```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
```

```python
# reading the data set
shop = pd.read_csv('shopping_trends_updated.csv')
```

```python
shop.shape
```

    (3900, 18)

```python
shop.head()
```

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size	Color	Season	Review Rating	Subscription Status	Shipping Type	Discount Applied
0	1	55	Male	Blouse	Clothing	53	Kentucky	L	Gray	Winter	3.1	Yes	Express	Yes
1	2	19	Male	Sweater	Clothing	64	Maine	L	Maroon	Winter	3.1	Yes	Express	Yes
2	3	50	Male	Jeans	Clothing	73	Massachusetts	S	Maroon	Spring	3.1	Yes	Free Shipping	Yes
3	4	21	Male	Sandals	Footwear	90	Rhode Island	M	Maroon	Spring	3.5	Yes	Next Day Air	Yes
4	5	45	Male	Blouse	Clothing	49	Oregon	M	Turquoise	Spring	2.7	Yes	Free Shipping	Yes

```python
shop.dtypes
```

	0
Customer ID	int64
Age	int64
Gender	object
Item Purchased	object
Category	object
Purchase Amount (USD)	int64
Location	object
Size	object
Color	object
Season	object
Review Rating	float64
Subscription Status	object
Shipping Type	object
Discount Applied	object
Promo Code Used	object
Previous Purchases	int64
Payment Method	object
Frequency of Purchases	object

dtype: object

```python
shop.columns
```

```
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')
```

```python
shop.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
```

```python
shop.isnull().sum()
```

	0
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

```
print(f"The unique values of the 'Gender' column are: {shop['Gender'].unique()}")
print()# This will print a blank line
print(f"The unique values of the 'Category' column are: {shop['Category'].unique()}")
print()# This will print a blank line
print(f"The unique values of the 'Size' column are: {shop['Size'].unique()}")
print()# This will print a blank line
print(f"The unique values of the 'Subscription Status' column are: {shop['Subscription Status'].unique()}")
print()# This will print a blank line
print(f"The unique values of the 'Shipping Type' column are: {shop['Shipping Type'].unique()}")
print()# This will print a blank line
print(f"The unique values of the 'Discount Applied' column are: {shop['Discount Applied'].unique()}")
print()# This will print a blank line
print(f"The unique values of the 'Promo Code Used' column are: {shop['Promo Code Used'].unique()}")
print()# This will print a blank line
print(f"The unique values of the 'Payment Method' column are: {shop['Payment Method'].unique()}")
```

The unique values of the 'Gender' column are: ['Male' 'Female']

The unique values of the 'Category' column are: ['Clothing' 'Footwear' 'Outerwear' 'Accessories']

The unique values of the 'Size' column are: ['L' 'S' 'M' 'XL']

The unique values of the 'Subscription Status' column are: ['Yes' 'No']

The unique values of the 'Shipping Type' column are: ['Express' 'Free Shipping' 'Next Day Air' 'Standard' '2-Day Shipping'
 'Store Pickup']

The unique values of the 'Discount Applied' column are: ['Yes' 'No']

The unique values of the 'Promo Code Used' column are: ['Yes' 'No']

The unique values of the 'Payment Method' column are: ['Venmo' 'Cash' 'Credit Card' 'PayPal' 'Bank Transfer' 'Debit Card']

1 What is the overall distribution of customer ages in the dataset?

```
shop['Age'].value_counts()
```

Age	count
69	88
57	87
41	86
25	85
49	84
50	83
54	83
27	83
62	83
32	82
19	81
58	81
42	80
43	79
28	79
31	79
37	77
46	76
29	76
68	75
59	75
63	75
56	74
36	74
55	73
52	73
64	73
35	72
51	72
65	72
40	72
45	72
47	71
66	71
30	71
23	71
38	70
53	70
18	69
21	69
26	69
34	68
48	68
24	68
39	68

70	67
22	66
61	65
60	65
33	63
20	62
67	54
44	51

dtype: int64

```python
shop['Age'].mean()
```

44.06846153846154

```python
shop['Gender'].unique()
```

array(['Male', 'Female'], dtype=object)

```python
shop['Age_category'] = pd.cut(shop['Age'], bins= [0,15, 18 , 30 , 50 , 70] , labels= ['child' , 'teen' , 'Young Adults' ,'Middle-Aged Adults
                                                                                        , 'old'] )
```

```python
shop['Age_category']
```

	Age_category
0	old
1	Young Adults
2	Middle-Aged Adults
3	Young Adults
4	Middle-Aged Adults
...	...
3895	Middle-Aged Adults
3896	old
3897	Middle-Aged Adults
3898	Middle-Aged Adults
3899	old

3900 rows × 1 columns

dtype: category

```python
fig = px.histogram(shop, y='Age', x = "Age_category")
fig.show()
```
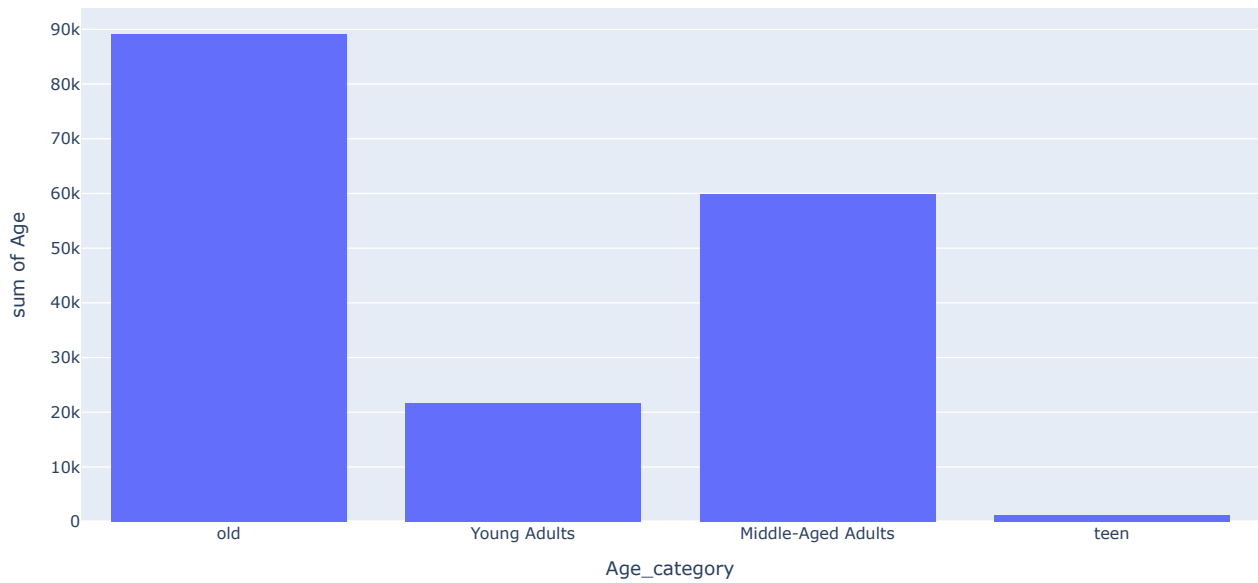
2 How does the average purchase amount vary across different product categories?

```
shop.columns
```

```
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', 'Age_category'],
      dtype='object')
```

```
shop['Category'].unique()
```

```
array(['Clothing', 'Footwear', 'Outerwear', 'Accessories'], dtype=object)
```

```
shop[['Category', 'Purchase Amount (USD)']]
```

	Category	Purchase Amount (USD)
0	Clothing	53
1	Clothing	64
2	Clothing	73
3	Footwear	90
4	Clothing	49
...	...	...
3895	Clothing	28
3896	Accessories	49
3897	Accessories	33
3898	Footwear	77
3899	Accessories	81

3900 rows × 2 columns

```
shop.groupby('Category')['Purchase Amount (USD)'].sum()
```

	Purchase Amount (USD)
Category	
Accessories	74200
Clothing	104264
Footwear	36093
Outerwear	18524

dtype: int64

```
shop.groupby('Category')['Purchase Amount (USD)'].mean()
```

	Purchase Amount (USD)
Category	
Accessories	59.838710
Clothing	60.025331
Footwear	60.255426
Outerwear	57.172840

dtype: float64

3 Which gender has the highest number of purchases?

```
sns.barplot(shop, x = 'Gender', y = 'Purchase Amount (USD)')
```

<Axes: xlabel='Gender', ylabel='Purchase Amount (USD)'>
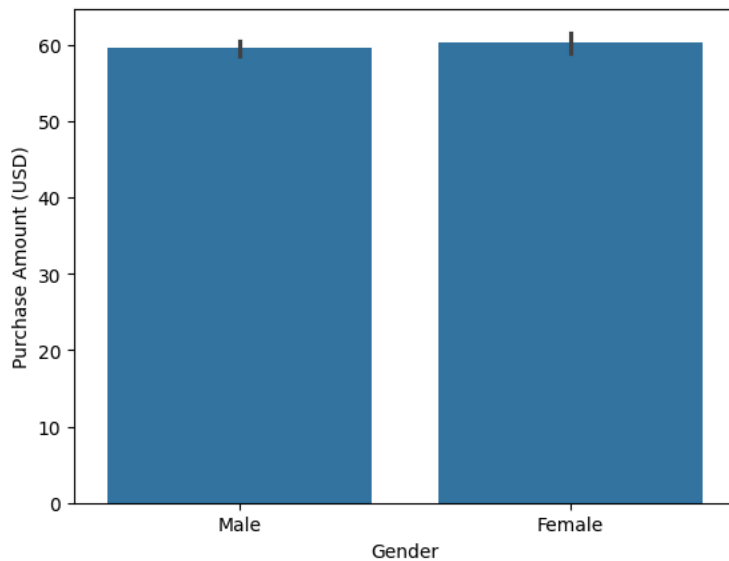
4 What are the most commonly purchased items in each category?

```
shop.groupby('Category')['Item Purchased'].value_counts()
```

		count
Category	**Item Purchased**	
Accessories	**Jewelry**	171
	Belt	161
	Sunglasses	161
	Scarf	157
	Hat	154
	Handbag	153
	Backpack	143
	Gloves	140
Clothing	**Blouse**	171
	Pants	171
	Shirt	169
	Dress	166
	Sweater	164
	Socks	159
	Skirt	158
	Shorts	157
	Hoodie	151
	T-shirt	147
	Jeans	124
Footwear	**Sandals**	160
	Shoes	150
	Sneakers	145
	Boots	144
Outerwear	**Jacket**	163
	Coat	161

dtype: int64

```
fig = px.histogram(shop , x = 'Item Purchased' , color = 'Category')
fig.show()
```
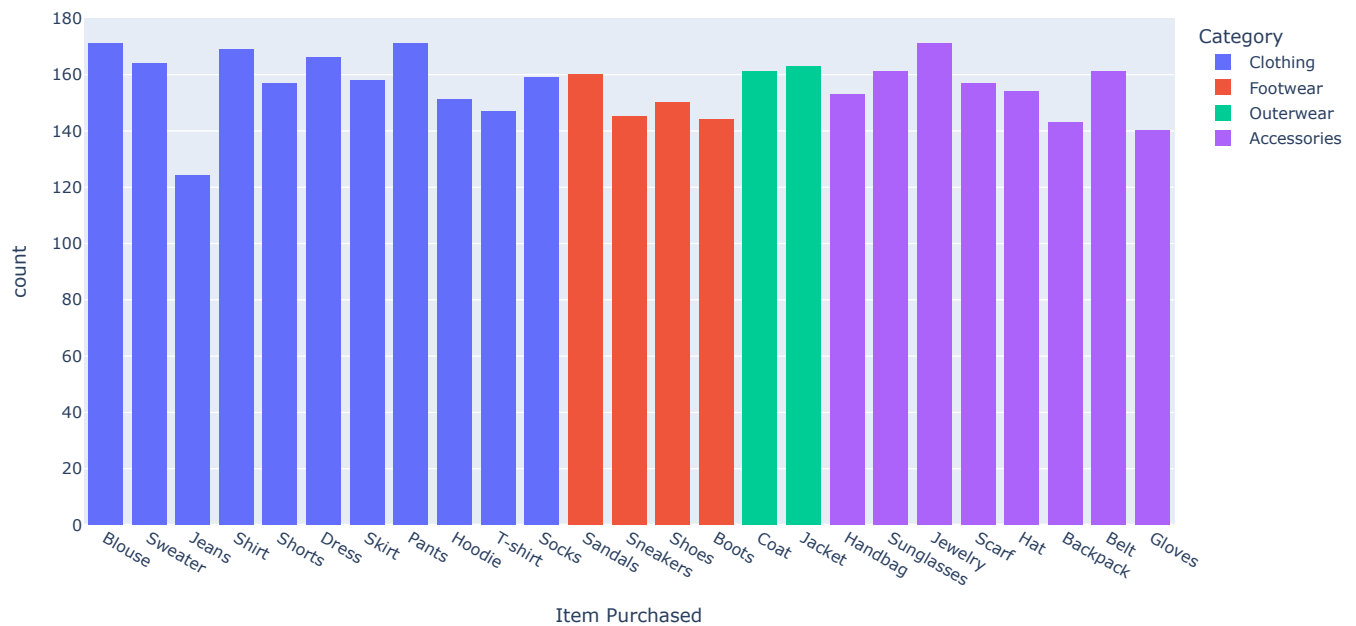
5 Are there any specific seasons or months where customer spending is significantly higher?

```
shop['Season'].unique()
```

```
array(['Winter', 'Spring', 'Summer', 'Fall'], dtype=object)
```
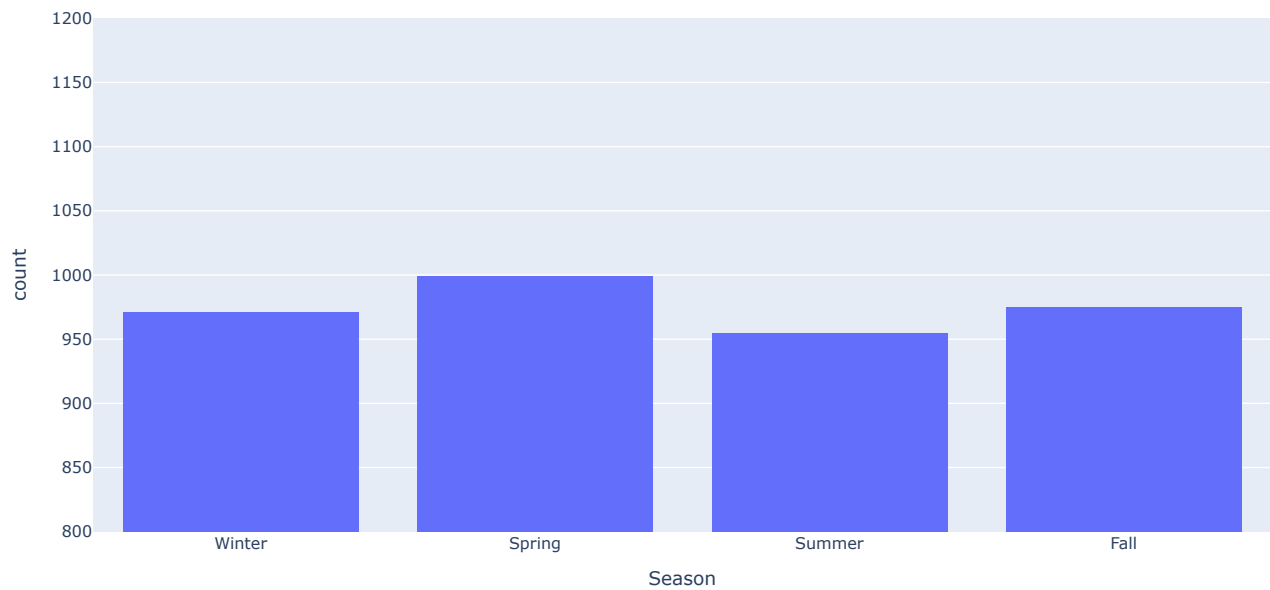
```
shop['Season'].value_counts()
```

	count
Season	
Spring	999
Fall	975
Winter	971
Summer	955

dtype: int64

```
fig = px.histogram(shop , x = 'Season' , range_y= [800 , 1200] )
fig.show()
```

6 What is the average rating given by customers for each product category?

```
shop.groupby('Category')['Review Rating'].mean()
```

|	Review Rating
Category	
Accessories	3.768629
Clothing	3.723143
Footwear	3.790651
Outerwear	3.746914

dtype: float64

```
shop_groupby = shop.groupby('Category')['Review Rating'].mean().reset_index()
print(shop_groupby)
```

```
     Category  Review Rating
0  Accessories       3.768629
1     Clothing       3.723143
2     Footwear       3.790651
3    Outerwear       3.746914
```

```
fig = px.bar(shop_groupby ,x= 'Category' , y = 'Review Rating' )
fig.show()
```
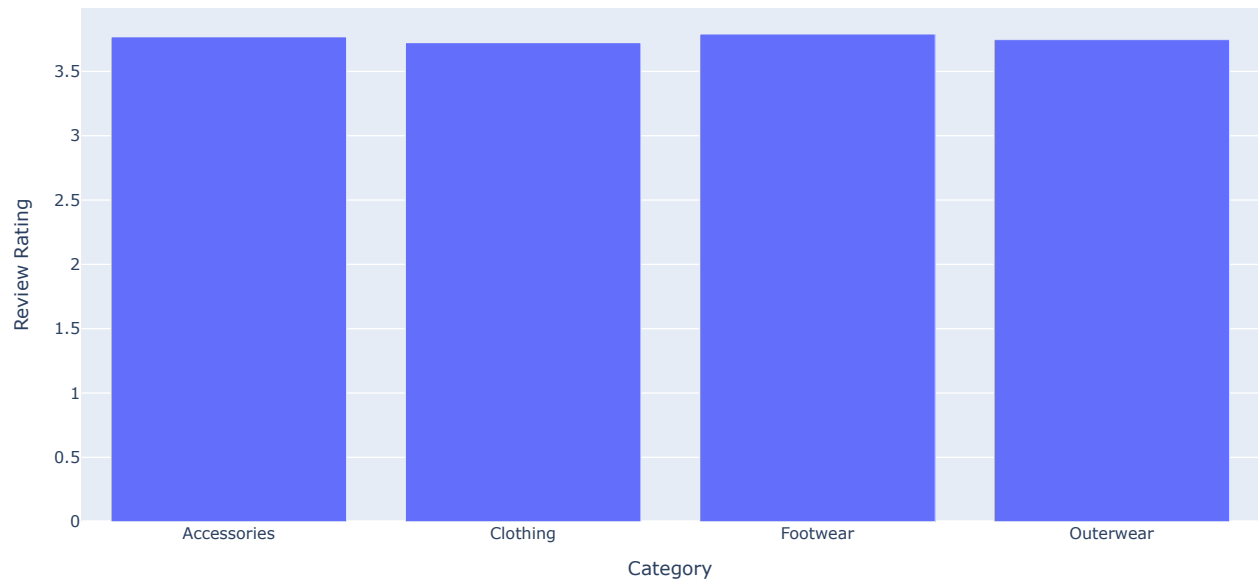
7 Are there any notable differences in purchase behavior between subscribed and non-subscribed customers?

```
shop['Subscription Status'].value_counts()
```
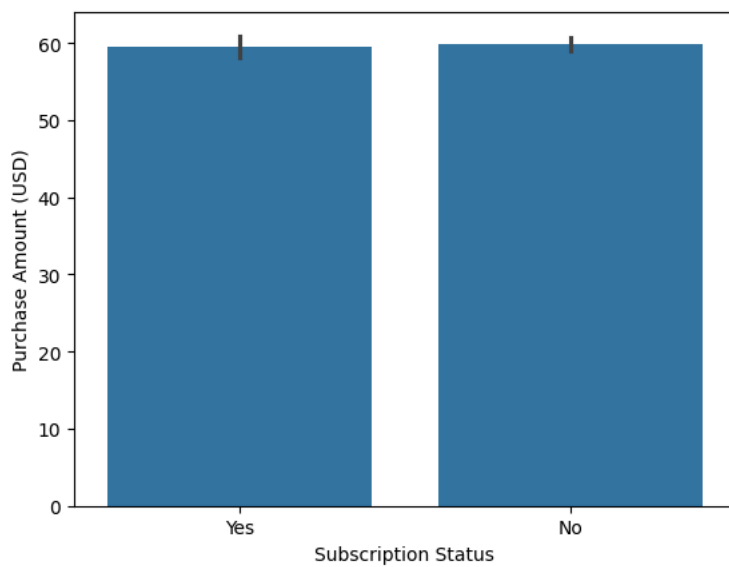
	count
Subscription Status	
No	2847
Yes	1053

dtype: int64

```
sns.barplot(shop  , x = 'Subscription Status' , y = 'Purchase Amount (USD)')
```

<Axes: xlabel='Subscription Status', ylabel='Purchase Amount (USD)'>


```
shop['Purchase Amount (USD)'].sum()
```

233081

```
shop.groupby('Subscription Status')['Purchase Amount (USD)'].mean()
```

	Purchase Amount (USD)
Subscription Status	
No	59.865121
Yes	59.491928

dtype: float64

8 Which payment method is the most popular among customers?

```
shop.groupby('Payment Method')['Purchase Amount (USD)'].mean().sort_values(ascending= False)
```
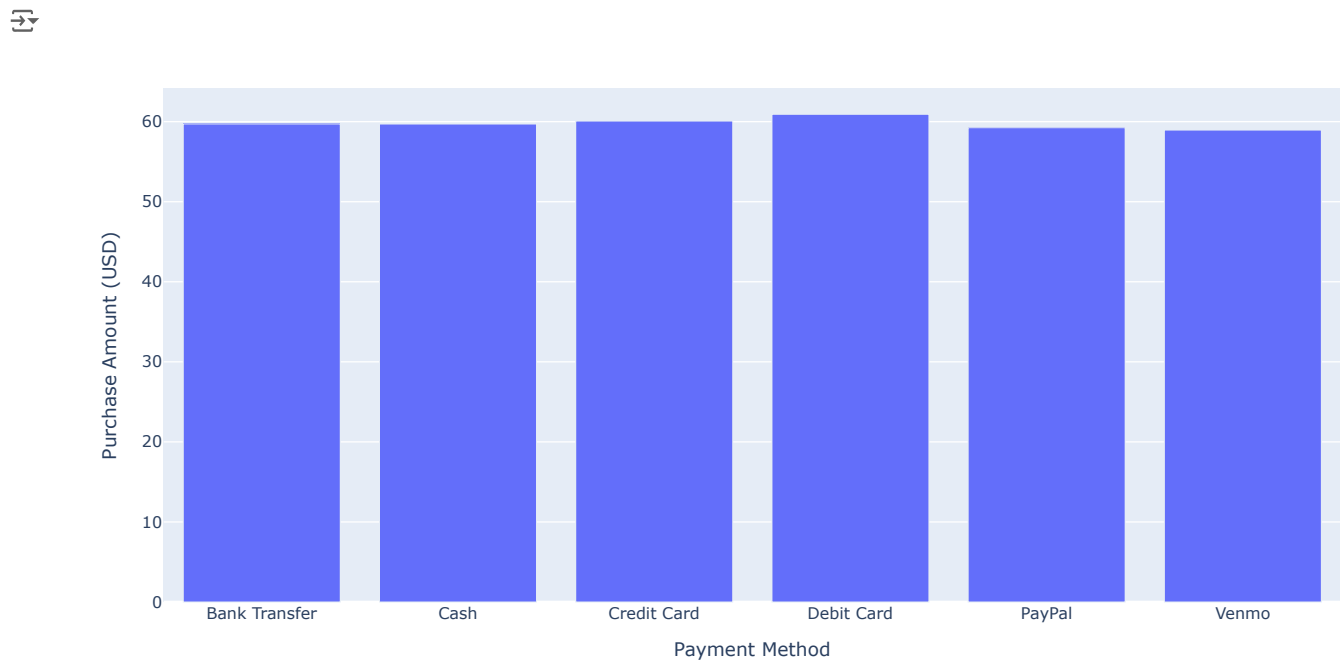
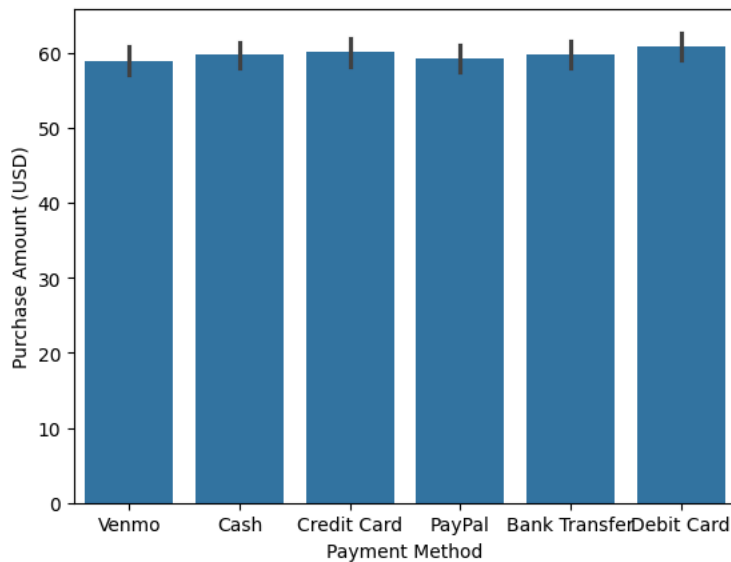|	Purchase Amount (USD)
Payment Method	
Debit Card	60.915094
Credit Card	60.074516
Bank Transfer	59.712418
Cash	59.704478
PayPal	59.245199
Venmo	58.949527

dtype: float64

```
shop_groupby = shop.groupby('Payment Method')['Purchase Amount (USD)'].mean().reset_index()
```

```
fig = px.bar(shop_groupby , x = 'Payment Method' , y = 'Purchase Amount (USD)')
fig.show()
```
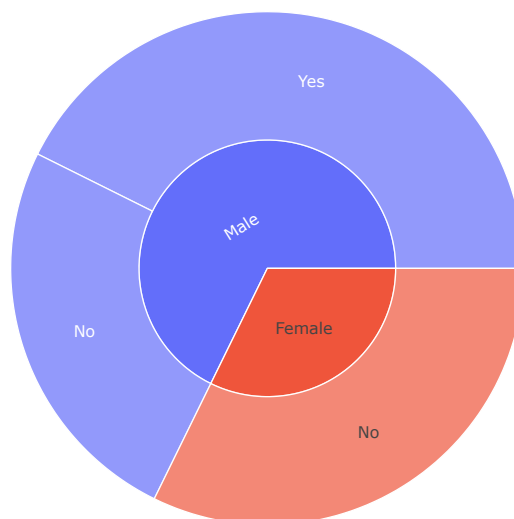


```
sns.barplot(shop ,x='Payment Method' , y = 'Purchase Amount (USD)')
plt.show()
```
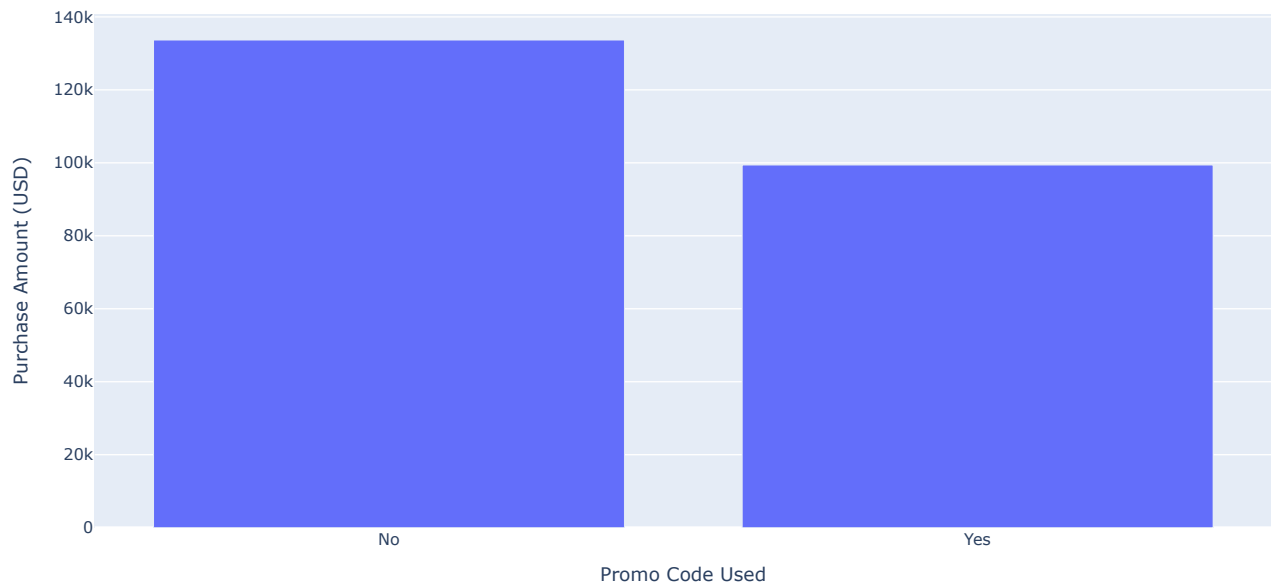
9 Do customers who use promo codes tend to spend more than those who don't?

```
shop_groupby  = shop.groupby('Promo Code Used')['Purchase Amount (USD)'].sum().reset_index()
```

```
fig = px.sunburst(shop , path=['Gender' , 'Promo Code Used'] , values='Purchase Amount (USD)')
fig.show()
```



```
fig  =  px.bar(shop_groupby , x= 'Promo Code Used' , y = 'Purchase Amount (USD)')
fig.show()
```

10 How does the frequency of purchases vary across different age groups?

```
shop[['Age' , 'Age_category']]
```

	Age	Age_category
0	55	old
1	19	Young Adults
2	50	Middle-Aged Adults
3	21	Young Adults
4	45	Middle-Aged Adults
...	...	...
3895	40	Middle-Aged Adults
3896	52	old
3897	46	Middle-Aged Adults
3898	44	Middle-Aged Adults
3899	52	old

3900 rows × 2 columns

```
shop['Age_category'].unique()
```

```
['old', 'Young Adults', 'Middle-Aged Adults', 'teen']
Categories (5, object): ['child' < 'teen' < 'Young Adults' < 'Middle-Aged Adults' < 'old']
```

```
shop_group = shop.groupby('Frequency of Purchases')['Age'].sum()
```

```
px.sunburst(shop , path=['Frequency of Purchases','Age_category'] , values='Age')
```

```
/usr/local/lib/python3.10/dist-packages/plotly/express/_core.py:1727: FutureWarning:

The default of observed=False is deprecated and will be changed to True in a future version of pandas. Pass observed=False to retain cur
```
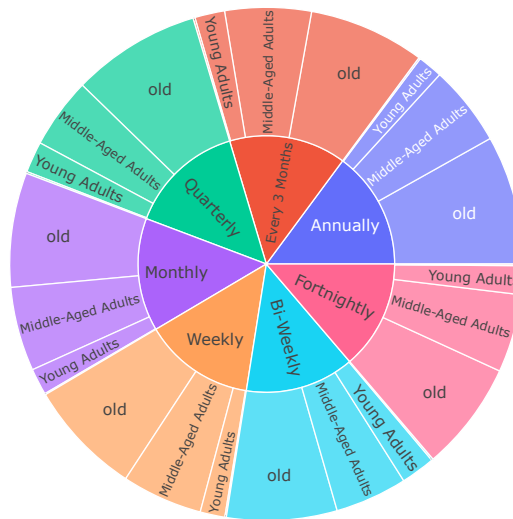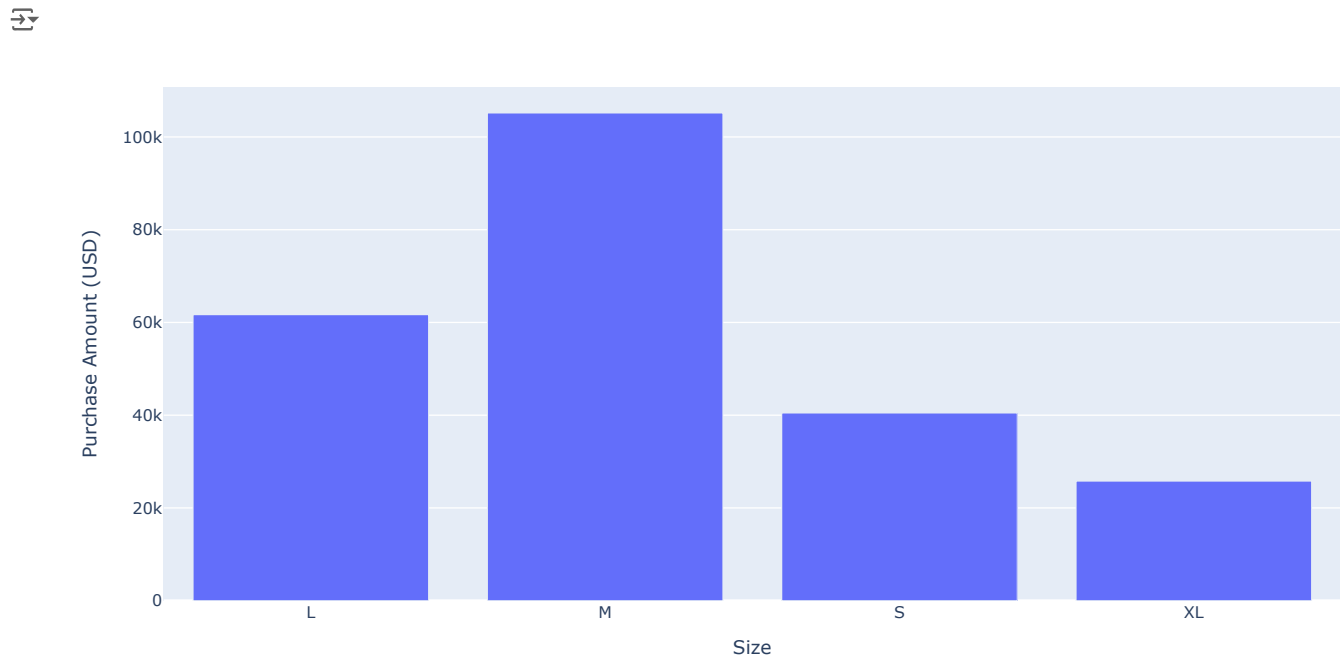

11 Are there any correlations between the size of the product and the purchase amount?

```python
shop_group = shop.groupby('Size')['Purchase Amount (USD)'].sum().reset_index()
```

```python
fig  = px.bar(shop_group , x = 'Size' , y ='Purchase Amount (USD)'  )
fig.show()
```

12 Which shipping type is preferred by customers for different product categories?

```
shop.groupby('Category')['Shipping Type'].value_counts().sort_values(ascending= False)
```
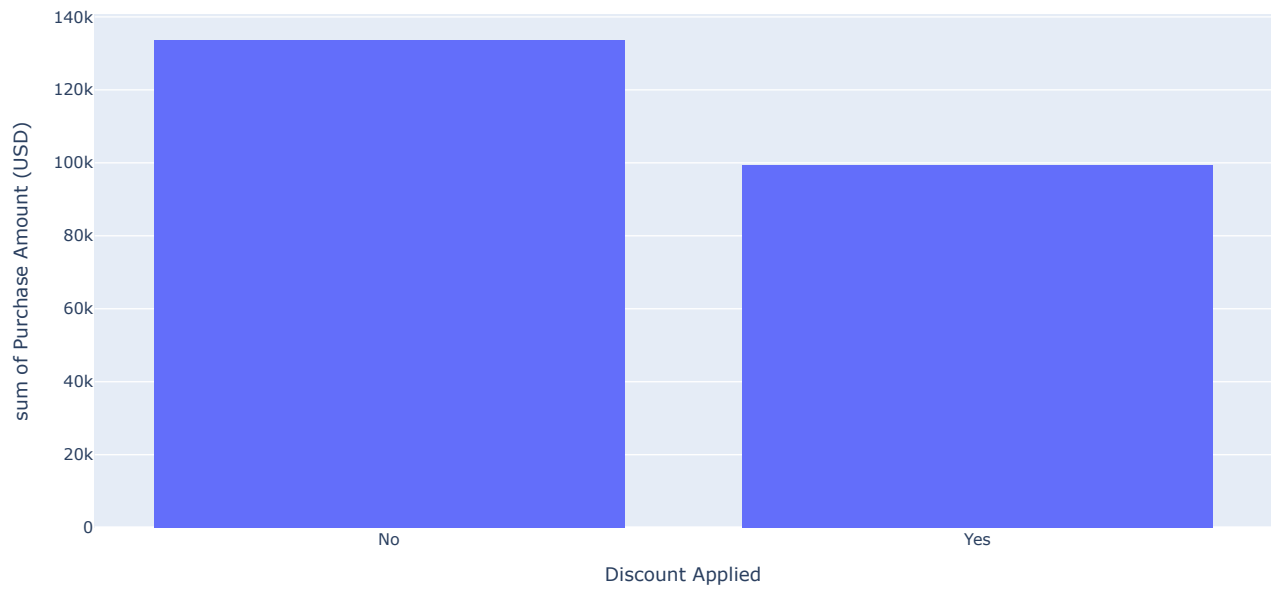
		count
Category	**Shipping Type**	
Clothing	**Standard**	297
	Free Shipping	294
	Next Day Air	293
	Express	290
	Store Pickup	282
	2-Day Shipping	281
Accessories	**Store Pickup**	217
	Next Day Air	211
	Standard	208
	2-Day Shipping	206
	Express	203
	Free Shipping	195
Footwear	**Free Shipping**	122
	Standard	100
	Store Pickup	98
	Express	96
	Next Day Air	93
	2-Day Shipping	90
Outerwear	**Free Shipping**	64
	Express	57
	Store Pickup	53
	Next Day Air	51
	2-Day Shipping	50
	Standard	49

dtype: int64

13 How does the presence of a discount affect the purchase decision of customers?

```
shop_group = shop.groupby('Discount Applied')['Purchase Amount (USD)'].sum().reset_index()
```

```
px.histogram(shop_group , x = 'Discount Applied' , y = 'Purchase Amount (USD)')
```

```
fig = px.sunburst(shop , path = ['Gender' , 'Discount Applied'], values='Purchase Amount (USD)' , color= 'Gender')
fig.show()
```
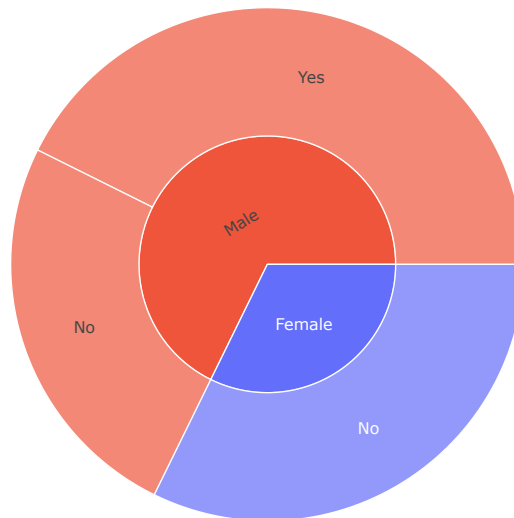

⌄ 14 Are there any specific colors that are more popular among customers?

```
shop['Color'].value_counts().nlargest(5)
```

	count
Color	
Olive	177
Yellow	174
Silver	173
Teal	172
Green	169

dtype: int64

```
px.histogram(shop , x = 'Color')
```
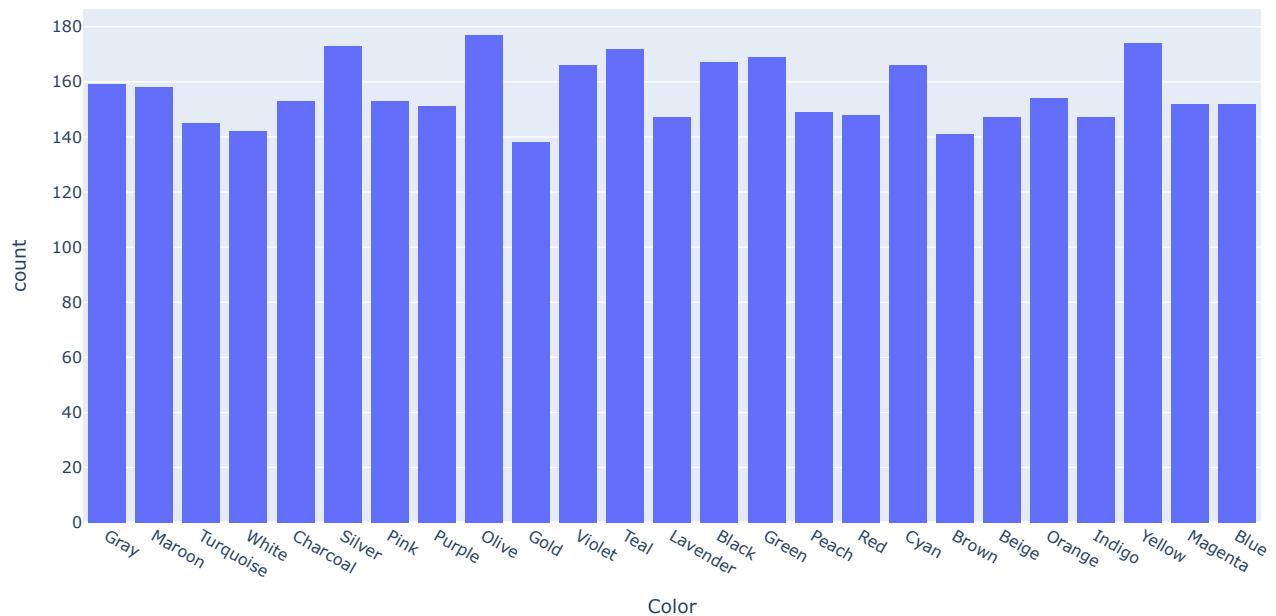

15 What is the average number of previous purchases made by customers?

```
shop['Previous Purchases'].mean()
```

25.35153846153846

16 Are there any noticeable differences in purchase behavior between different locations?

```
shop.groupby('Location')['Purchase Amount (USD)'].mean().sort_values(ascending = False)
```

	Purchase Amount (USD)
Location	
Alaska	67.597222
Pennsylvania	66.567568
Arizona	66.553846
West Virginia	63.876543
Nevada	63.379310
Washington	63.328767
North Dakota	62.891566
Virginia	62.883117
Utah	62.577465
Michigan	62.095890
Tennessee	61.974026
New Mexico	61.901235
Rhode Island	61.444444
Texas	61.194805
Arkansas	61.113924
Illinois	61.054348
Mississippi	61.037500
Massachusetts	60.888889
Iowa	60.884058
North Carolina	60.794872
Wyoming	60.690141
South Dakota	60.514286
New York	60.425287
Ohio	60.376623
Montana	60.250000
Idaho	60.075269
Nebraska	59.448276
New Hampshire	59.422535
Alabama	59.112360
California	59.000000
Indiana	58.924051
Georgia	58.797468
South Carolina	58.407895
Oklahoma	58.346667
Missouri	57.913580
Hawaii	57.723077
Louisiana	57.714286
Oregon	57.337838
Vermont	57.176471
Maine	56.987013
New Jersey	56.746269
Minnesota	56.556818
Colorado	56.293333
Wisconsin	55.946667
Florida	55.852941