

# Ramesh Kumar Prajapati



## E-commerce Data Analysis



### Insights & Trends in Online Shopping

In [170...]

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [171...]

```
df=pd.read_csv("C://Users//ALWAYSRAJESH//Documents//customer_details.csv")
df
```

Out[171...]

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size	Color
0	1	55	Male	Blouse	Clothing	53	Kentucky	L	Black
1	2	19	Male	Sweater	Clothing	64	Maine	L	Maroon
2	3	50	Male	Jeans	Clothing	73	Massachusetts	S	Blue
3	4	21	Male	Sandals	Footwear	90	Rhode Island	M	Maroon
4	5	45	Male	Blouse	Clothing	49	Oregon	M	Turquoise
...	...	...	...	...	...	...	...	...	...
3895	3896	40	Female	Hoodie	Clothing	28	Virginia	L	Turquoise
3896	3897	52	Female	Backpack	Accessories	49	Iowa	L	Yellow
3897	3898	46	Female	Belt	Accessories	33	New Jersey	L	Cream
3898	3899	44	Female	Shoes	Footwear	77	Minnesota	S	Black
3899	3900	52	Female	Handbag	Accessories	81	California	M	Grey

3900 rows × 18 columns



In [172...]

```
df.head()
```

Out[172...]

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size	Color
0	1	55	Male	Blouse	Clothing	53	Kentucky	L	Gray
1	2	19	Male	Sweater	Clothing	64	Maine	L	Maroon
2	3	50	Male	Jeans	Clothing	73	Massachusetts	S	Maroon
3	4	21	Male	Sandals	Footwear	90	Rhode Island	M	Maroon
4	5	45	Male	Blouse	Clothing	49	Oregon	M	Turquoise



In [173...]

df.shape

Out[173...]

(3900, 18)

In [174...]

df.info()

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3900 entries, 0 to 3899
Data columns (total 18 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Customer ID      3900 non-null   int64  
 1   Age               3900 non-null   int64  
 2   Gender            3900 non-null   object  
 3   Item Purchased   3900 non-null   object  
 4   Category          3900 non-null   object  
 5   Purchase Amount (USD) 3900 non-null   int64  
 6   Location          3900 non-null   object  
 7   Size              3900 non-null   object  
 8   Color              3900 non-null   object  
 9   Season             3900 non-null   object  
 10  Review Rating    3900 non-null   float64 
 11  Subscription Status 3900 non-null   object  
 12  Shipping Type    3900 non-null   object  
 13  Discount Applied 3900 non-null   object  
 14  Promo Code Used  3900 non-null   object  
 15  Previous Purchases 3900 non-null   int64  
 16  Payment Method   3900 non-null   object  
 17  Frequency of Purchases 3900 non-null   object  
dtypes: float64(1), int64(4), object(13)
memory usage: 548.6+ KB

```

In [175...]

df.tail(1)

Out[175...]

Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size	Color	Season
3899	3900	52	Female	Handbag	Accessories	81	California	M	Beige



In [176...]

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

Out[176...]

```
Customer ID          4
Age                  21
Gender               Male
Item Purchased      Sandals
Category             Footwear
Purchase Amount (USD) 90
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 [177...]

```
df.describe()
```

Out[177...]

	Customer ID	Age	Purchase Amount (USD)	Review Rating	Previous Purchases
<b>count</b>	3900.000000	3900.000000	3900.000000	3900.000000	3900.000000
<b>mean</b>	1950.500000	44.068462	59.764359	3.749949	25.351538
<b>std</b>	1125.977353	15.207589	23.685392	0.716223	14.447125
<b>min</b>	1.000000	18.000000	20.000000	2.500000	1.000000
<b>25%</b>	975.750000	31.000000	39.000000	3.100000	13.000000
<b>50%</b>	1950.500000	44.000000	60.000000	3.700000	25.000000
<b>75%</b>	2925.250000	57.000000	81.000000	4.400000	38.000000
<b>max</b>	3900.000000	70.000000	100.000000	5.000000	50.000000

In [178...]

```
df.columns
```

```
Out[178]: 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 [179]: df.isnull().sum()
```

```
Out[179]: Customer ID      0  
Age            0  
Gender         0  
Item Purchased 0  
Category        0  
Purchase Amount (USD) 0  
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
```

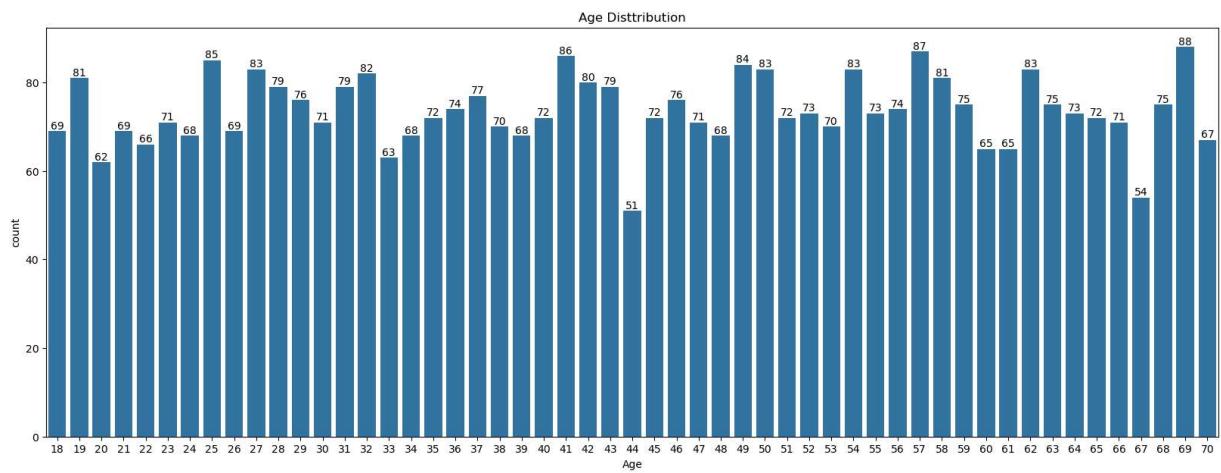
```
In [180]: df['Purchase Amount (USD)'] = df['Purchase Amount (USD)'].fillna(0)
```

```
In [ ]:
```

**No any null value.**

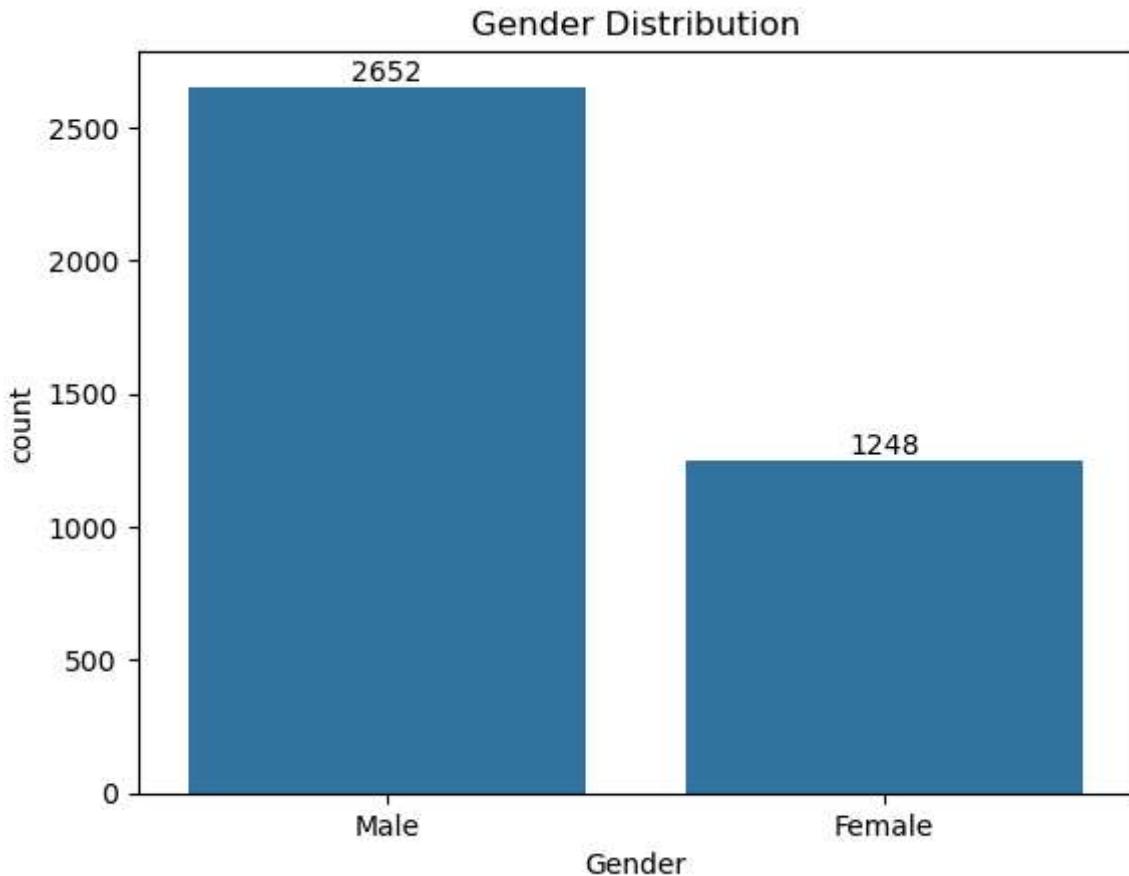
```
In [181]: ## EDA
```

```
In [182]: 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**

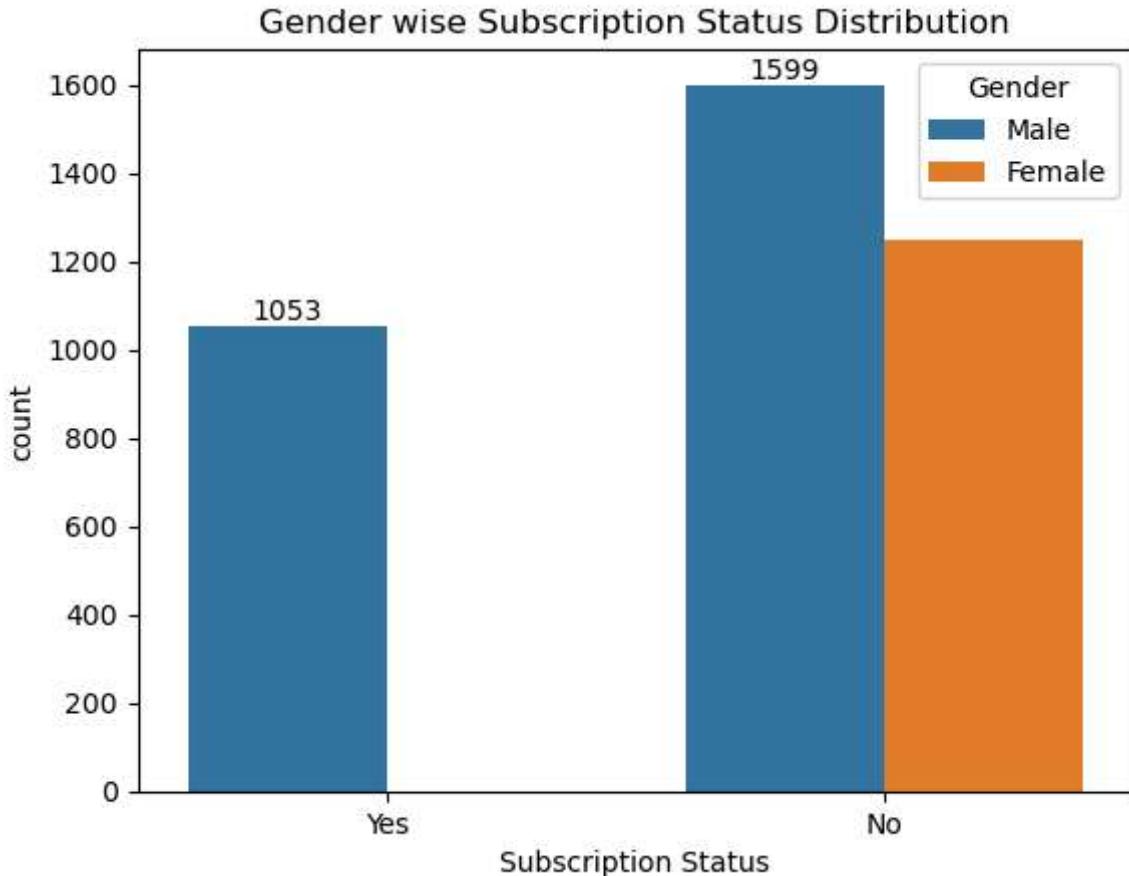
```
In [183...]: ax = sns.countplot(x='Gender', data=df)
ax.bar_label(ax.containers[0])
plt.title('Gender Distribution')
plt.show()
```



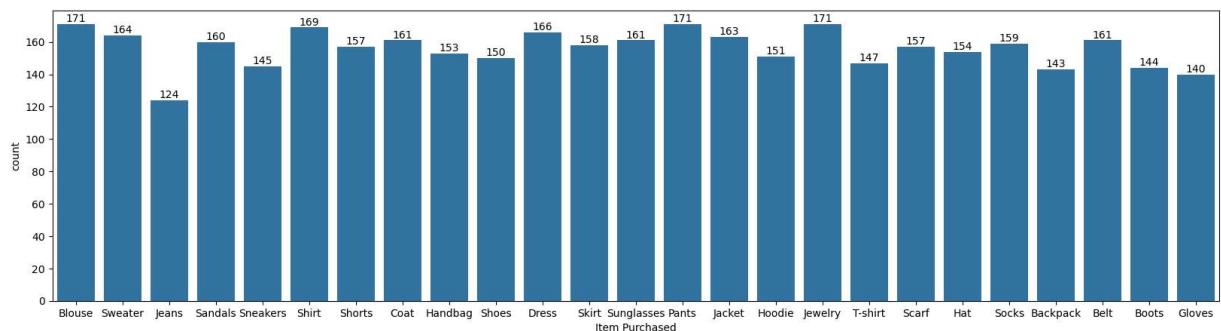
**The number of males are more than the females**

