

Ramesh Kumar Pr ajapati



E-commerce Data Analysis



Insights & Trends in Online Shopping

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

```
In [2]: df=pd.read_csv("Downloads/customer_details.csv")  
df
```

Out[2]:

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size	Color
0	1	55	Male	Blouse	Clothing	53.0	Kentucky	L	Grey
1	2	19	Male	Sweater	Clothing	64.0	Maine	L	Maroon
2	3	50	Male	Jeans	Clothing	NaN	Massachusetts	S	Maroon
3	4	21	Male	Sandals	Footwear	90.0	Rhode Island	M	Maroon
4	5	45	Male	Blouse	Clothing	49.0	Oregon	M	Turquoise
5	6	46	Male	Sneakers	Footwear	20.0	Wyoming	M	White
6	7	63	Male	Shirt	Clothing	85.0	Montana	M	Grey
7	8	27	Male	Shorts	Clothing	34.0	Louisiana	L	Charcoal
8	9	26	Male	Coat	Outerwear	97.0	West Virginia	L	Silver
9	10	57	Male	Handbag	Accessories	31.0	Missouri	M	Pink



```
In [3]: df.head()
```

Out[3]:

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size	Color
0	1	55	Male	Blouse	Clothing	53.0	Kentucky	L	Gray
1	2	19	Male	Sweater	Clothing	64.0	Maine	L	Maroon
2	3	50	Male	Jeans	Clothing	NaN	Massachusetts	S	Maroon
3	4	21	Male	Sandals	Footwear	90.0	Rhode Island	M	Maroon
4	5	45	Male	Blouse	Clothing	49.0	Oregon	M	Turquoise



In [4]: df.shape

Out[4]: (10, 18)

In [5]: df.info()

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10 entries, 0 to 9
Data columns (total 18 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Customer ID      10 non-null    int64  
 1   Age              10 non-null    int64  
 2   Gender            10 non-null    object  
 3   Item Purchased   10 non-null    object  
 4   Category          10 non-null    object  
 5   Purchase Amount (USD) 9 non-null    float64
 6   Location          10 non-null    object  
 7   Size              10 non-null    object  
 8   Color              10 non-null    object  
 9   Season             10 non-null    object  
 10  Review Rating    10 non-null    float64
 11  Subscription Status 10 non-null    object  
 12  Shipping Type     10 non-null    object  
 13  Discount Applied  10 non-null    object  
 14  Promo Code Used   10 non-null    object  
 15  Previous Purchases 10 non-null    int64  
 16  Payment Method    10 non-null    object  
 17  Frequency of Purchases 10 non-null    object  
dtypes: float64(2), int64(3), object(13)
memory usage: 1.5+ KB

```

In [6]: df.tail(1)

Out[6]:

Customer ID	Age	Gender	Item Purchased	Category	Purchase					
					Amount (USD)	Location	Size	Color	Season	
9	10	57	Male	Handbag	Accessories	31.0	Missouri	M	Pink	Spring



In [7]:

```
#df.dropna()
df.iloc[3]
```

Out[7]:

```
Customer ID           4
Age                  21
Gender               Male
Item Purchased       Sandals
Category             Footwear
Purchase Amount (USD) 90.0
Location             Rhode Island
Size                 M
Color                Maroon
Season               Spring
Review Rating        3.5
Subscription Status  Yes
Shipping Type        Next Day Air
Discount Applied     Yes
Promo Code Used      Yes
Previous Purchases   49
Payment Method       PayPal
Frequency of Purchases Weekly
Name: 3, dtype: object
```

In []:

In [8]:

```
df.describe()
```

Out[8]:

	Customer ID	Age	Purchase Amount (USD)	Review Rating	Previous Purchases
count	10.00000	10.000000	9.000000	10.000000	10.00000
mean	5.50000	40.900000	58.111111	3.220000	21.30000
std	3.02765	16.203223	27.759883	0.612463	16.97089
min	1.00000	19.000000	20.000000	2.600000	2.00000
25%	3.25000	26.250000	34.000000	2.950000	9.50000
50%	5.50000	45.500000	53.000000	3.100000	16.50000
75%	7.75000	53.750000	85.000000	3.200000	29.00000
max	10.00000	63.000000	97.000000	4.800000	49.00000

In [9]:

```
df.columns
```

```
Out[9]: Index(['Customer ID', 'Age', 'Gender', 'Item Purchased', 'Category',  
   'Purchase Amount (USD)', 'Location', 'Size', 'Color', 'Season',  
   'Review Rating', 'Subscription Status', 'Shipping Type',  
   'Discount Applied', 'Promo Code Used', 'Previous Purchases',  
   'Payment Method', 'Frequency of Purchases'],  
  dtype='object')
```

Checking null values

```
In [10]: df.isnull().sum()
```

```
Out[10]: Customer ID      0  
Age            0  
Gender         0  
Item Purchased 0  
Category        0  
Purchase Amount (USD) 1  
Location        0  
Size            0  
Color            0  
Season           0  
Review Rating    0  
Subscription Status 0  
Shipping Type    0  
Discount Applied 0  
Promo Code Used 0  
Previous Purchases 0  
Payment Method    0  
Frequency of Purchases 0  
dtype: int64
```

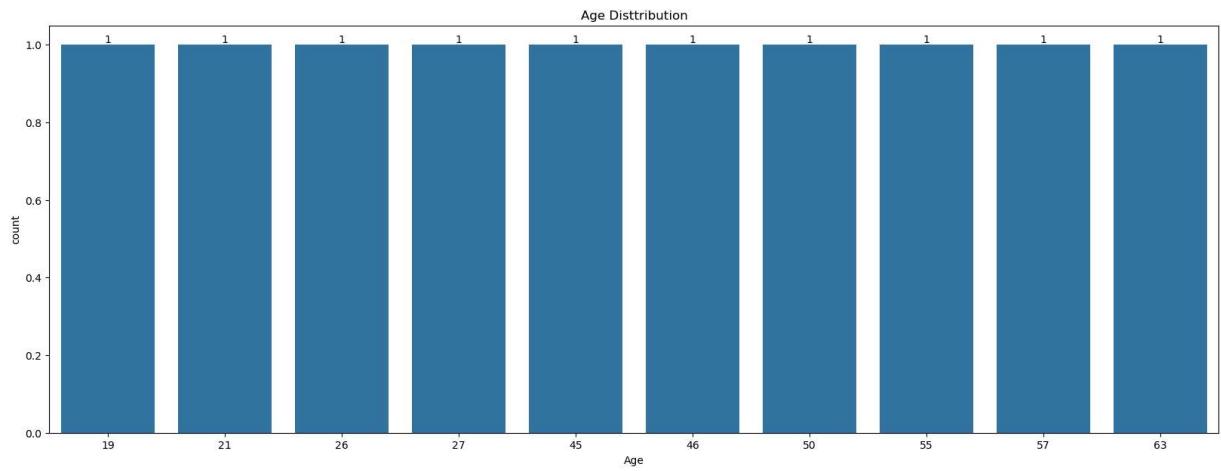
```
In [11]: df['Purchase Amount (USD)'] = df['Purchase Amount (USD)'].fillna(0)
```

```
In [ ]:
```

No any null value.

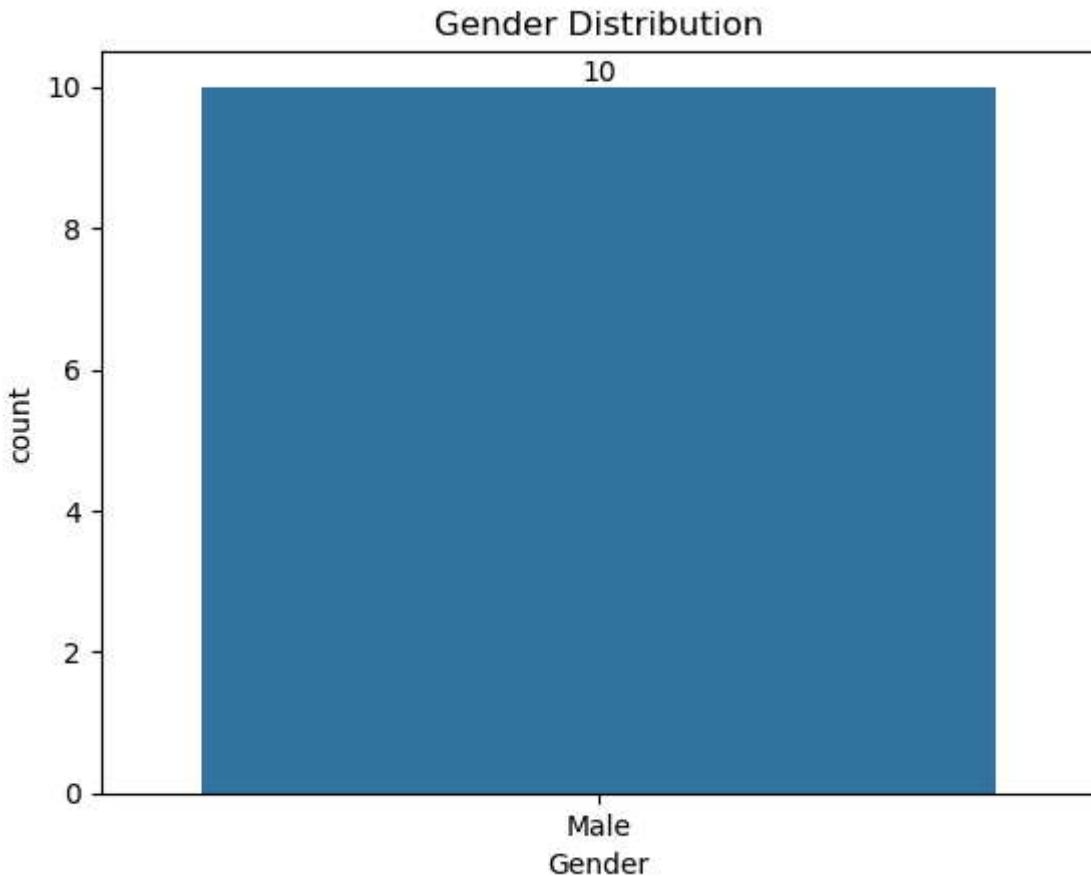
```
In [12]: ## EDA
```

```
In [13]: plt.figure(figsize=(20,7))  
ax= sns.countplot(x='Age',data=df)  
ax.bar_label(ax.containers[0])  
plt.title('Age Distribution')  
plt.show()
```



Most of the peoples are of 69 age

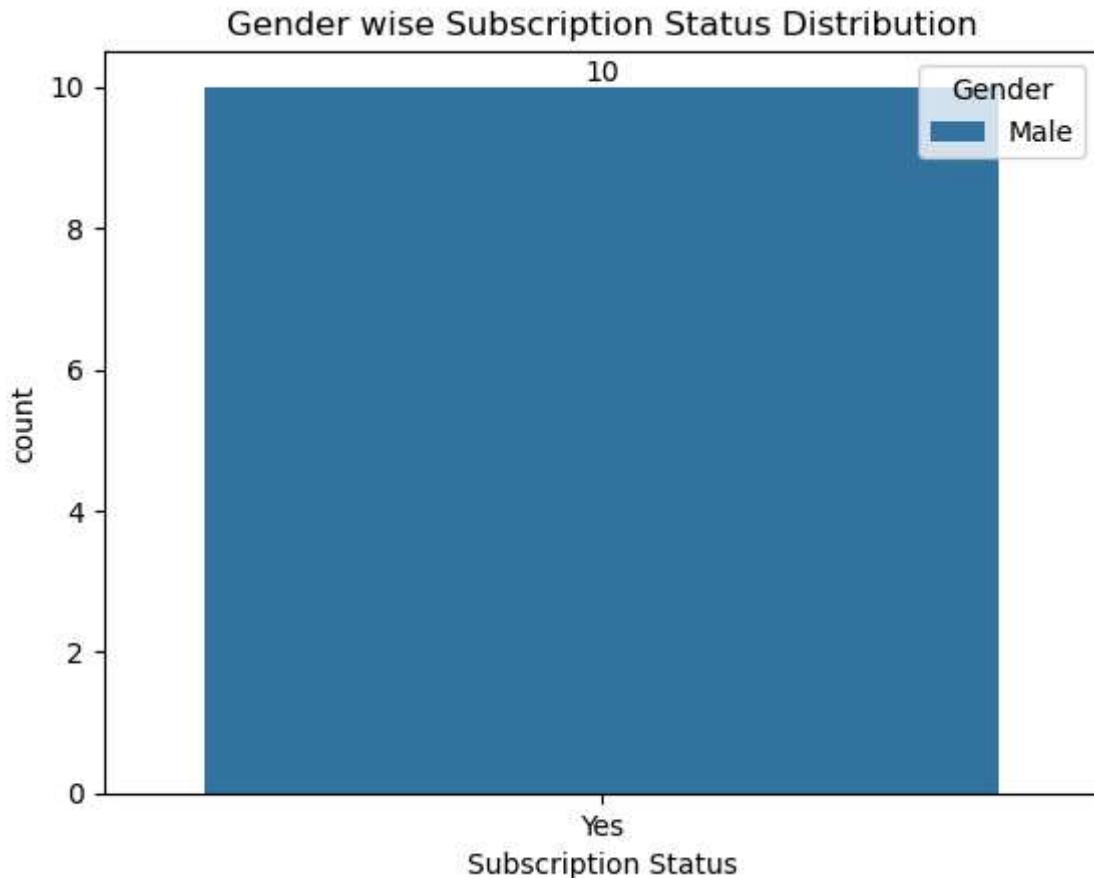
```
In [14]: ax = sns.countplot(x='Gender', data=df)
ax.bar_label(ax.containers[0])
plt.title('Gender Distribution')
plt.show()
```



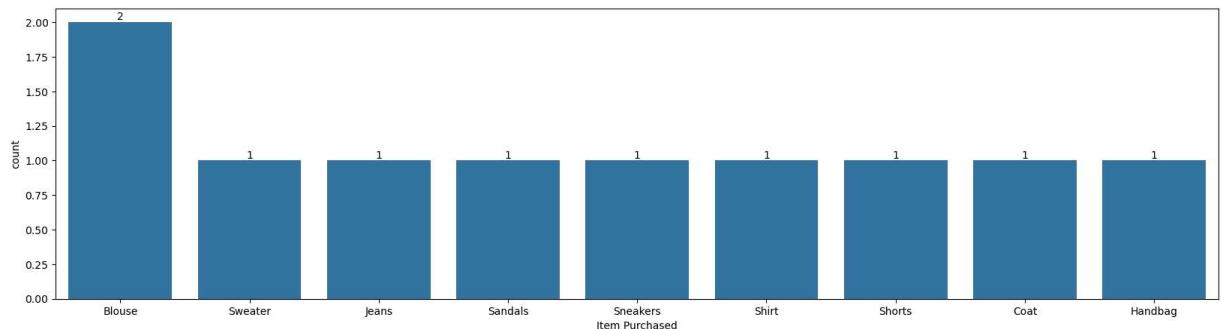
The number of males are more than the females

```
In [15]: #2
ax = sns.countplot(hue='Gender', x='Subscription Status', data=df)
ax.bar_label(ax.containers[0])
```

```
plt.title('Gender wise Subscription Status Distribution')
plt.show()
```

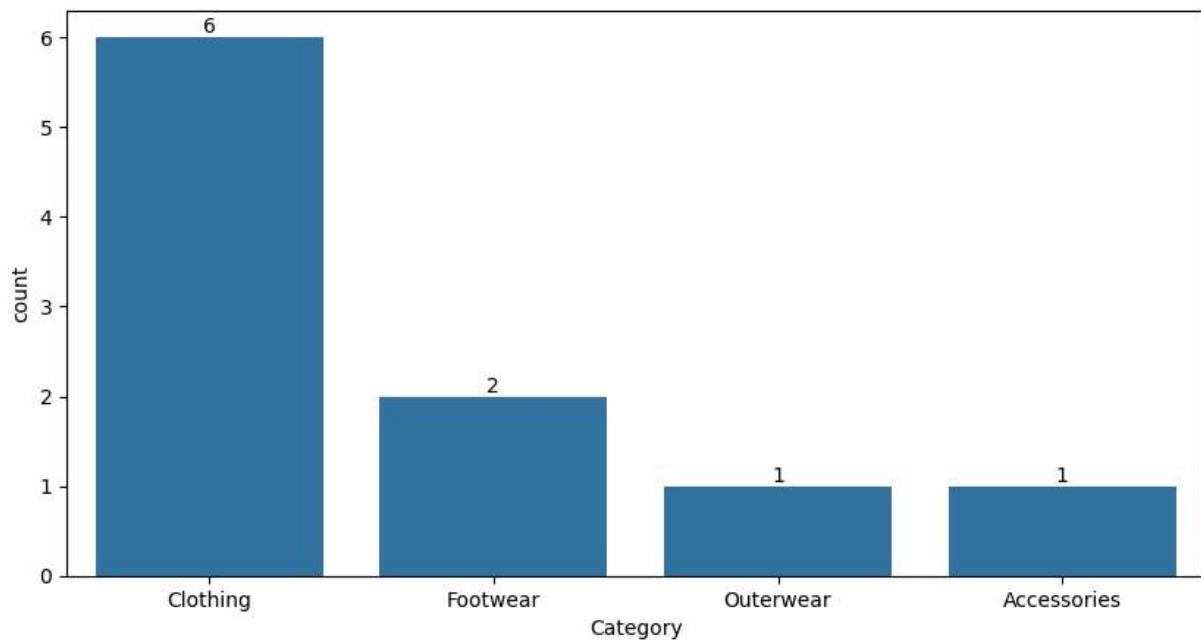


```
In [16]: plt.figure(figsize=(20,5))
ax = sns.countplot(x='Item Purchased', data=df)
ax.bar_label(ax.containers[0])
plt.show()
```



Most sales are of Blouse, Pants, and Jewelry.

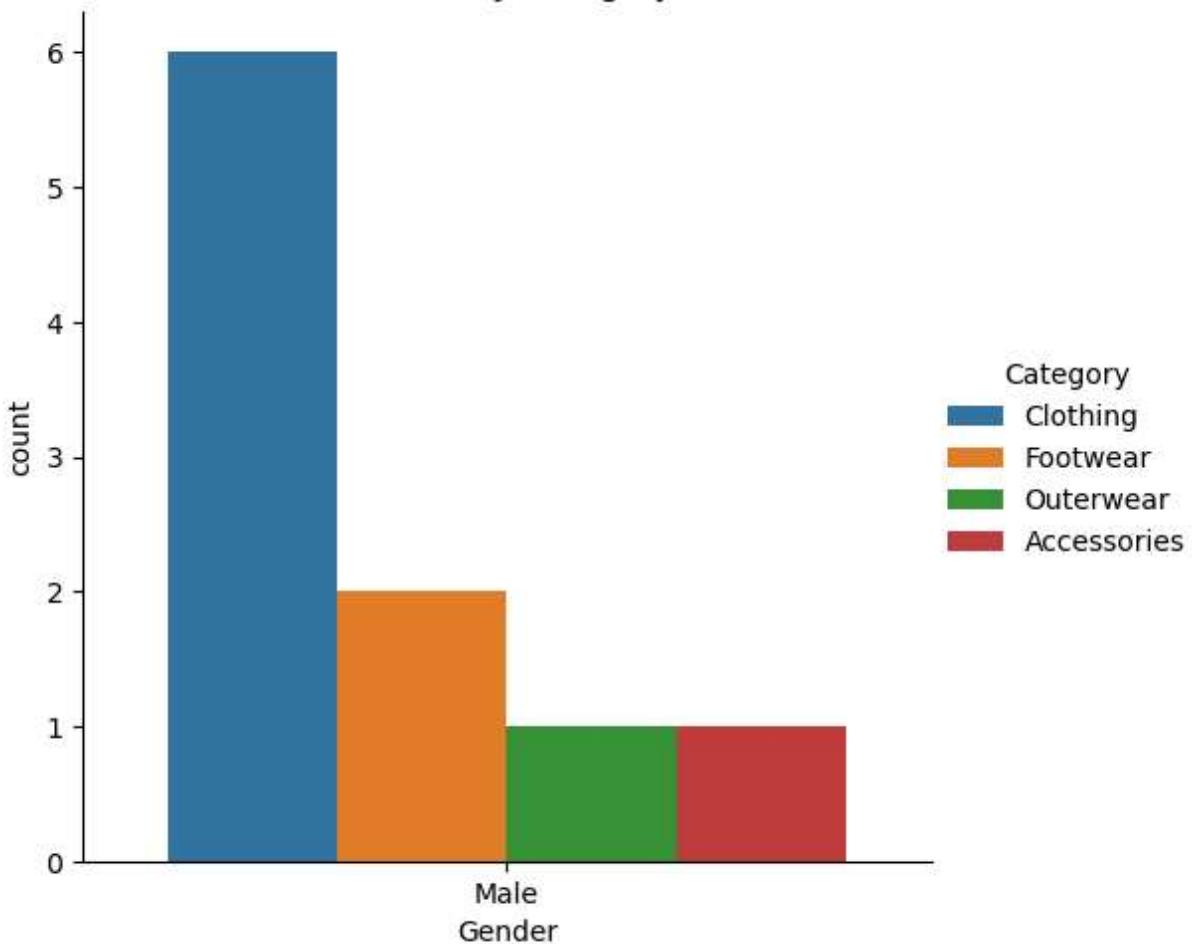
```
In [69]: plt.figure(figsize=(10,5))
ax = sns.countplot(x='Category', data=df)
ax.bar_label(ax.containers[0])
plt.show()
```



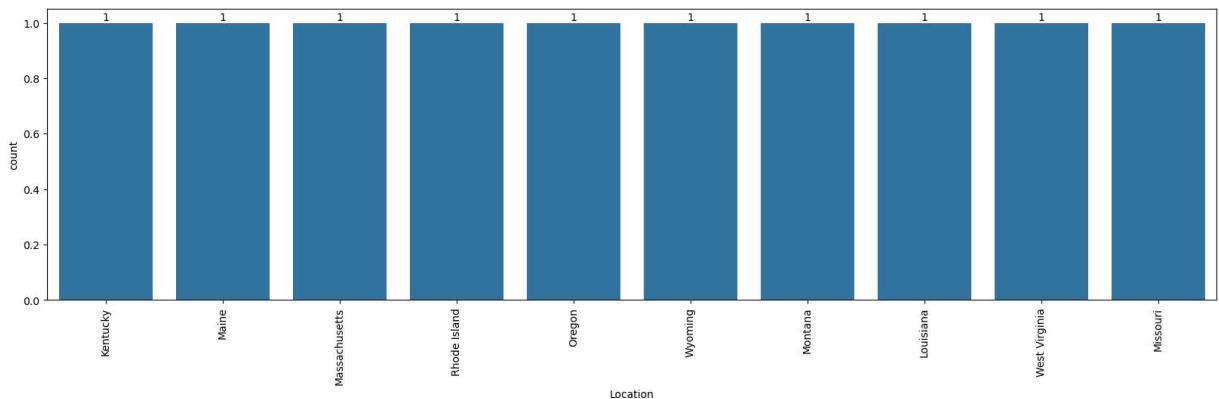
Clothing category is most among all others.

```
In [18]: #2
sns.catplot(hue='Category', kind='count', x='Gender', data=df)
plt.title('Gender by Category')
plt.show()
```

Gender by Category



```
In [19]: plt.figure(figsize=(20,5))
ax = sns.countplot(x='Location', data=df)
plt.xticks(rotation=90)
ax.bar_label(ax.containers[0])
plt.show()
```

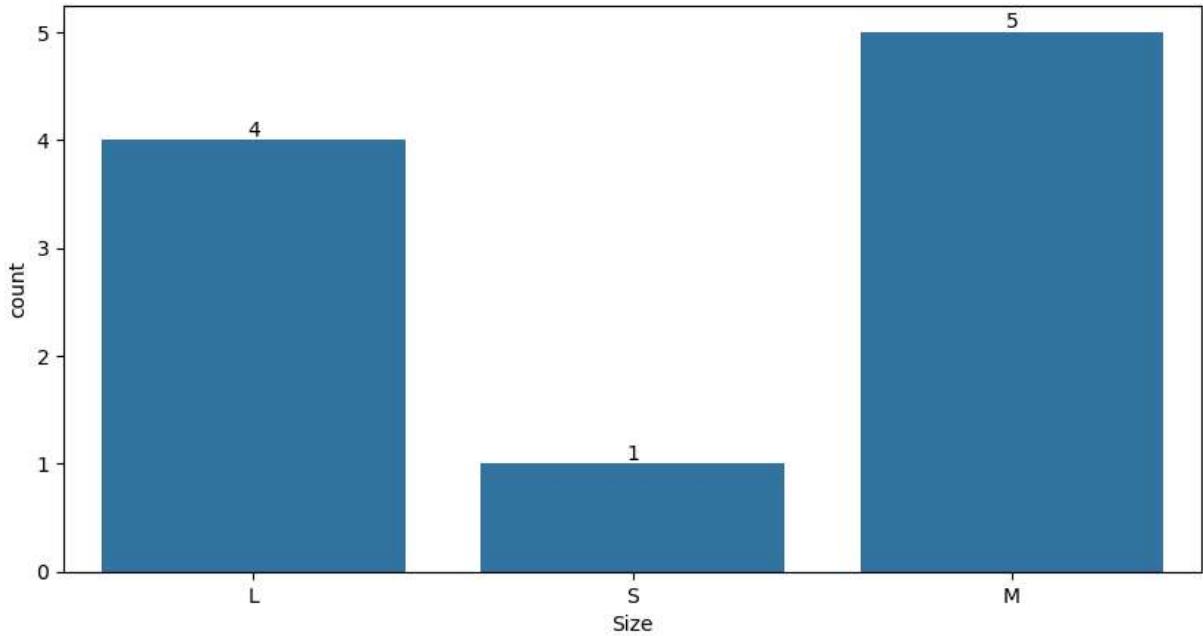


Most buyers are from Montana, California, and Idaho.

```
In [ ]:
```

```
In [20]: plt.figure(figsize=(10,5))
ax = sns.countplot(x='Size', data=df)
```

```
ax.bar_label(ax.containers[0])
plt.show()
```

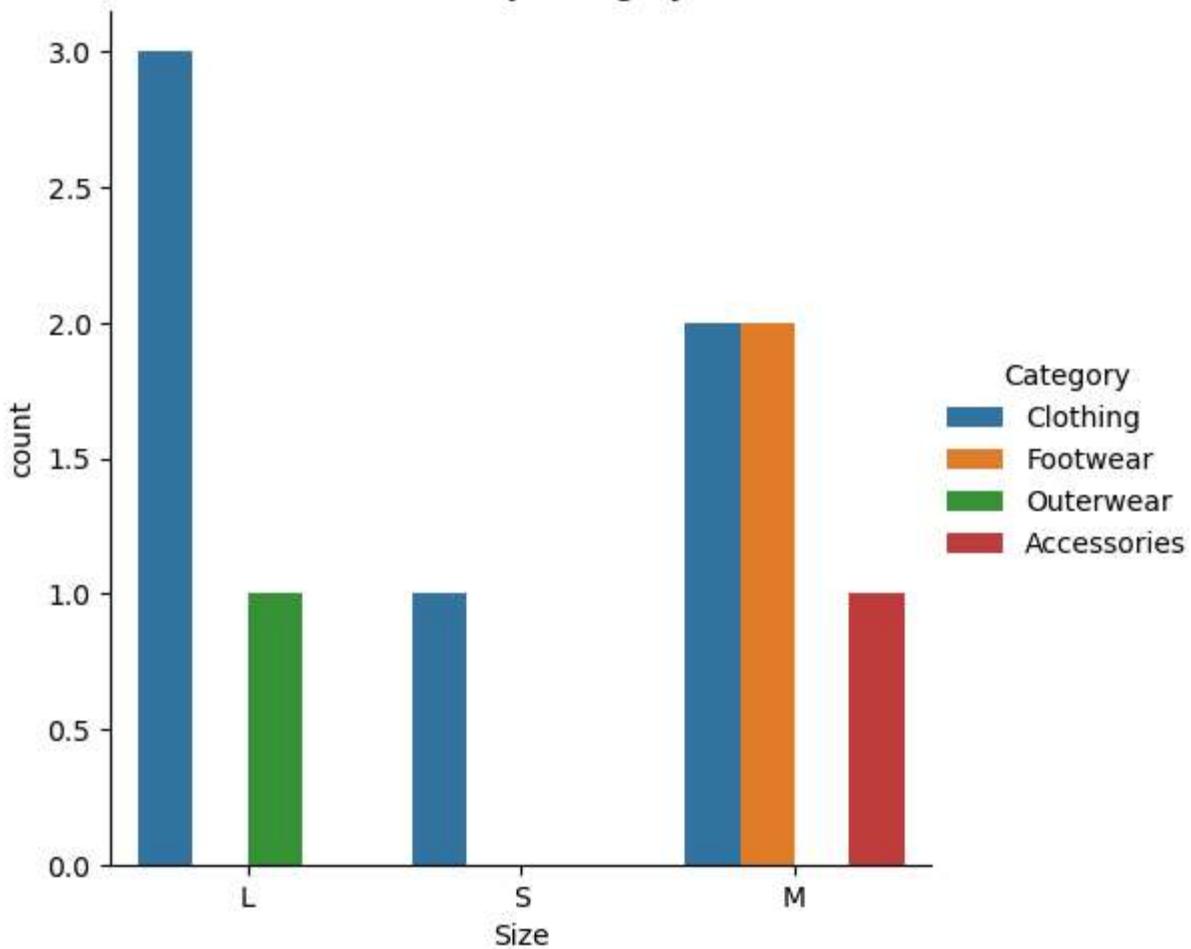


Medium size products have the highest range of sell.

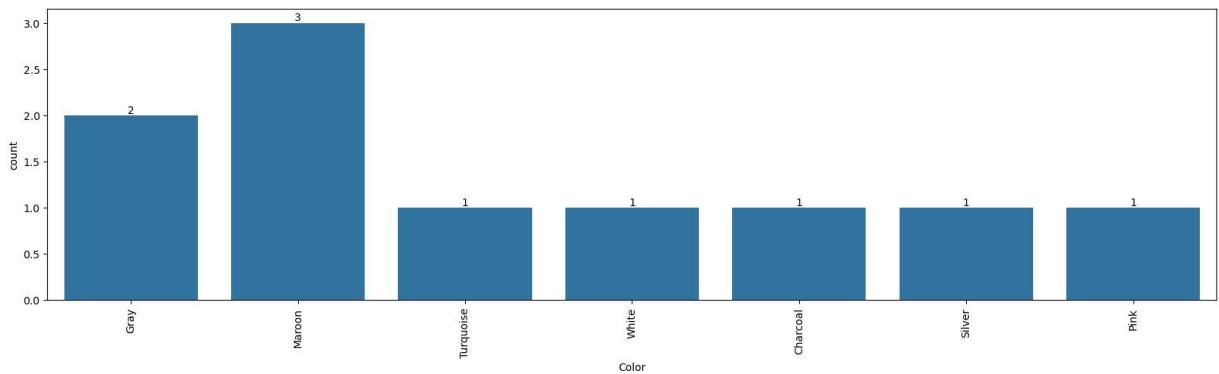
In [21]:

```
#2
sns.catplot(hue='Category',kind='count',x='Size',data=df)
plt.title('Gender by Category')
plt.show()
```

Gender by Category

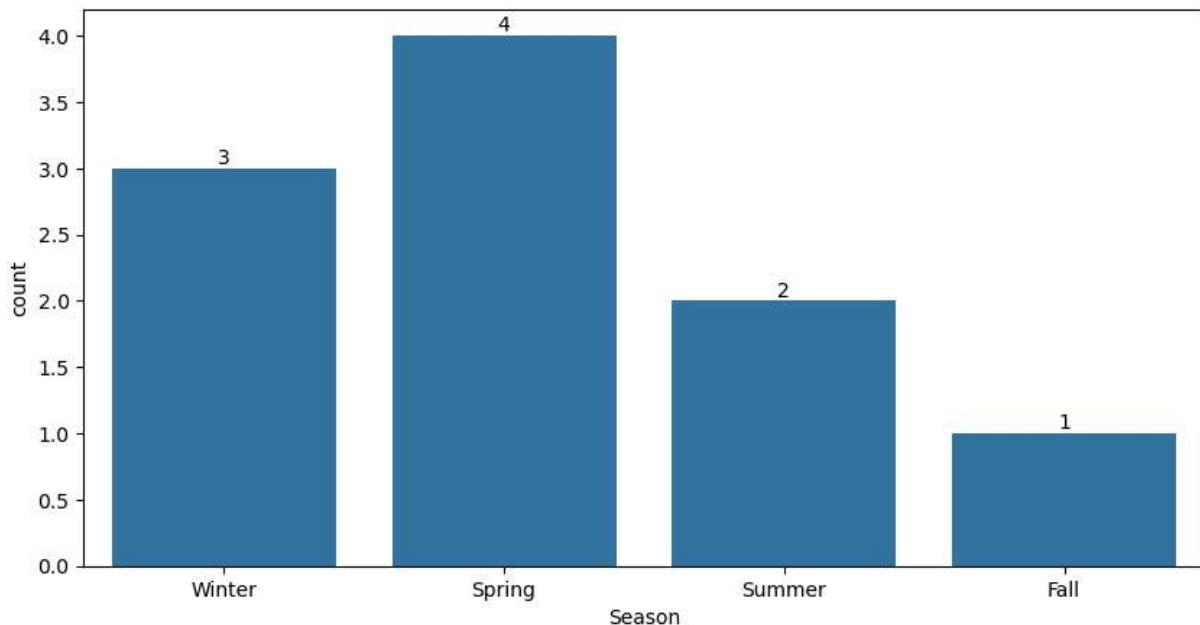


```
In [22]: plt.figure(figsize=(20,5))
ax = sns.countplot(x='Color', data=df)
plt.xticks(rotation=90)
ax.bar_label(ax.containers[0])
plt.show()
```



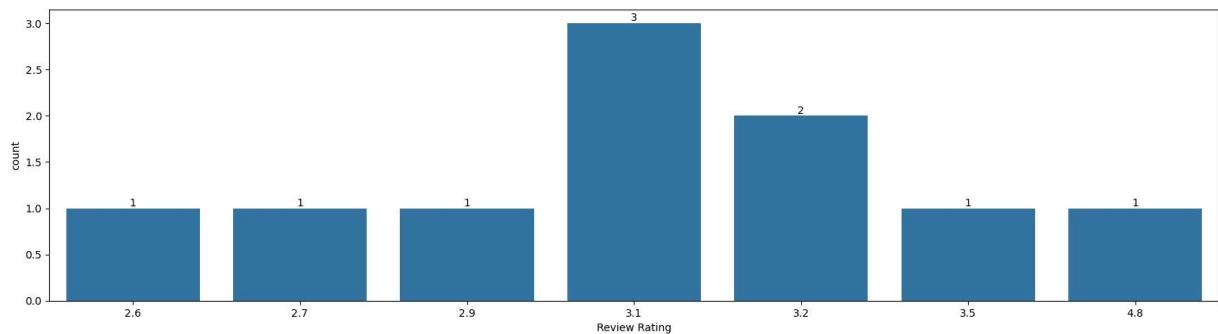
Buyers bought Olive, Yellow, and Silver colors most.

```
In [23]: plt.figure(figsize=(10,5))
ax = sns.countplot(x='Season', data=df)
ax.bar_label(ax.containers[0])
plt.show()
```



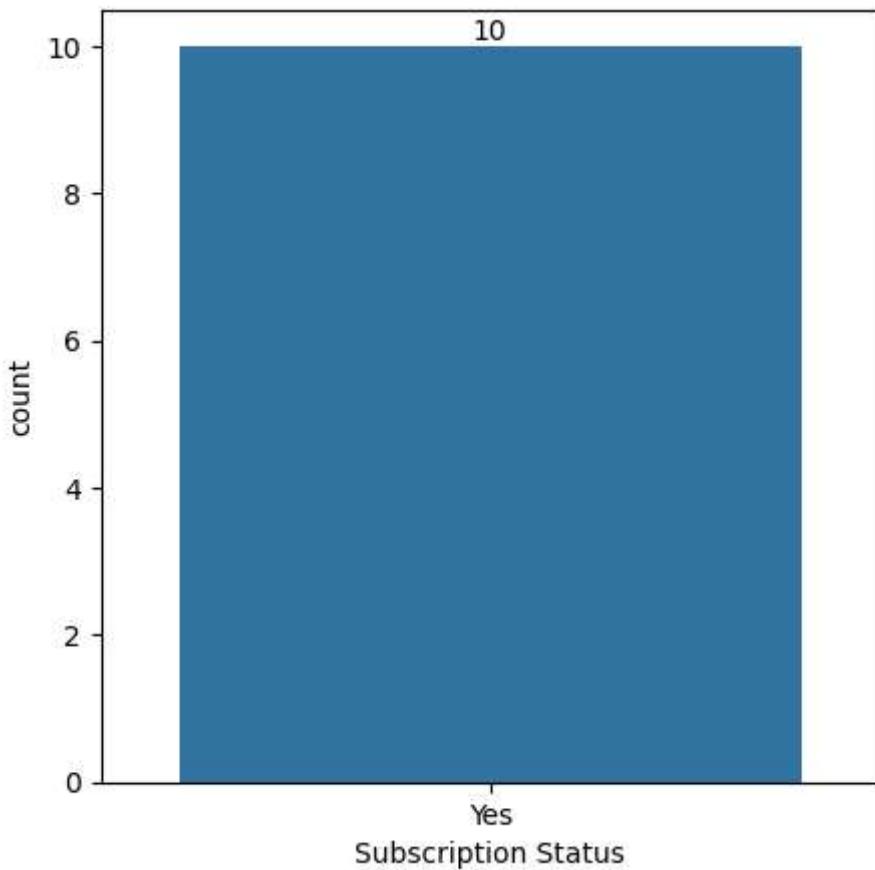
In Spring, Fall, Winter and then Summer seasons the sell out is most.

```
In [24]: plt.figure(figsize=(20,5))
ax = sns.countplot(x='Review Rating', data=df)
ax.bar_label(ax.containers[0])
plt.show()
```

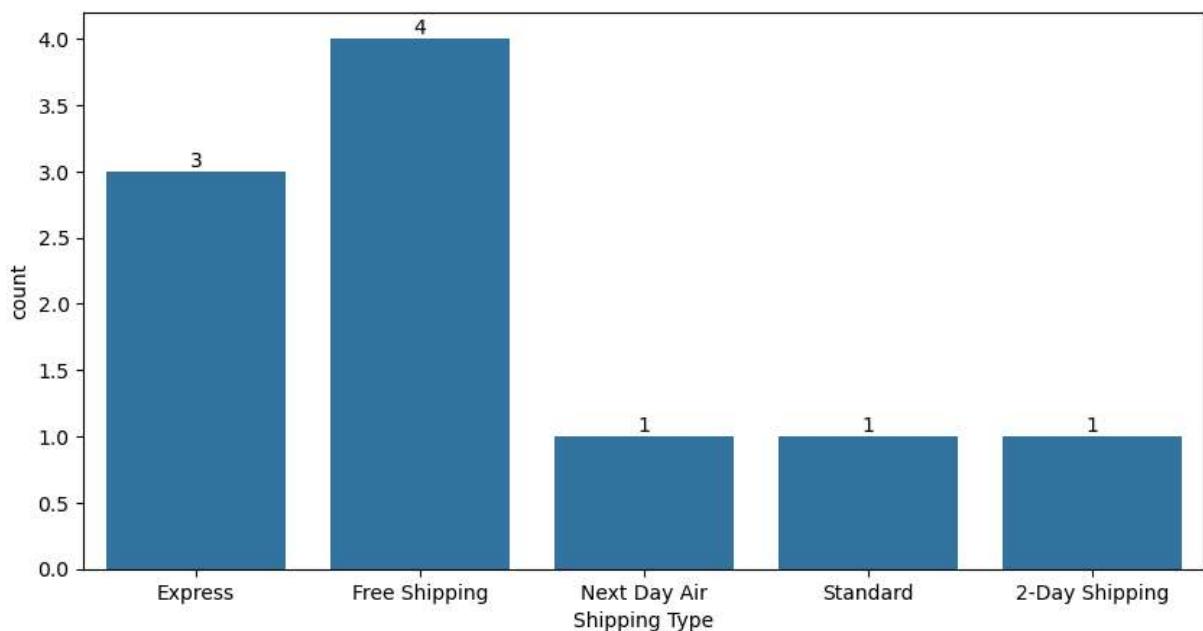


The rating of 3.4, 4.0, and 4.6 were given.

```
In [25]: plt.figure(figsize=(5,5))
ax = sns.countplot(x='Subscription Status', data=df)
ax.bar_label(ax.containers[0])
plt.show()
```

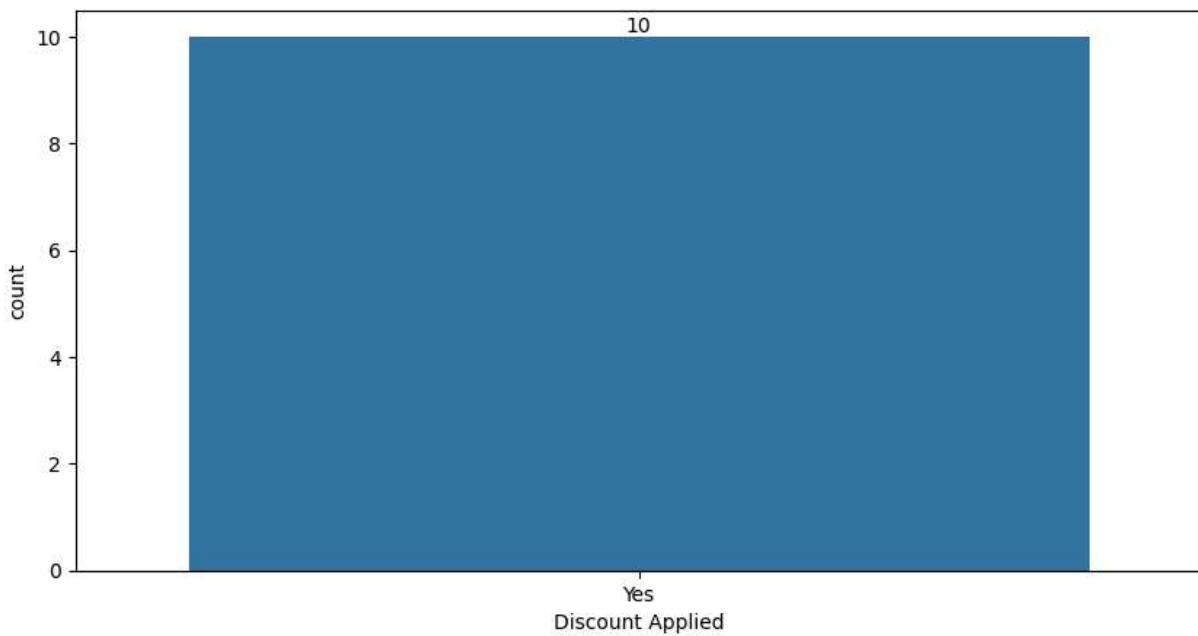


```
In [26]: plt.figure(figsize=(10,5))
ax = sns.countplot(x='Shipping Type', data=df)
ax.bar_label(ax.containers[0])
plt.show()
```

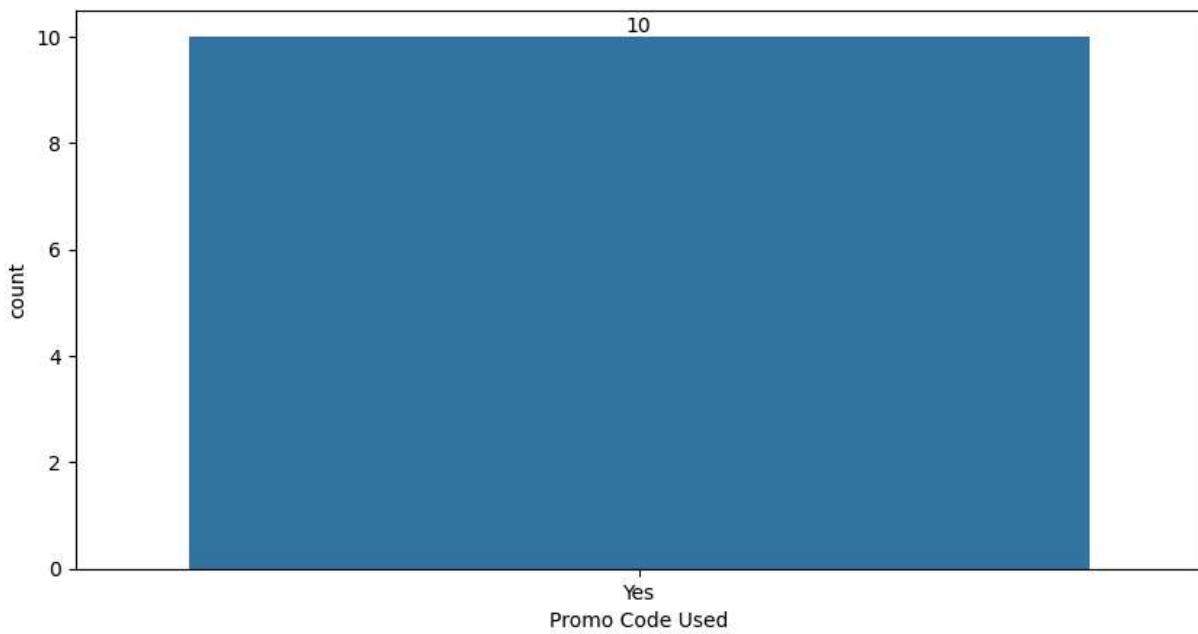


Free shipping has the largest amount of buyers.

```
In [70]: plt.figure(figsize=(10,5))
ax = sns.countplot(x='Discount Applied', data=df)
ax.bar_label(ax.containers[0])
plt.show()
```



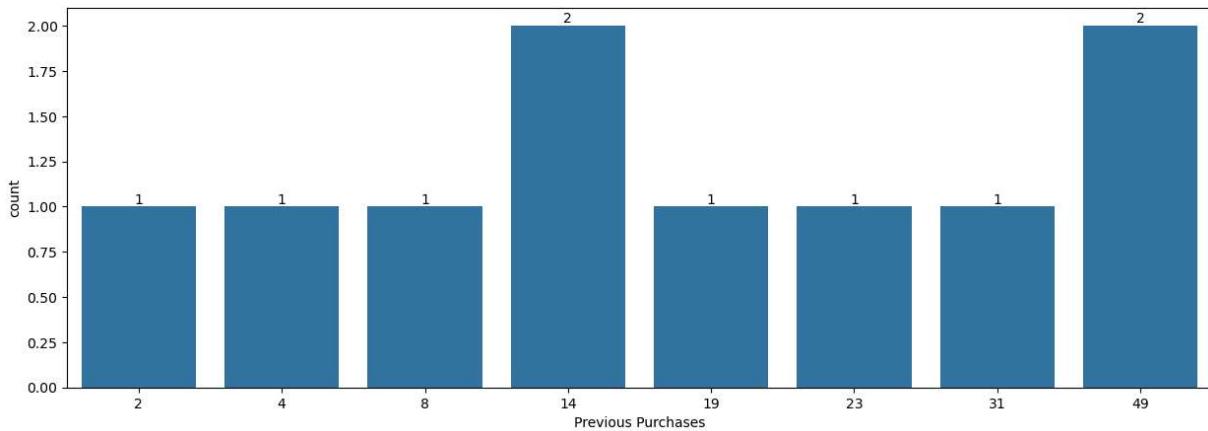
```
In [71]: plt.figure(figsize=(10,5))
ax = sns.countplot(x='Promo Code Used', data=df)
ax.bar_label(ax.containers[0])
plt.show()
```



2223 buyers have not used Promo Code.

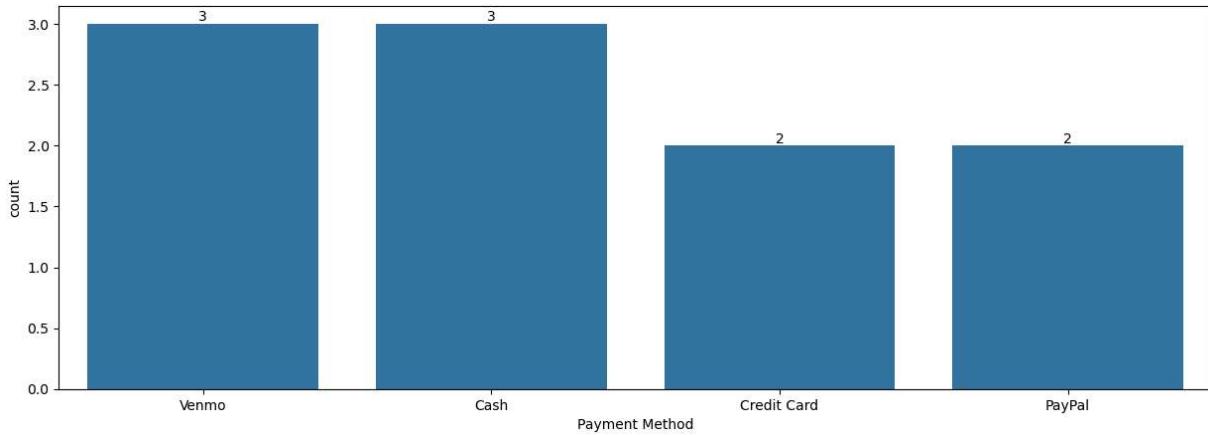
```
In [29]: plt.figure(figsize=(15,5))
ax = sns.countplot(x='Previous Purchases', data=df)
```

```
ax.bar_label(ax.containers[0])
plt.show()
```



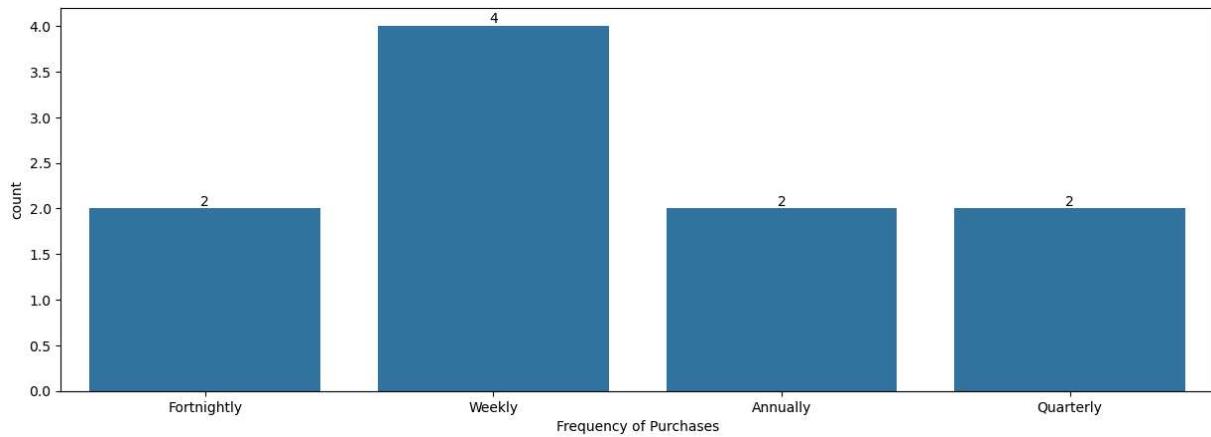
97 buyers have 31 previous purchases.

```
In [30]: plt.figure(figsize=(15,5))
ax = sns.countplot(x='Payment Method', data=df)
ax.bar_label(ax.containers[0])
plt.show()
```

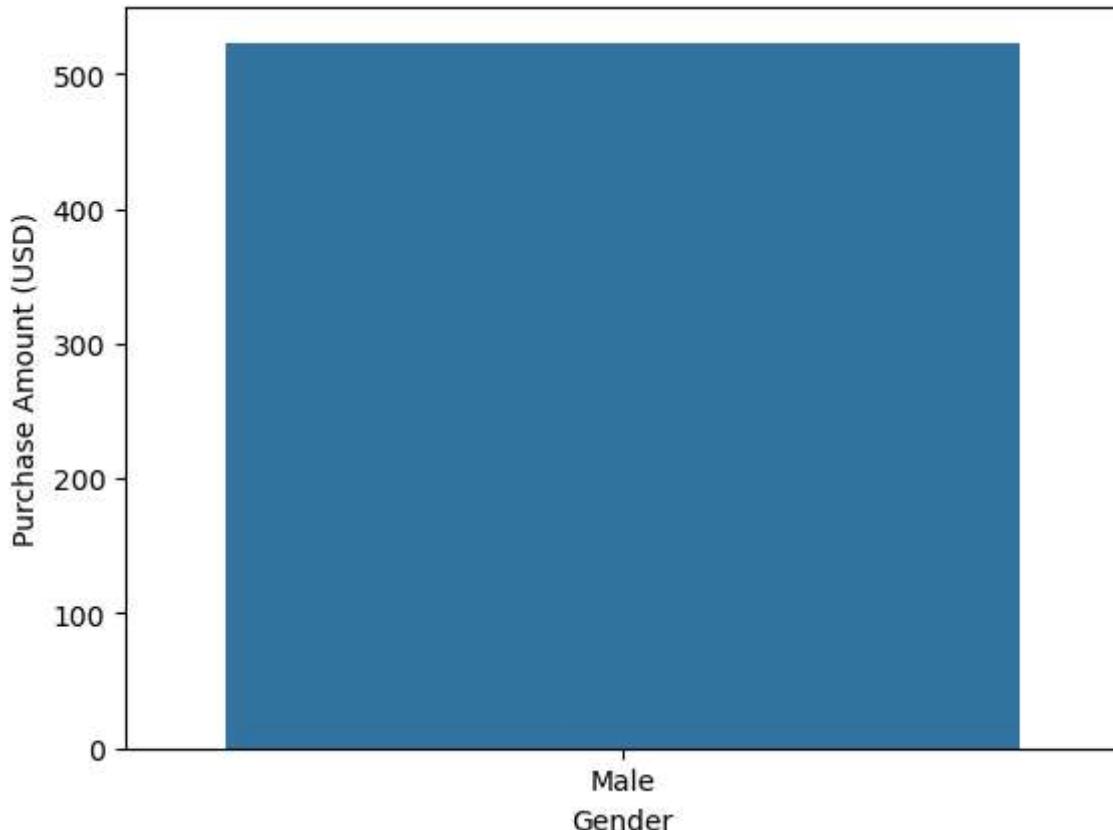


677 buyers used PayPal, 671 used Credit Card, and 670 paid Cash.

```
In [31]: plt.figure(figsize=(15,5))
ax = sns.countplot(x='Frequency of Purchases', data=df)
ax.bar_label(ax.containers[0])
plt.show()
```



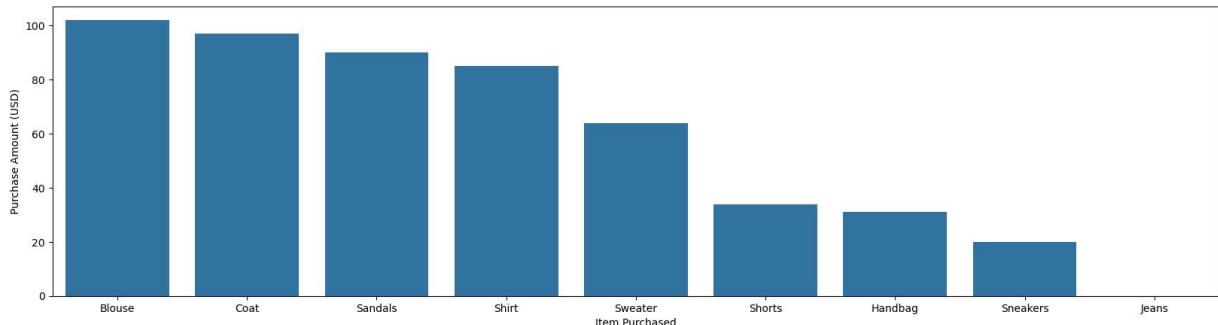
```
In [72]: sales = df.groupby('Gender')[ 'Purchase Amount (USD)' ].sum().reset_index().sort_values(ascending=False)
sns.barplot(data=sales, x='Gender', y='Purchase Amount (USD)')
plt.show()
```



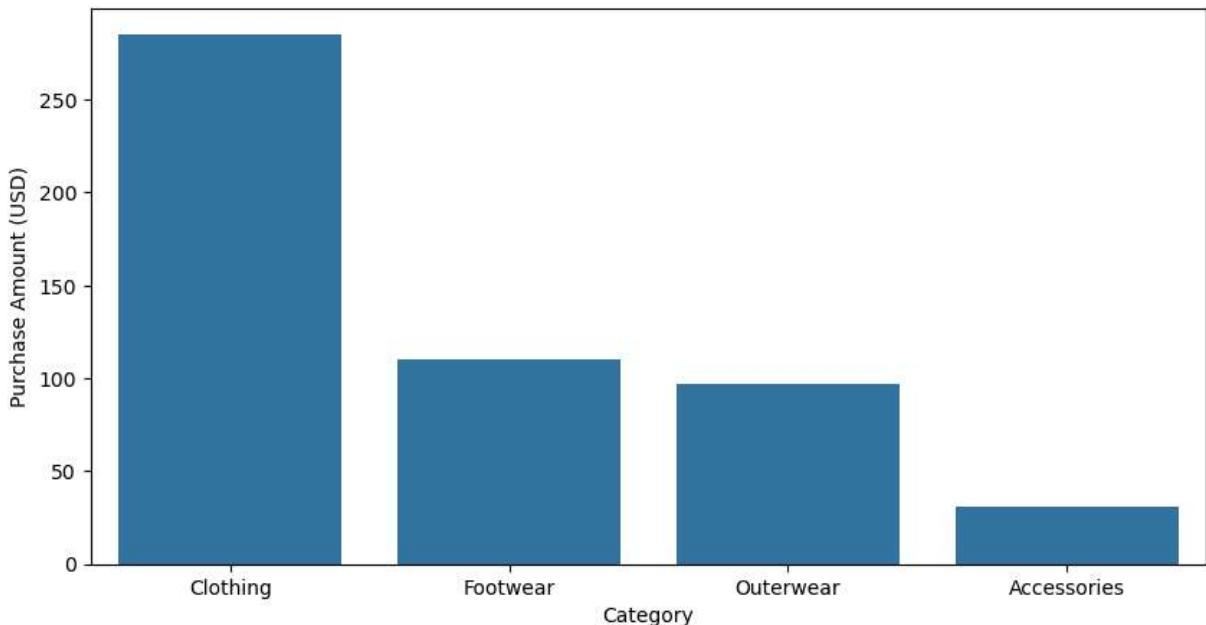
```
In [33]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10 entries, 0 to 9
Data columns (total 18 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Customer ID      10 non-null    int64  
 1   Age               10 non-null    int64  
 2   Gender            10 non-null    object  
 3   Item Purchased   10 non-null    object  
 4   Category          10 non-null    object  
 5   Purchase Amount (USD) 10 non-null    float64 
 6   Location          10 non-null    object  
 7   Size              10 non-null    object  
 8   Color              10 non-null    object  
 9   Season             10 non-null    object  
 10  Review Rating    10 non-null    float64 
 11  Subscription Status 10 non-null    object  
 12  Shipping Type    10 non-null    object  
 13  Discount Applied 10 non-null    object  
 14  Promo Code Used  10 non-null    object  
 15  Previous Purchases 10 non-null    int64  
 16  Payment Method   10 non-null    object  
 17  Frequency of Purchases 10 non-null    object  
dtypes: float64(2), int64(3), object(13)
memory usage: 1.5+ KB
```

```
In [73]: plt.figure(figsize=(20,5))
sales = df.groupby(['Item Purchased'],as_index=False)[['Purchase Amount (USD)']].sum()
sns.barplot(data=sales,x='Item Purchased',y='Purchase Amount (USD)')
plt.show()
```

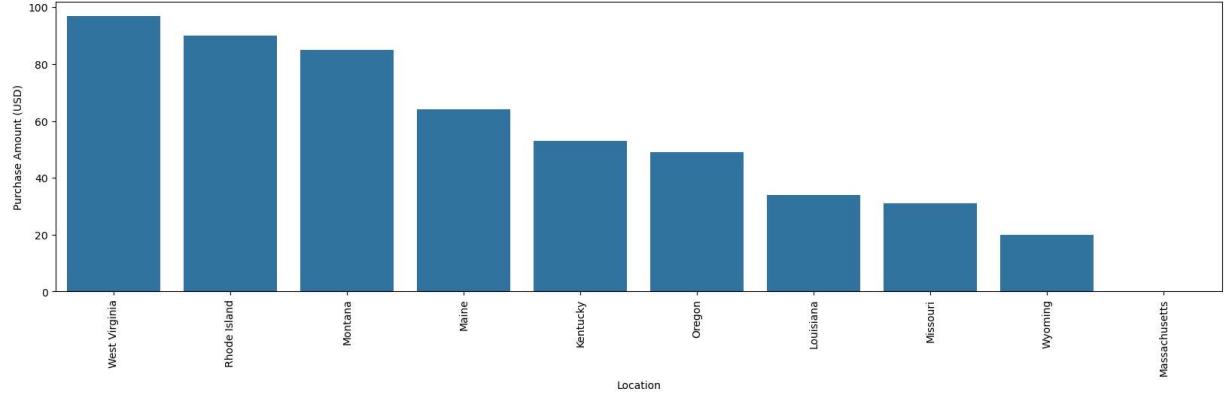


```
In [74]: plt.figure(figsize=(10,5))
sales = df.groupby(['Category'],as_index=False)[['Purchase Amount (USD)']].sum().sort
sns.barplot(data=sales,x='Category',y='Purchase Amount (USD)')
plt.show()
```



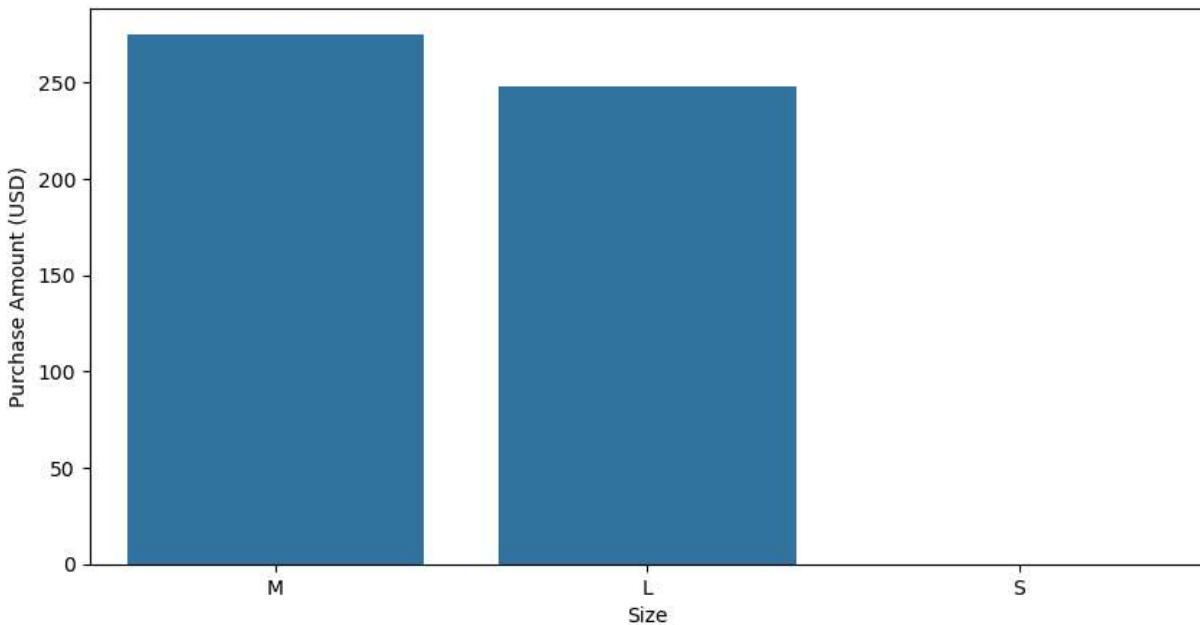
Clothing category has the most Purchased amount.

```
In [36]: plt.figure(figsize=(20,5))
sales = df.groupby(['Location'],as_index=False)[ 'Purchase Amount (USD)' ].sum().sort
sns.barplot(data=sales,x='Location',y='Purchase Amount (USD)')
plt.xticks(rotation=90)
plt.show()
```



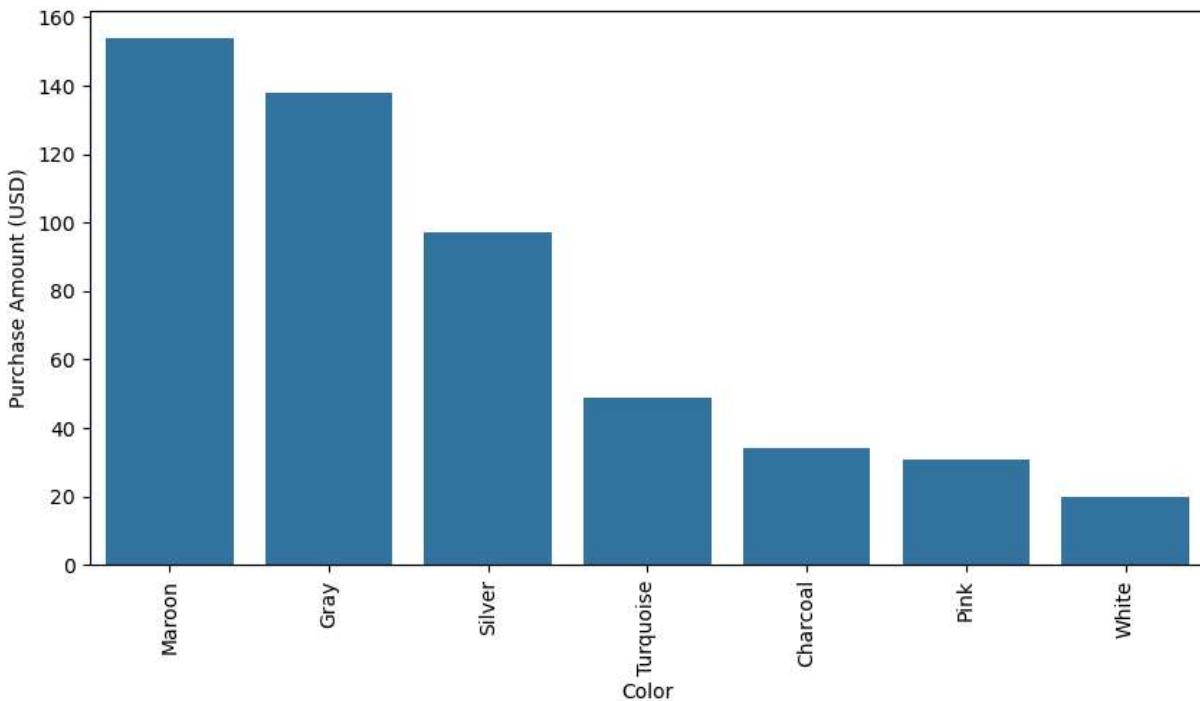
Montana has produced the most amount.

```
In [75]: plt.figure(figsize=(10,5))
sales = df.groupby(['Size'],as_index=False)[ 'Purchase Amount (USD)' ].sum().sort_val
sns.barplot(data=sales,x='Size',y='Purchase Amount (USD)')
plt.show()
```



Medium size has generated the most profit.

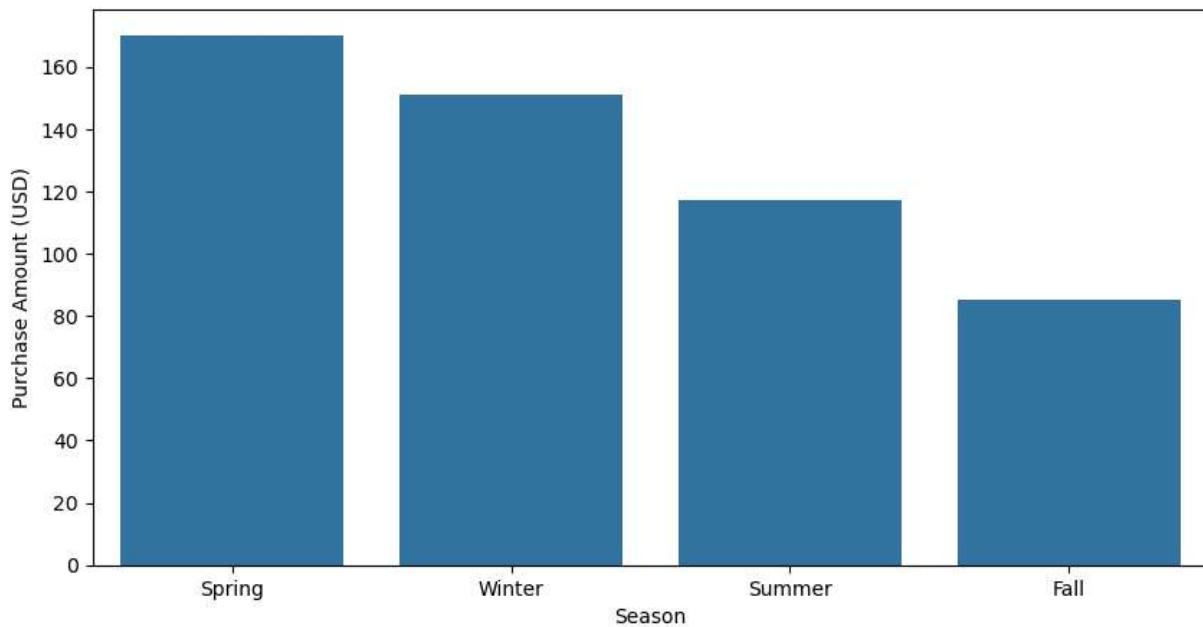
```
In [38]: plt.figure(figsize=(10,5))
sales = df.groupby(['Color'],as_index=False)[ 'Purchase Amount (USD)'].sum().sort_values(ascending=False)
sns.barplot(data=sales,x='Color',y='Purchase Amount (USD)')
plt.xticks(rotation=90)
plt.show()
```



Buyers bought products of green color.

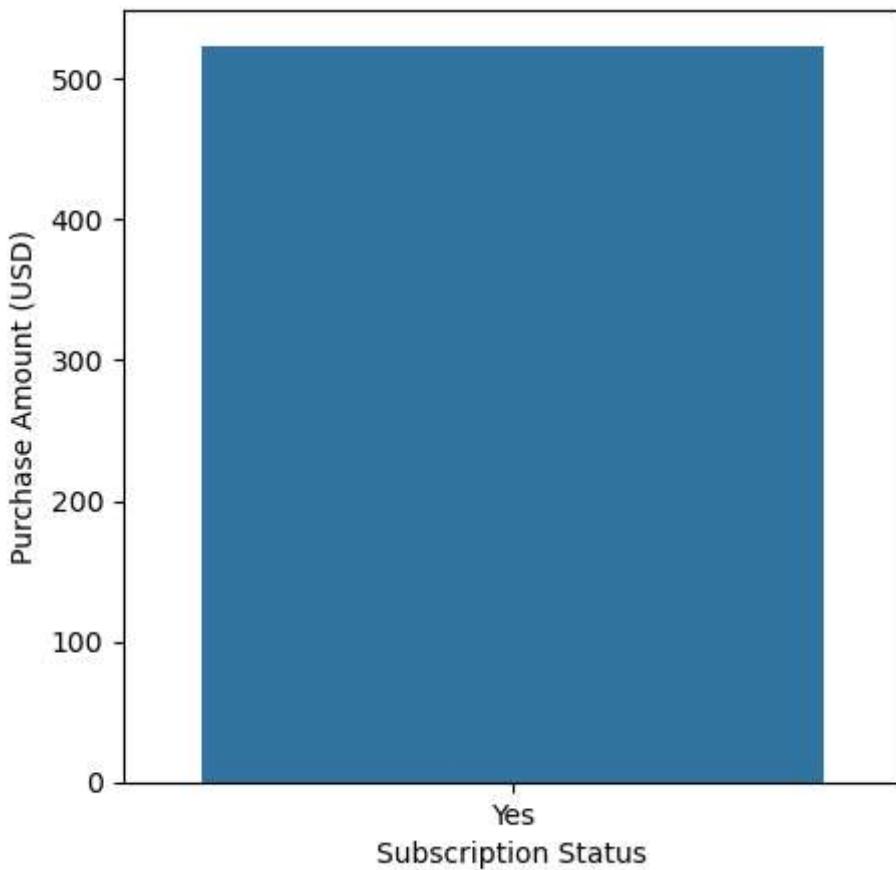
```
In [76]: plt.figure(figsize=(10,5))
sales = df.groupby(['Season'],as_index=False)[ 'Purchase Amount (USD)'].sum().sort_values(ascending=False)
```

```
sns.barplot(data=sales,x='Season',y='Purchase Amount (USD)')  
plt.show()
```



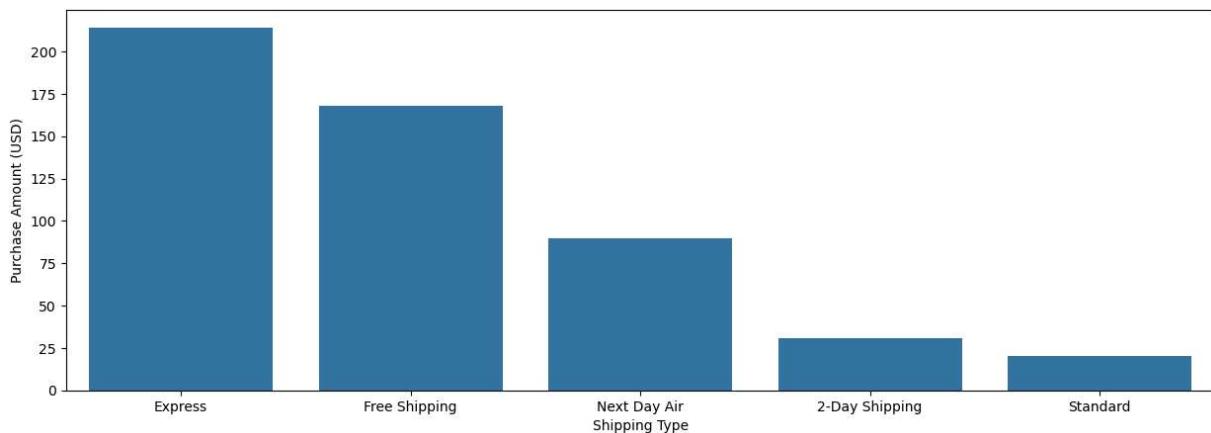
Fall season has produced the most profit.

```
In [78]: plt.figure(figsize=(5,5))  
sales = df.groupby(['Subscription Status'],as_index=False)[['Purchase Amount (USD)']]  
sns.barplot(data=sales,x='Subscription Status',y='Purchase Amount (USD)')  
plt.show()
```



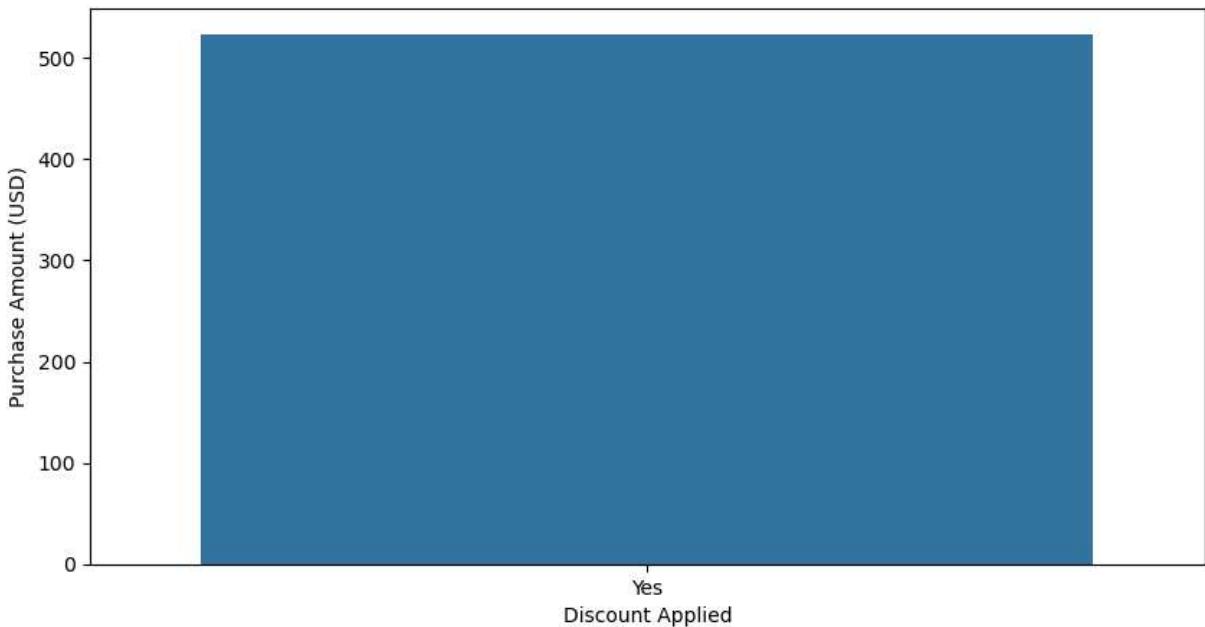
```
In [41]: # Most Profit came from those who haven't subscribed.
```

```
In [79]: plt.figure(figsize=(15,5))
sales = df.groupby(['Shipping Type'],as_index=False)[['Purchase Amount (USD)']].sum()
sns.barplot(data=sales,x='Shipping Type',y='Purchase Amount (USD)')
plt.show()
```

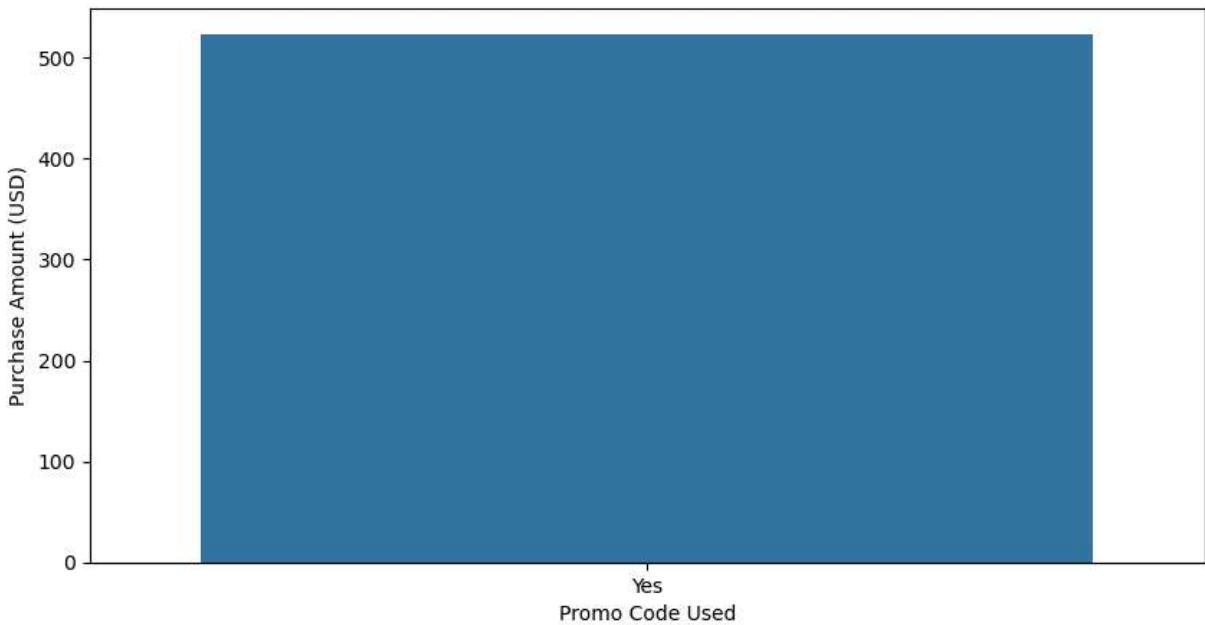


Free shipping producd most profit.

```
In [80]: plt.figure(figsize=(10,5))
sales = df.groupby(['Discount Applied'],as_index=False)[['Purchase Amount (USD)']].sum()
sns.barplot(data=sales,x='Discount Applied',y='Purchase Amount (USD)')
plt.show()
```

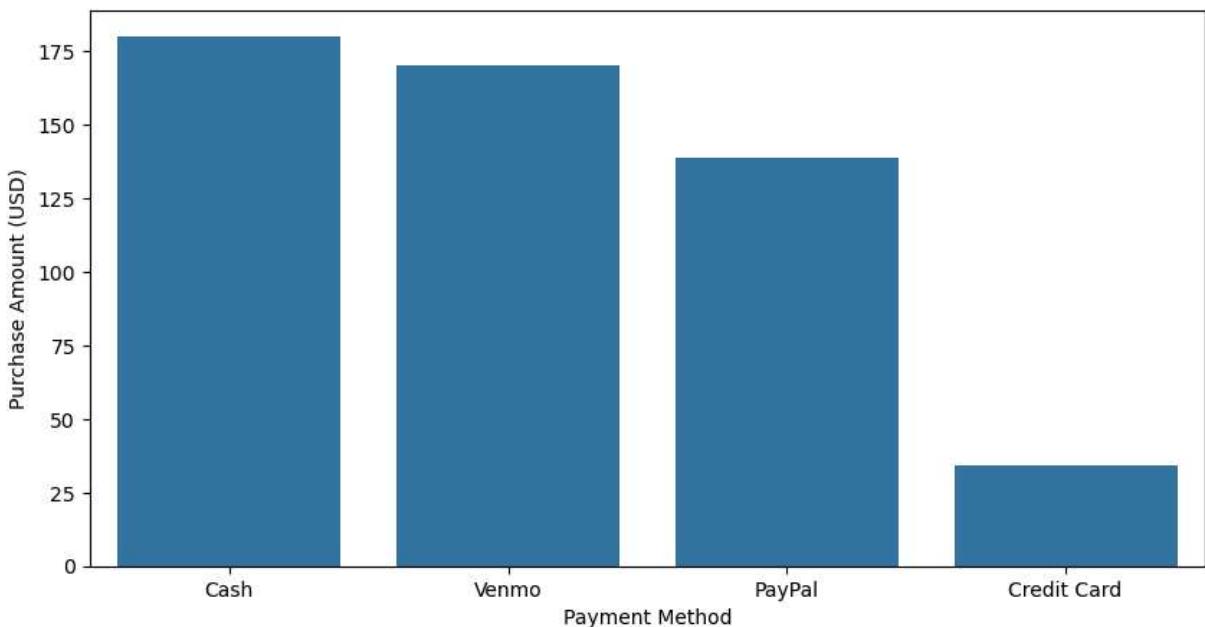


```
In [81]: plt.figure(figsize=(10,5))
sales = df.groupby(['Promo Code Used'],as_index=False)[['Purchase Amount (USD)']].sum()
sns.barplot(data=sales,x='Promo Code Used',y='Purchase Amount (USD)')
plt.show()
```



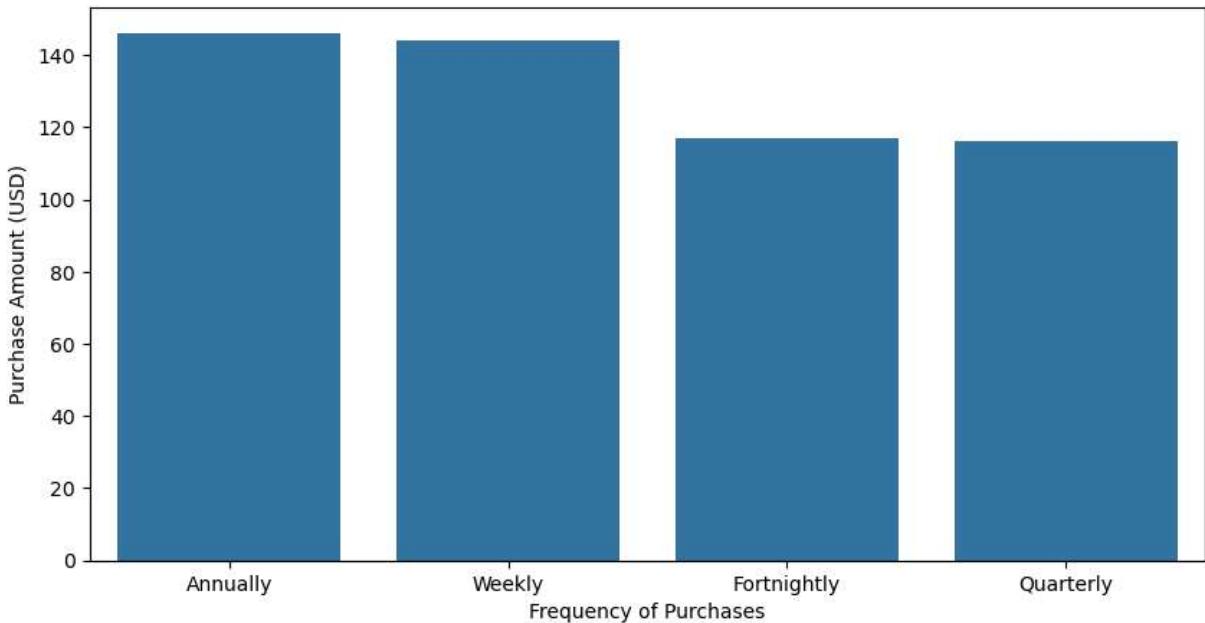
Most of the profit came from those who hadn't used Promo Code.

```
In [82]: plt.figure(figsize=(10,5))
sales = df.groupby(['Payment Method'],as_index=False)[['Purchase Amount (USD)']].sum()
sns.barplot(data=sales,x='Payment Method',y='Purchase Amount (USD)')
plt.show()
```

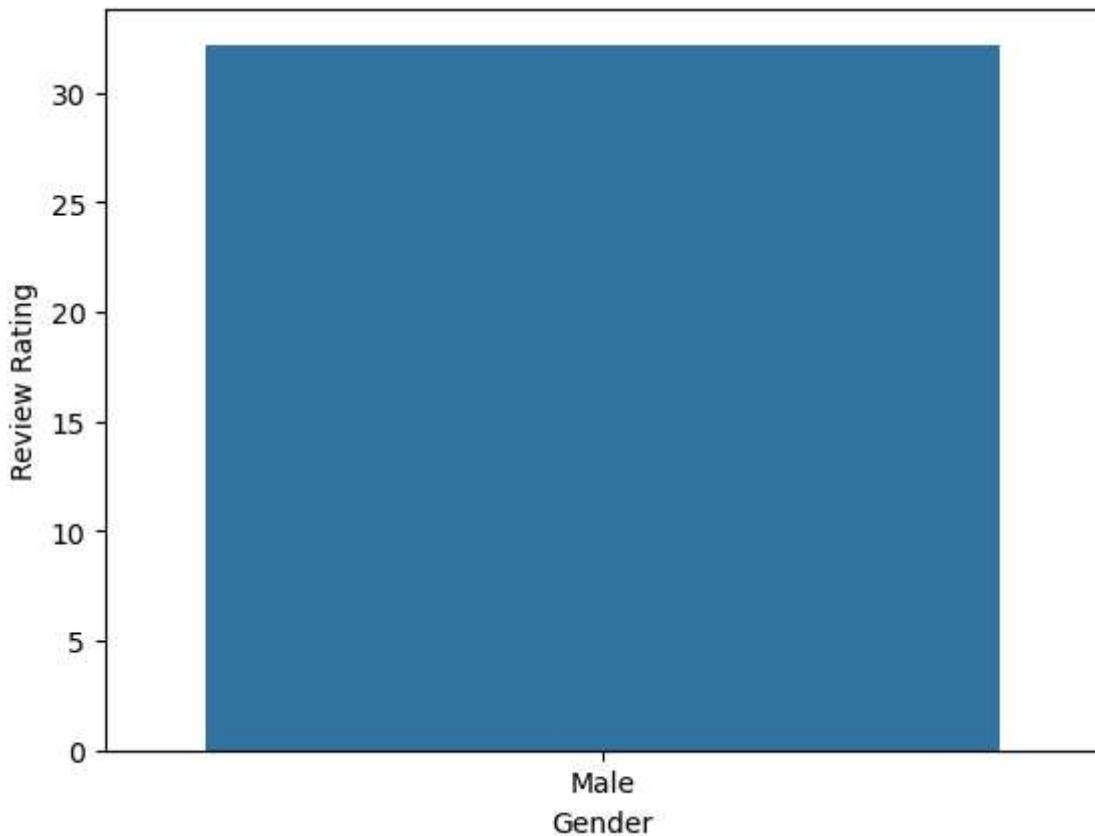


Most amount was created by Credit card, and then PayPal.

```
In [83]: plt.figure(figsize=(10,5))
sales = df.groupby(['Frequency of Purchases'],as_index=False)[['Purchase Amount (USD)']]
sns.barplot(data=sales,x='Frequency of Purchases',y='Purchase Amount (USD)')
plt.show()
```

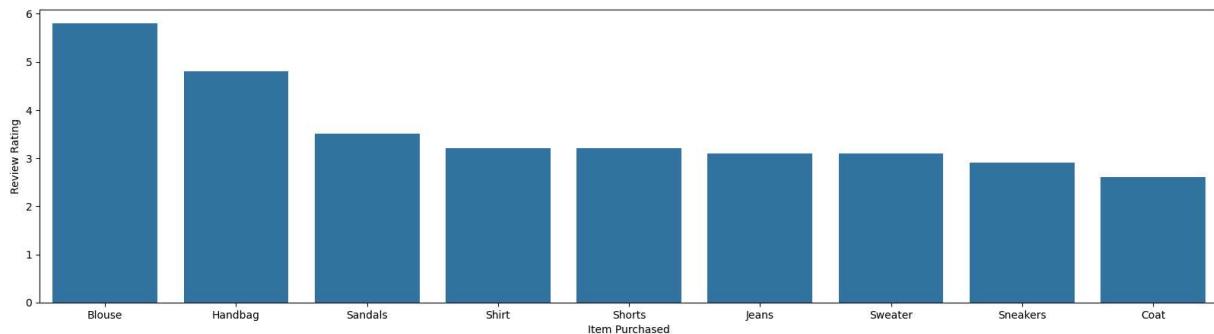


```
In [87]: sales = df.groupby(['Gender'],as_index=False)[['Review Rating']].sum().sort_values(by='Review Rating')
sns.barplot(data=sales,x='Gender',y='Review Rating')
plt.show()
```



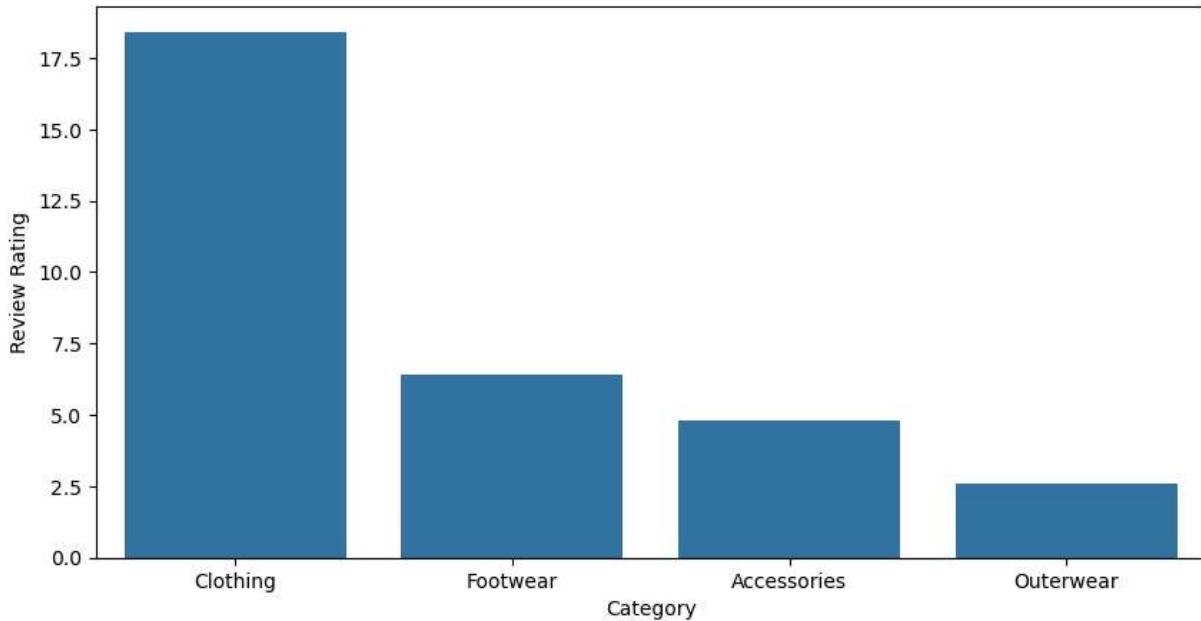
Males have given the most Review Rating.

```
In [88]: plt.figure(figsize=(20,5))
sales = df.groupby(['Item Purchased'],as_index=False)[['Review Rating']].sum().sort_values(['Review Rating'], ascending=False)
sns.barplot(data=sales,x='Item Purchased',y='Review Rating')
plt.show()
```



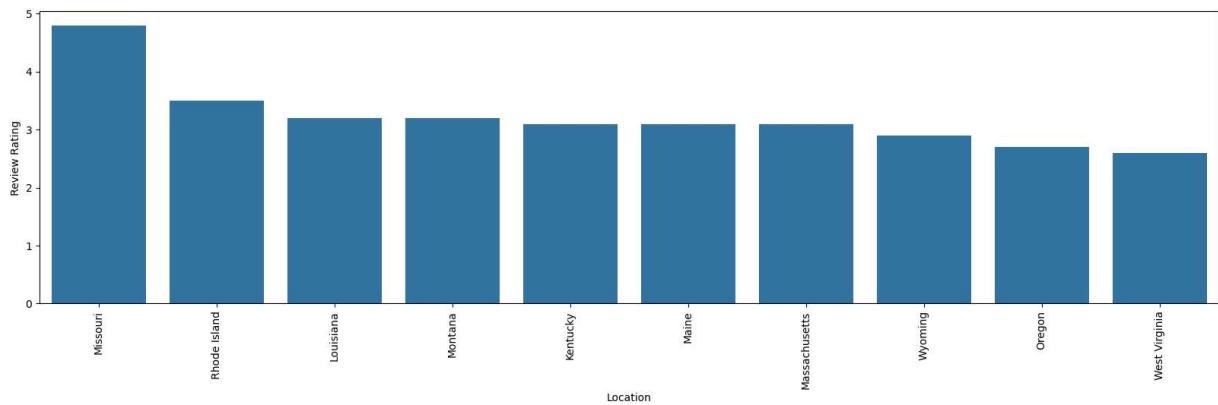
Jewelry buyers have given the a lot of reviews.

```
In [89]: plt.figure(figsize=(10,5))
sales = df.groupby(['Category'],as_index=False)[['Review Rating']].sum().sort_values(['Review Rating'], ascending=False)
sns.barplot(data=sales,x='Category',y='Review Rating')
plt.show()
```



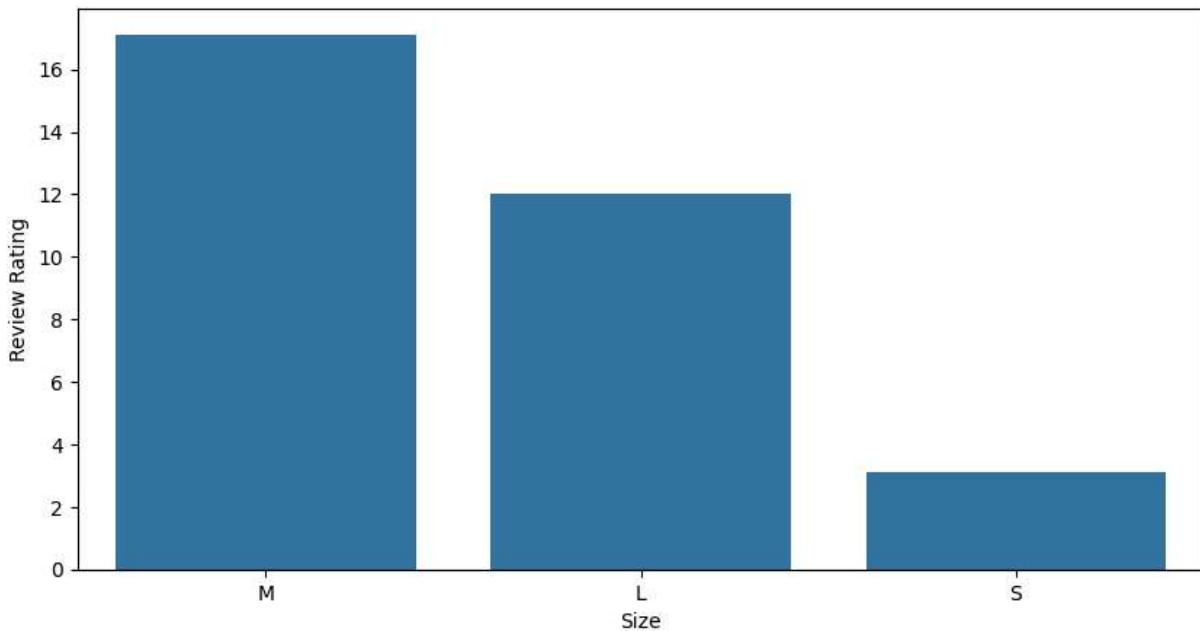
Clothing buyers have given the most reviews.

```
In [50]: plt.figure(figsize=(20,5))
sales = df.groupby(['Location'],as_index=False)[['Review Rating']].sum().sort_values(
sns.barplot(data=sales,x='Location',y='Review Rating')
plt.xticks(rotation=90)
plt.show()
```



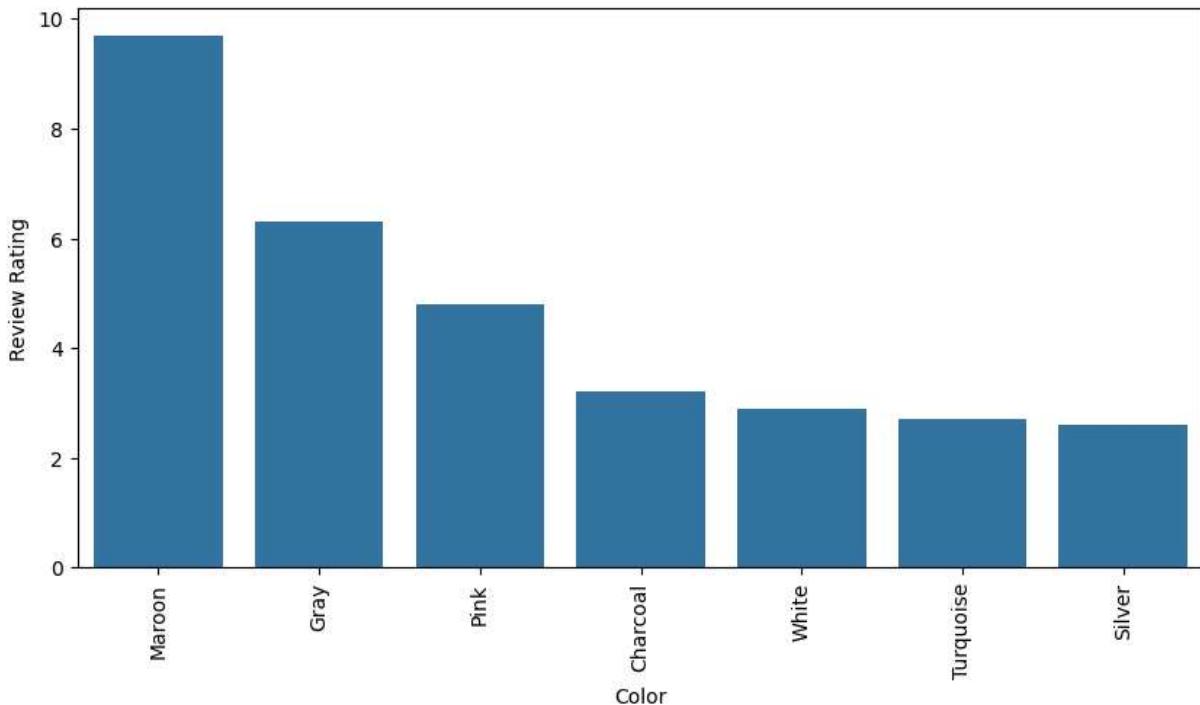
Mantola's buyers have given most reviews.

```
In [90]: plt.figure(figsize=(10,5))
sales = df.groupby(['Size'],as_index=False)[['Review Rating']].sum().sort_values(by='
sns.barplot(data=sales,x='Size',y='Review Rating')
plt.show()
```



Medium size buyers have given most reviews.

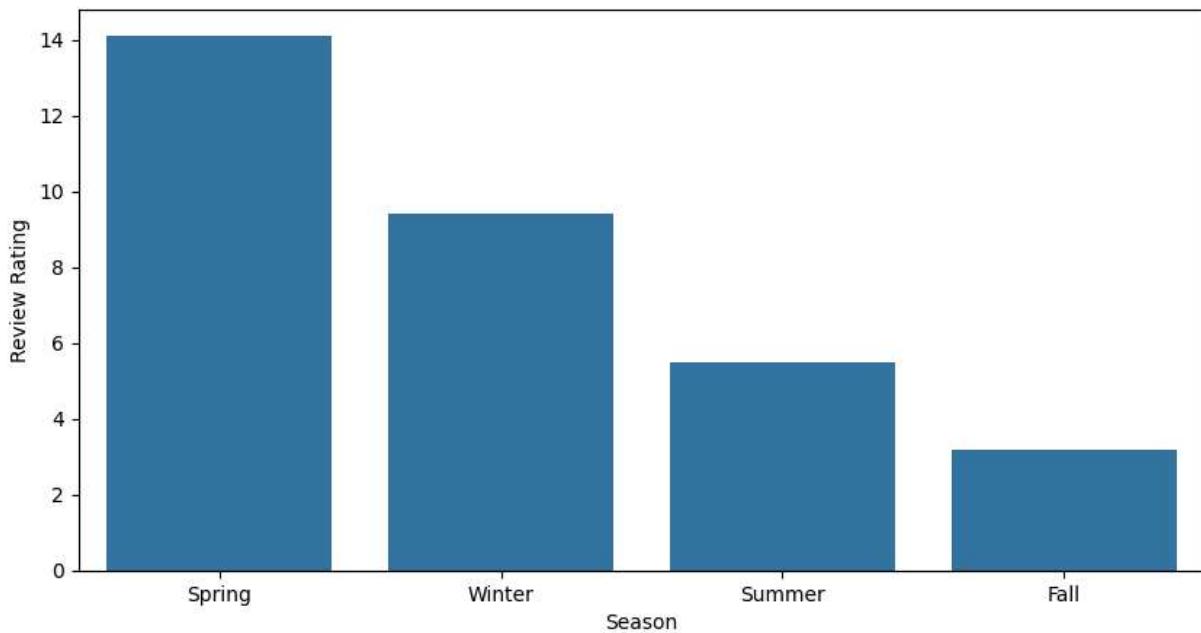
```
In [52]: plt.figure(figsize=(10,5))
sales = df.groupby(['Color'],as_index=False)[['Review Rating']].sum().sort_values(by='Review Rating', ascending=False)
sns.barplot(data=sales,x='Color',y='Review Rating')
plt.xticks(rotation=90)
plt.show()
```



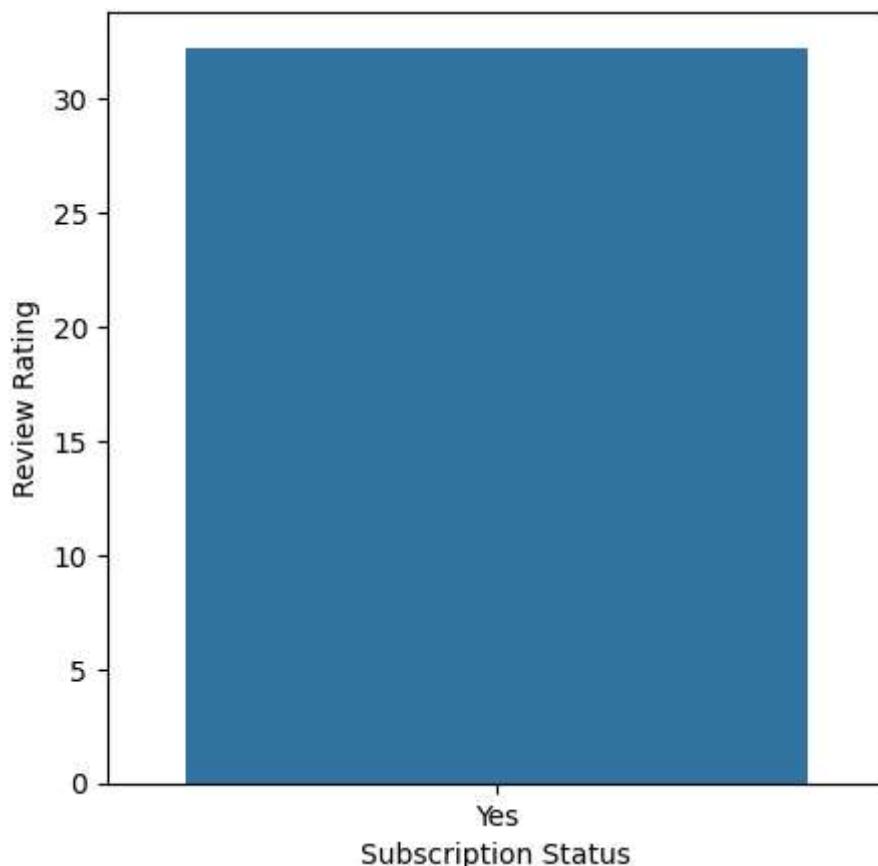
Yellow color buyers have given most reviews.

```
In [91]: plt.figure(figsize=(10,5))
sales = df.groupby(['Season'],as_index=False)[['Review Rating']].sum().sort_values(by='Review Rating', ascending=False)
```

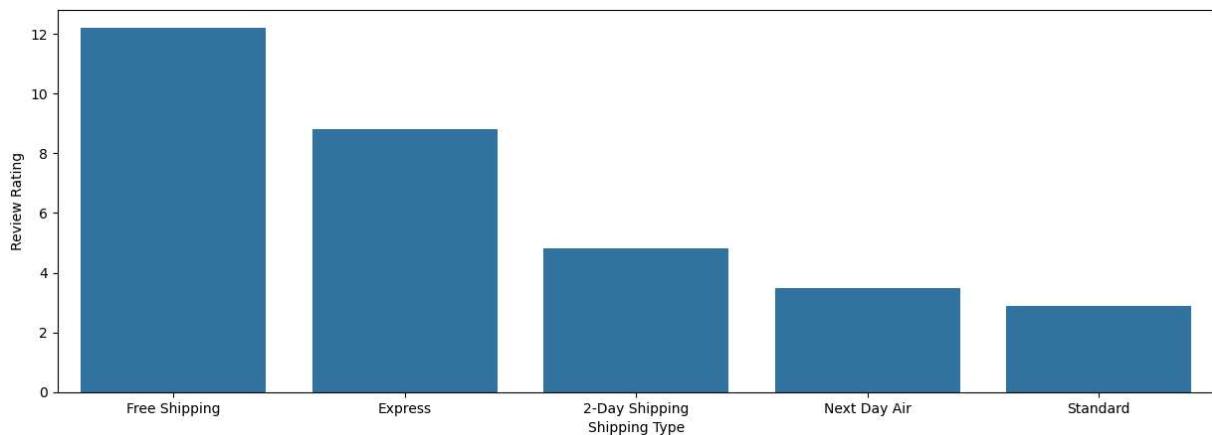
```
sns.barplot(data=sales,x='Season',y='Review Rating')
plt.show()
```



```
In [92]: plt.figure(figsize=(5,5))
sales = df.groupby(['Subscription Status'],as_index=False)[['Review Rating']].sum().s
sns.barplot(data=sales,x='Subscription Status',y='Review Rating')
plt.show()
```

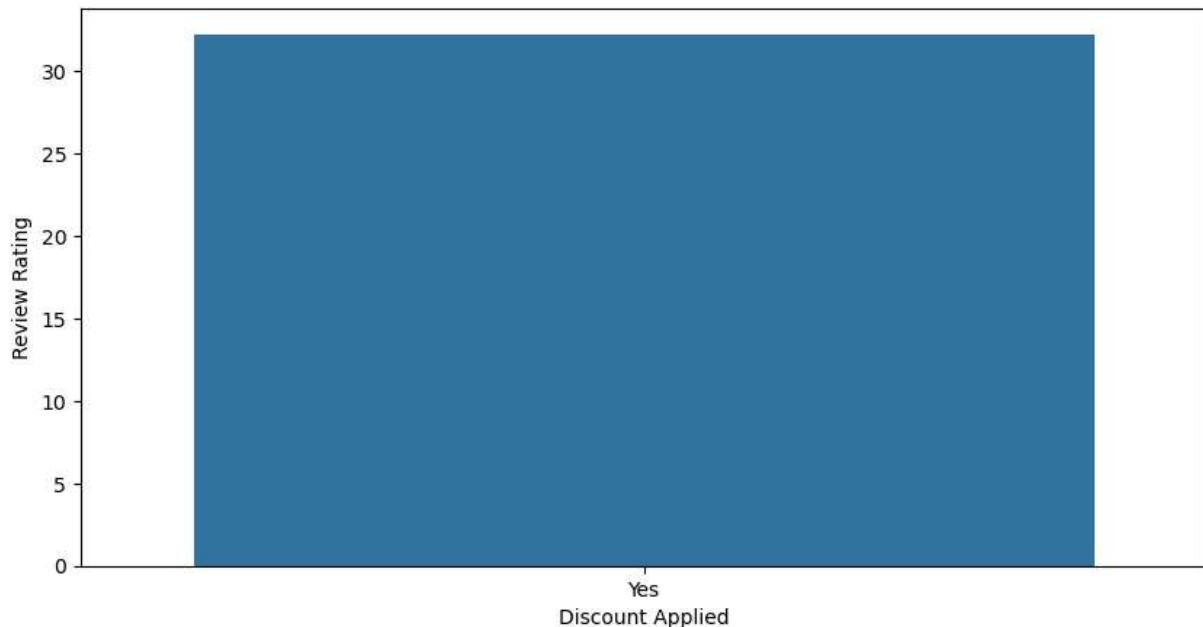


```
In [94]: plt.figure(figsize=(15,5))
sales = df.groupby(['Shipping Type'],as_index=False)[['Review Rating']].sum().sort_values()
sns.barplot(data=sales,x='Shipping Type',y='Review Rating')
plt.show()
```

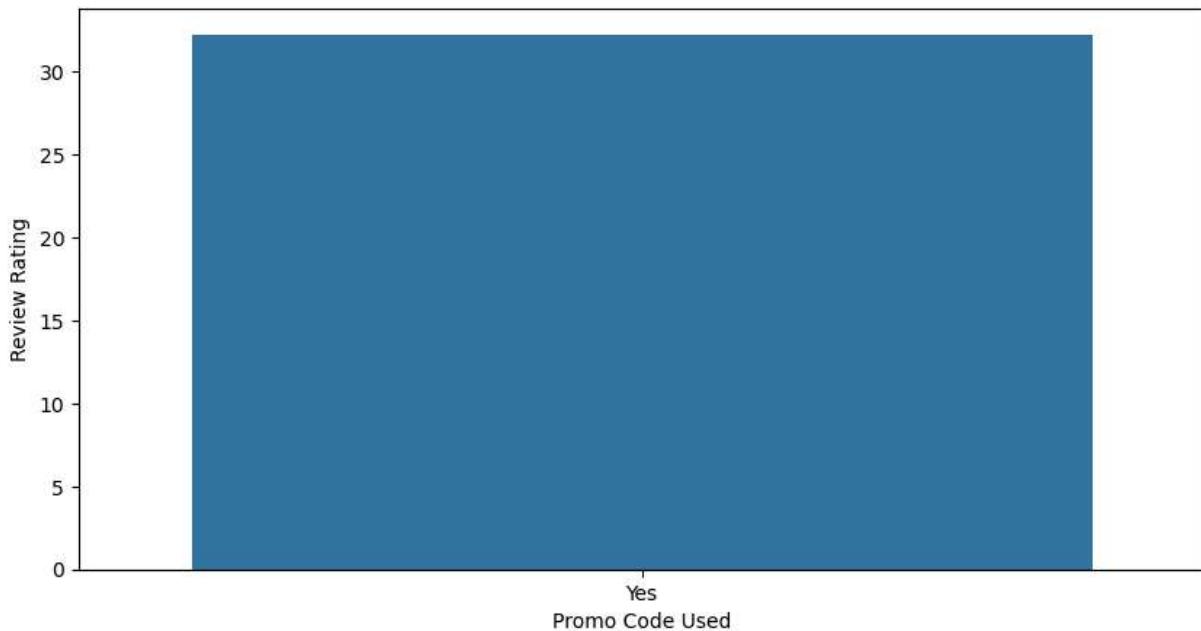


Free shipping buyers have given the most reviews.

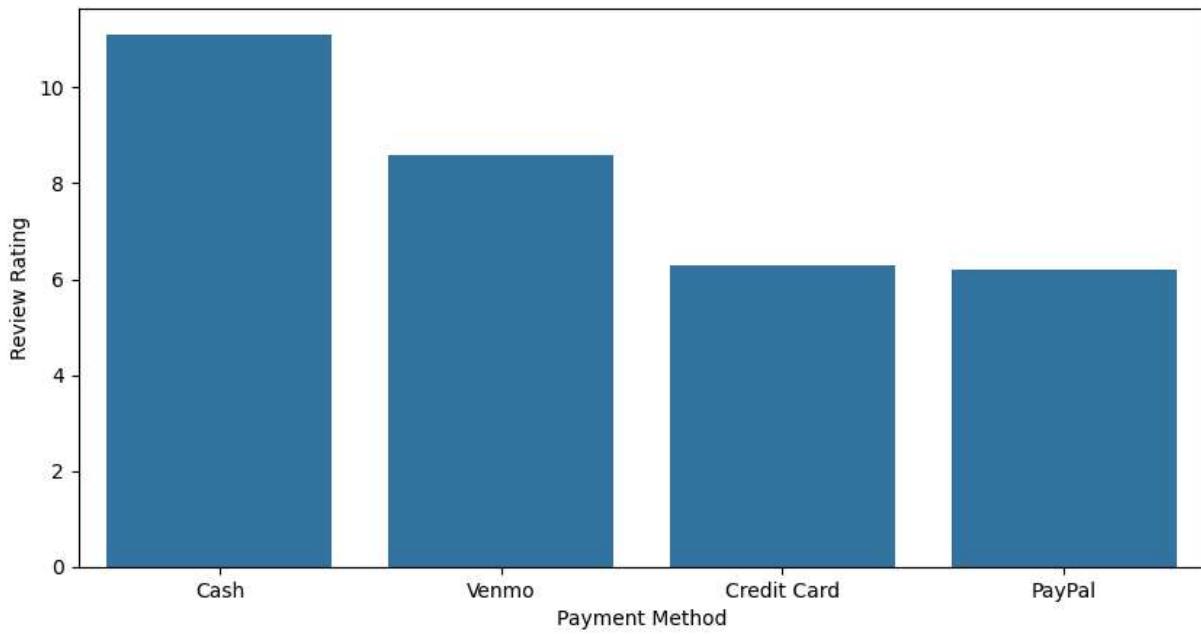
```
In [95]: plt.figure(figsize=(10,5))
sales = df.groupby(['Discount Applied'],as_index=False)[['Review Rating']].sum().sort_values()
sns.barplot(data=sales,x='Discount Applied',y='Review Rating')
plt.show()
```



```
In [96]: plt.figure(figsize=(10,5))
sales = df.groupby(['Promo Code Used'],as_index=False)[['Review Rating']].sum().sort_values()
sns.barplot(data=sales,x='Promo Code Used',y='Review Rating')
plt.show()
```

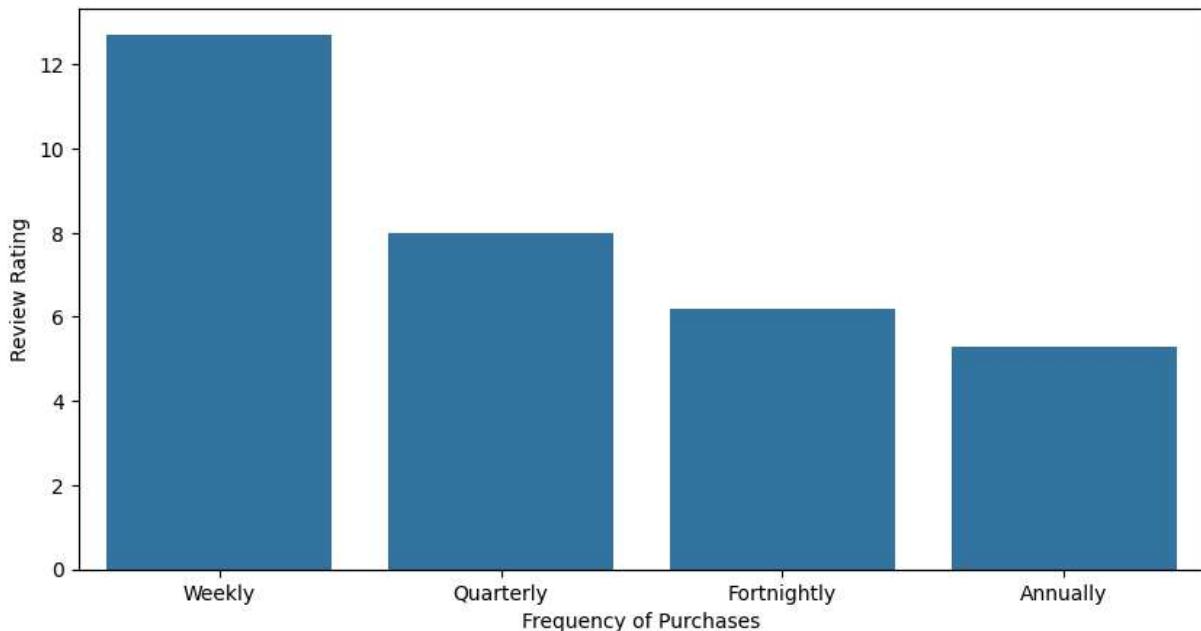


```
In [97]: plt.figure(figsize=(10,5))
sales = df.groupby(['Payment Method'],as_index=False)[['Review Rating']].sum().sort_v
sns.barplot(data=sales,x='Payment Method',y='Review Rating')
plt.show()
```

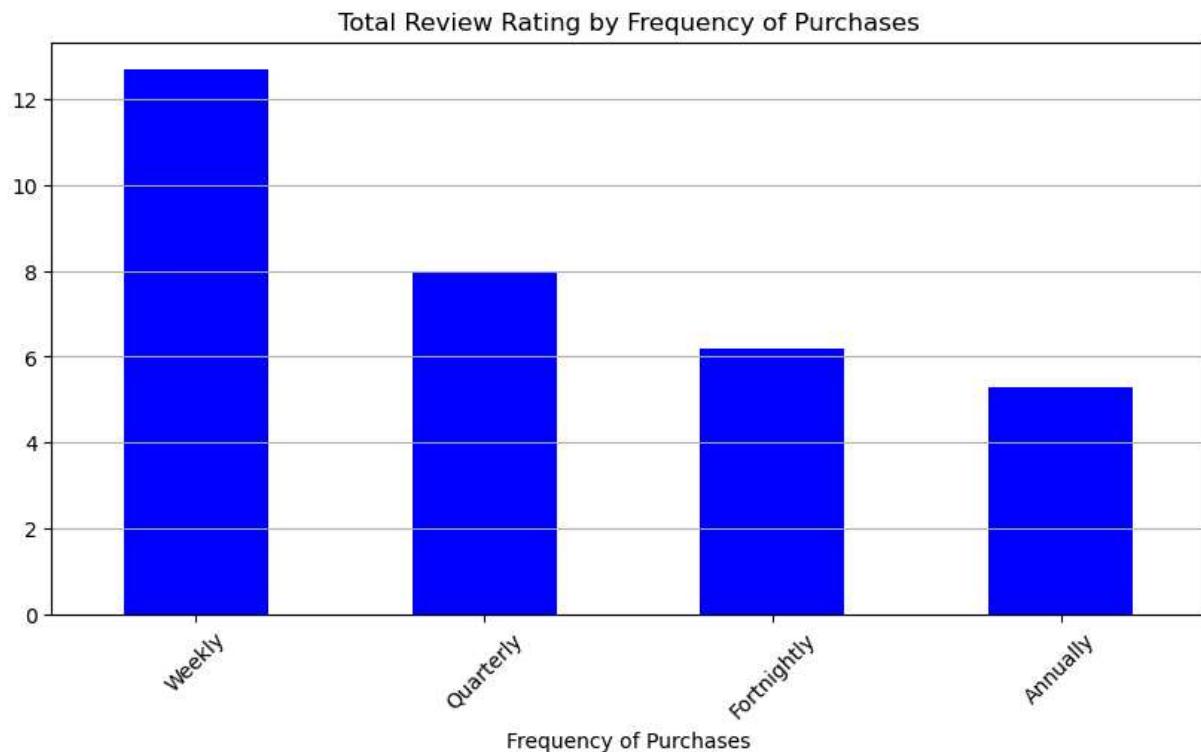


PayPal users have given the most reviews.

```
In [59]: plt.figure(figsize=(10,5))
sales = df.groupby(['Frequency of Purchases'],as_index=False)[['Review Rating']].sum()
sns.barplot(data=sales,x='Frequency of Purchases',y='Review Rating')
plt.show()
```



```
In [60]: sales = df.groupby('Frequency of Purchases')['Review Rating'].sum().sort_values(ascending=True)
sales.plot(kind='bar', color='blue', figsize=(10, 5))
plt.title('Total Review Rating by Frequency of Purchases')
plt.xticks(rotation=45)
plt.grid(axis='y')
plt.show()
```



*=====



Thank You

In []: