Booking	Hotel wifi service	Departure/Arrival convenience	Ease of Online booking	Hotel location	Food and drink	Stay comfort	Common Room entertainment	Checkin
0	70172	Male	13	aviation	Personal Travel	Not defined	3	4	3	1	5	5	5	
1	5047	Male	25	tourism	Group Travel	Group bookings	3	2	3	3	1	1	1	
2	110028	Female	26	tourism	Group Travel	Group bookings	2	2	2	2	5	5	5	
3	24026	Female	25	tourism	Group Travel	Group bookings	2	5	5	5	2	2	2	
4	119299	Male	61	aviation	Group Travel	Group bookings	3	3	3	3	4	5	3	

```
In [4]: data.columns
# columns of the data
```

```
Out[4]: Index(['id', 'Gender', 'Age', 'purpose_of_travel', 'Type of Travel',
              'Type Of Booking', 'Hotel wifi service',
              'Departure/Arrival convenience', 'Ease of Online booking',
              'Hotel location', 'Food and drink', 'Stay comfort',
              'Common Room entertainment', 'Checkin/Checkout service',
              'Other service', 'Cleanliness', 'satisfaction'],
              dtype='object')
```

```
In [5]: data.tail()
# to count the last five enteries of the data
```

```
Out[5]:
```

	Type of ravel	Type Of Booking	Hotel wifi service	Departure/Arrival convenience	Ease of Online booking	Hotel location	Food and drink	Stay comfort	Common Room entertainment	Checkin/Checkout service	Other service	Cleanliness	satisfaction
Group ravel	Individual/Couple		2	1	2	3	2	2	2	2	3	2	neutral o dissatisfie
Group ravel	Group bookings		4	4	4	4	2	5	5	5	5	4	satisfie
Group ravel	Group bookings		1	1	1	3	4	5	4	5	5	4	neutral o dissatisfie
Group ravel	Individual/Couple		1	1	1	5	1	1	1	5	4	1	neutral o dissatisfie
Group ravel	Group bookings		1	3	3	3	1	1	1	4	3	1	neutral o dissatisfie

```
In [6]: data.dtypes  
#datatypes of the data
```

```
Out[6]: id                int64  
Gender                object  
Age                  int64  
purpose_of_travel    object  
Type of Travel       object  
Type Of Booking      object  
Hotel wifi service    int64  
Departure/Arrival convenience int64  
Ease of Online booking int64  
Hotel location        int64  
Food and drink        int64  
Stay comfort          int64  
Common Room entertainment int64  
Checkin/Checkout service int64  
Other service         int64  
Cleanliness           int64  
satisfaction          object  
dtype: object
```

```
In [7]: data.info()
#information about the data
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 103904 entries, 0 to 103903
Data columns (total 17 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   id                                         103904 non-null  int64
1   Gender                                    103904 non-null  object
2   Age                                       103904 non-null  int64
3   purpose_of_travel                        103904 non-null  object
4   Type of Travel                           103904 non-null  object
5   Type Of Booking                          103904 non-null  object
6   Hotel wifi service                       103904 non-null  int64
7   Departure/Arrival convenience            103904 non-null  int64
8   Ease of Online booking                   103904 non-null  int64
9   Hotel location                           103904 non-null  int64
10  Food and drink                           103904 non-null  int64
11  Stay comfort                             103904 non-null  int64
12  Common Room entertainment                103904 non-null  int64
13  Checkin/Checkout service                 103904 non-null  int64
14  Other service                            103904 non-null  int64
15  Cleanliness                             103904 non-null  int64
16  satisfaction                             103904 non-null  object
dtypes: int64(12), object(5)
memory usage: 13.5+ MB
```

```
In [8]: data.describe()
#describe about the data
```

Out[8]:

	id	Age	Hotel wifi service	Departure/Arrival convenience	Ease of Online booking	Hotel location	Food and drink	Stay comfort	Common Room entertainment
count	103904.000000	103904.000000	103904.000000	103904.000000	103904.000000	103904.000000	103904.000000	103904.000000	103904.000000
mean	64924.210502	39.379706	2.729683	3.060296	2.756901	2.976883	3.202129	3.439396	3.358158
std	37463.812252	15.114964	1.327829	1.525075	1.398929	1.277621	1.329533	1.319088	1.332991
min	1.000000	7.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	32533.750000	27.000000	2.000000	2.000000	2.000000	2.000000	2.000000	2.000000	2.000000
50%	64856.500000	40.000000	3.000000	3.000000	3.000000	3.000000	3.000000	4.000000	4.000000
75%	97368.250000	51.000000	4.000000	4.000000	4.000000	4.000000	4.000000	5.000000	4.000000
max	129880.000000	85.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000



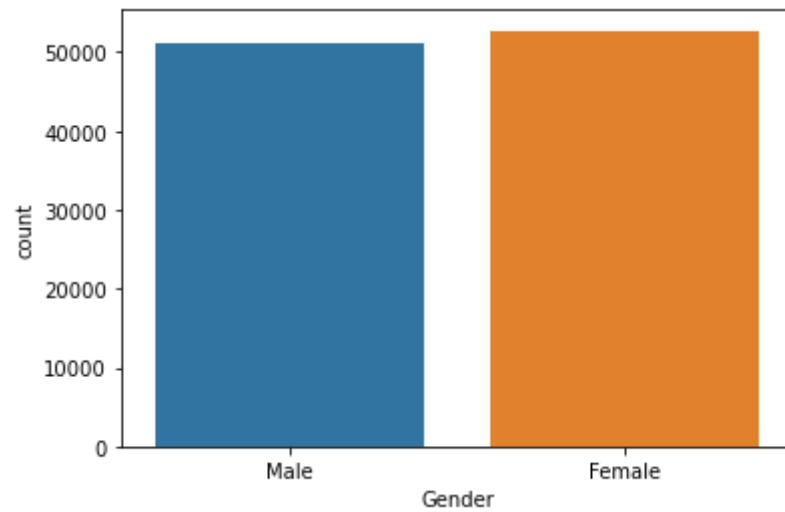
```
In [9]: data.isnull().sum()  
#to count the numbr of null values prsent in each column
```

```
Out[9]: id          0  
Gender          0  
Age            0  
purpose_of_travel 0  
Type of Travel  0  
Type Of Booking 0  
Hotel wifi service 0  
Departure/Arrival convenience 0  
Ease of Online booking 0  
Hotel location  0  
Food and drink  0  
Stay comfort    0  
Common Room entertainment 0  
Checkin/Checkout service 0  
Other service   0  
Cleanliness     0  
satisfaction    0  
dtype: int64
```

```
In [10]: data.duplicated().sum()
```

```
Out[10]: 0
```

```
In [11]: #number of male and female in this data  
sns.countplot(x=data["Gender"],data=data)  
plt.show()
```



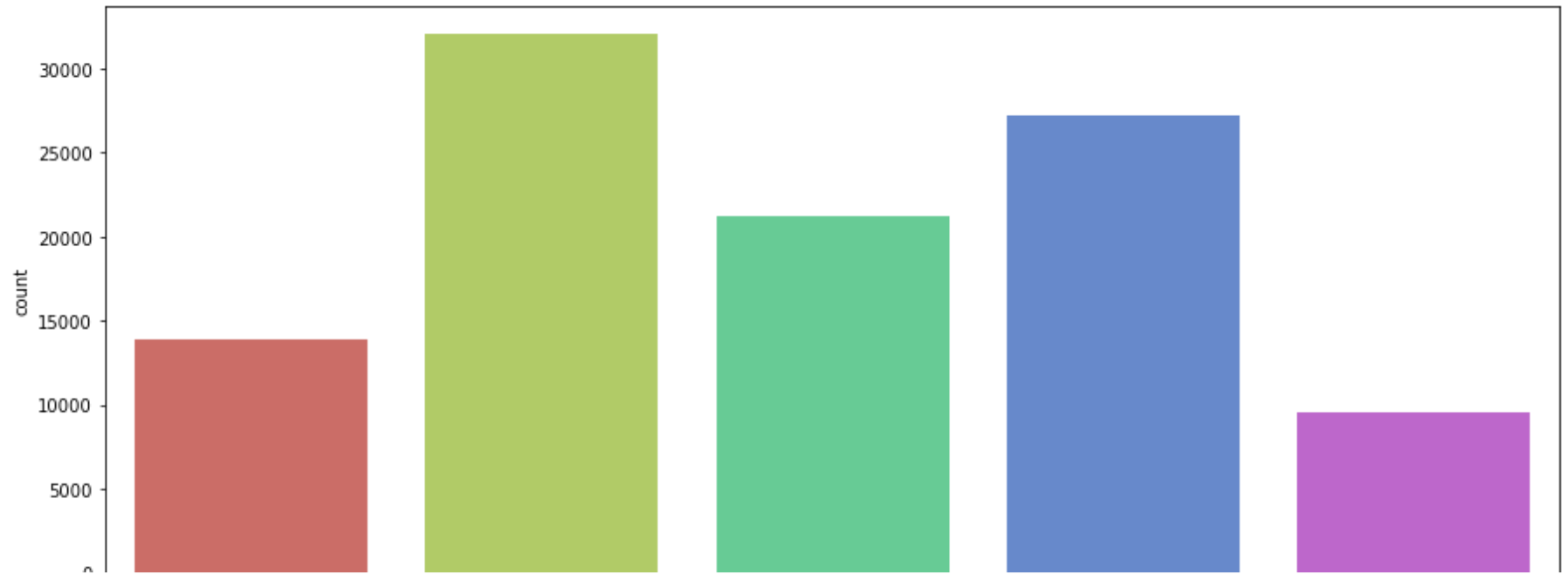
```
In [12]: data.columns
```

```
Out[12]: Index(['id', 'Gender', 'Age', 'purpose_of_travel', 'Type of Travel',  
              'Type Of Booking', 'Hotel wifi service',  
              'Departure/Arrival convenience', 'Ease of Online booking',  
              'Hotel location', 'Food and drink', 'Stay comfort',  
              'Common Room entertainment', 'Checkin/Checkout service',  
              'Other service', 'Cleanliness', 'satisfaction'],  
              dtype='object')
```

```
In [13]: data1=data[['purpose_of_travel', 'Type of Travel',  
                    'Type Of Booking', 'Hotel wifi service',  
                    'Departure/Arrival convenience', 'Ease of Online booking',  
                    'Hotel location', 'Food and drink', 'Stay comfort',  
                    'Common Room entertainment', 'Checkin/Checkout service',  
                    'Other service', 'Cleanliness', 'satisfaction']]
```

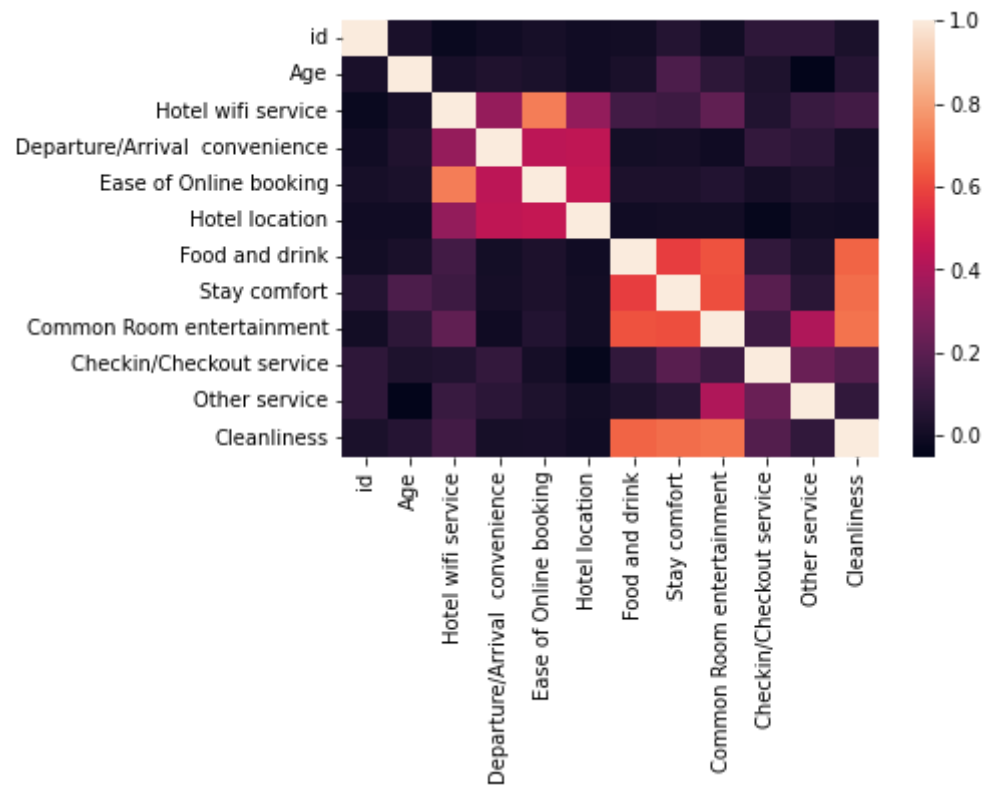


```
In [14]: for i in data1.columns:
plt.figure(figsize=(15,6))
sns.countplot(data1[i],data=data1,palette='hls')
plt.xticks(rotation=90)
plt.show()
#to count the number of categories whih is present in each column
```

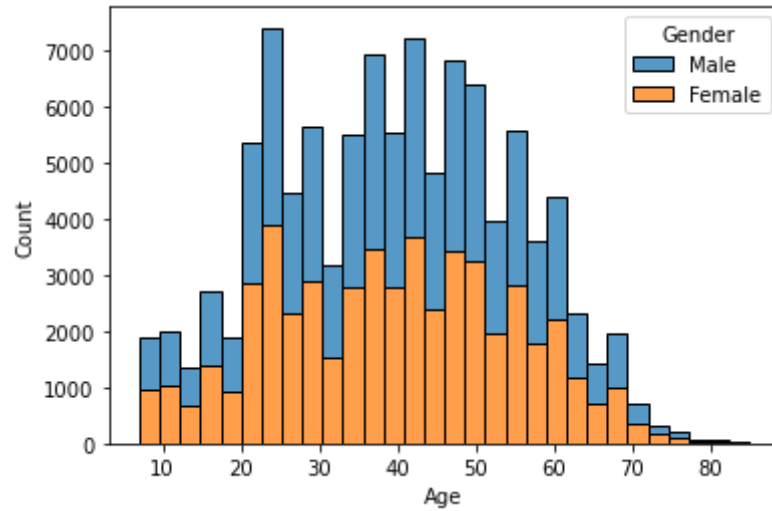


In [15]:

```
sns.heatmap(data.corr())  
plt.show()  
# to display the correlation between the data
```

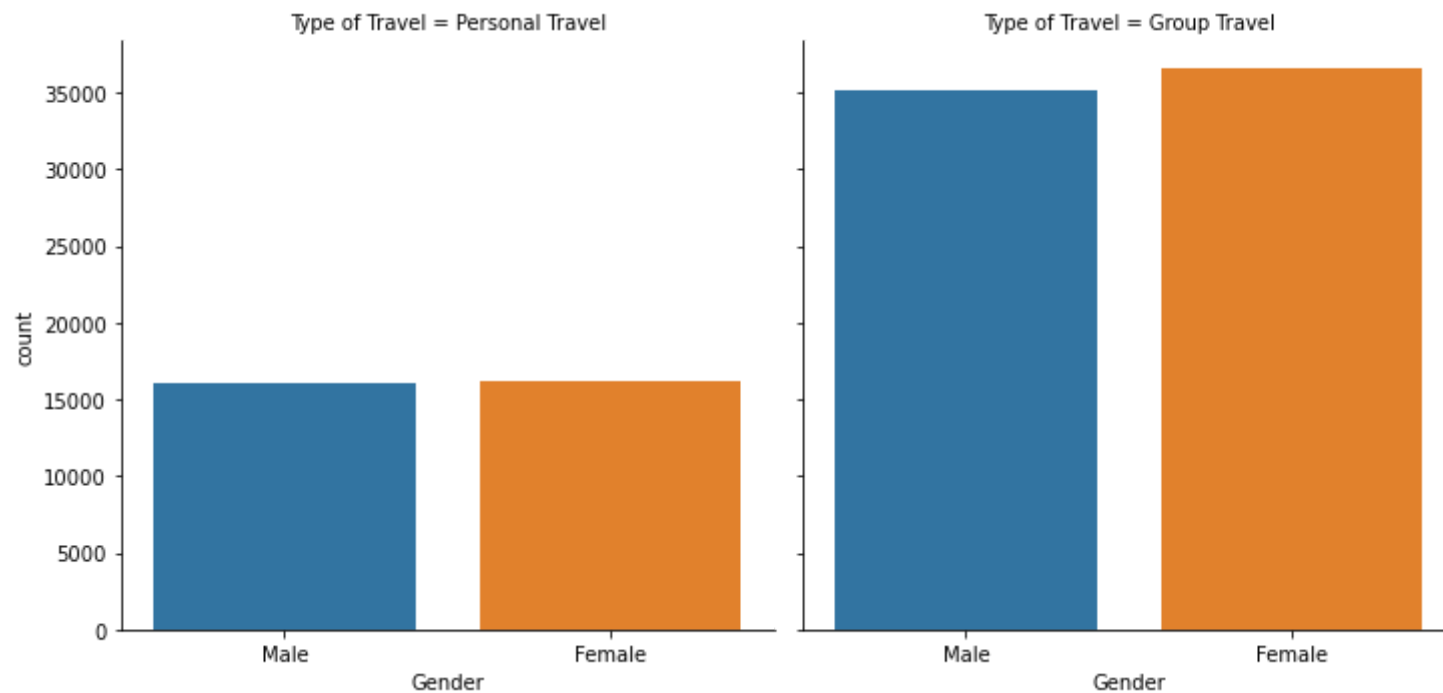


```
In [16]: sns.histplot(data=data, x="Age", bins=30, hue="Gender", multiple="stack")  
plt.show()  
#To show the number of males and females in each category of age
```



```
In [17]: import pandas as pd
import seaborn as sb
from matplotlib import pyplot as plt

sb.factorplot("Gender", col = "Type of Travel", col_wrap = 3, data = data, kind = "count")
plt.show()
#To count the number of males and females in each category of type and travel
```



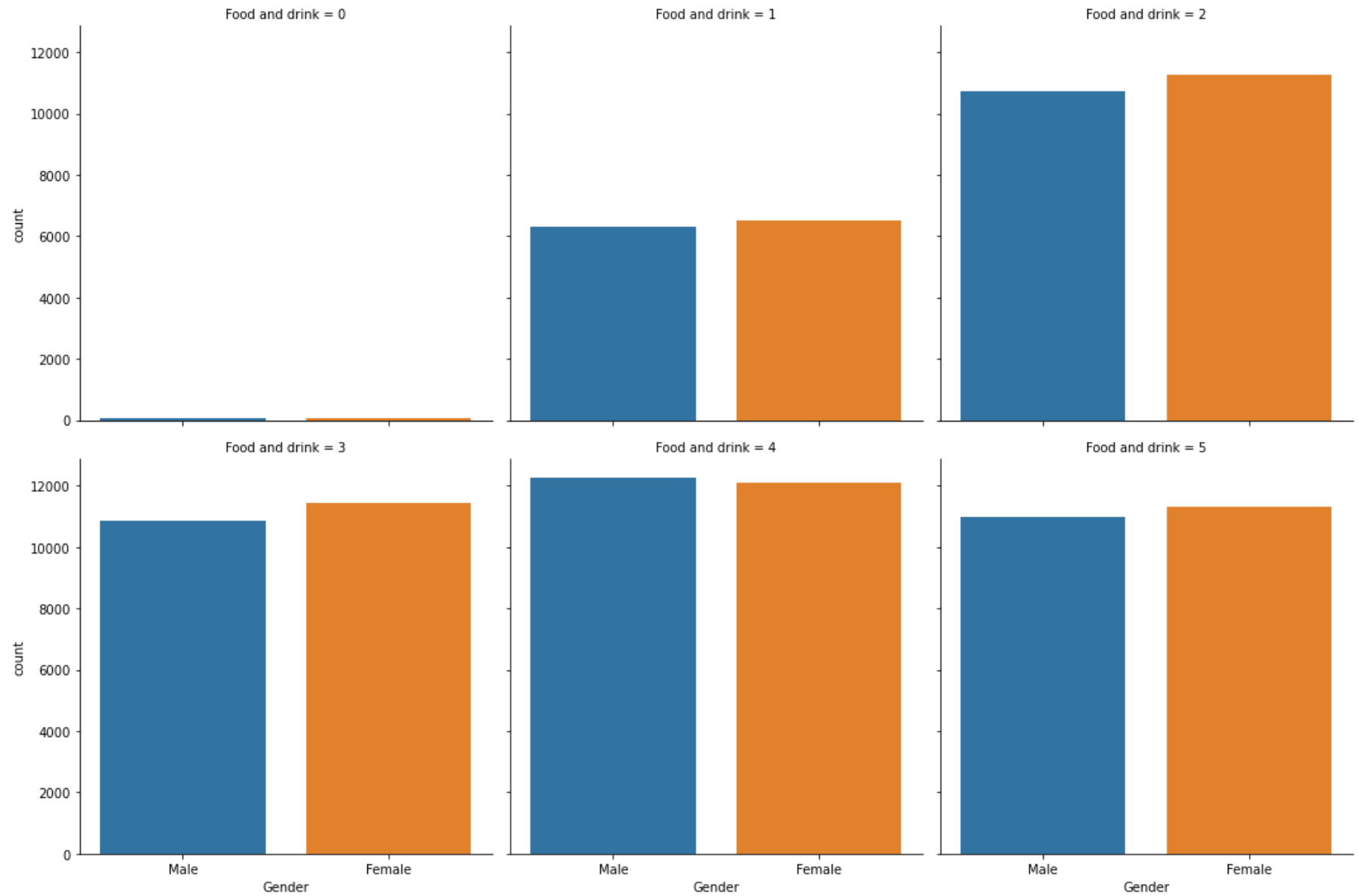
```
In [18]: import pandas as pd
import seaborn as sb
from matplotlib import pyplot as plt

sb.factorplot("Gender", col = "Type Of Booking", col_wrap = 3, data = data, kind = "count")
plt.show()
#To count the number of males and females in each category of type of booking
```



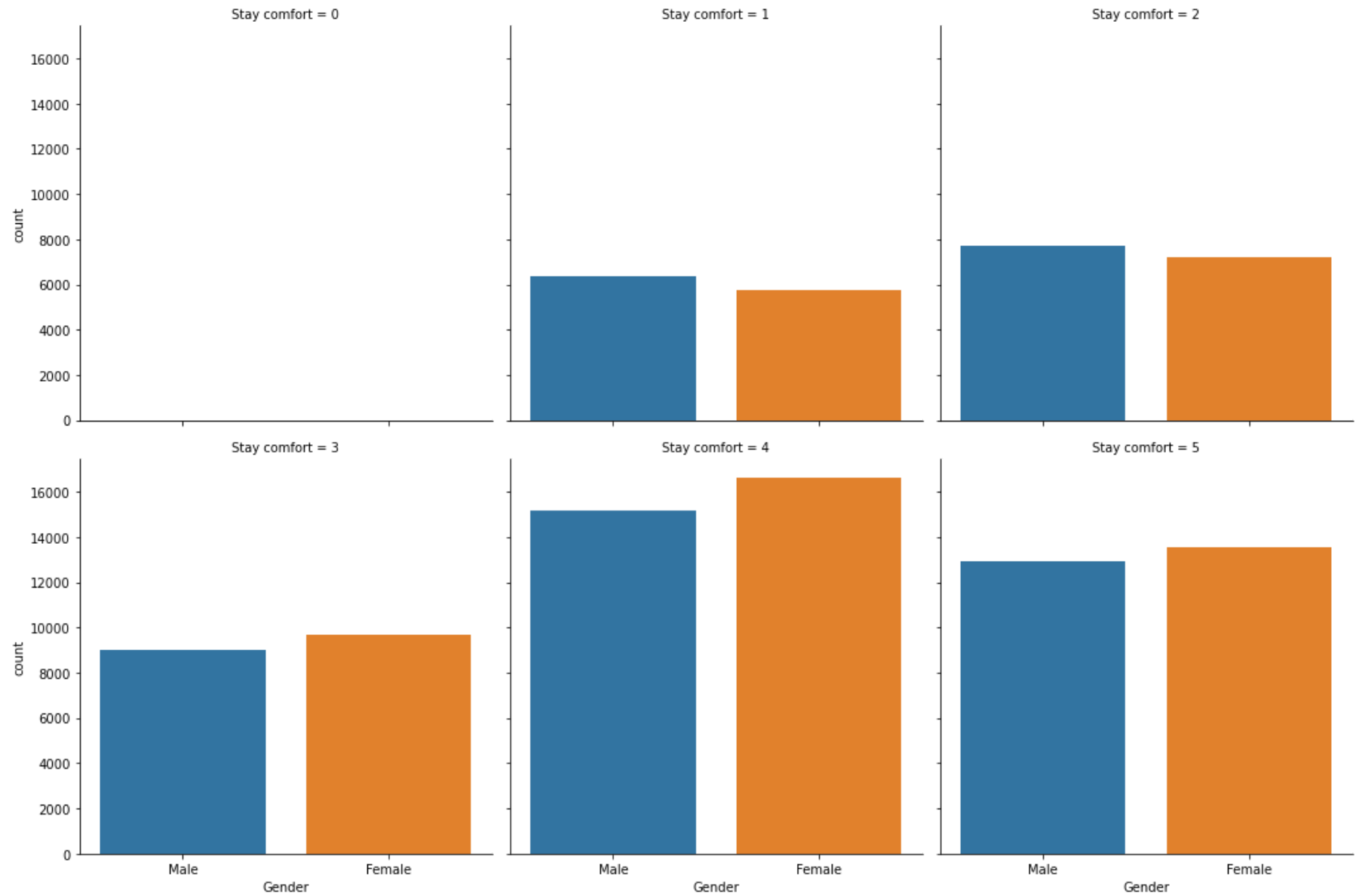
```
In [21]: import pandas as pd
import seaborn as sb
from matplotlib import pyplot as plt

sb.factorplot("Gender", col = "Food and drink", col_wrap = 3,data = data,kind = "count")
plt.show()
#To count the number of males and females in each category of food and drink
```



```
In [23]: import pandas as pd
import seaborn as sb
from matplotlib import pyplot as plt

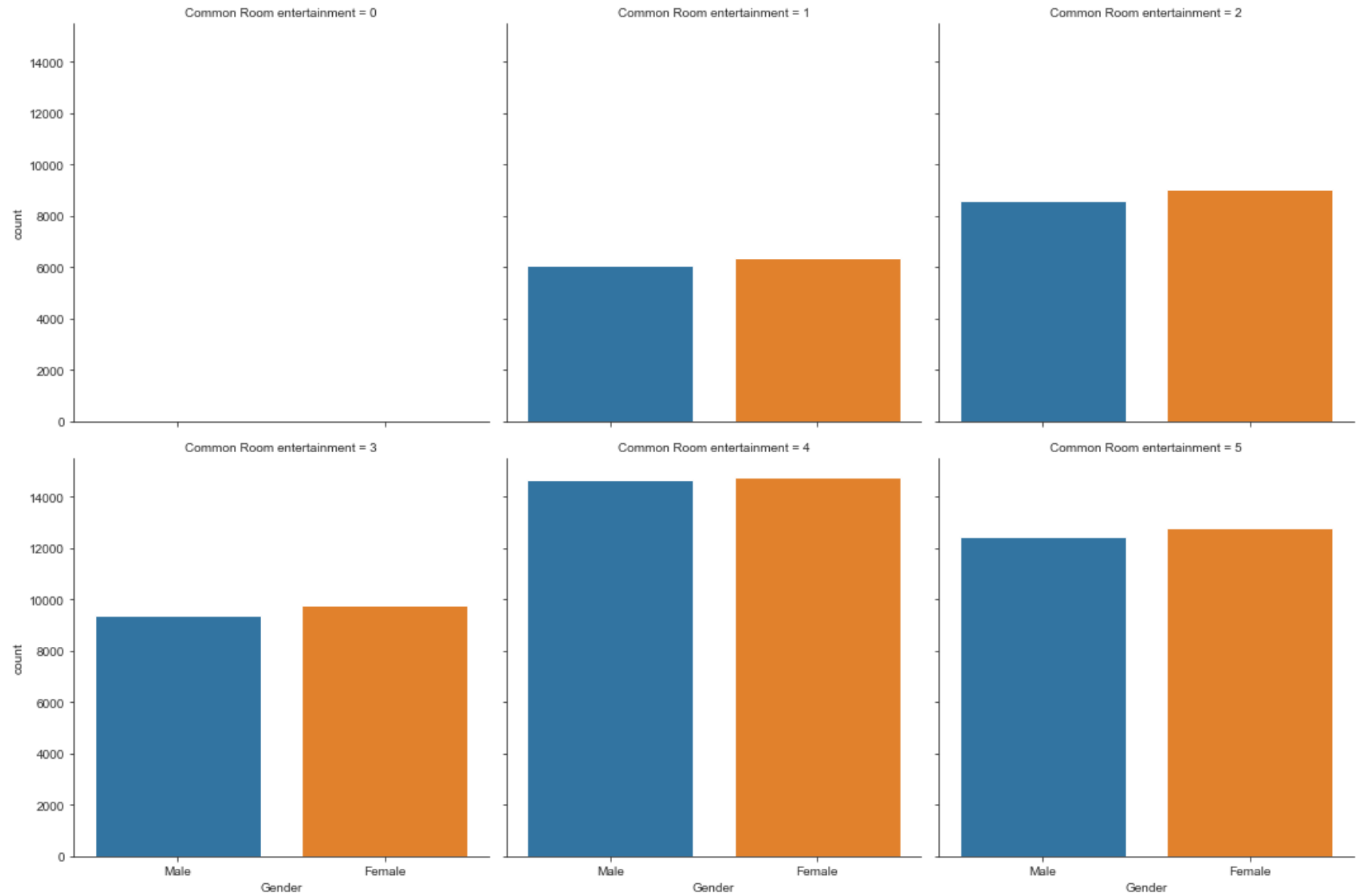
sb.factorplot("Gender", col = "Stay comfort", col_wrap = 3, data = data, kind = "count")
plt.show()
#To count the number of males and females in each category of stay and comfort
```

In [116]:

```
import pandas as pd
import seaborn as sb
from matplotlib import pyplot as plt

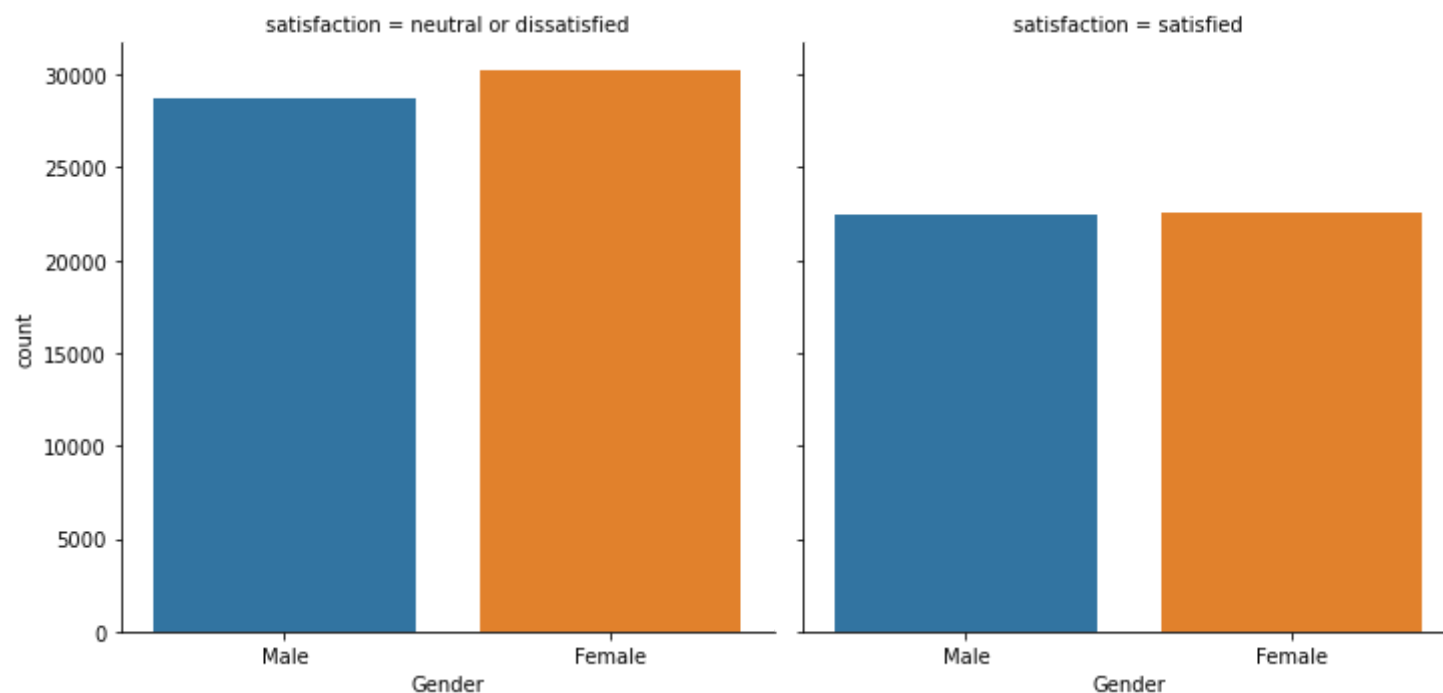
sb.factorplot("Gender", col ="Common Room entertainment", col_wrap = 3,data = data,kind = "count")
plt.show()
#To count the number of males and females in each category of common room and entertainment
```



```
In [25]: data.columns
```

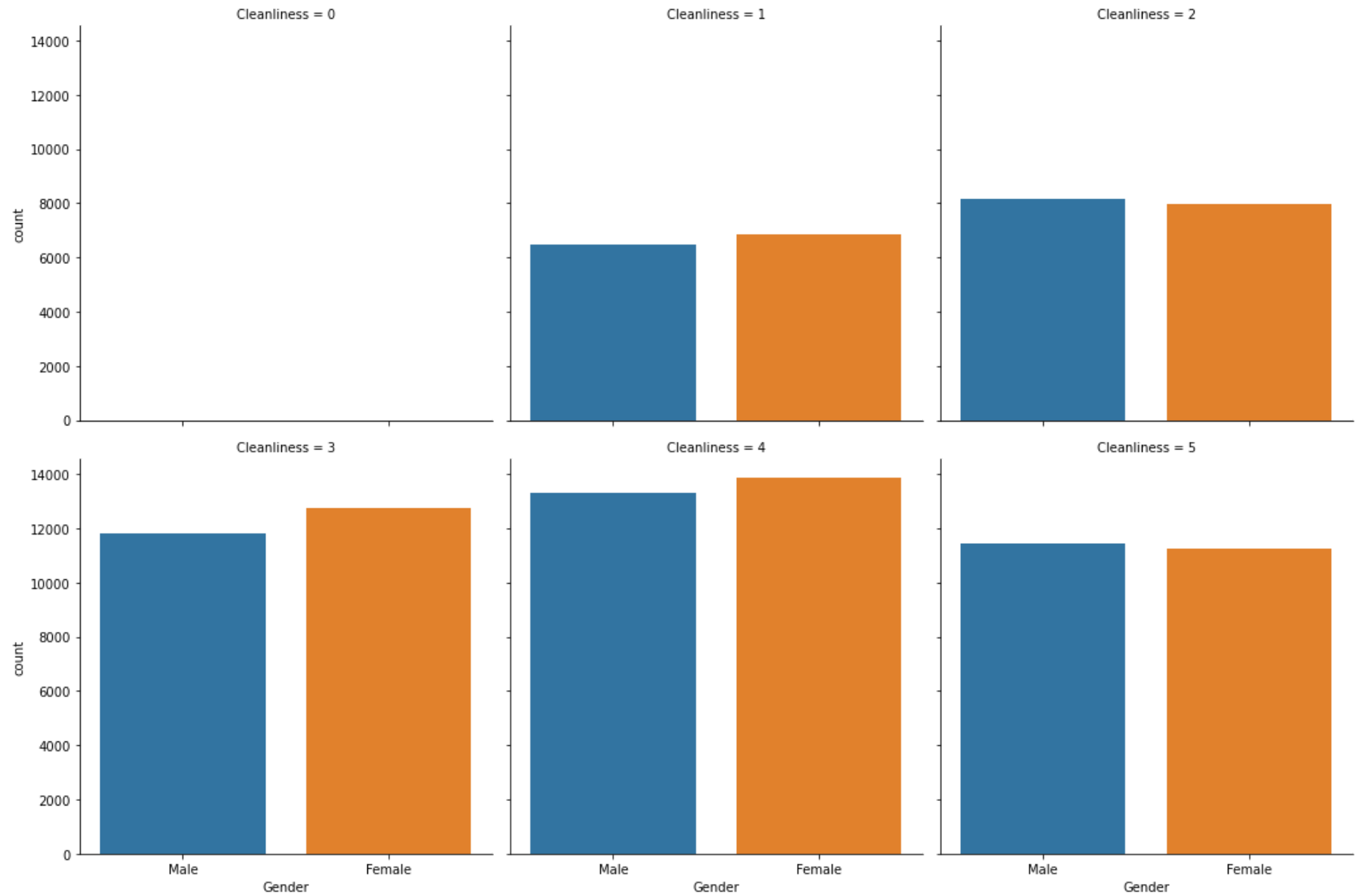
```
Out[25]: Index(['id', 'Gender', 'Age', 'purpose_of_travel', 'Type of Travel',  
              'Type Of Booking', 'Hotel wifi service',  
              'Departure/Arrival convenience', 'Ease of Online booking',  
              'Hotel location', 'Food and drink', 'Stay comfort',  
              'Common Room entertainment', 'Checkin/Checkout service',  
              'Other service', 'Cleanliness', 'satisfaction'],  
              dtype='object')
```

```
In [30]: import pandas as pd  
import seaborn as sb  
from matplotlib import pyplot as plt  
  
sb.factorplot("Gender", col ="satisfaction", col_wrap = 3,data = data,kind = "count")  
plt.show()  
#cleanliness for different category
```



```
In [29]: import pandas as pd
import seaborn as sb
from matplotlib import pyplot as plt

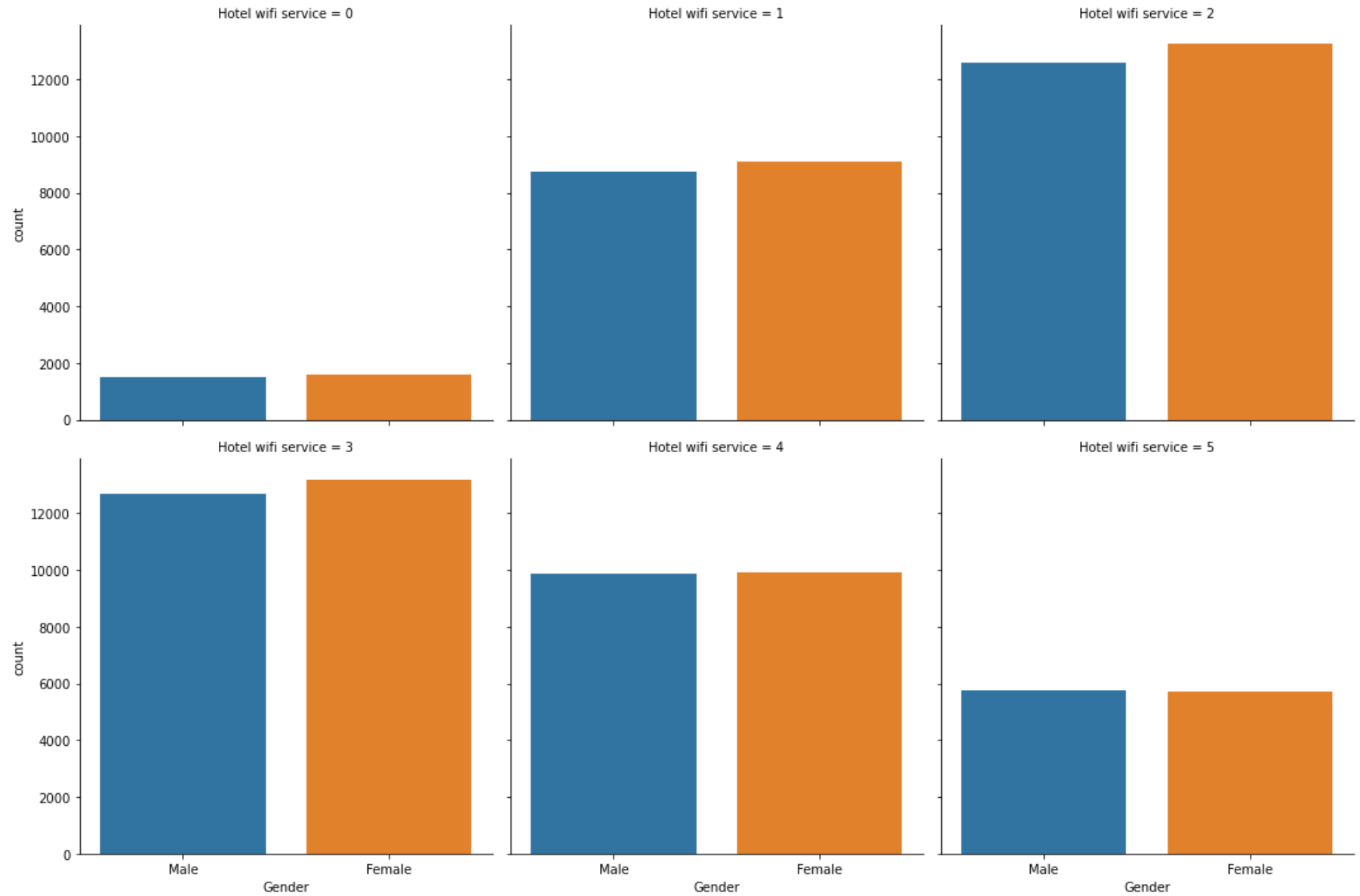
sb.factorplot("Gender", col = "Cleanliness", col_wrap = 3, data = data, kind = "count")
plt.show()
#cleanliness for different category
```



In [28]:

```
import pandas as pd
import seaborn as sb
from matplotlib import pyplot as plt

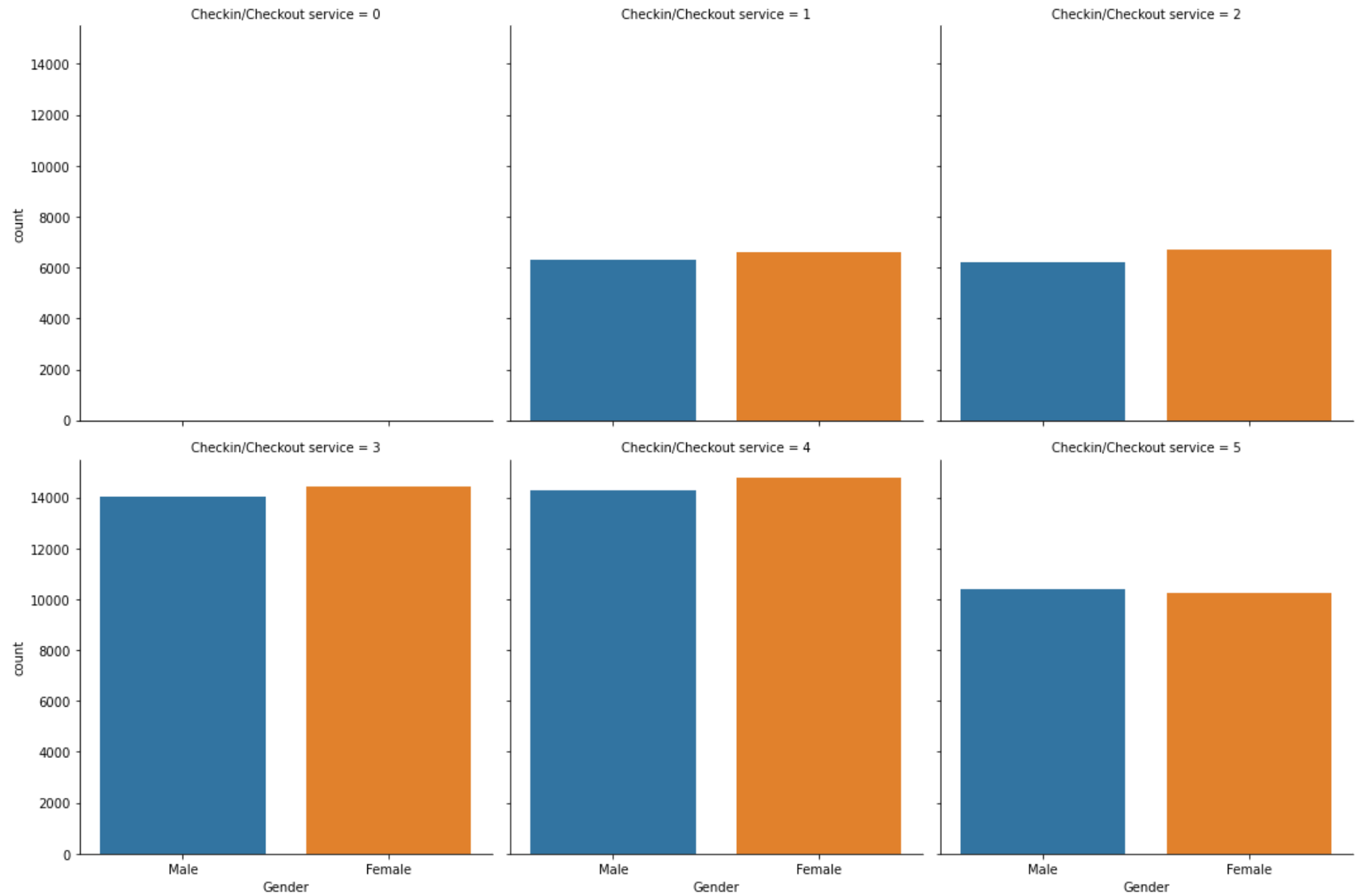
sb.factorplot("Gender", col = "Hotel wifi service", col_wrap = 3, data = data, kind = "count")
plt.show()
#hotelwofo service satisfaction
```



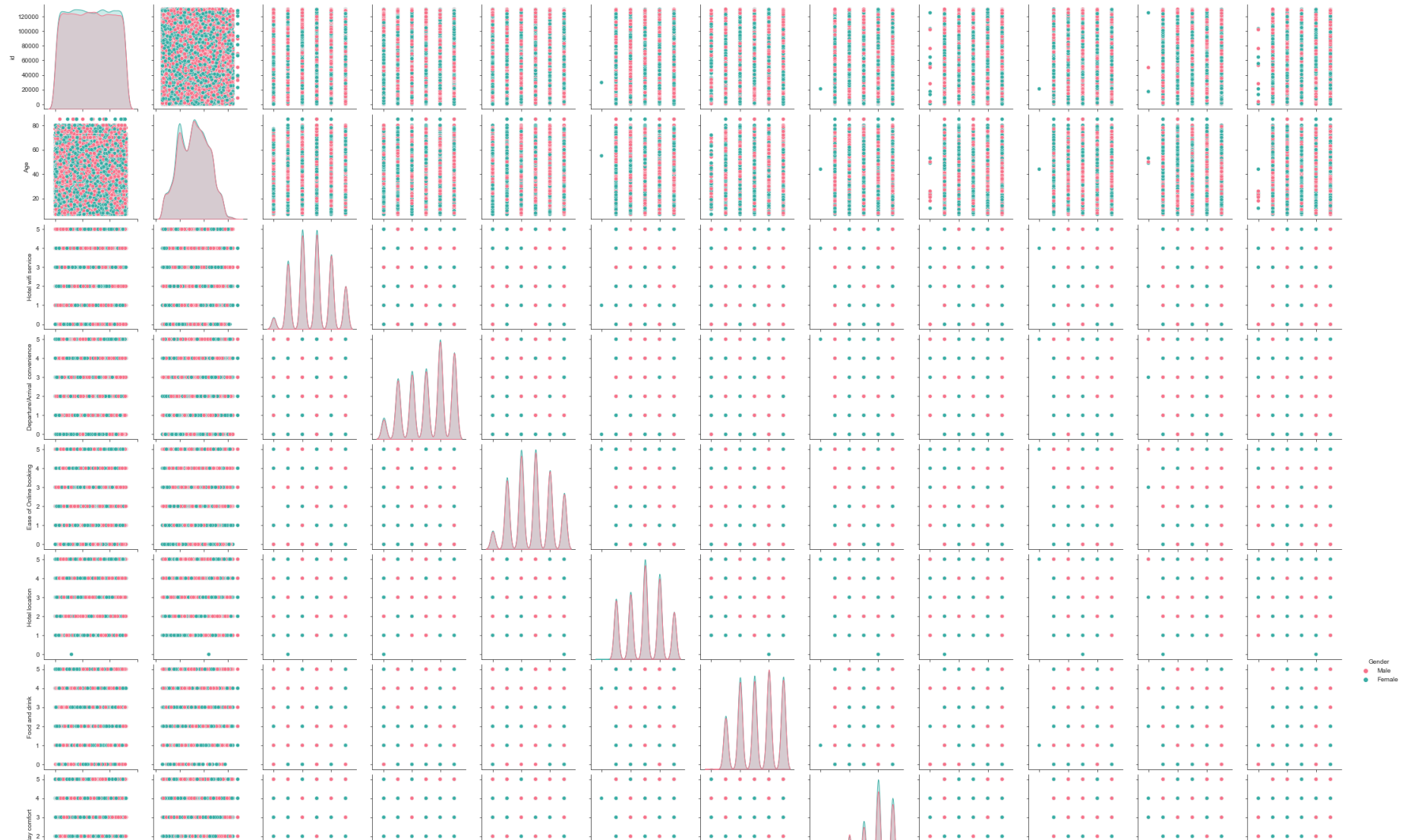
In [26]:

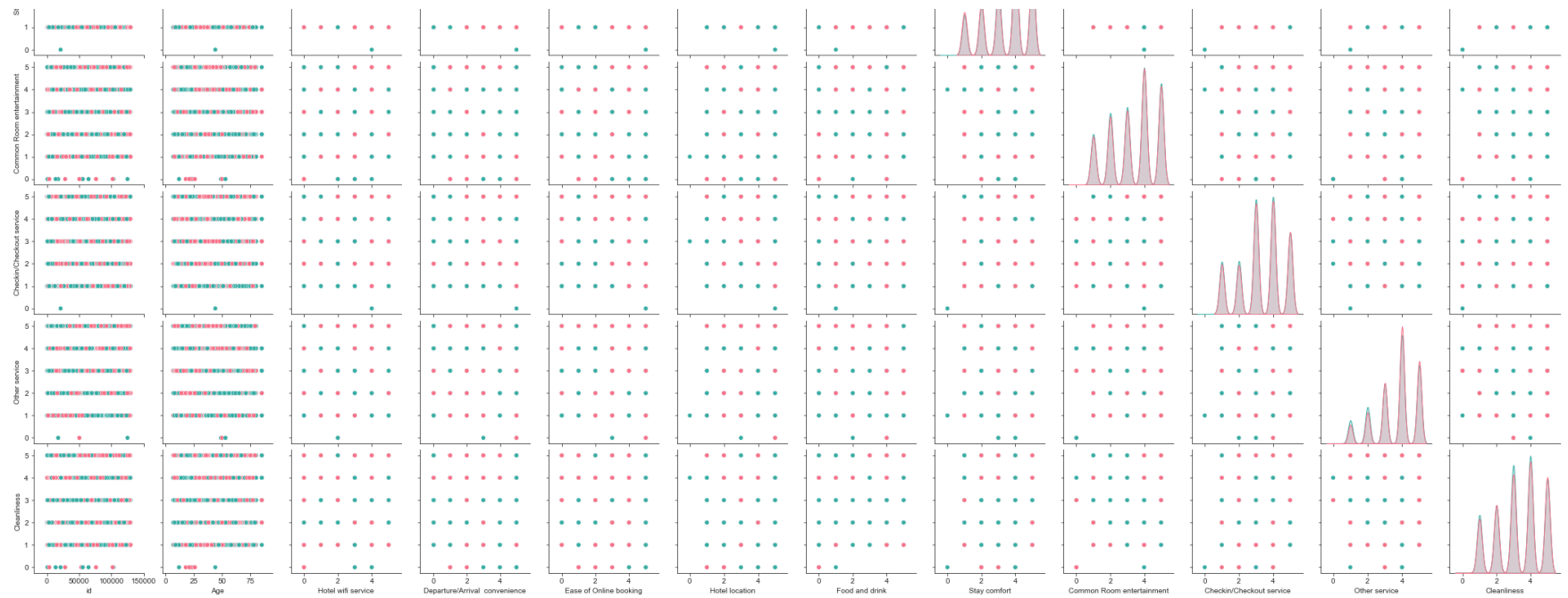
```
import pandas as pd
import seaborn as sb
from matplotlib import pyplot as plt

sb.factorplot("Gender", col = "Checkin/Checkout service", col_wrap = 3, data = data, kind = "count")
plt.show()
#Checkincheck out service
```



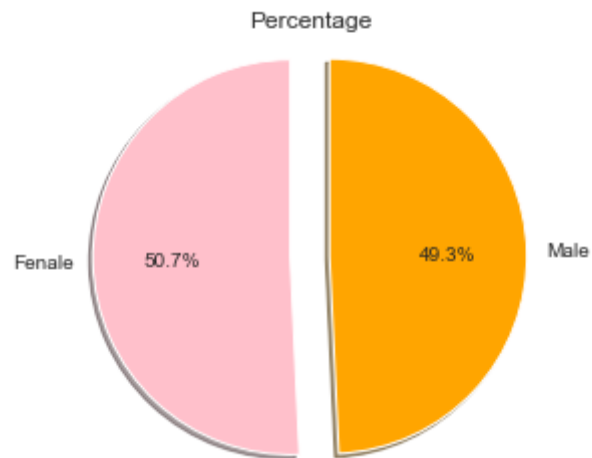
```
In [73]: import pandas as pd
import seaborn as sb
from matplotlib import pyplot
sb.set_style("ticks")
sb.pairplot(data, hue = 'Gender', diag_kind = "kde", kind = "scatter", palette = "husl")
plt.show()
# to find the relationship between many variables
```





```
In [80]: count=data["Gender"].value_counts()
```

```
In [107]: colors = ["pink", "orange"]
explode = (0.1, 0.1)
lables=["Fenale", "Male"]
plt.pie(count, explode=explode, colors=colors,
autopct='%1.1f%%', shadow=True, startangle=90, labels=lables)
plt.title("Percentage")
plt.axis("equal")
plt.show()
#to count the percentage of male and female
```



```
In [97]: def age(Age):
        if Age > 18:
            return "Adult"

        else:
            return "Child"
data['age1']=data.apply(lambda x: age(x['Age']),axis=1)
#define a function
```

In [98]: data

Out[98]:

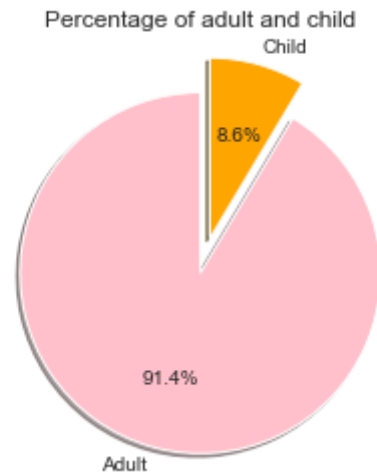
	id	Gender	Age	purpose_of_travel	Type of Travel	Type Of Booking	Hotel wifi service	Departure/Arrival convenience	Ease of Online booking	Hotel location	Food and drink	Stay comfort	Comn Ro entertainm
0	70172	Male	13	aviation	Personal Travel	Not defined	3	4	3	1	5	5	
1	5047	Male	25	tourism	Group Travel	Group bookings	3	2	3	3	1	1	
2	110028	Female	26	tourism	Group Travel	Group bookings	2	2	2	2	5	5	
3	24026	Female	25	tourism	Group Travel	Group bookings	2	5	5	5	2	2	
4	119299	Male	61	aviation	Group Travel	Group bookings	3	3	3	3	4	5	
...	
103899	94171	Female	23	business	Group Travel	Individual/Couple	2	1	2	3	2	2	
103900	73097	Male	49	tourism	Group Travel	Group bookings	4	4	4	4	2	5	
103901	68825	Male	30	tourism	Group Travel	Group bookings	1	1	1	3	4	5	
103902	54173	Female	22	business	Group Travel	Individual/Couple	1	1	1	5	1	1	
103903	62567	Male	27	academic	Group Travel	Group bookings	1	3	3	3	1	1	

103904 rows × 18 columns

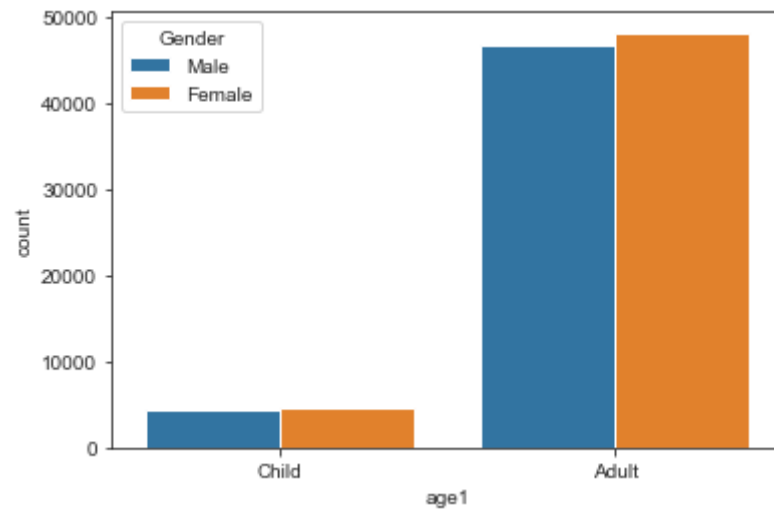


In [110]: count_age1=data['age1'].value_counts()

```
In [112]: colors = ["pink", "orange"]
explode = (0.1, 0.1)
lables=["Adult","Child"]
plt.pie(count_age1, explode=explode, colors=colors,
autopct='%1.1f%%', shadow=True, startangle=90,labels=lables)
plt.title("Percentage of adult and child")
plt.axis("equal")
plt.show()
# to count percentage of adult and child
```



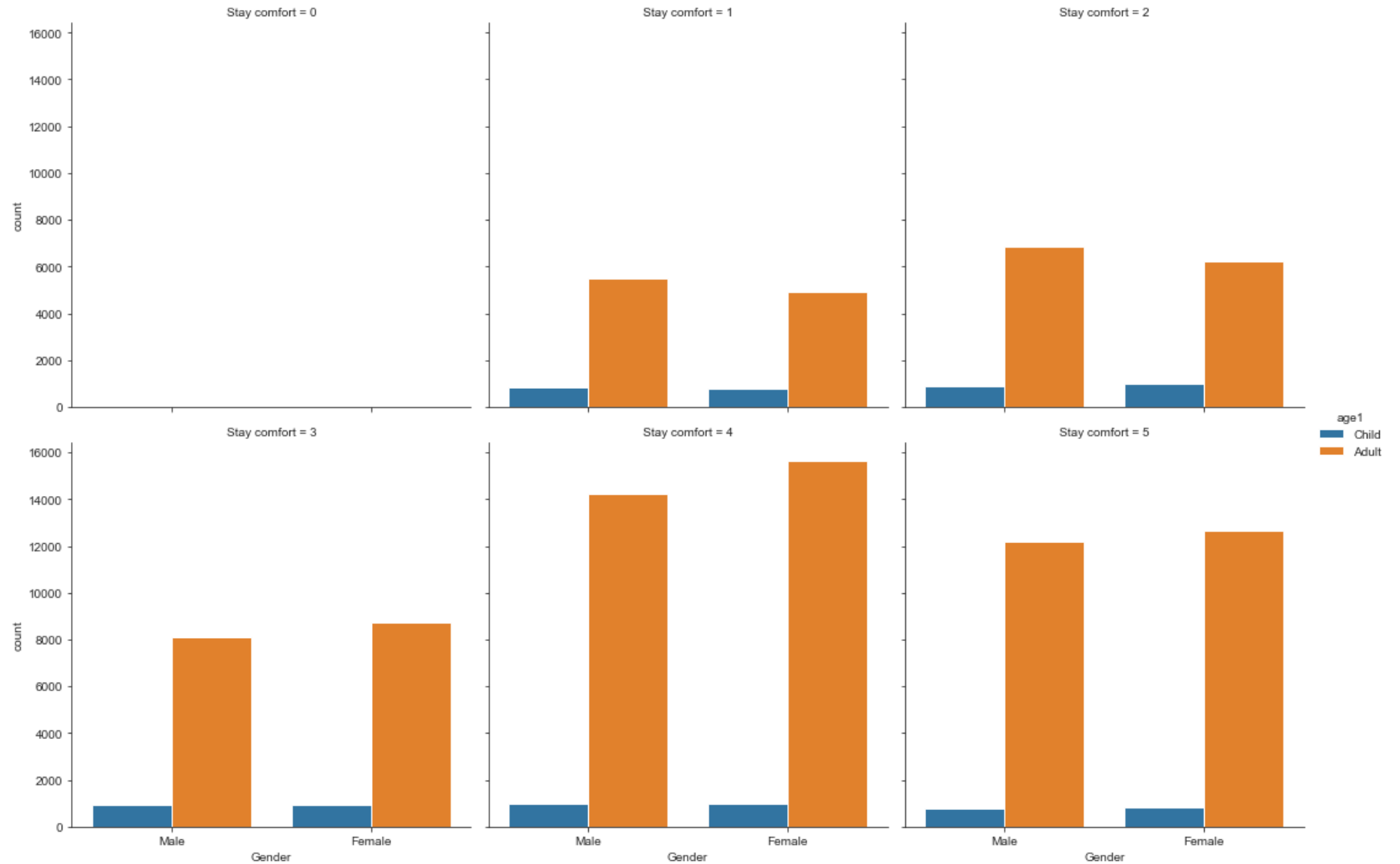
```
In [114]: sns.countplot(x=data["age1"],hue=data["Gender"])  
plt.show()  
# to count number of male and female in child and adult
```




```
In [118]: import pandas as pd
import seaborn as sb
from matplotlib import pyplot as plt

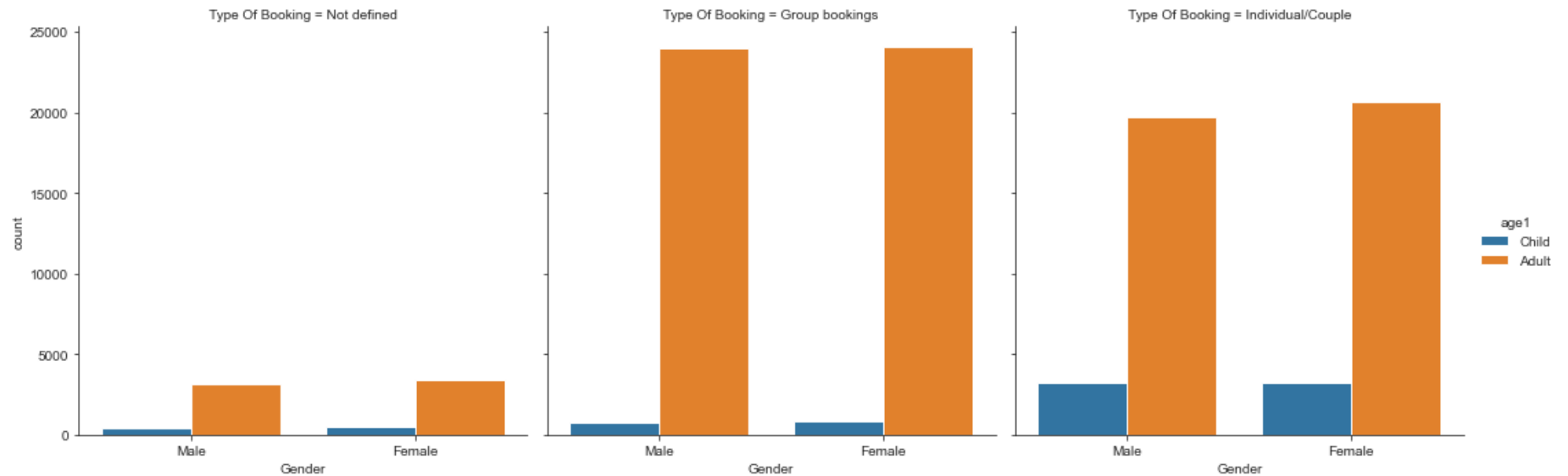
sb.factorplot("Gender", col = "Stay comfort", col_wrap = 3,data = data,kind = "count",hue="age1")
plt.show()

#To count the number of males and females which are adult or child in each category of stay comfort
```



```
In [120]: import pandas as pd
import seaborn as sb
from matplotlib import pyplot as plt

sb.factorplot("Gender", col = "Type Of Booking", col_wrap = 3, data = data, kind = "count", hue="age1")
plt.show()
#To count the number of males and females which are adult or child in each category of type of booking
```



```
In [121]: average_rating=['Hotel wifi service',
    'Departure/Arrival convenience', 'Ease of Online booking',
    'Hotel location', 'Food and drink', 'Stay comfort',
    'Common Room entertainment', 'Checkin/Checkout service',
    'Other service', 'Cleanliness', 'satisfaction']
```

```
In [127]: average=data[average_rating].mean().round(2)
average
```

```
Out[127]: Hotel wifi service          2.73
Departure/Arrival convenience  3.06
Ease of Online booking        2.76
Hotel location                 2.98
Food and drink                 3.20
Stay comfort                   3.44
Common Room entertainment     3.36
Checkin/Checkout service      3.30
Other service                  3.64
Cleanliness                    3.29
dtype: float64
```

```
In [130]: average.plot(kind='barh',color='#A020F0',figsize=(8,10))
```

```
Out[130]: <AxesSubplot:>
```