```
In [184...]: #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()
```

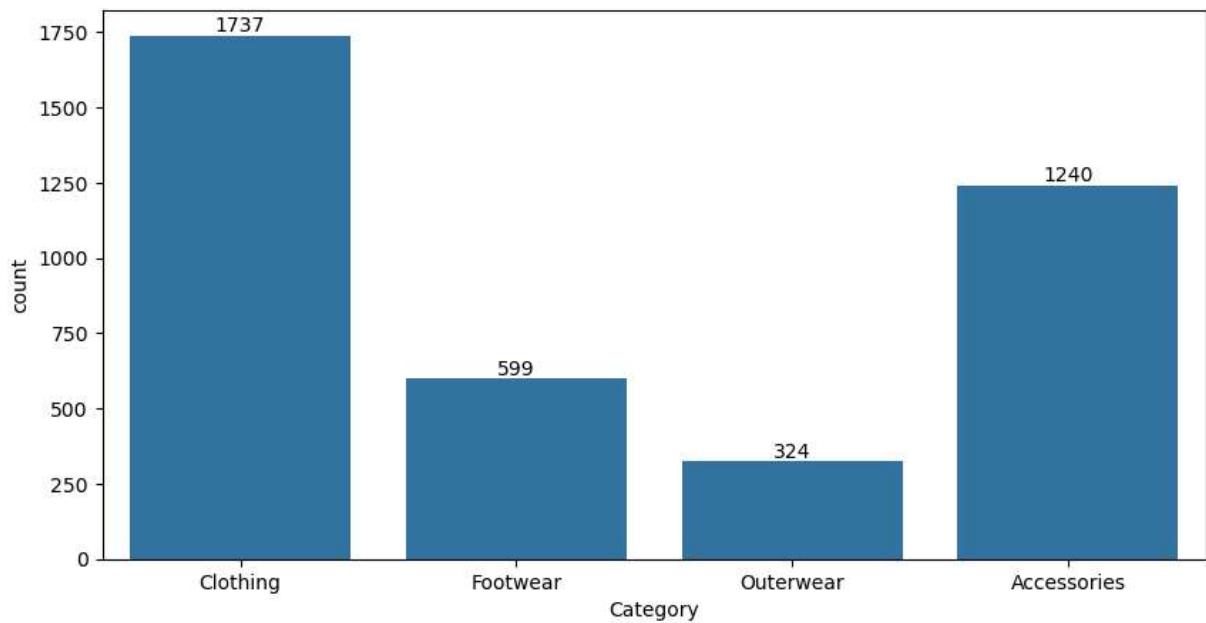


```
In [185...]
plt.figure(figsize=(20,5))
ax = sns.countplot(x='Item Purchased', data=df)
ax.bar_label(ax.containers[0])
plt.show()
```



**Most sales are of Blouse, Pants, and Jewelry.**

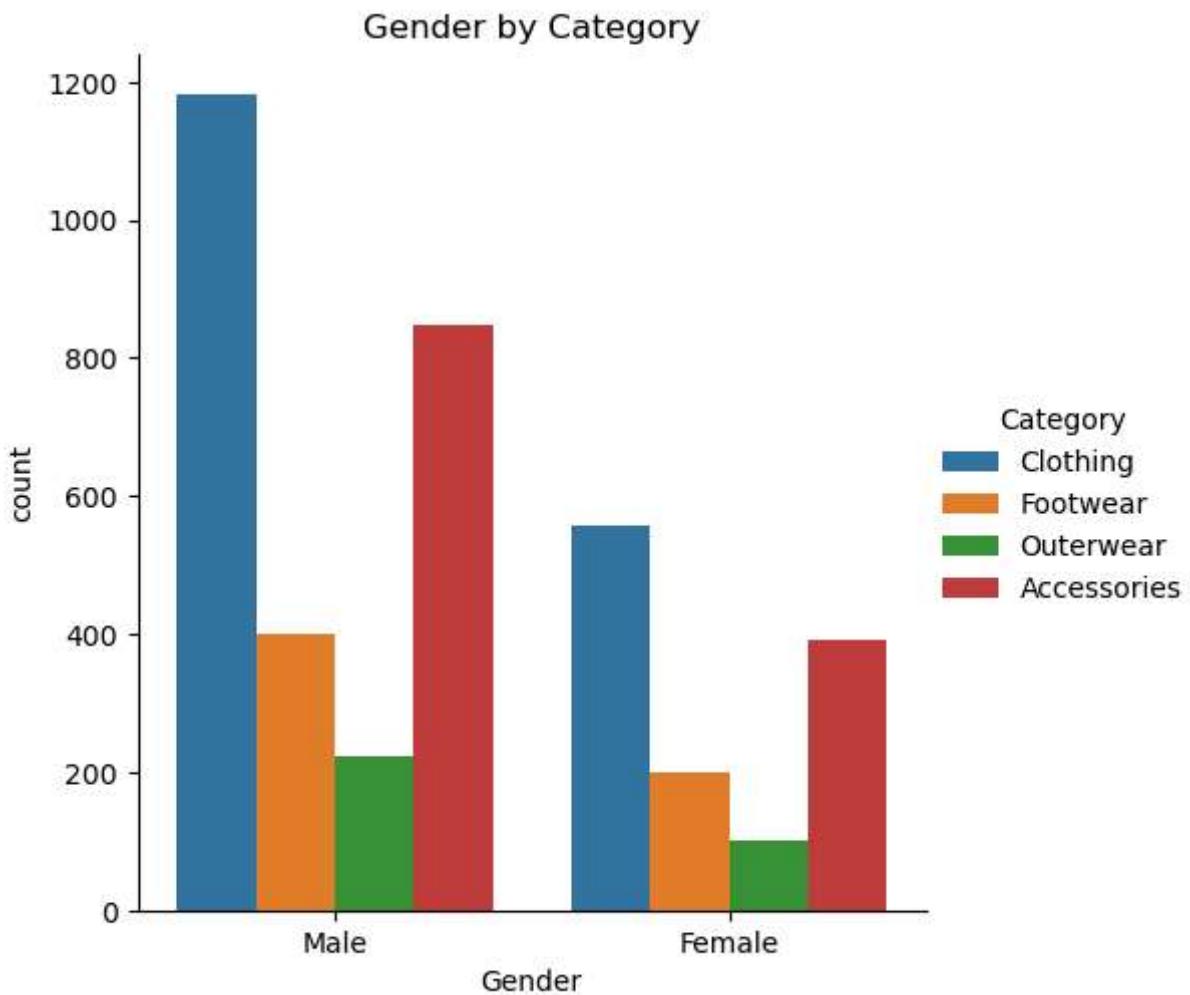
```
In [186...]
plt.figure(figsize=(10,5))
ax = sns.countplot(x='Category', data=df)
ax.bar_label(ax.containers[0])
plt.show()
```



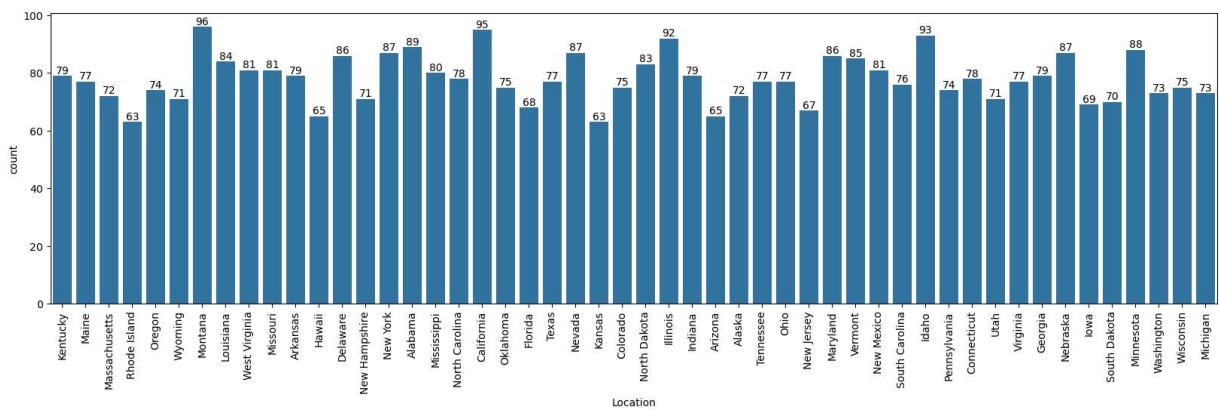
**Clothing category is most among all others.**

In [187...]

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



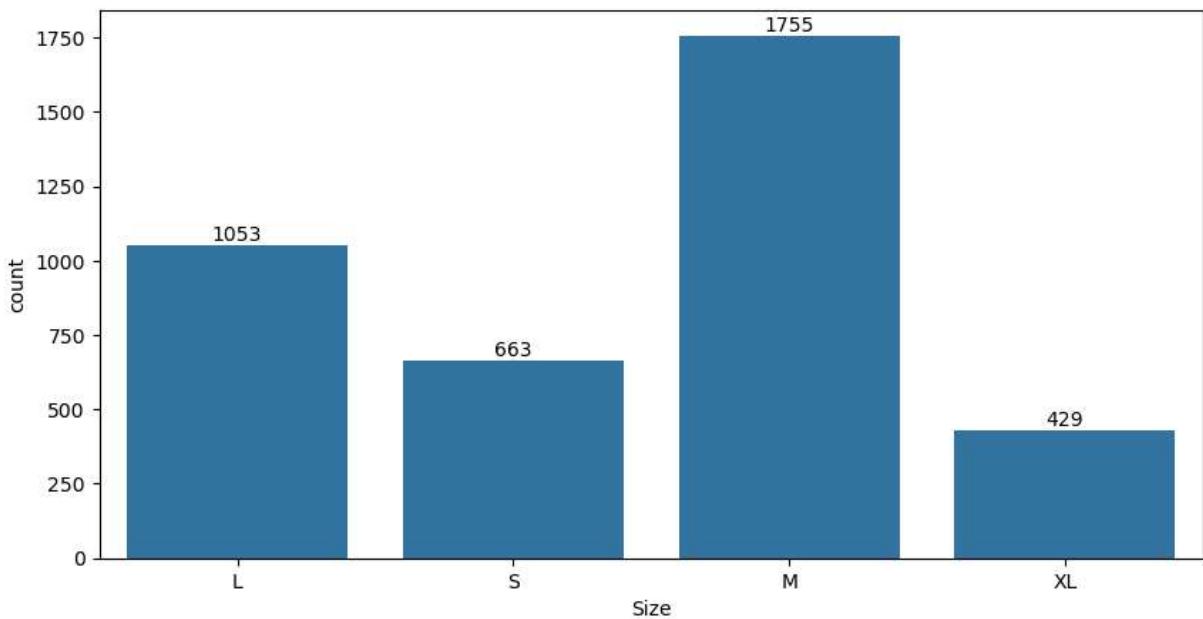
```
In [188]: plt.figure(figsize=(20,5))
ax = sns.countplot(x='Location', data=df)
plt.xticks(rotation=90)
ax.bar_label(ax.containers[0])
plt.show()
```



**Most buyers are from Montana, California, and Idaho.**

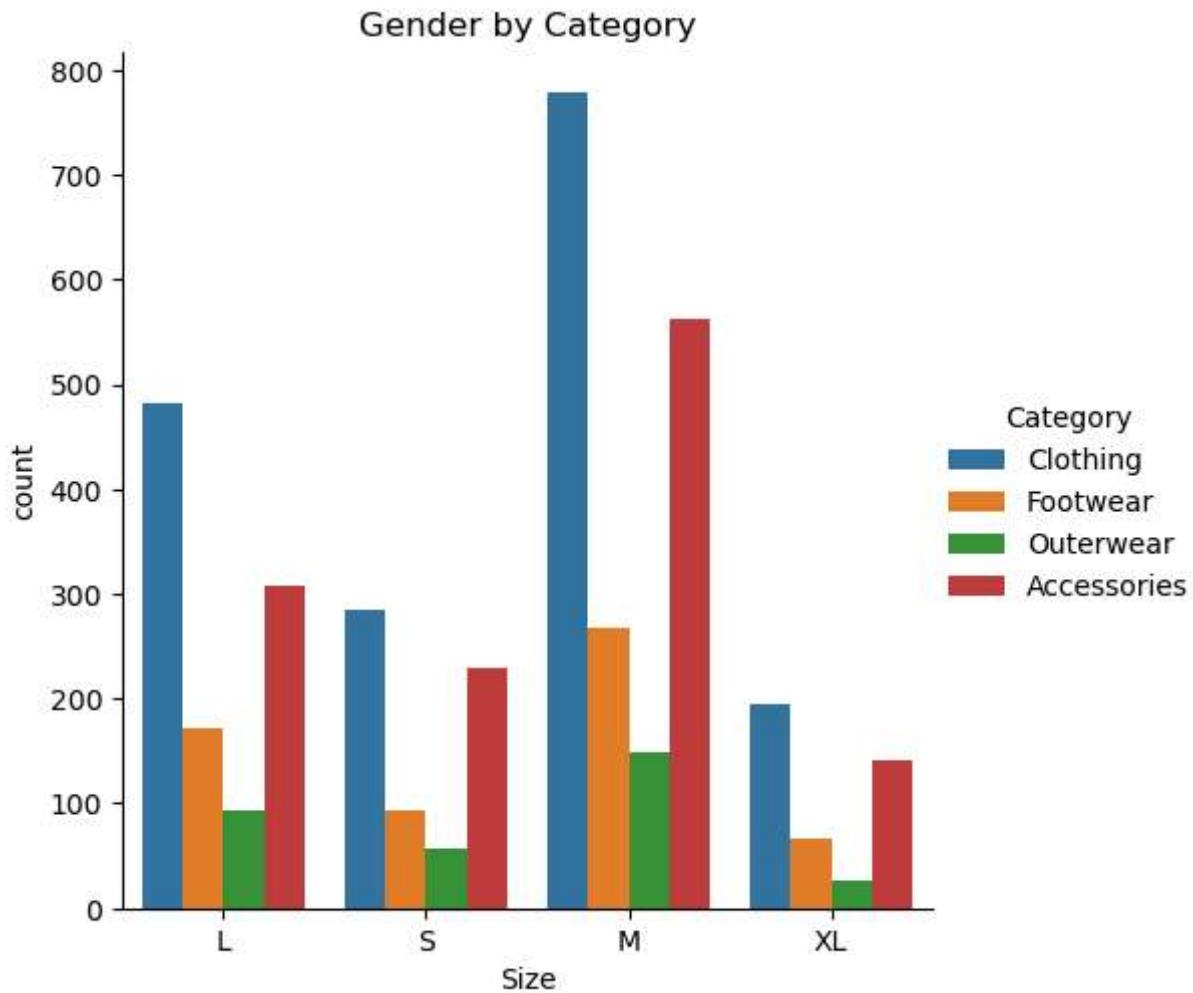
```
In [ ]:
```

```
In [189...  
plt.figure(figsize=(10,5))  
ax = sns.countplot(x='Size', data=df)  
ax.bar_label(ax.containers[0])  
plt.show()
```



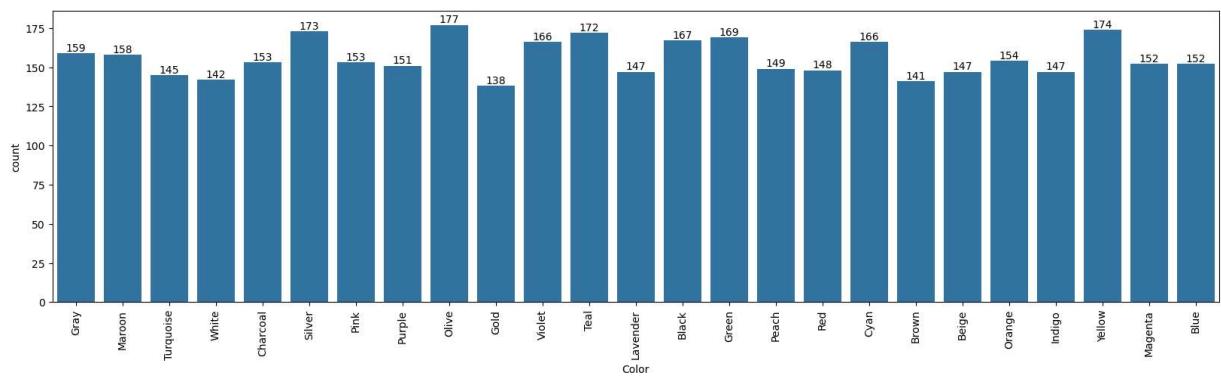
**Medium size products have the highest range of sell.**

```
In [190...  
#2  
sns.catplot(hue='Category', kind='count', x='Size', data=df)  
plt.title('Gender by Category')  
plt.show()
```



In [191...]

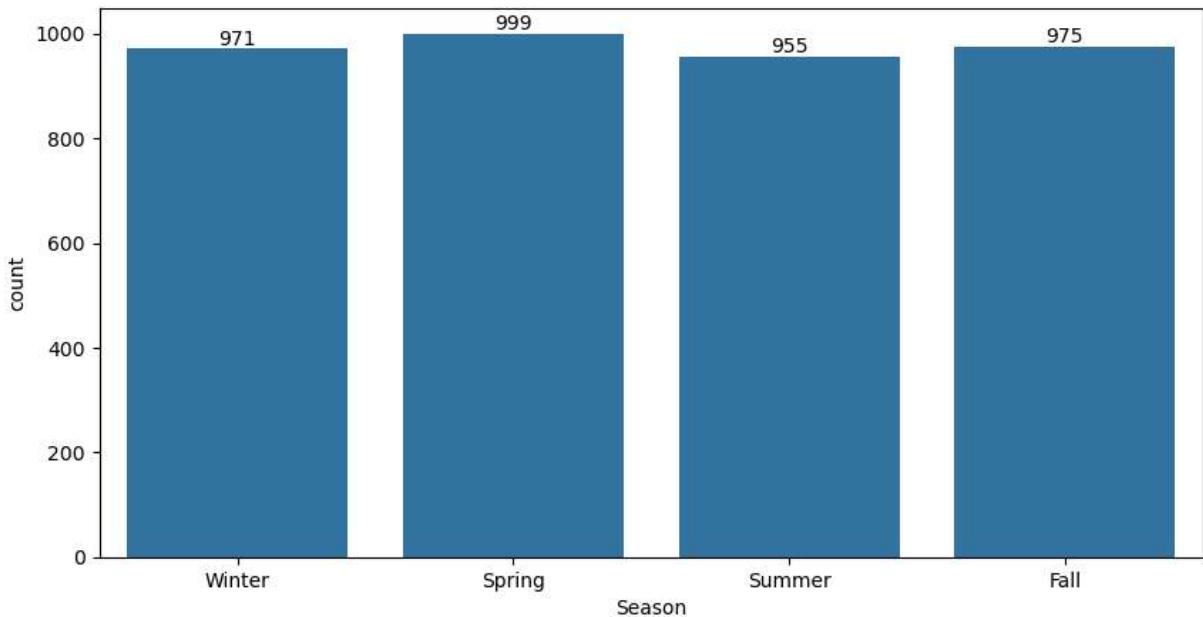
```
plt.figure(figsize=(20,5))
ax = sns.countplot(x='Color', data=df)
plt.xticks(rotation=90)
ax.bar_label(ax.containers[0])
plt.show()
```



**Buyers bought Olive, Yellow, and Silver colors most.**

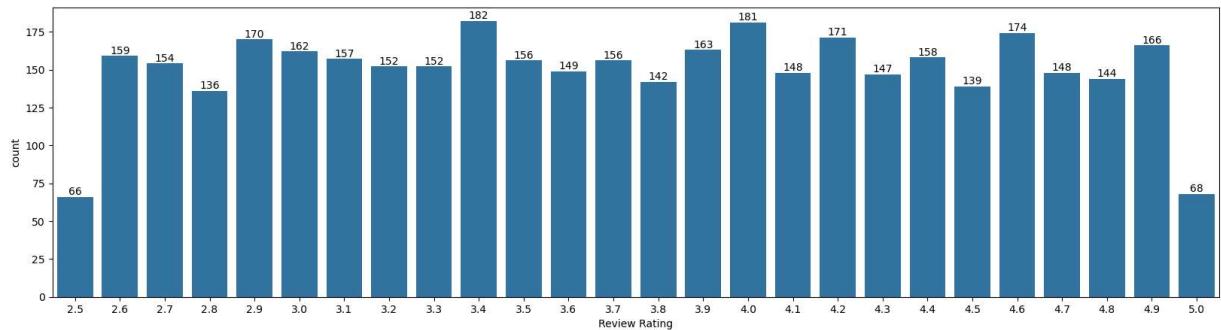
In [192...]

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



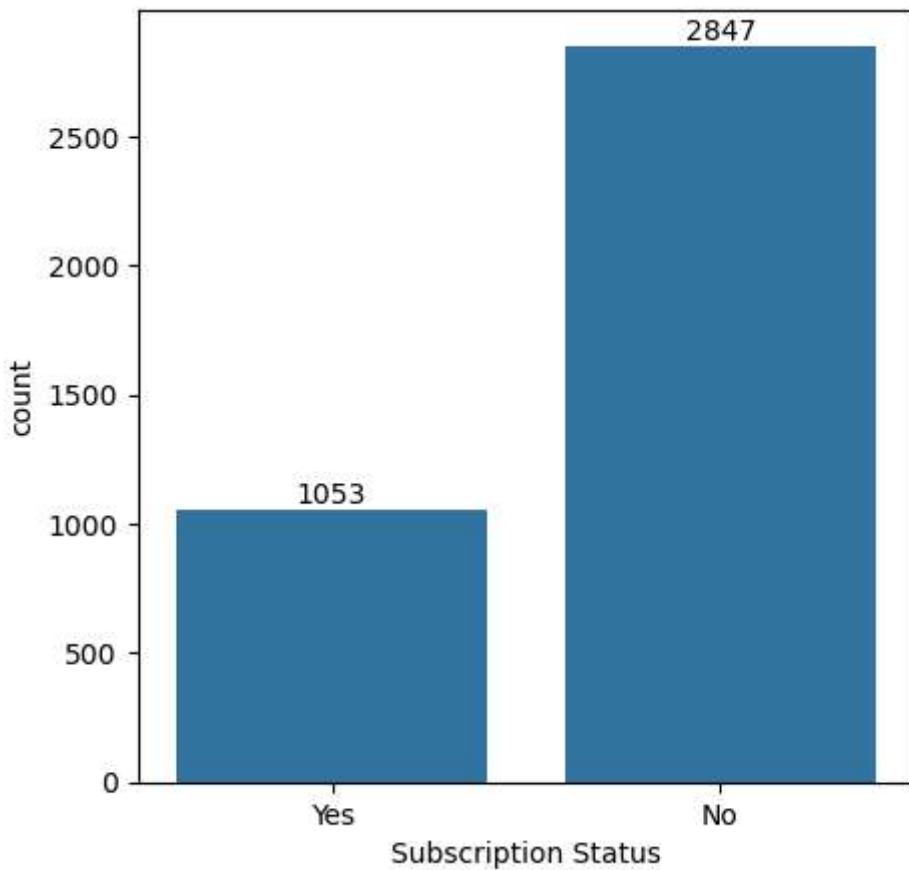
**In Spring, Fall, Winter and then Summer seasons the sell out is most.**

```
In [193...]: plt.figure(figsize=(20,5))
ax = sns.countplot(x='Review Rating', data=df)
ax.bar_label(ax.containers[0])
plt.show()
```



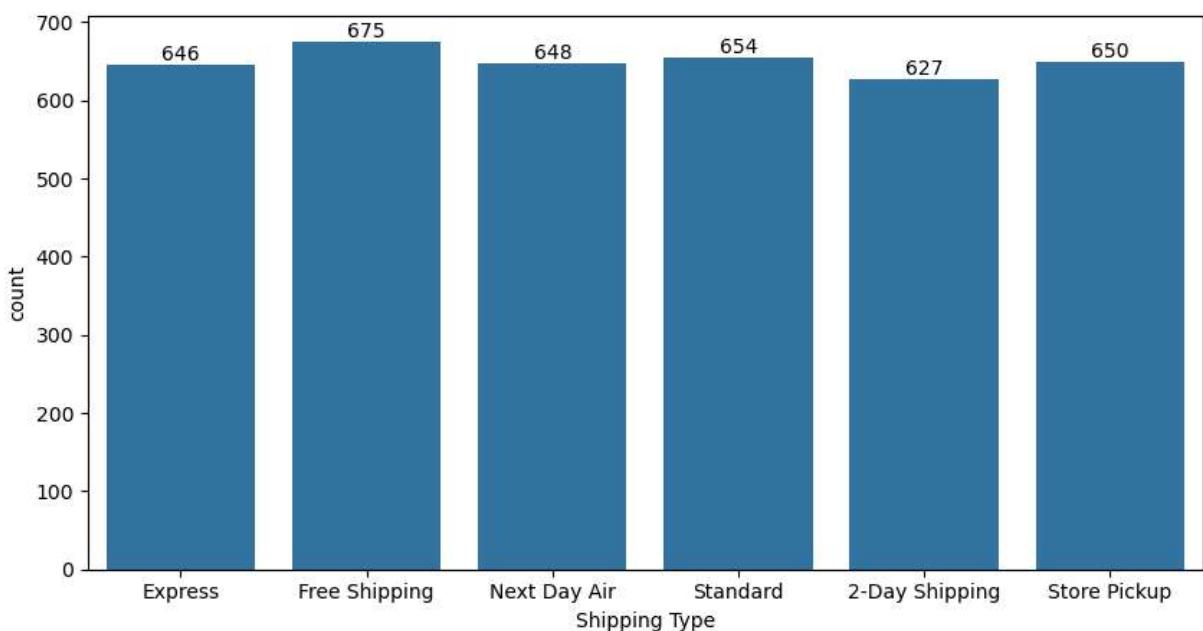
**The rating of 3.4, 4.0, and 4.6 were given.**

```
In [194...]: plt.figure(figsize=(5,5))
ax = sns.countplot(x='Subscription Status', data=df)
ax.bar_label(ax.containers[0])
plt.show()
```



In [195...]

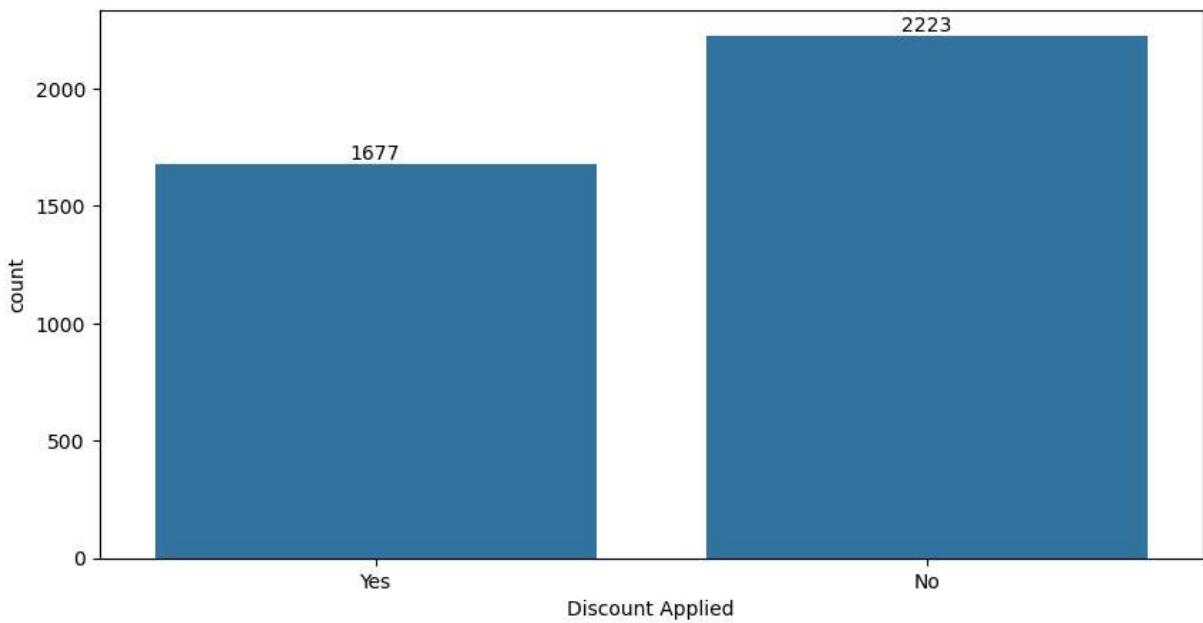
```
plt.figure(figsize=(10,5))
ax = sns.countplot(x='Shipping Type', data=df)
ax.bar_label(ax.containers[0])
plt.show()
```



**Free shipping has the largest amount of buyers.**

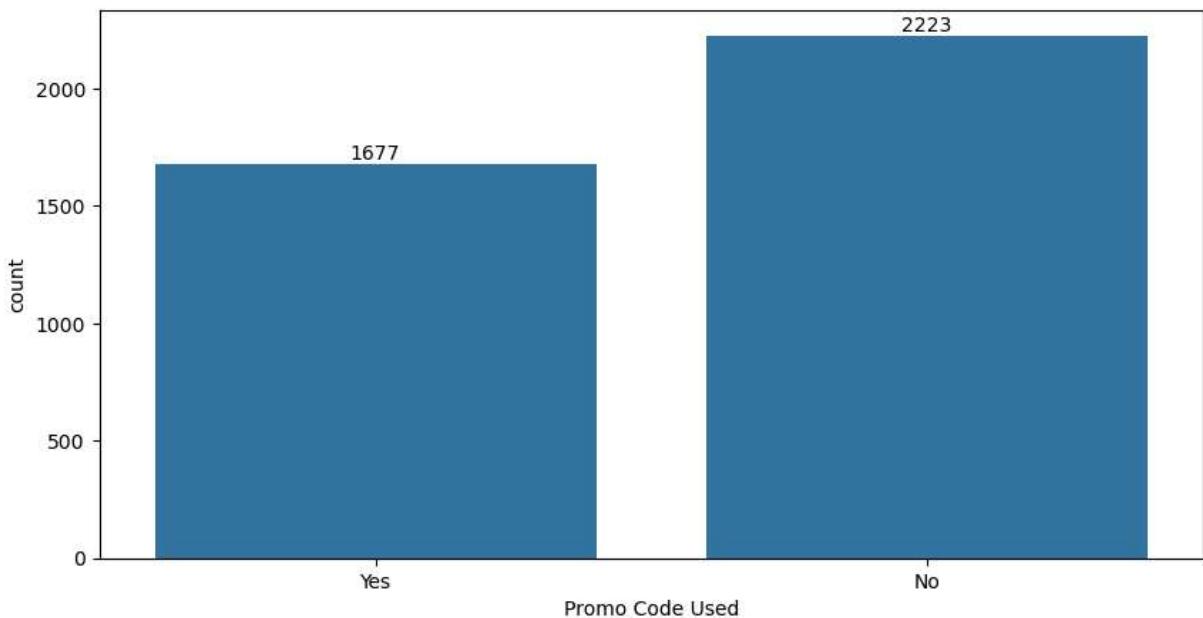
In [196...]

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



In [197...]

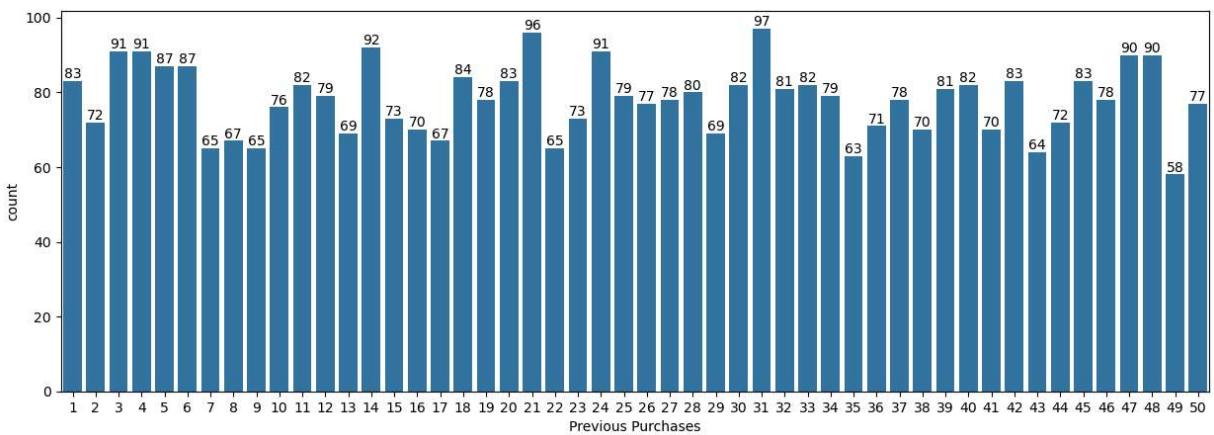
```
plt.figure(figsize=(10,5))
ax = sns.countplot(x='Promo Code Used', data=df)
ax.bar_label(ax.containers[0])
plt.show()
```



**2223 buyers have not used Promo Code.**

In [198...]

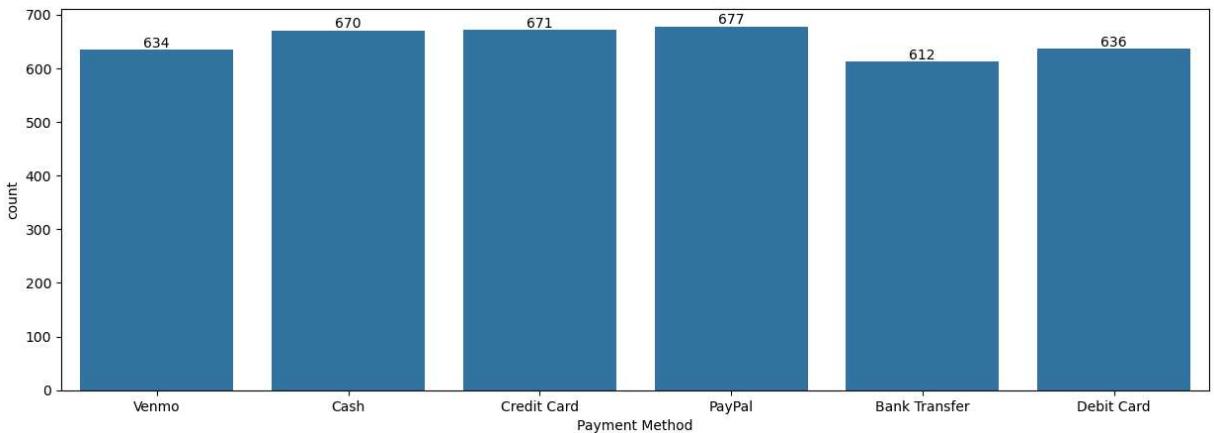
```
plt.figure(figsize=(15,5))
ax = sns.countplot(x='Previous Purchases', data=df)
ax.bar_label(ax.containers[0])
plt.show()
```



**97 buyers have 31 previous purchases.**

In [199]:

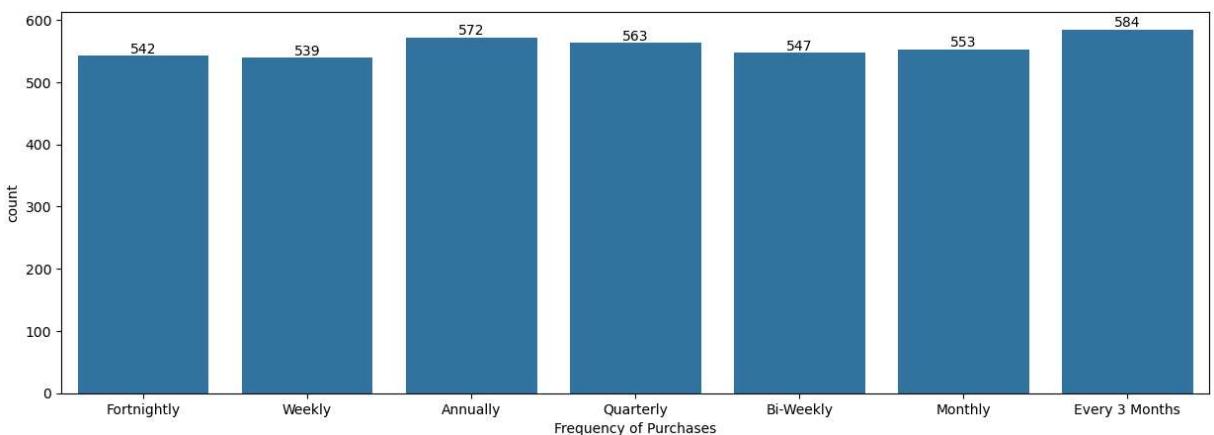
```
plt.figure(figsize=(15,5))
ax = sns.countplot(x='Payment Method', data=df)
ax.bar_label(ax.containers[0])
plt.show()
```



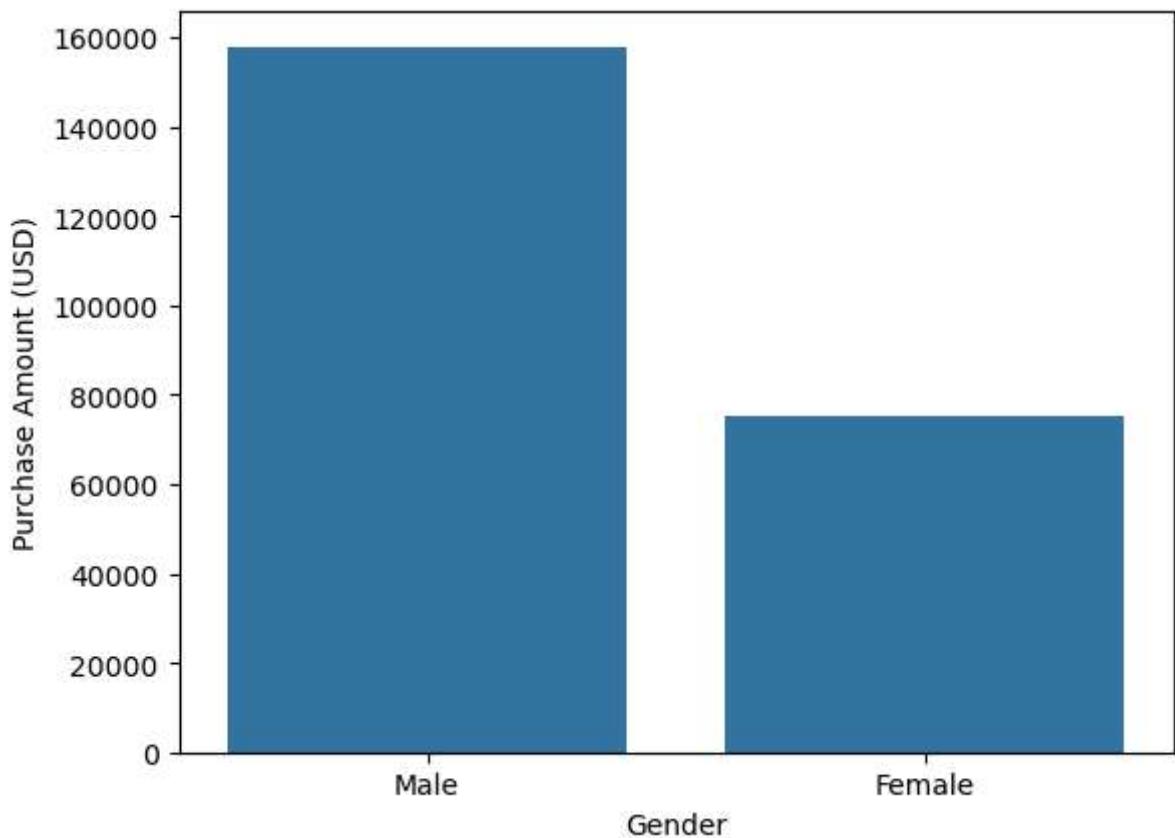
**677 buyers used PayPal, 671 used Credit Card, and 670 paid Cash.**

In [200]:

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



```
In [201... 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()
```

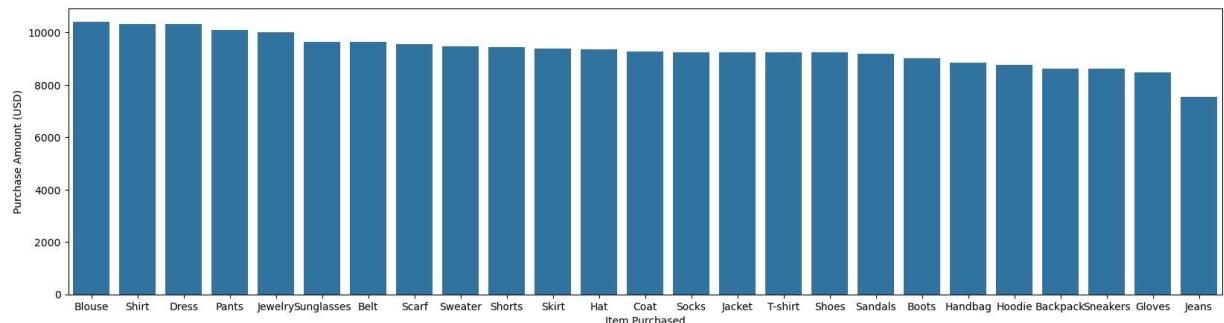


```
In [202... df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3900 entries, 0 to 3899
Data columns (total 18 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Customer ID      3900 non-null   int64  
 1   Age               3900 non-null   int64  
 2   Gender            3900 non-null   object  
 3   Item Purchased   3900 non-null   object  
 4   Category          3900 non-null   object  
 5   Purchase Amount (USD) 3900 non-null   int64  
 6   Location          3900 non-null   object  
 7   Size               3900 non-null   object  
 8   Color              3900 non-null   object  
 9   Season             3900 non-null   object  
 10  Review Rating    3900 non-null   float64 
 11  Subscription Status 3900 non-null   object  
 12  Shipping Type    3900 non-null   object  
 13  Discount Applied 3900 non-null   object  
 14  Promo Code Used  3900 non-null   object  
 15  Previous Purchases 3900 non-null   int64  
 16  Payment Method   3900 non-null   object  
 17  Frequency of Purchases 3900 non-null   object  
dtypes: float64(1), int64(4), object(13)
memory usage: 548.6+ KB
```

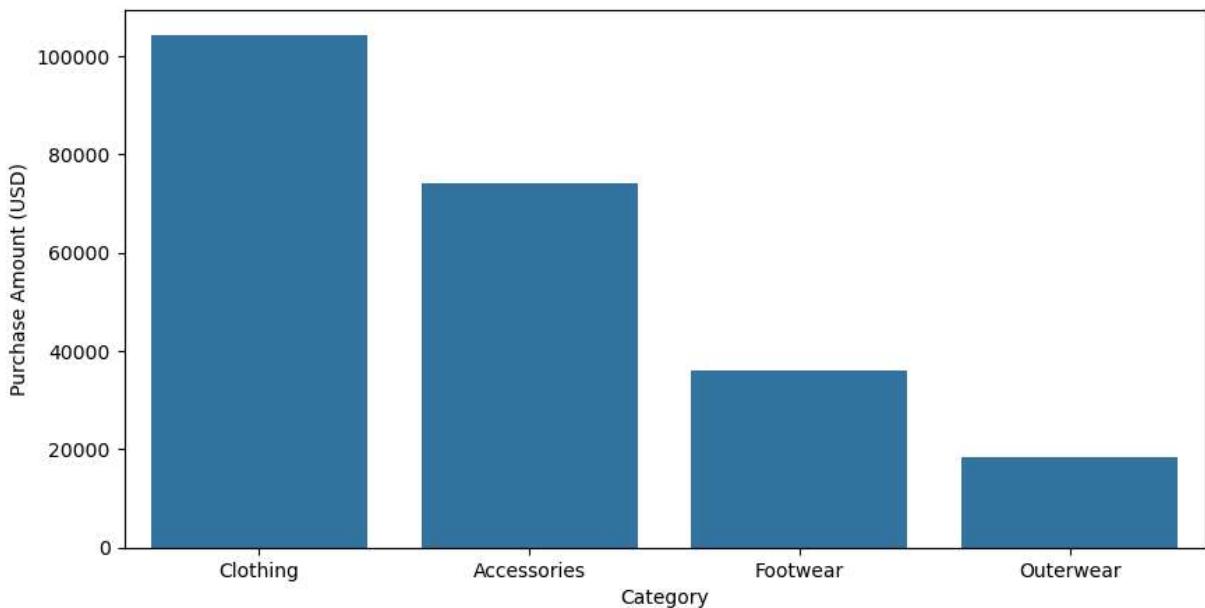
In [203...]

```
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 [204...]

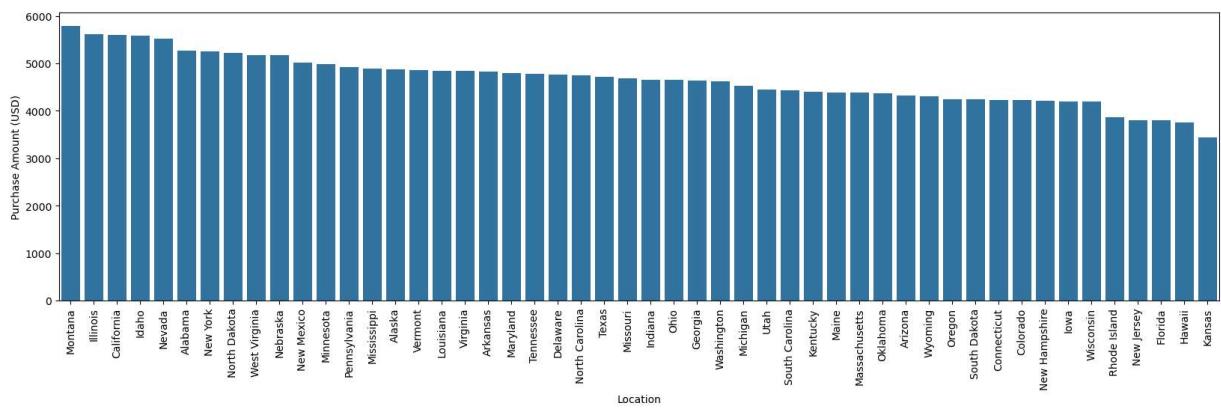
```
plt.figure(figsize=(10,5))
sales = df.groupby(['Category'],as_index=False)[['Purchase Amount (USD)']].sum().sort_values('Purchase Amount (USD)', ascending=False)
sns.barplot(data=sales,x='Category',y='Purchase Amount (USD)')
plt.show()
```



**Clothing category has the most Purchased amount.**

In [205...]

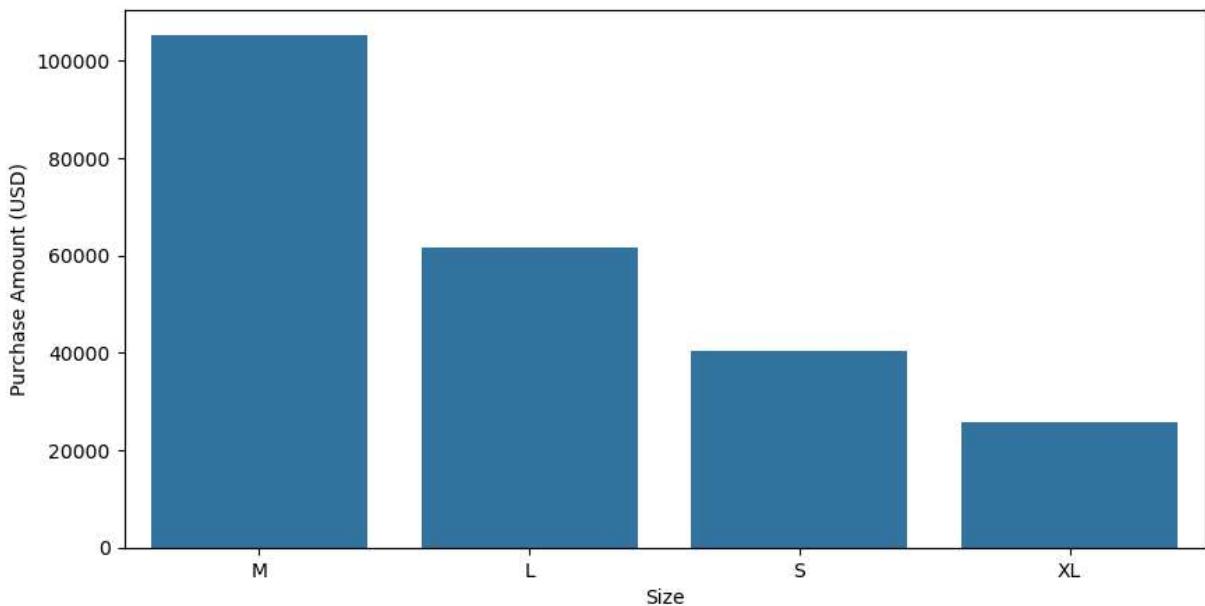
```
plt.figure(figsize=(20,5))
sales = df.groupby(['Location'],as_index=False)[ 'Purchase Amount (USD)'].sum().sort_values(ascending=False)
sns.barplot(data=sales,x='Location',y='Purchase Amount (USD)')
plt.xticks(rotation=90)
plt.show()
```



**Montana has produced the most amount.**

In [206...]

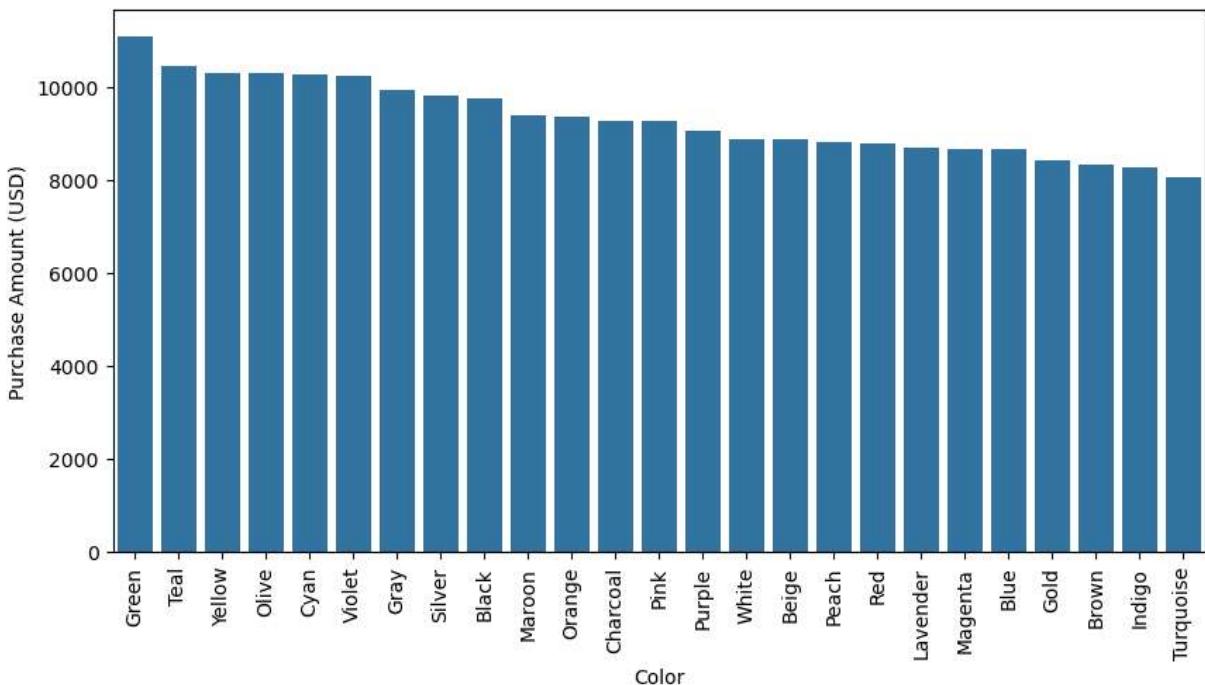
```
plt.figure(figsize=(10,5))
sales = df.groupby(['Size'],as_index=False)[ 'Purchase Amount (USD)'].sum().sort_values(ascending=False)
sns.barplot(data=sales,x='Size',y='Purchase Amount (USD)')
plt.show()
```



**Medium size has generated the most profit.**

In [207...]

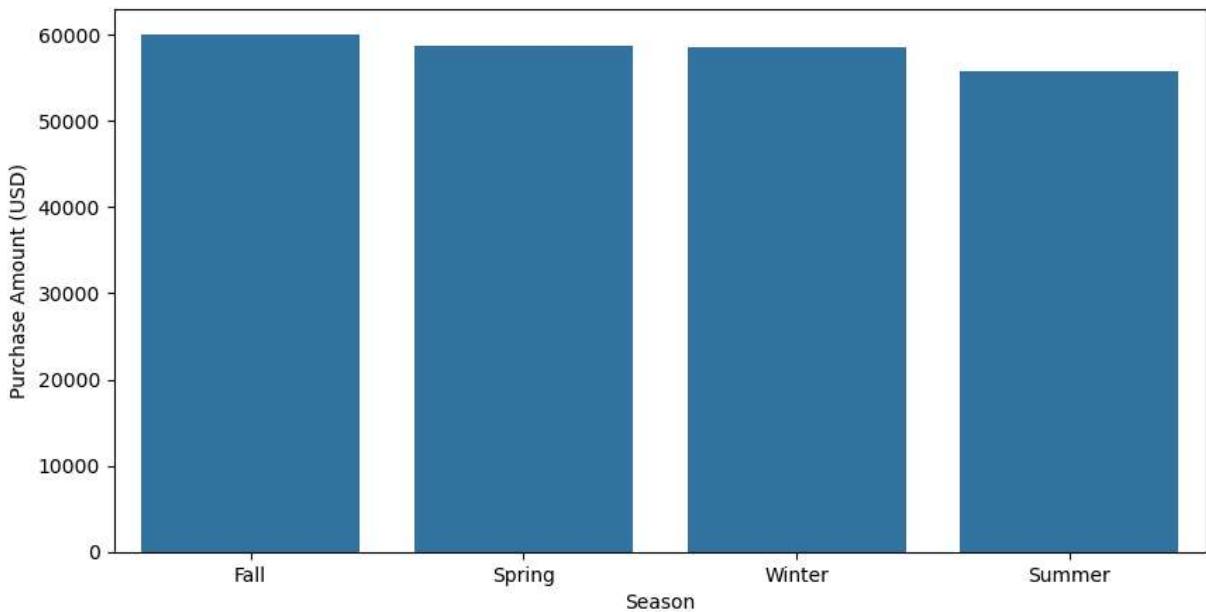
```
plt.figure(figsize=(10,5))
sales = df.groupby(['Color'],as_index=False)[['Purchase Amount (USD)']].sum().sort_values(by='Purchase Amount (USD)', ascending=False)
sns.barplot(data=sales,x='Color',y='Purchase Amount (USD)')
plt.xticks(rotation=90)
plt.show()
```



**Buyers bought products of green color.**

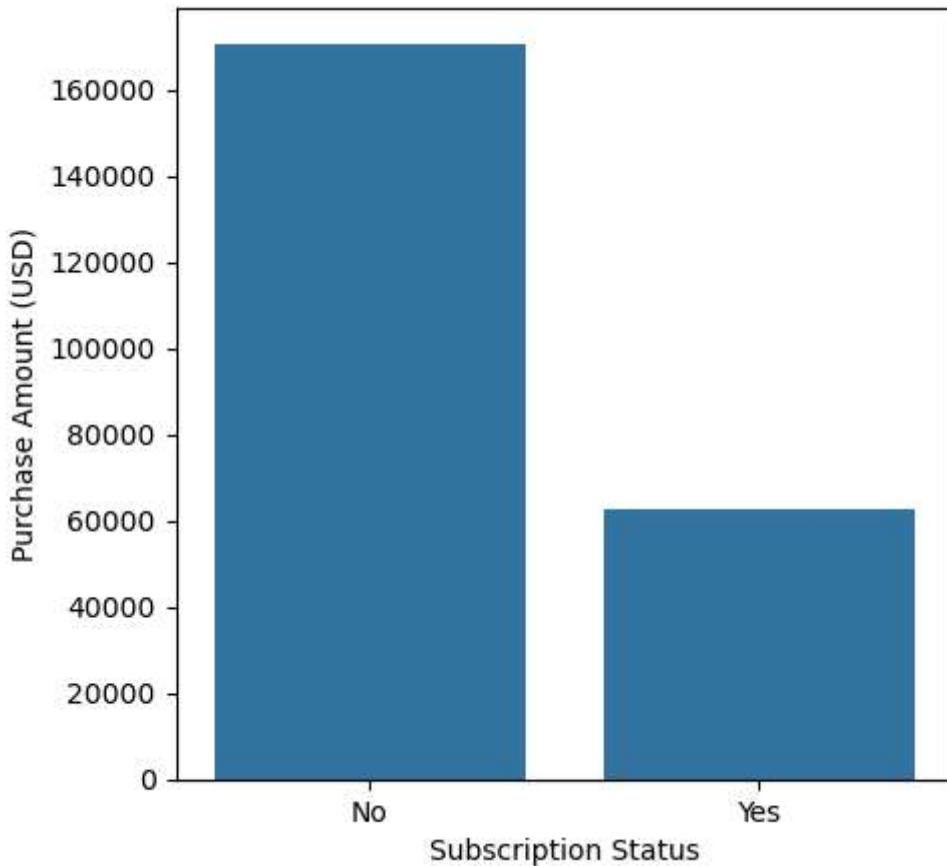
In [208...]

```
plt.figure(figsize=(10,5))
sales = df.groupby(['Season'],as_index=False)[['Purchase Amount (USD)']].sum().sort_values(by='Purchase Amount (USD)', ascending=False)
sns.barplot(data=sales,x='Season',y='Purchase Amount (USD)')
plt.show()
```



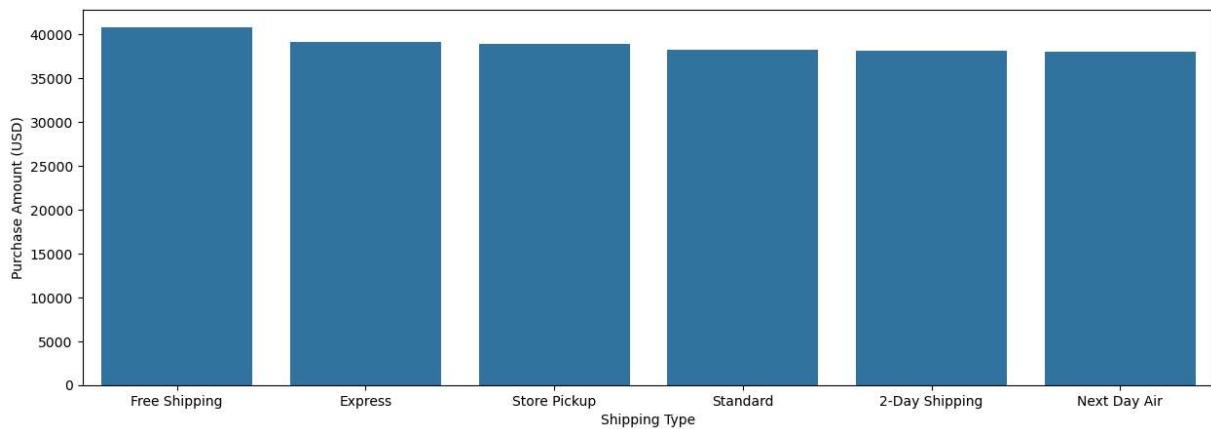
**Fall season has produced the most profit.**

```
In [209]: 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()
```



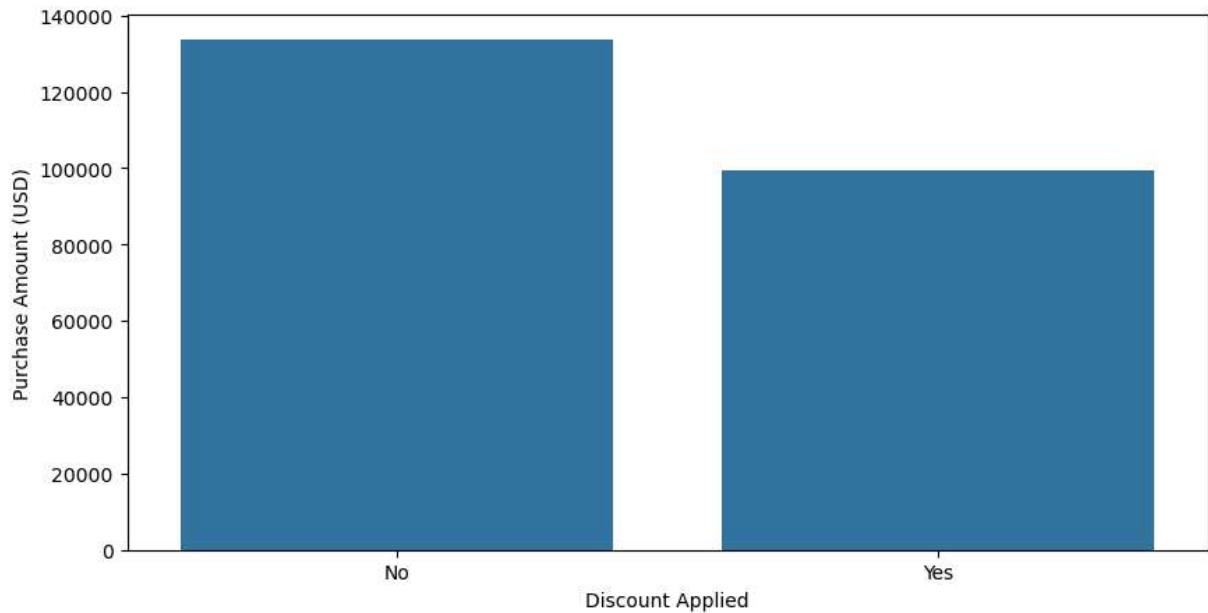
```
In [210]: # Most Profit came from those who haven't subscribed.
```

```
In [211... 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()
```

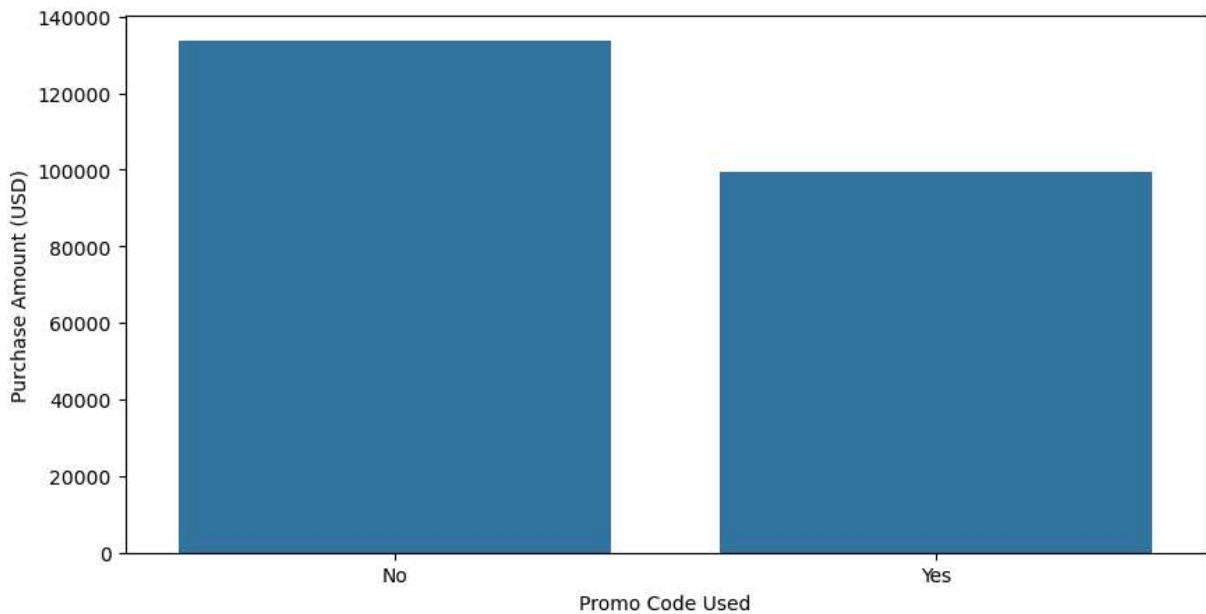


**Free shipping producd most profit.**

```
In [212... 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()
```

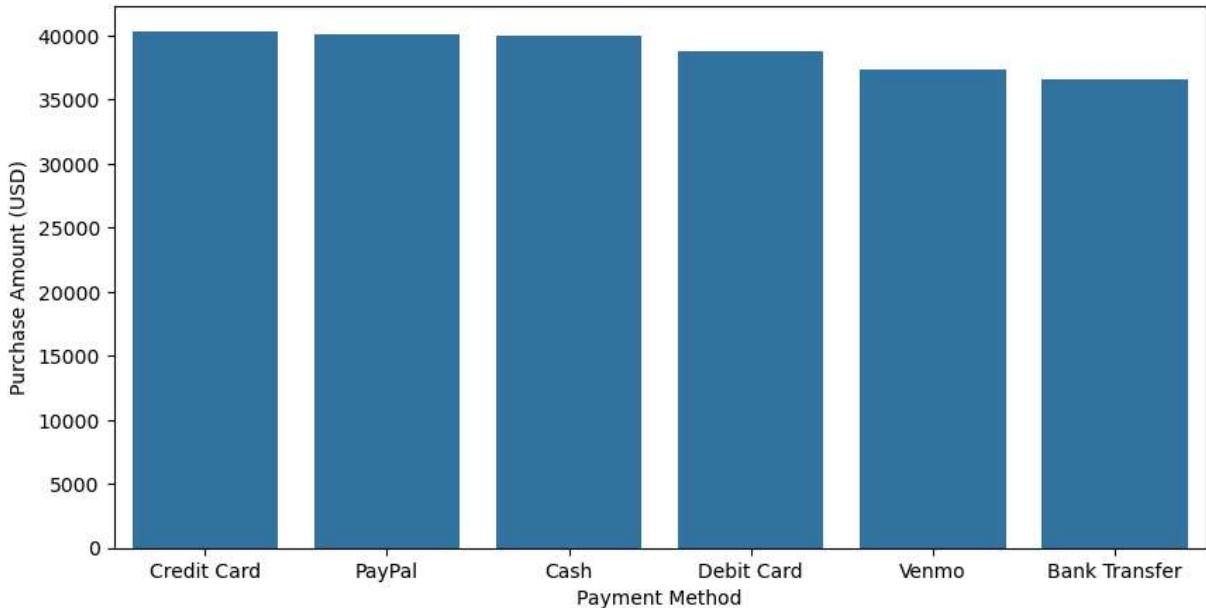


```
In [213... plt.figure(figsize=(10,5))
sales = df.groupby(['Promo Code Used'],as_index=False)[['Purchase Amount (USD)']].sum()
sns.barplot(data=sales,x='Promo Code Used',y='Purchase Amount (USD)')
plt.show()
```



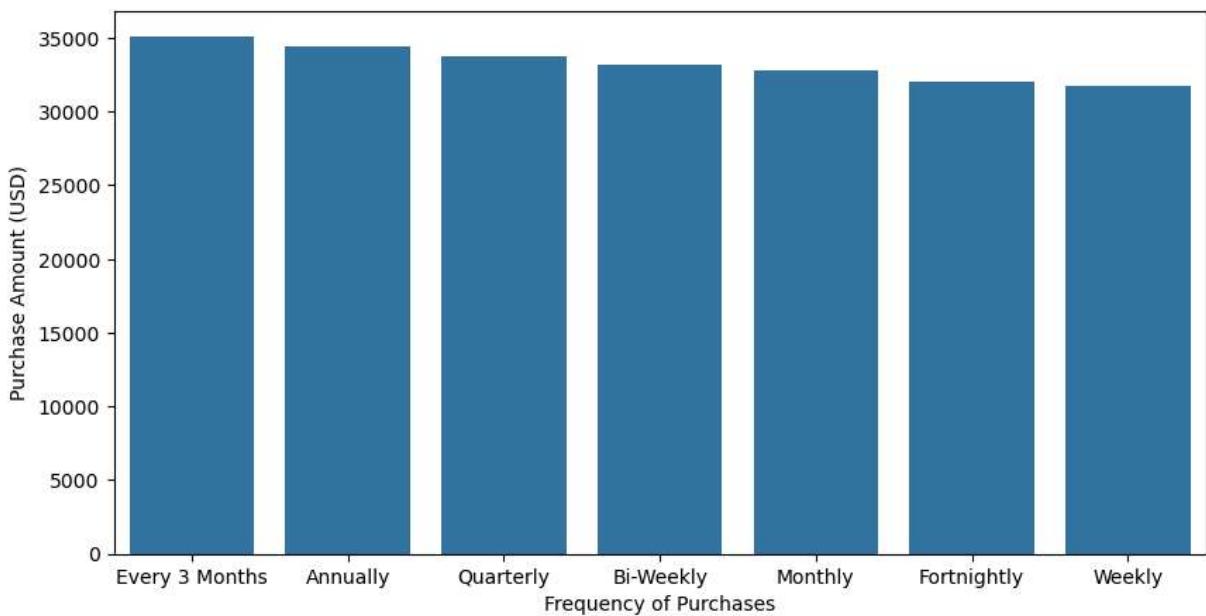
**Most of the profit came from those who hadn't used Promo Code.**

```
In [214...]: 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()
```

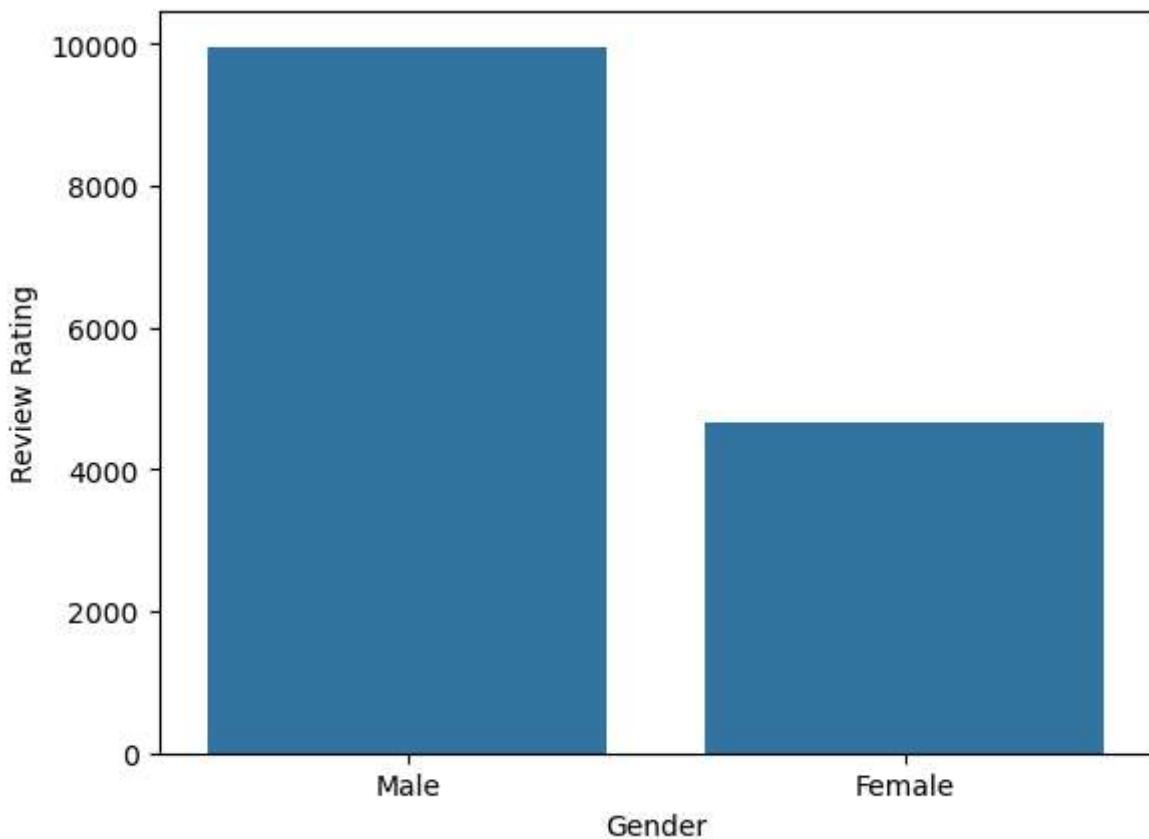


**Most amount was created by Credit card, and then PayPal.**

```
In [215...]: plt.figure(figsize=(10,5))
sales = df.groupby(['Frequency of Purchases'],as_index=False)[['Purchase Amount (USD)']].sum()
sns.barplot(data=sales,x='Frequency of Purchases',y='Purchase Amount (USD)')
plt.show()
```



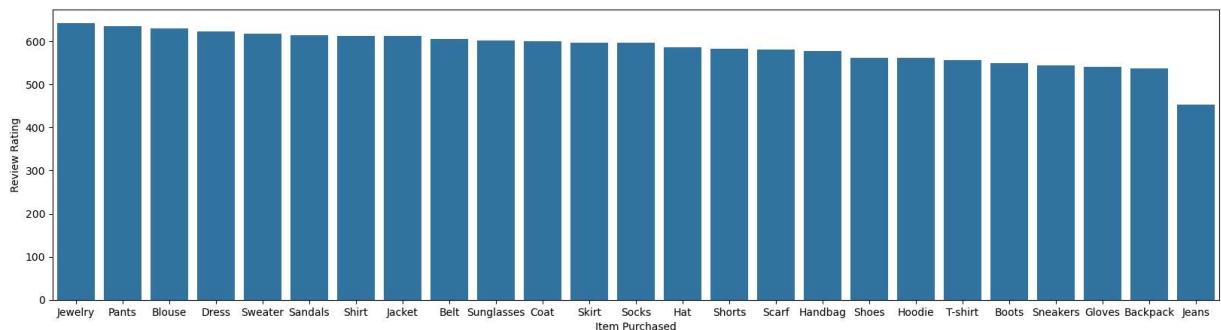
```
In [216...]: sales = df.groupby(['Gender'], as_index=False)[['Review Rating']].sum().sort_values(by='Review Rating', ascending=False)
sns.barplot(data=sales, x='Gender', y='Review Rating')
plt.show()
```



**Males have given the most Review Rating.**

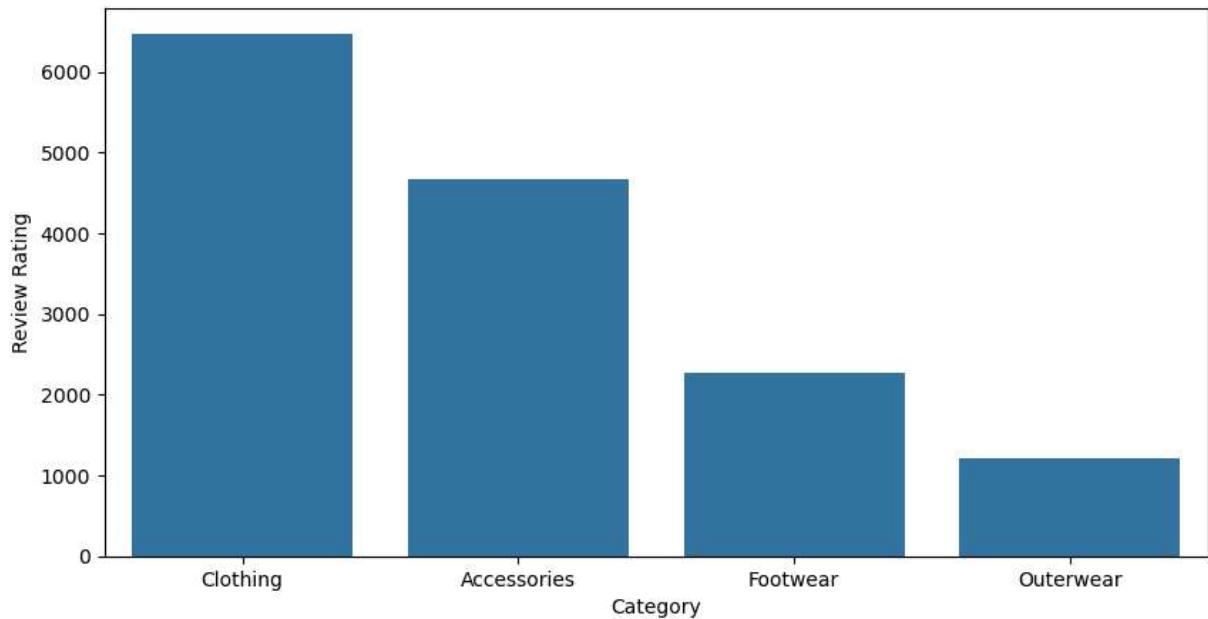
```
In [217...]: plt.figure(figsize=(20,5))
sales = df.groupby(['Item Purchased'], as_index=False)[['Review Rating']].sum().sort_v
```

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



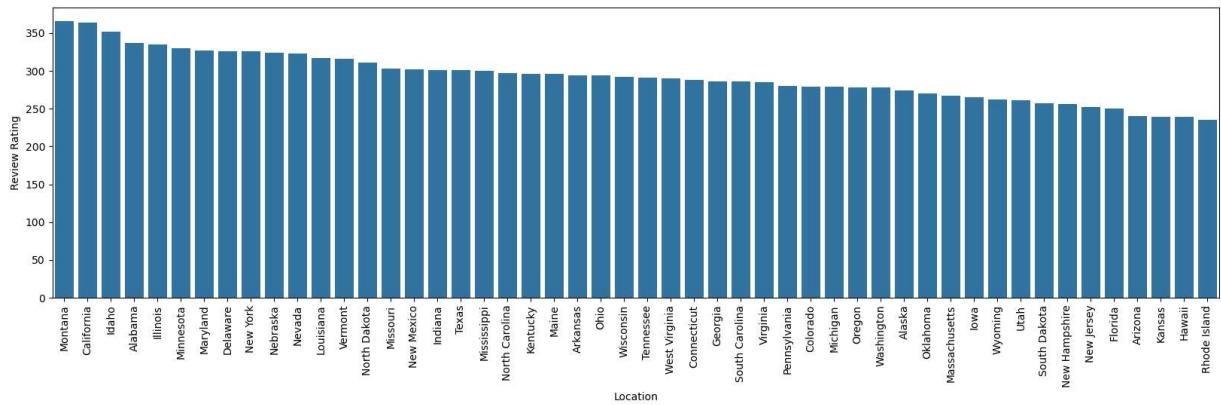
**Jewelry buyers have given the a lot of reviews.**

```
In [218...]
plt.figure(figsize=(10,5))
sales = df.groupby(['Category'],as_index=False)[['Review Rating']].sum().sort_values(['Review Rating'])
sns.barplot(data=sales,x='Category',y='Review Rating')
plt.show()
```



**Clothing buyers have given the most reviews.**

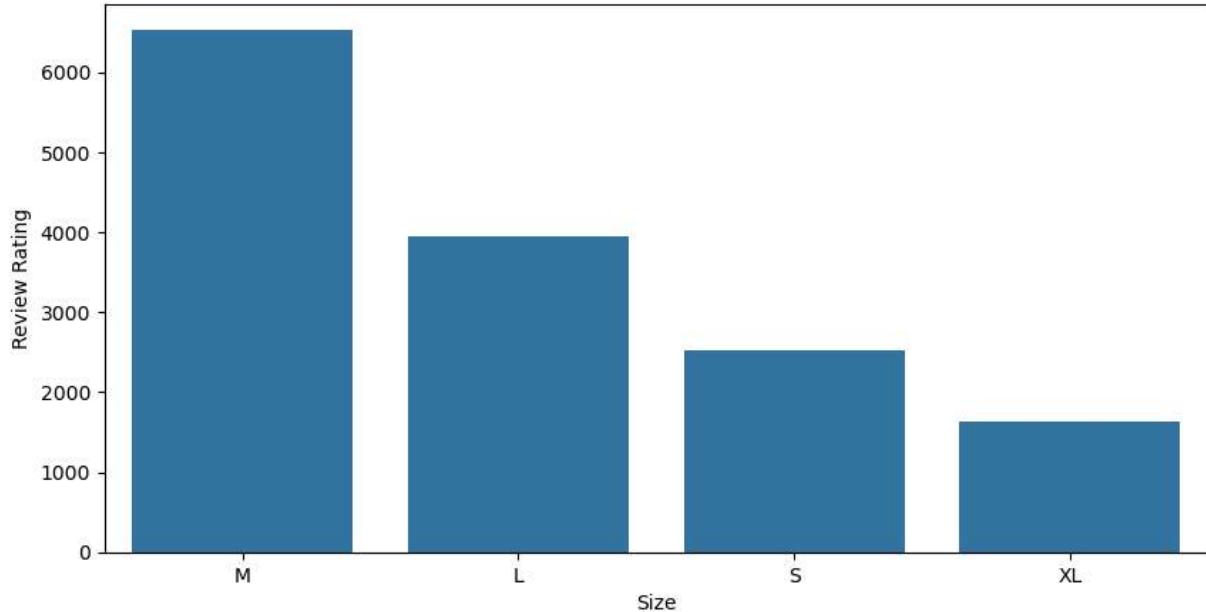
```
In [219...]
plt.figure(figsize=(20,5))
sales = df.groupby(['Location'],as_index=False)[['Review Rating']].sum().sort_values(['Review Rating'])
sns.barplot(data=sales,x='Location',y='Review Rating')
plt.xticks(rotation=90)
plt.show()
```



**Mantola's buyers have given most reviews.**

In [220...]

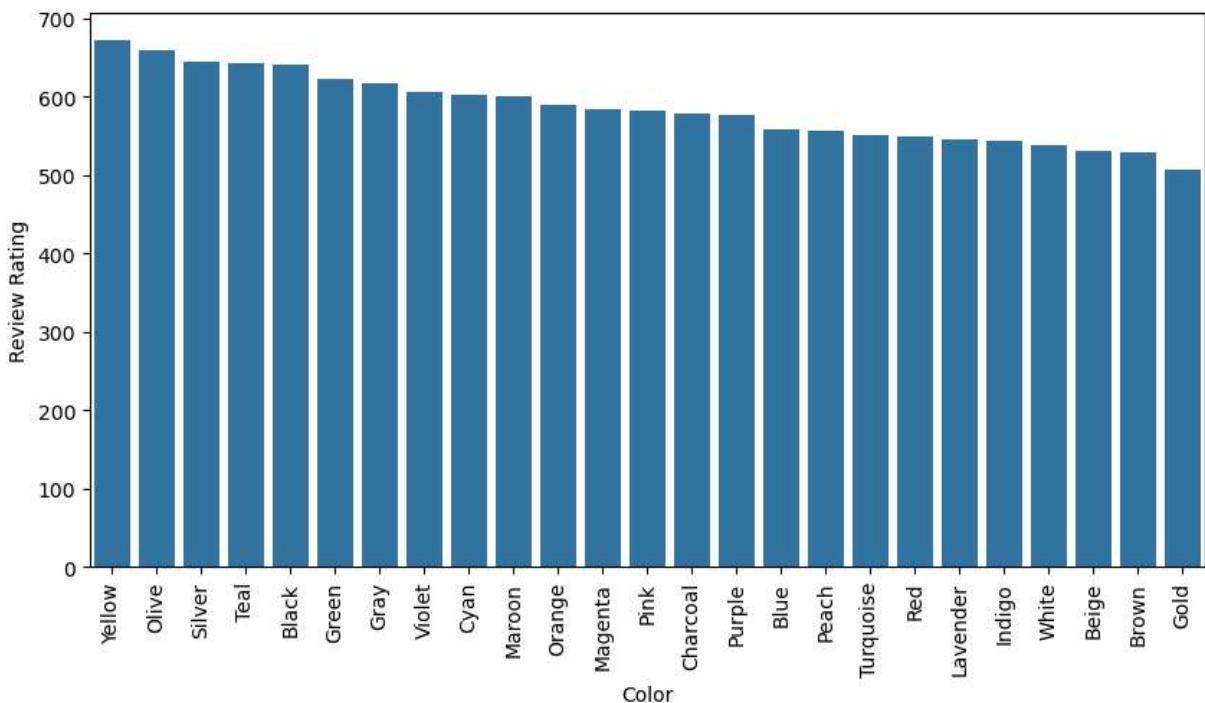
```
plt.figure(figsize=(10,5))
sales = df.groupby(['Size'],as_index=False)[['Review Rating']].sum().sort_values(by='Review Rating', ascending=False)
sns.barplot(data=sales,x='Size',y='Review Rating')
plt.show()
```



**Medium size buyers have given most reviews.**

In [221...]

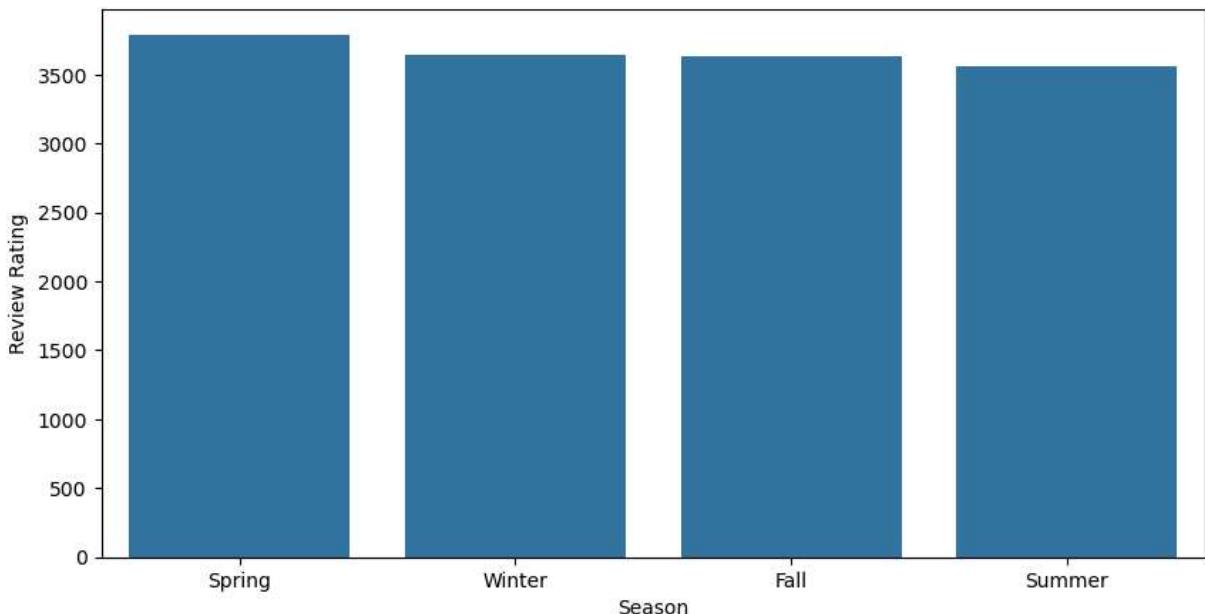
```
plt.figure(figsize=(10,5))
sales = df.groupby(['Color'],as_index=False)[['Review Rating']].sum().sort_values(by='Review Rating', ascending=False)
sns.barplot(data=sales,x='Color',y='Review Rating')
plt.xticks(rotation=90)
plt.show()
```



**Yellow color buyers have given most reviews.**

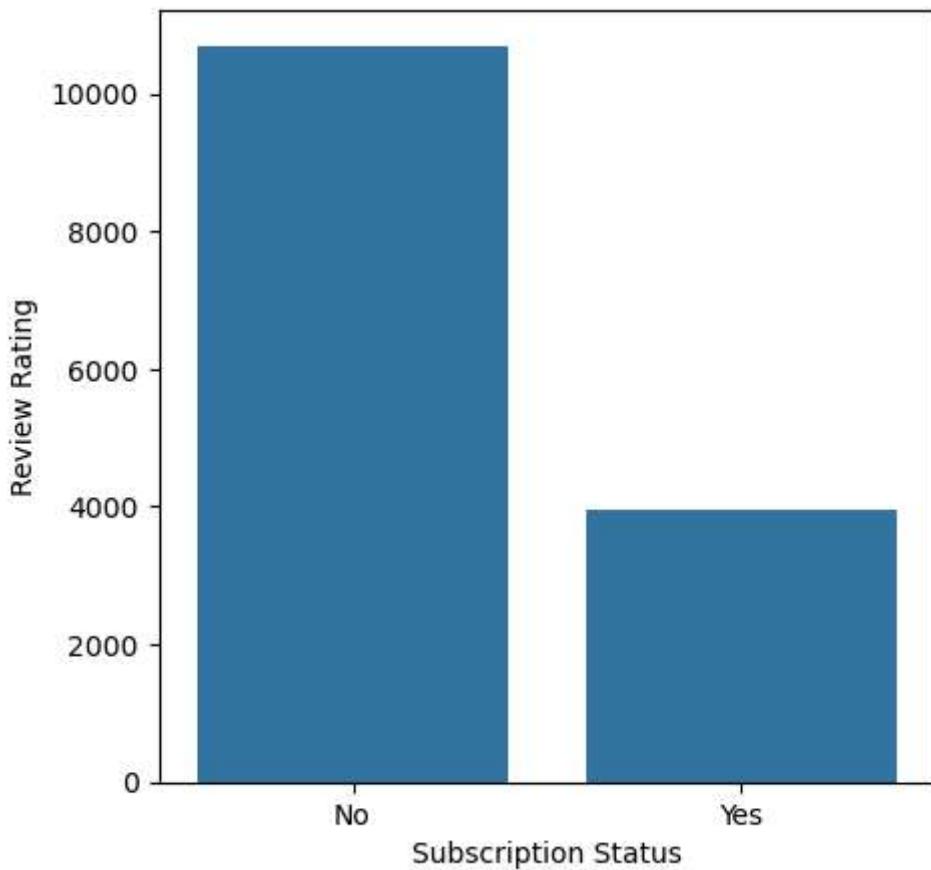
In [222...]

```
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 [223...]

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

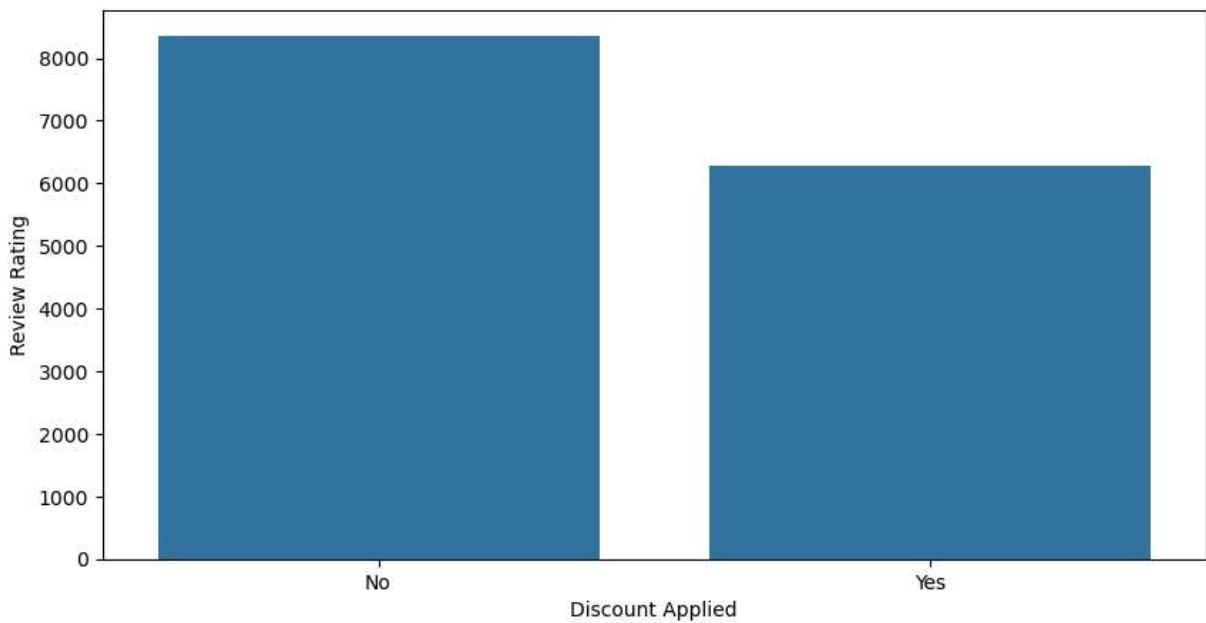


```
In [224...]:  
plt.figure(figsize=(15,5))  
sales = df.groupby(['Shipping Type'],as_index=False)[['Review Rating']].sum().sort_values  
sns.barplot(data=sales,x='Shipping Type',y='Review Rating')  
plt.show()
```



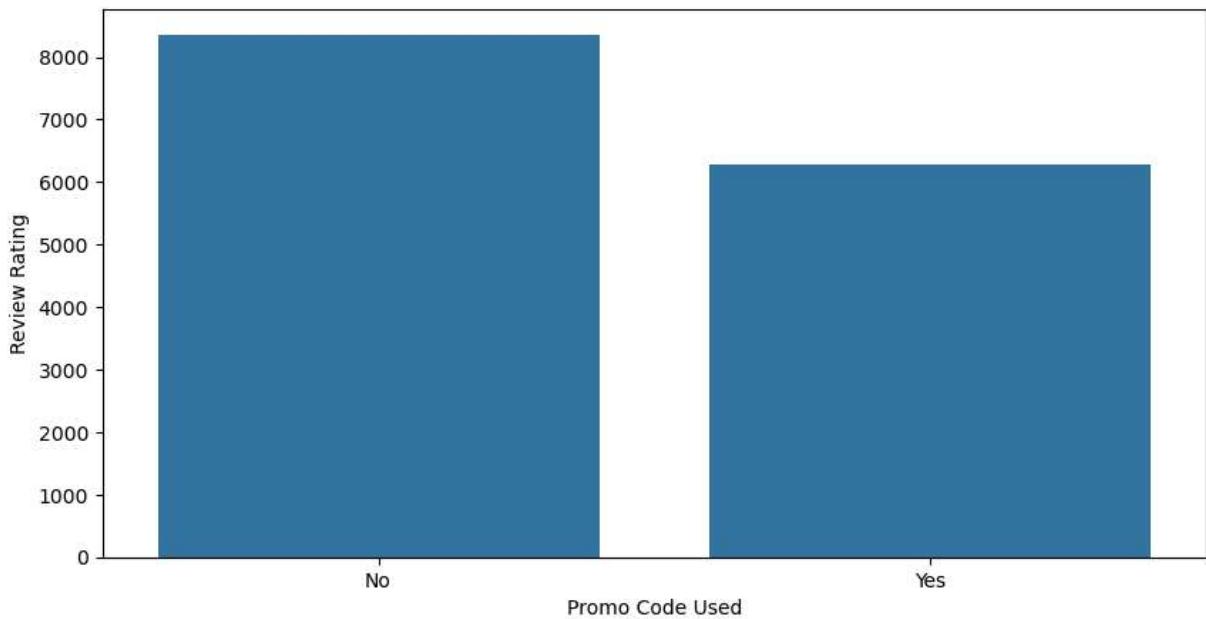
**Free shipping buyers have given the most reviews.**

```
In [225...]:  
plt.figure(figsize=(10,5))  
sales = df.groupby(['Discount Applied'],as_index=False)[['Review Rating']].sum().sort  
sns.barplot(data=sales,x='Discount Applied',y='Review Rating')  
plt.show()
```



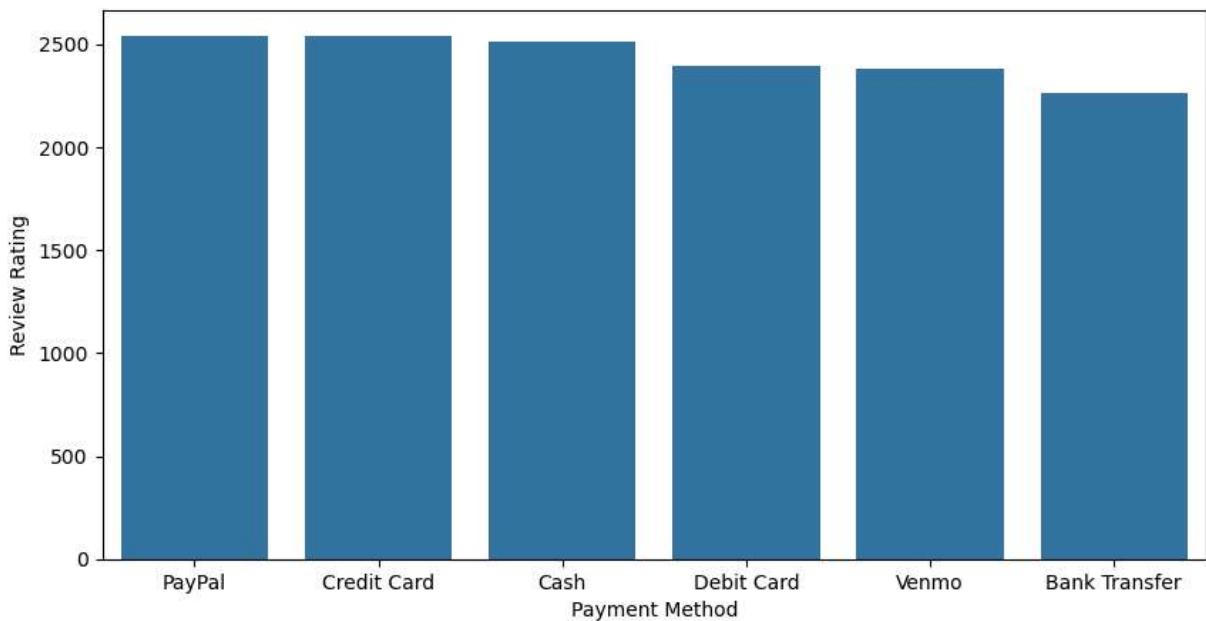
In [226...]

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



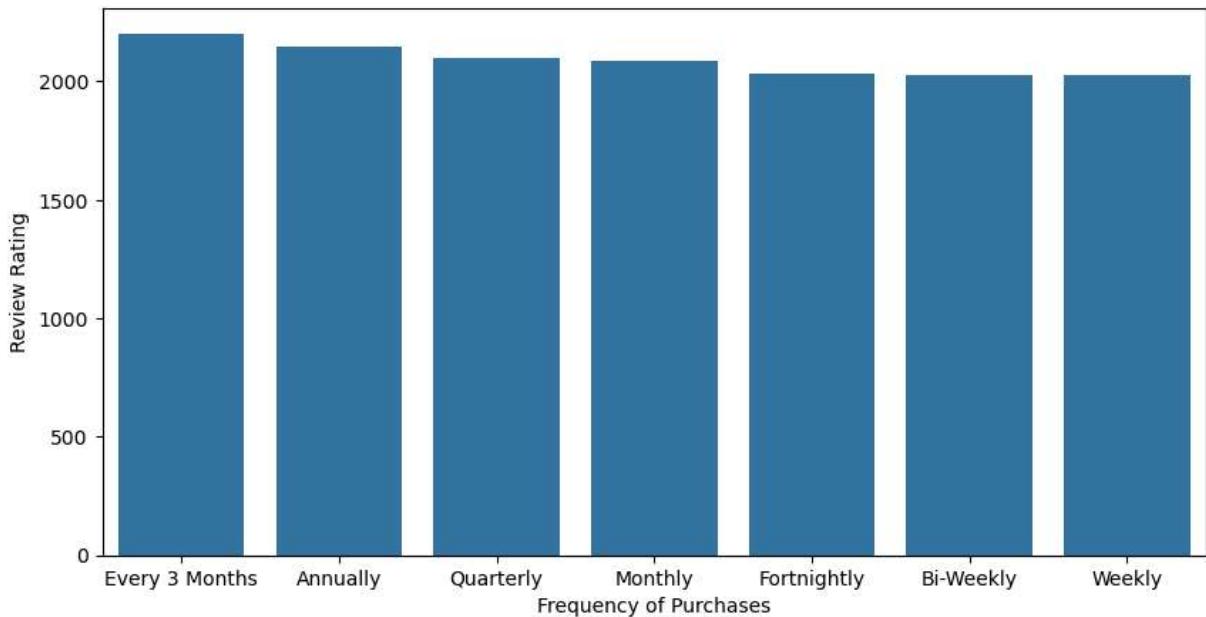
In [227...]

```
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()
```

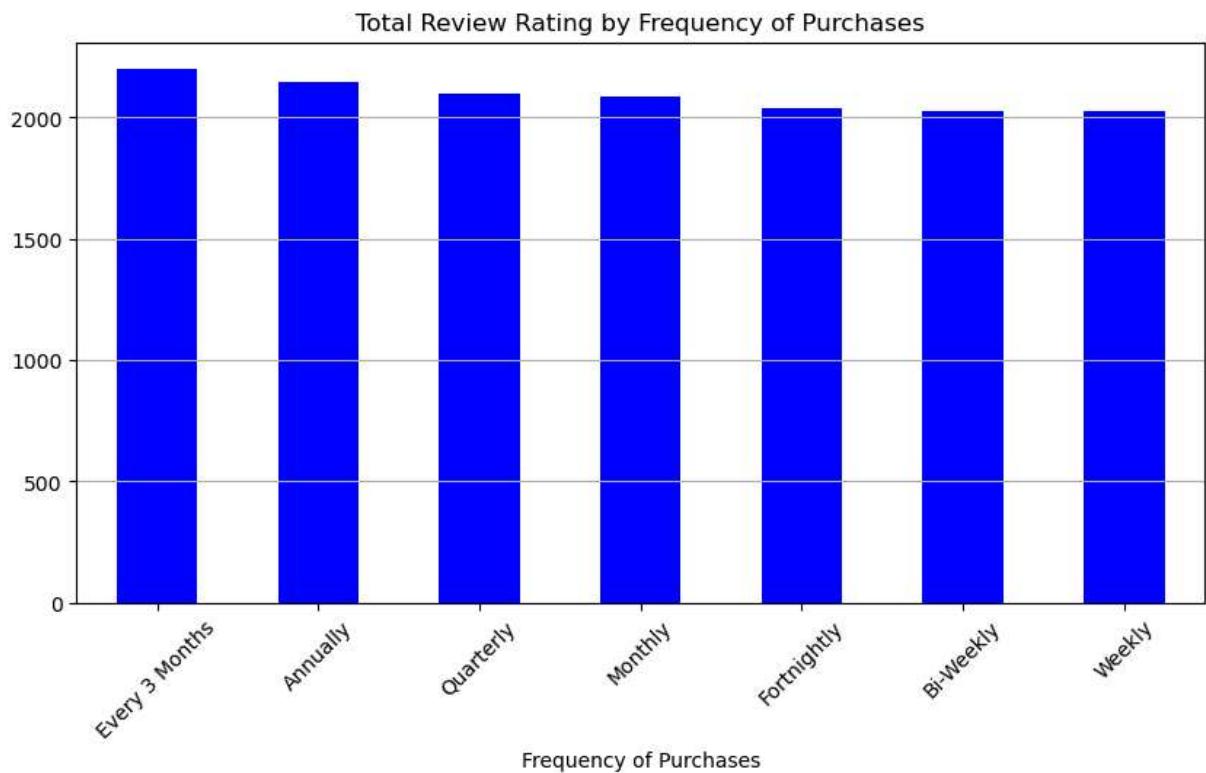


**PayPal users have given the most reviews.**

```
In [228...]:  
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 [229...]:  
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 [ ]: