

# Decoding Customer Shopping Trends

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
```

```
In [2]: shop=pd.read_csv('shopping_trends.csv')
shop.head(5)
```

Out[2]:

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size	Color	Season	Review Rating
0	1	55	Male	Blouse	Clothing	53	Kentucky	L	Gray	Winter	4.5
1	2	19	Male	Sweater	Clothing	64	Maine	L	Maroon	Winter	4.0
2	3	50	Male	Jeans	Clothing	73	Massachusetts	S	Maroon	Spring	4.5
3	4	21	Male	Sandals	Footwear	90	Rhode Island	M	Maroon	Spring	4.0
4	5	45	Male	Blouse	Clothing	49	Oregon	M	Turquoise	Spring	4.5

```
In [3]: shop.shape
```

Out[3]: (3900, 18)

```
In [4]: shop.dtypes
```

Out[4]: 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

```
In [5]: shop.columns
```

```
Out[5]: 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')
```

```
In [6]: 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
```

```
In [7]: shop.describe()
```

Out[7]:

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

```
In [8]: shop.isnull().sum()
```

```
Out[8]: 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 [9]: print(f"The unique values of the 'Gender' column are: {shop['Gender'].unique()}")
        print()
        print(f"The unique values of the 'Category' column are: {shop['Category'].unique()}")
        print()
        print(f"The unique values of the 'Size' column are: {shop['Size'].unique()}")
        print()
        print(f"The unique values of the 'Subscription Status' column are: {shop['Subscription Status'].unique()}")
        print()
        print(f"The unique values of the 'Shipping Type' column are: {shop['Shipping Type'].unique()}")
        print()
        print(f"The unique values of the 'Discount Applied' column are: {shop['Discount Applied'].unique()}")
        print()
        print(f"The unique values of the 'Promo Code Used' column are: {shop['Promo Code Used'].unique()}")
        print()
        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?

```
In [10]: shop['Age'].value_counts()
```

```
Out[10]: Age
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
Name: count, dtype: int64
```

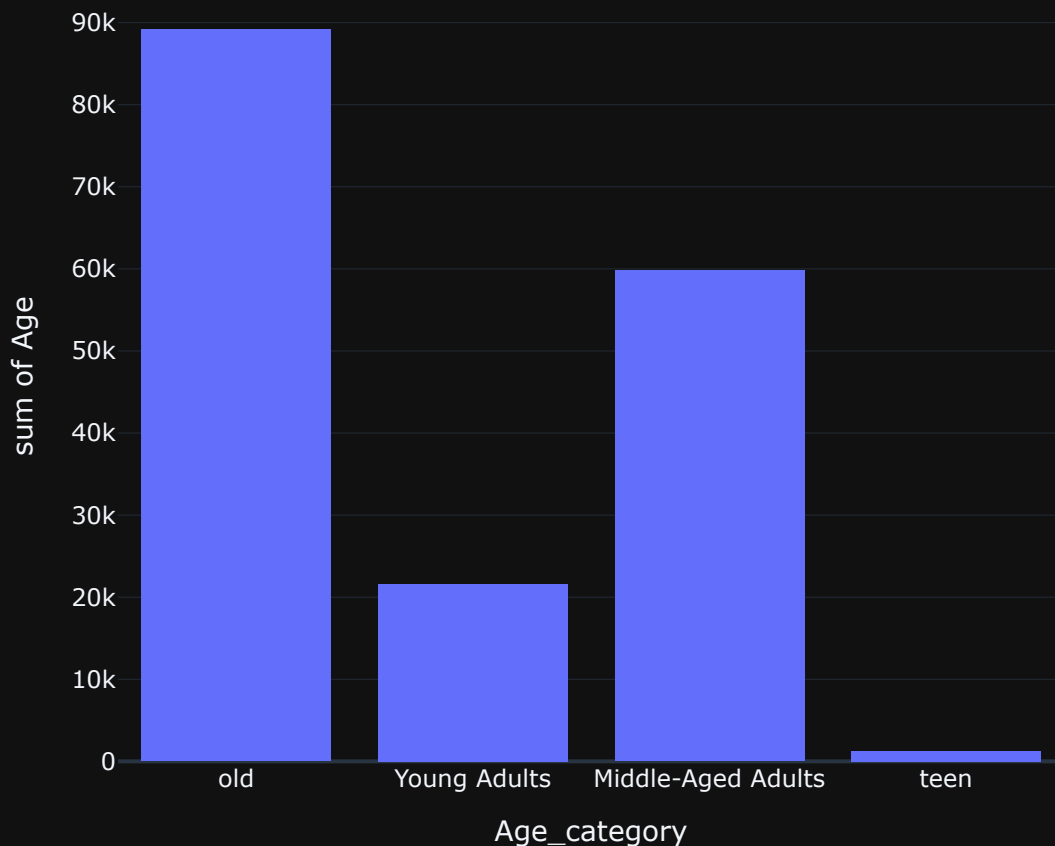
```
In [11]: shop['Age'].mean()
```

```
Out[11]: 44.06846153846154
```

```
In [12]: shop['Age_category'] = pd.cut(shop['Age'], bins= [0,15, 18 , 30 , 50 , 70] , labels= ['child'
```

```
In [13]: fig = px.histogram(shop , y = 'Age' , x = 'Age_category')
```

```
fig.show()
```



## 2 How does the average purchase amount vary across different product categories?

```
In [14]: shop.columns
```

```
Out[14]: 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')
```

```
In [15]: shop['Category'].unique()
```

```
Out[15]: array(['Clothing', 'Footwear', 'Outerwear', 'Accessories'], dtype=object)
```

```
In [16]: shop.groupby('Category')['Purchase Amount (USD)'].mean()
```

```
Out[16]: Category  
Accessories    59.838710  
Clothing       60.025331  
Footwear       60.255426  
Outerwear      57.172840  
Name: Purchase Amount (USD), dtype: float64
```

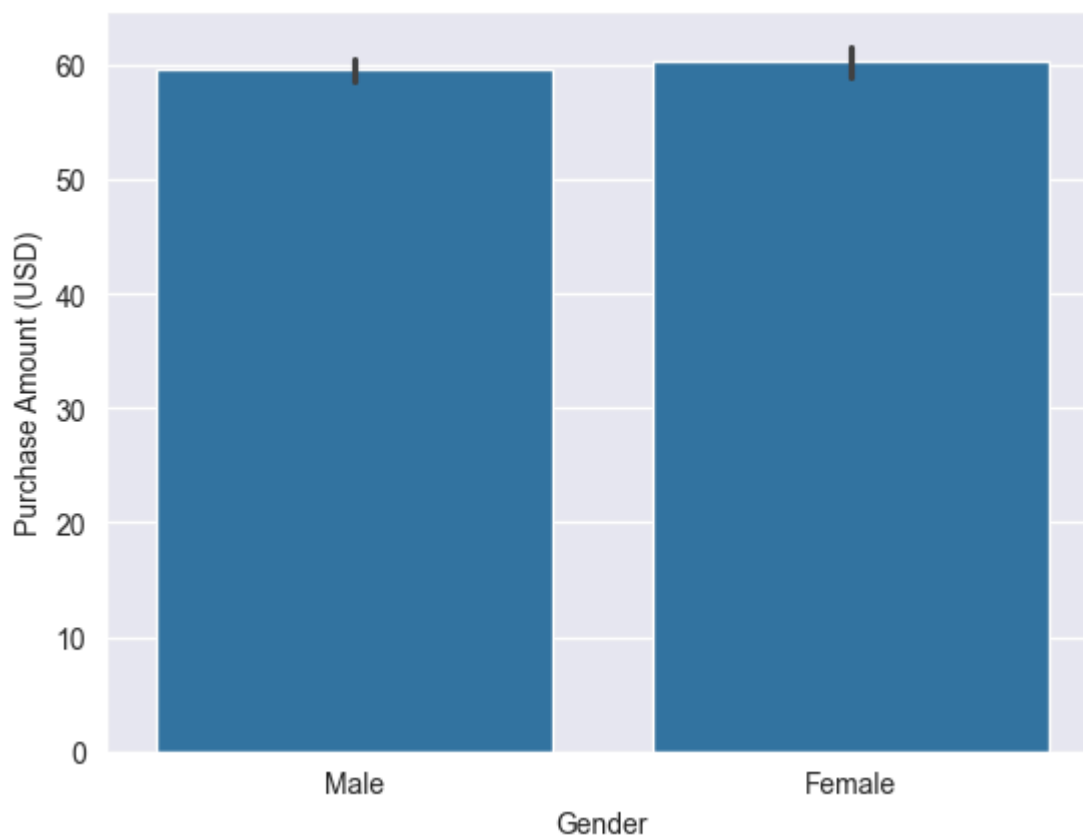
## 3 Which gender has the highest number of purchases?

```
In [17]: shop.columns
```

```
Out[17]: 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')
```

```
In [18]: sns.barplot(shop , x = 'Gender' , y = 'Purchase Amount (USD)')
```

```
Out[18]: <Axes: xlabel='Gender', ylabel='Purchase Amount (USD)'>
```



## 4 What are the most commonly purchased items in each category?

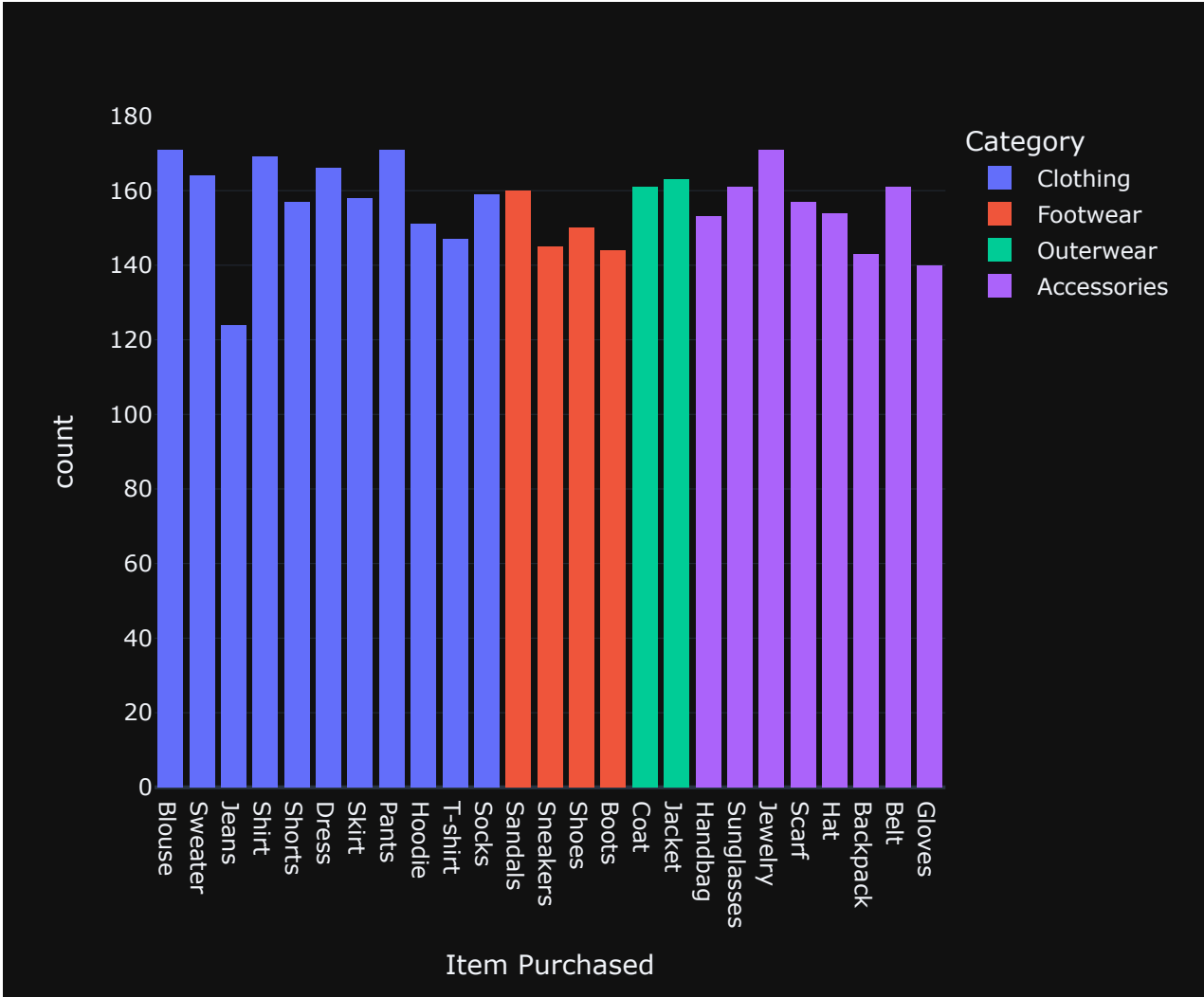
```
In [19]: shop.columns
```

```
Out[19]: 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')
```

```
In [20]: shop.groupby('Category')['Item Purchased'].value_counts()
```

```
Out[20]: 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
Name: count, dtype: int64
```

```
In [21]: 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?

```
In [22]: shop['Season'].unique()
```

```
Out[22]: array(['Winter', 'Spring', 'Summer', 'Fall'], dtype=object)
```

```
In [23]: shop[shop['Season'] == 'Summer'].value_counts().sum()
```

```
Out[23]: 955
```

```
In [24]: shop[shop['Season'] == 'Winter'].value_counts().sum()
```

```
Out[24]: 971
```

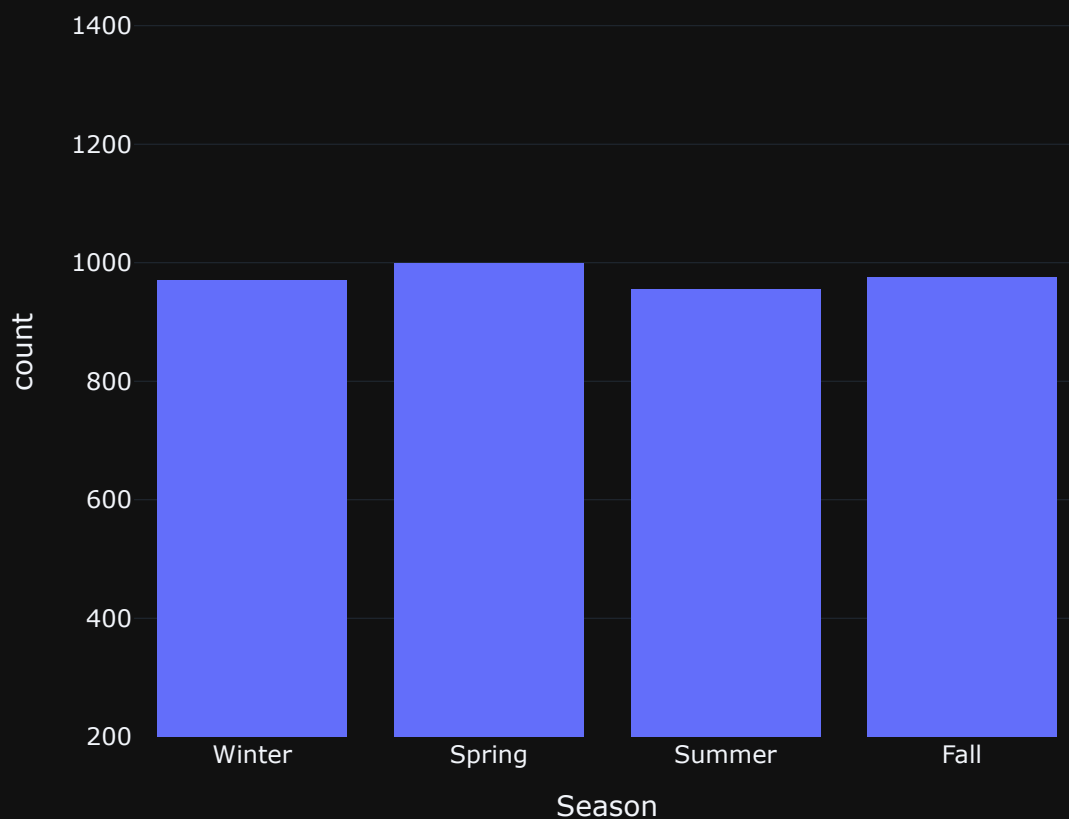
```
In [25]: shop[shop['Season'] == 'Spring'].value_counts().sum()
```

```
Out[25]: 999
```

```
In [26]: shop[shop['Season'] == 'Fall'].value_counts().sum()
```

```
Out[26]: 975
```

```
In [27]: fig = px.histogram(shop , x = 'Season' , range_y= [200 , 1500] )  
fig.show()
```

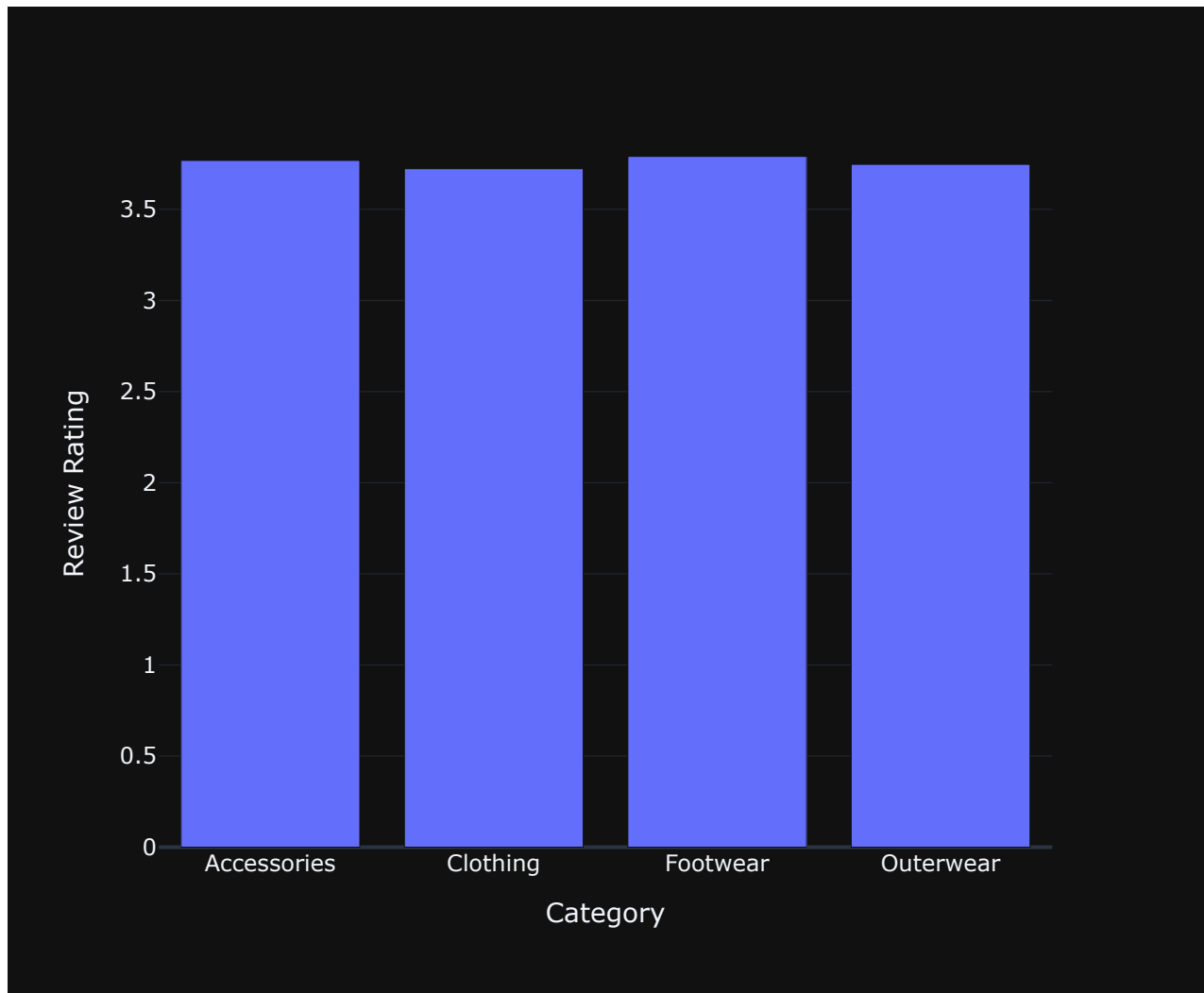




## 6 What is the average rating given by customers for each product category?

```
In [28]: shop.groupby('Category')['Review Rating'].mean().reset_index()
```

```
In [29]: 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?

```
In [30]: shop.columns
```

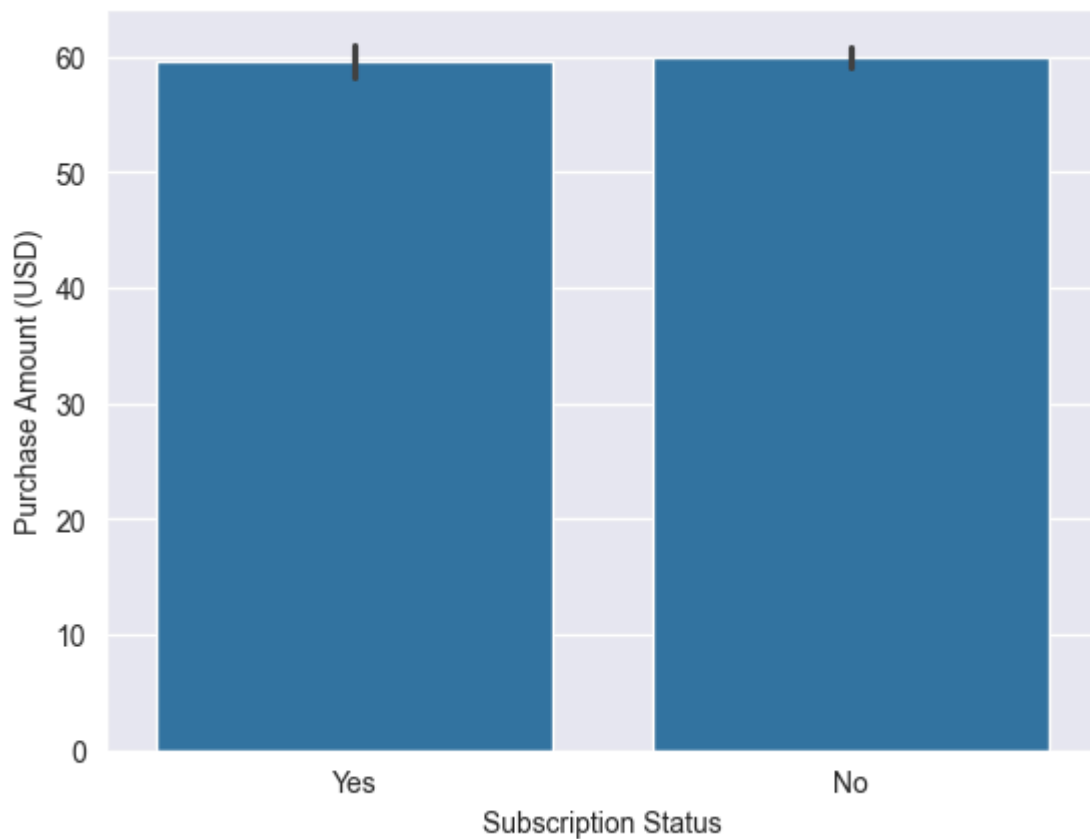
```
Out[30]: Index(['Customer ID', 'Age', 'Gender', 'Item Purchased', 'Category',  
              'Purchase Amount (USD)', 'Location', 'Size', 'Color', 'Season',  
              'Review Rating', 'Subscription Status', 'Shipping Type',  
              'Discount Applied', 'Promo Code Used', 'Previous Purchases',  
              'Payment Method', 'Frequency of Purchases', 'Age_category'],  
             dtype='object')
```

```
In [31]: shop['Subscription Status'].unique()
```

```
Out[31]: array(['Yes', 'No'], dtype=object)
```

```
In [32]: sns.barplot(shop , x = 'Subscription Status' , y = 'Purchase Amount (USD)')
```

```
Out[32]: <Axes: xlabel='Subscription Status', ylabel='Purchase Amount (USD)'>
```



```
In [33]: shop['Purchase Amount (USD)'].sum()
```

```
Out[33]: 233081
```

```
In [34]: shop.groupby('Subscription Status')['Purchase Amount (USD)'].mean()
```

```
Out[34]: Subscription Status
No      59.865121
Yes     59.491928
Name: Purchase Amount (USD), dtype: float64
```

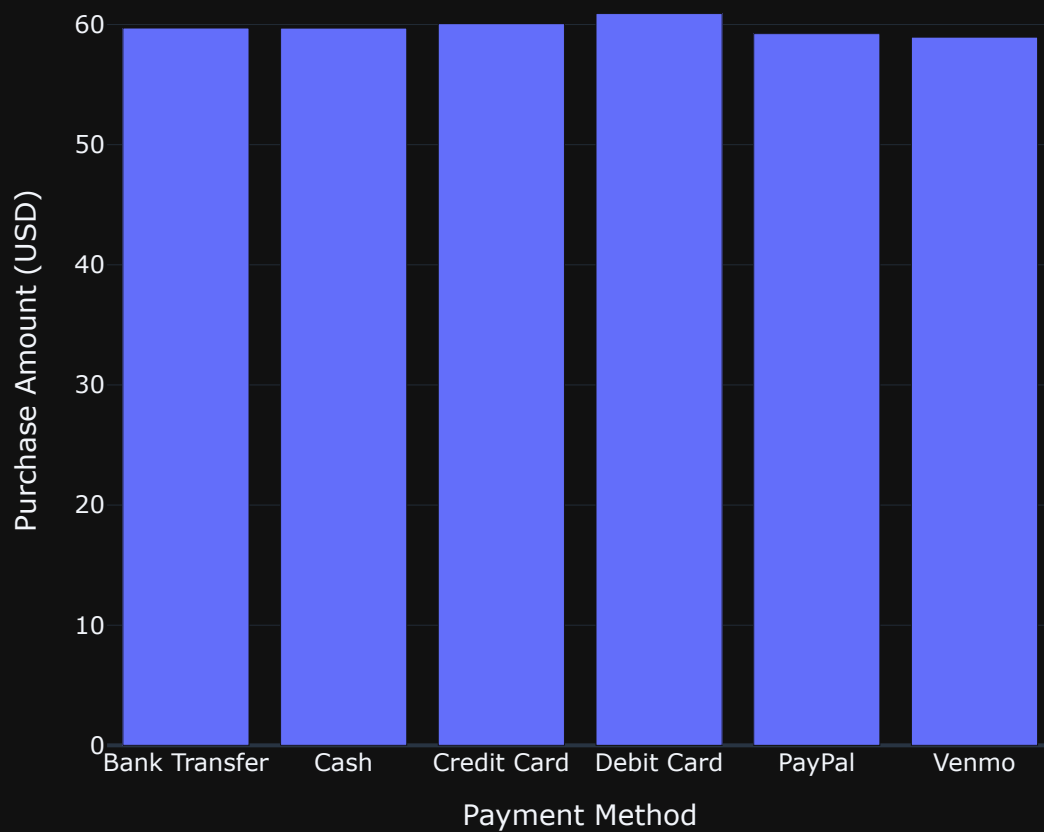
## 8 Which payment method is the most popular among customers?

```
In [35]: shop.groupby('Payment Method')['Purchase Amount (USD)'].mean().sort_values(ascending= False)
```

```
Out[35]: Payment Method
Debit Card      60.915094
Credit Card    60.074516
Bank Transfer   59.712418
Cash            59.704478
PayPal          59.245199
Venmo           58.949527
Name: Purchase Amount (USD), dtype: float64
```

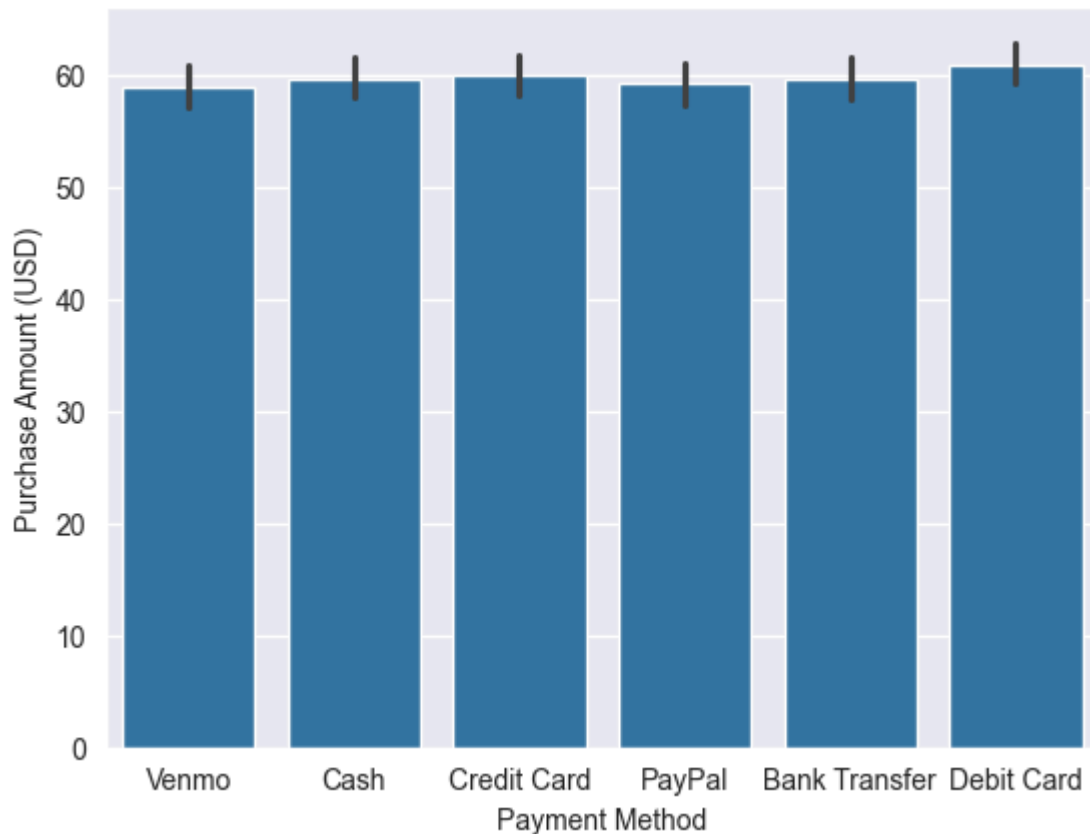
```
In [36]: shop_groupby = shop.groupby('Payment Method')['Purchase Amount (USD)'].mean().reset_index()
```

```
In [37]: fig = px.bar(shop_groupby , x = 'Payment Method' , y = 'Purchase Amount (USD)')
fig.show()
```



```
In [38]: sns.barplot(shop ,x='Payment Method' , y = 'Purchase Amount (USD)')
```

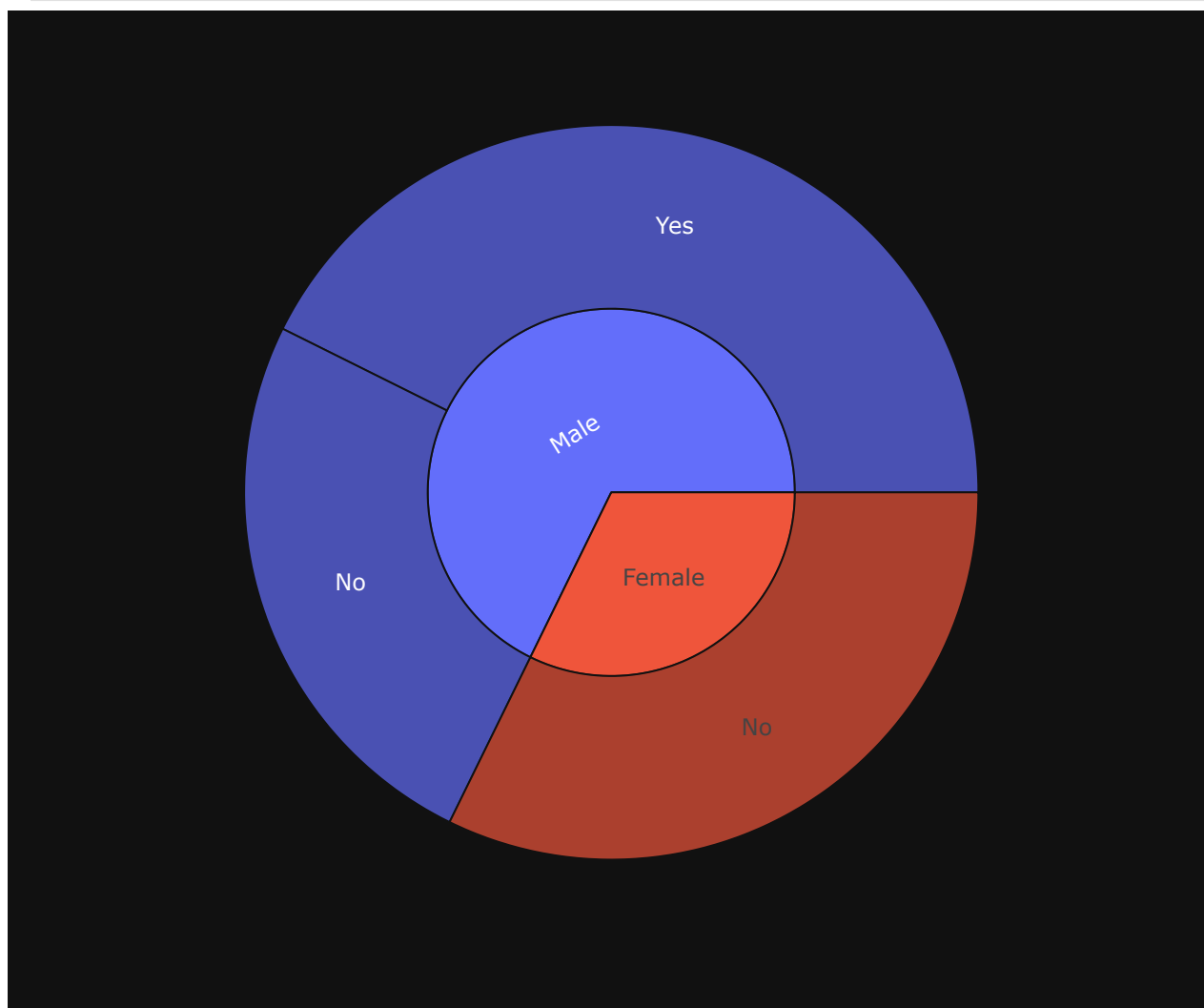
```
Out[38]: <Axes: xlabel='Payment Method', ylabel='Purchase Amount (USD)'>
```



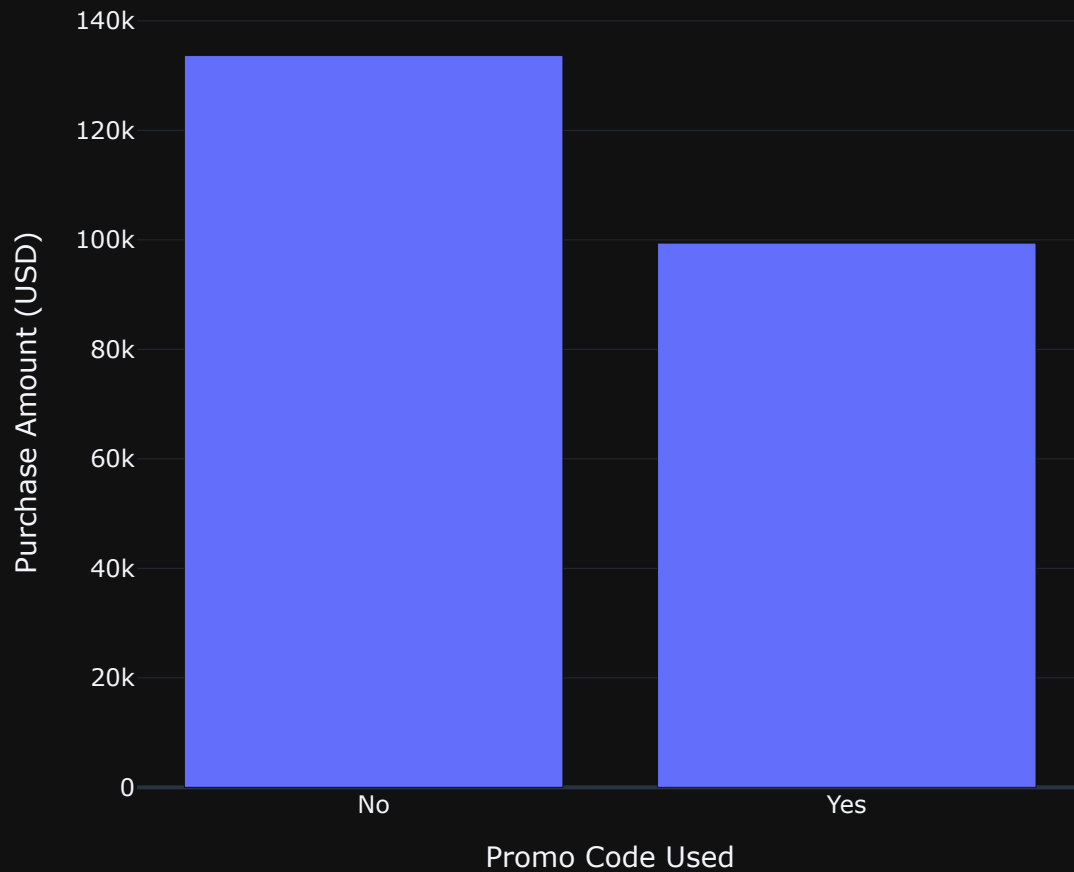
9 Do customers who use promo codes tend to spend more than those who don't?

```
In [39]: shop_groupby = shop.groupby('Promo Code Used')['Purchase Amount (USD)'].sum().reset_index()
```

```
In [40]: fig = px.sunburst(shop , path=['Gender' , 'Promo Code Used'] , values='Purchase Amount (USD)')  
fig.show()
```



```
In [41]: 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?

```
In [42]: shop[['Age' , 'Age_category']]
```

```
Out[42]:
```

	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

```
In [43]: shop['Age_category'].unique()
```

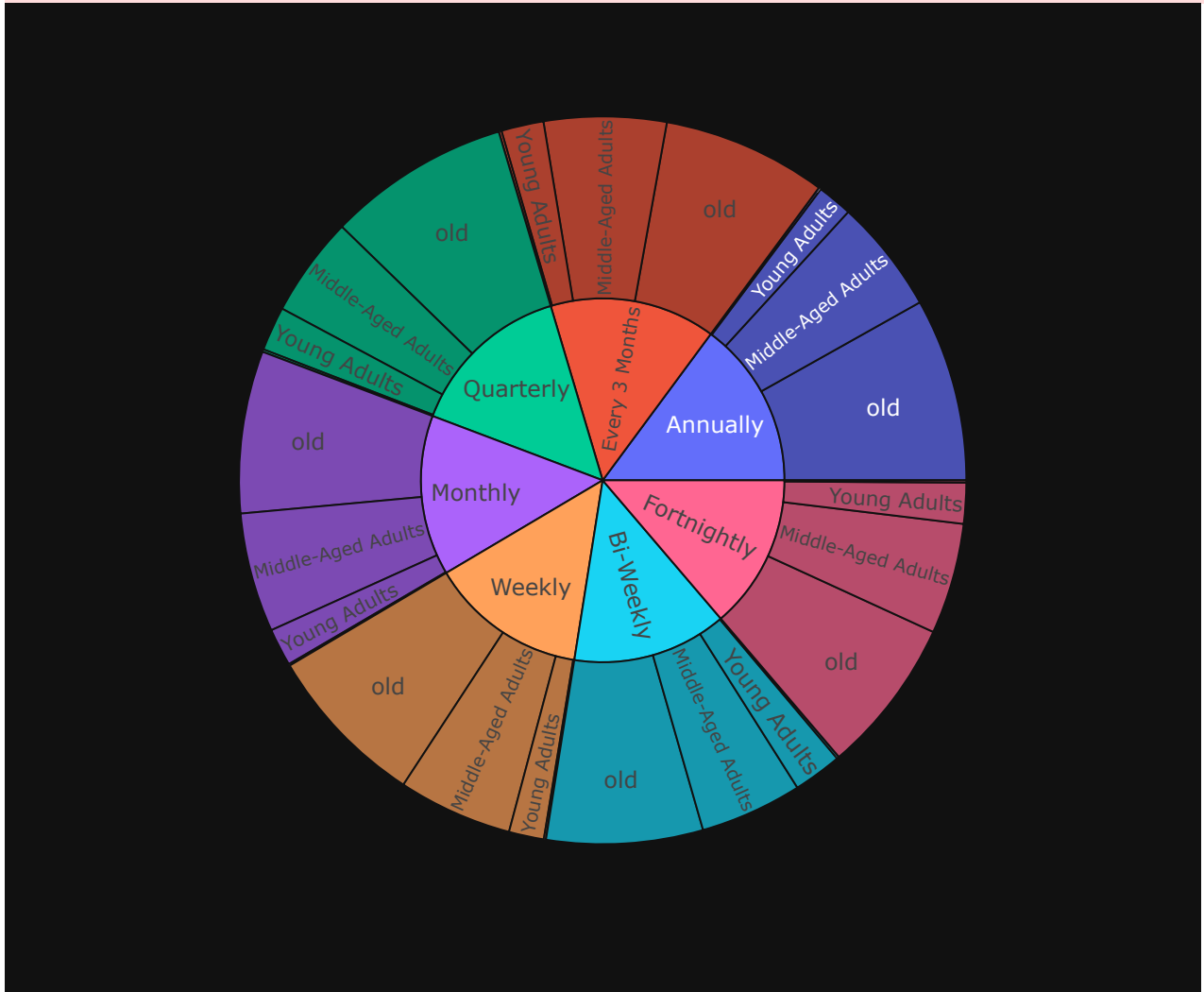
```
Out[43]: ['old', 'Young Adults', 'Middle-Aged Adults', 'teen']
Categories (5, object): ['child' < 'teen' < 'Young Adults' < 'Middle-Aged Adults' < 'old']
```

```
In [44]: shop_group = shop.groupby('Frequency of Purchases')['Age'].sum()
```

```
In [45]: px.sunburst(shop , path=['Frequency of Purchases', 'Age_category'] , values='Age')
```

C:\Users\sharv\AppData\Local\Programs\Python\Python310\lib\site-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 current behavior or observed=True to adopt the future default and silence this warning.



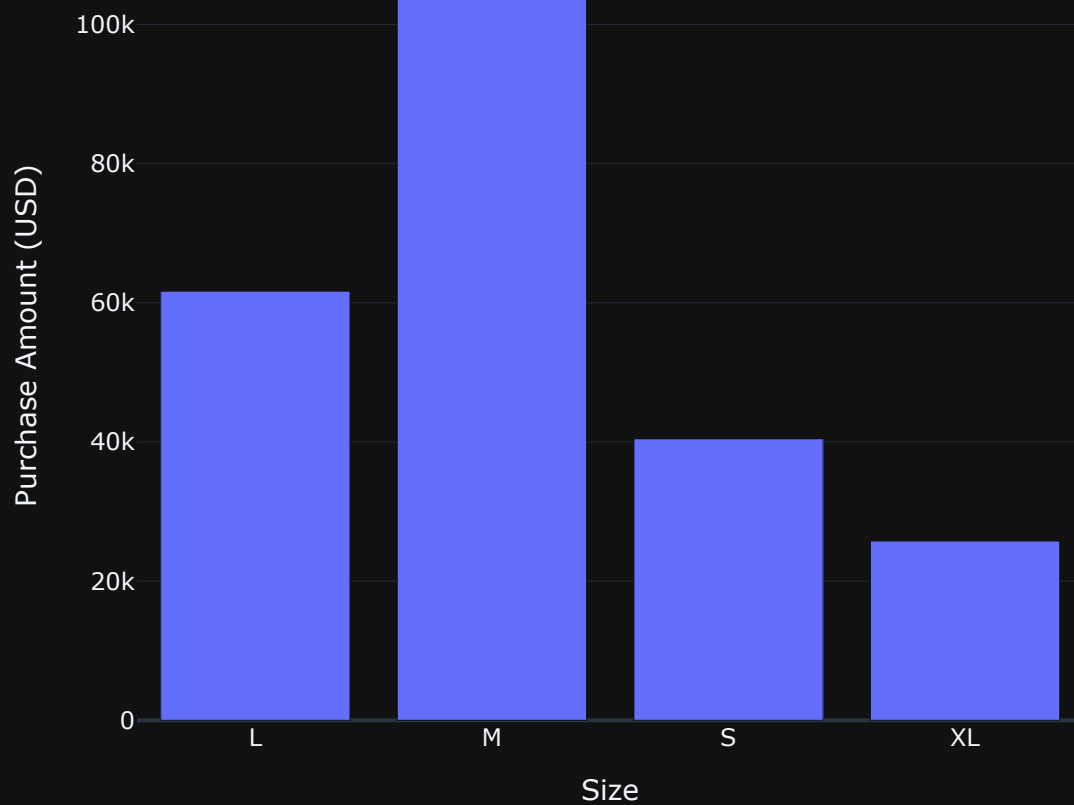
## 11 Are there any correlations between the size of the product and the purchase amount?

```
In [46]: shop.columns
```

```
Out[46]: 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')
```

```
In [47]: shop_group = shop.groupby('Size')['Purchase Amount (USD)'].sum().reset_index()
```

```
In [48]: 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?

```
In [49]: shop.groupby('Category')['Shipping Type'].value_counts().sort_values(ascending=False)
```

```
Out[49]: Category Shipping Type
Clothing Standard 297
Clothing Free Shipping 294
Clothing Next Day Air 293
Clothing Express 290
Clothing Store Pickup 282
Clothing 2-Day Shipping 281
Accessories Store Pickup 217
Accessories Next Day Air 211
Accessories Standard 208
Accessories 2-Day Shipping 206
Accessories Express 203
Footwear Free Shipping 195
Footwear Free Shipping 122
Footwear Standard 100
Footwear Store Pickup 98
Footwear Express 96
Footwear Next Day Air 93
Footwear 2-Day Shipping 90
Outerwear Free Shipping 64
Outerwear Express 57
Outerwear Store Pickup 53
Outerwear Next Day Air 51
Outerwear 2-Day Shipping 50
Outerwear Standard 49
Name: count, dtype: int64
```

```
In [50]: shop['Shipping_Category'] =shop['ShippingType'].map({'Express': 0, 'Free Shipping': 1, 'Next
```

```
In [51]: shop['Category'].unique()
```

```
Out[51]: array(['Clothing', 'Footwear', 'Outerwear', 'Accessories'], dtype=object)
```

```
In [52]: shop['Category_num'] =shop['Category'].map({'Clothing':1, 'Footwear':2, 'Outerwear':3, 'Acces
```

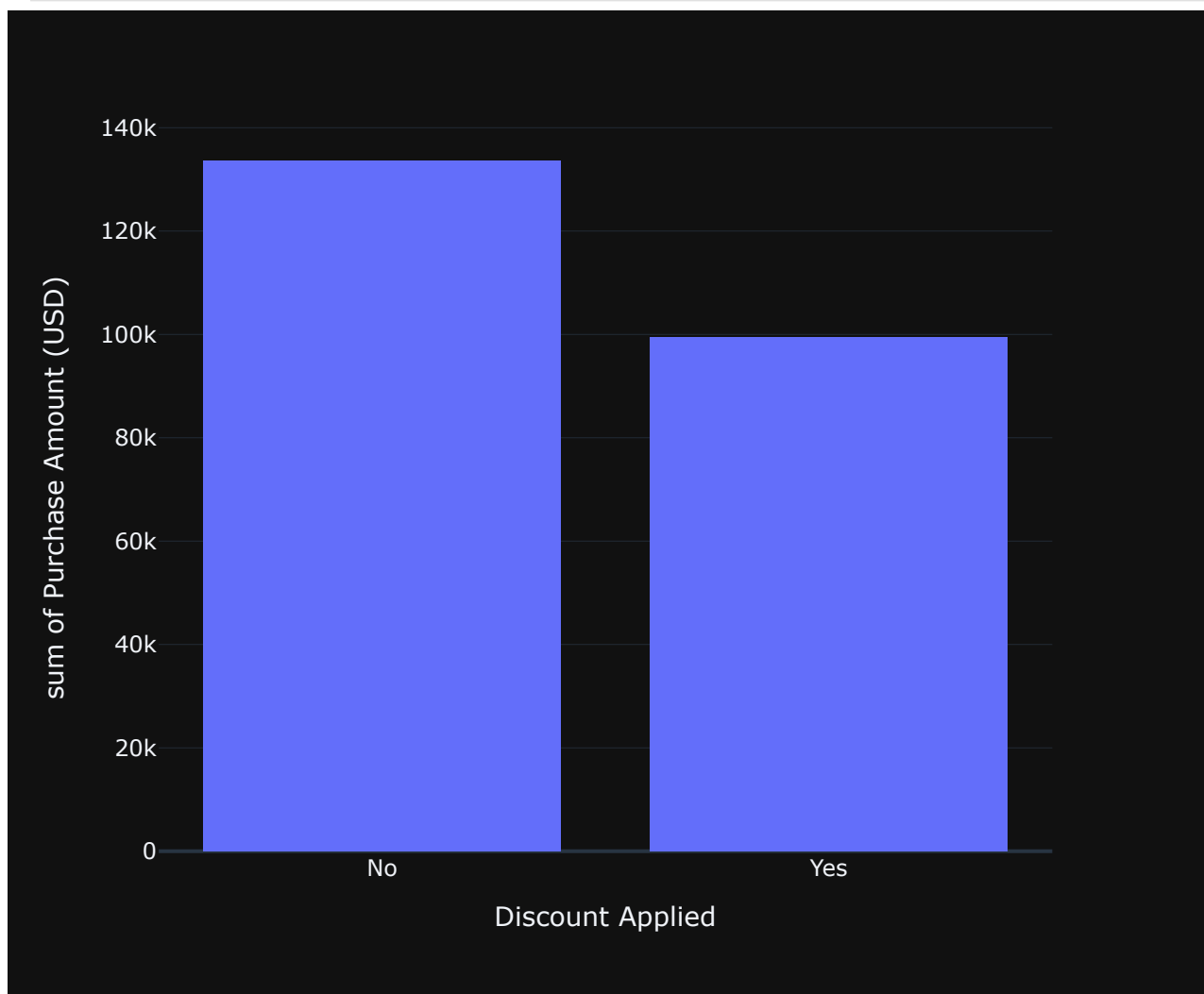
## 13 How does the presence of a discount affect the purchase decision of customers?

```
In [53]: shop.columns
```

```
Out[53]: 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',  
              'Shipping_Category', 'Category_num'],  
             dtype='object')
```

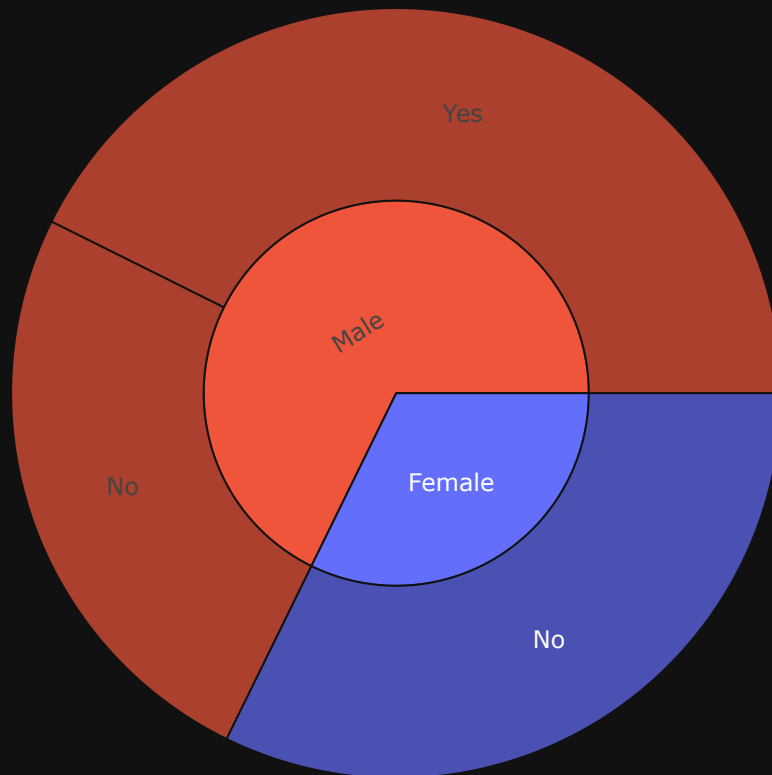
```
In [54]: shop_group = shop.groupby('Discount Applied')['Purchase Amount (USD)'].sum().reset_index()
```

```
In [55]: px.histogram(shop_group , x = 'Discount Applied' , y = 'Purchase Amount (USD)')
```



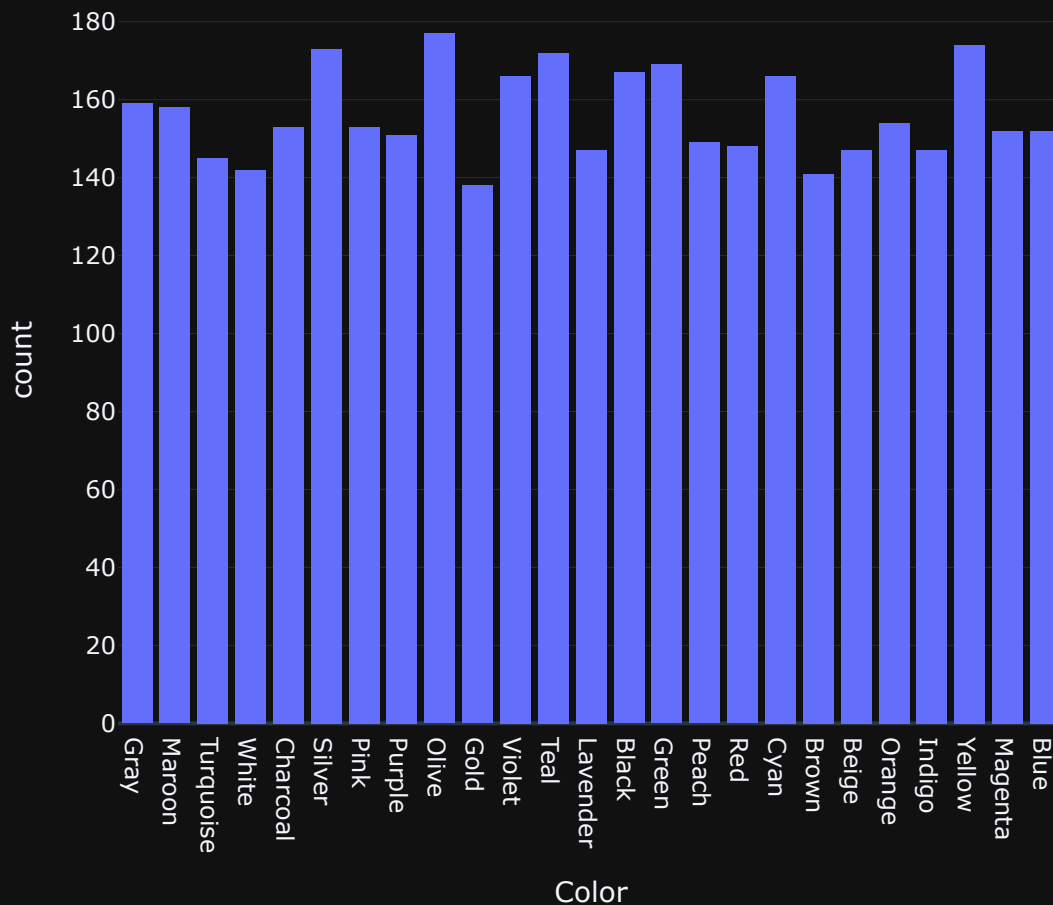
```
In [56]: fig = px.sunburst(shop , path = ['Gender' , 'Discount Applied'], values='Purchase Amount (USD)  
fig.show()
```





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

```
In [57]: px.histogram(shop , x = 'Color')
```



```
In [58]: shop['Color'].value_counts().nlargest(5)
```

```
Out[58]: Color
Olive      177
Yellow     174
Silver     173
Teal       172
Green      169
Name: count, dtype: int64
```

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

```
In [59]: shop['Previous Purchases'].mean()
```

```
Out[59]: 25.35153846153846
```

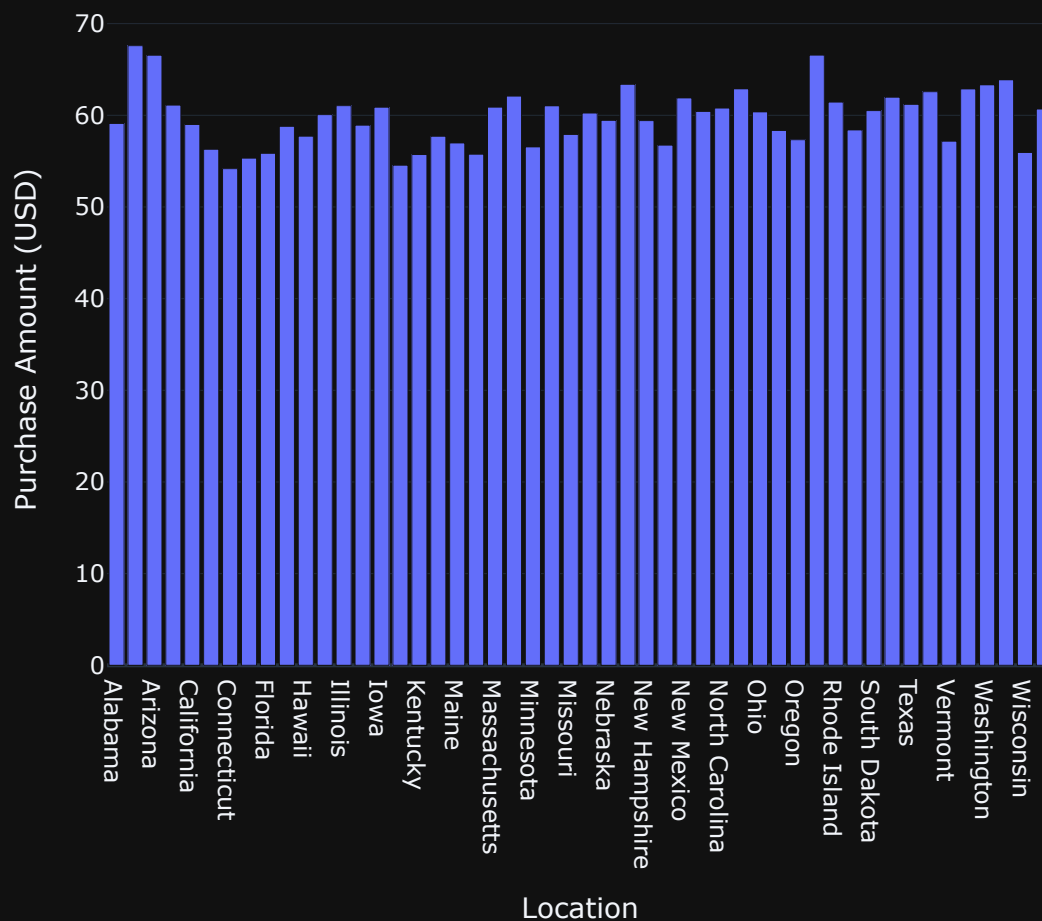
**16 Are there any noticeable differences in purchase behavior between different locations?**

```
In [60]: shop.groupby('Location')['Purchase Amount (USD)'].mean().sort_values(ascending = False)
```

```
Out[60]: 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
Maryland     55.755814
Kentucky     55.721519
Delaware     55.325581
Kansas       54.555556
Connecticut  54.179487
Name: Purchase Amount (USD), dtype: float64
```

```
In [61]: shop_group = shop.groupby('Location')['Purchase Amount (USD)'].mean().reset_index()
```

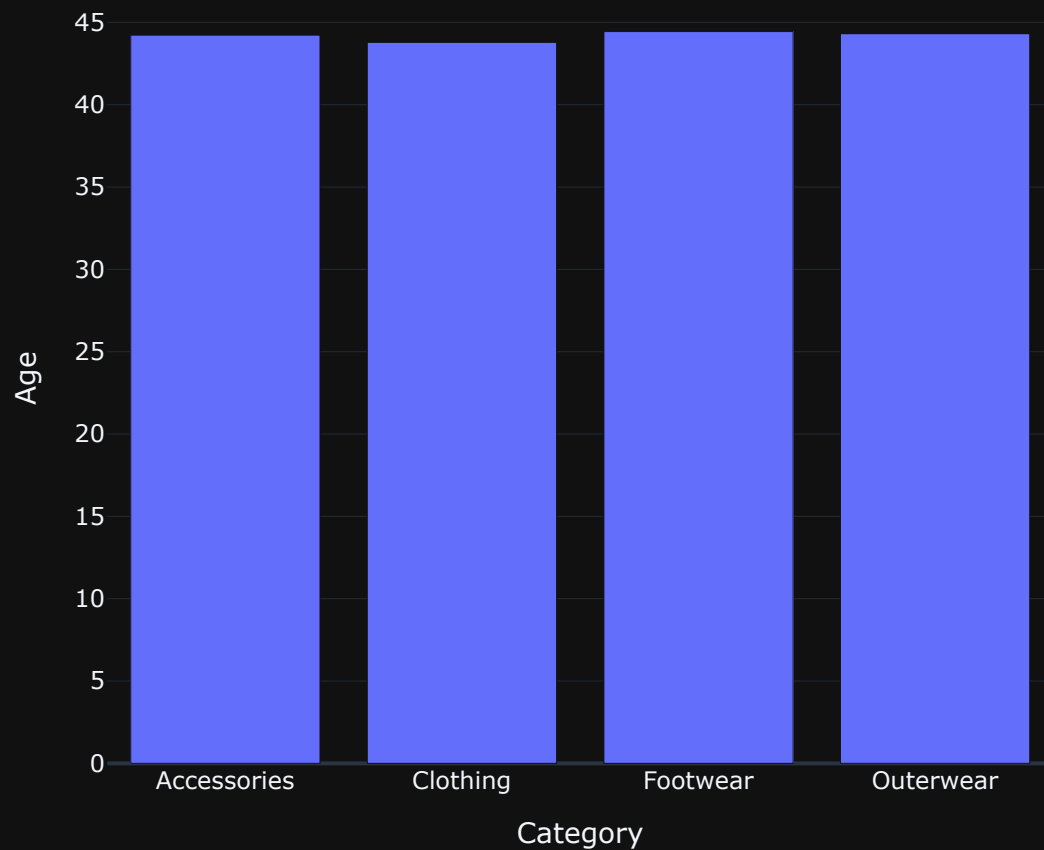
```
In [62]: fig = px.bar(shop_group, x = 'Location' , y = 'Purchase Amount (USD)')
fig.show()
```



## 17 Is there a relationship between customer age and the category of products they purchase?

```
In [63]: shop_group = shop.groupby('Category')['Age'].mean().reset_index()
```

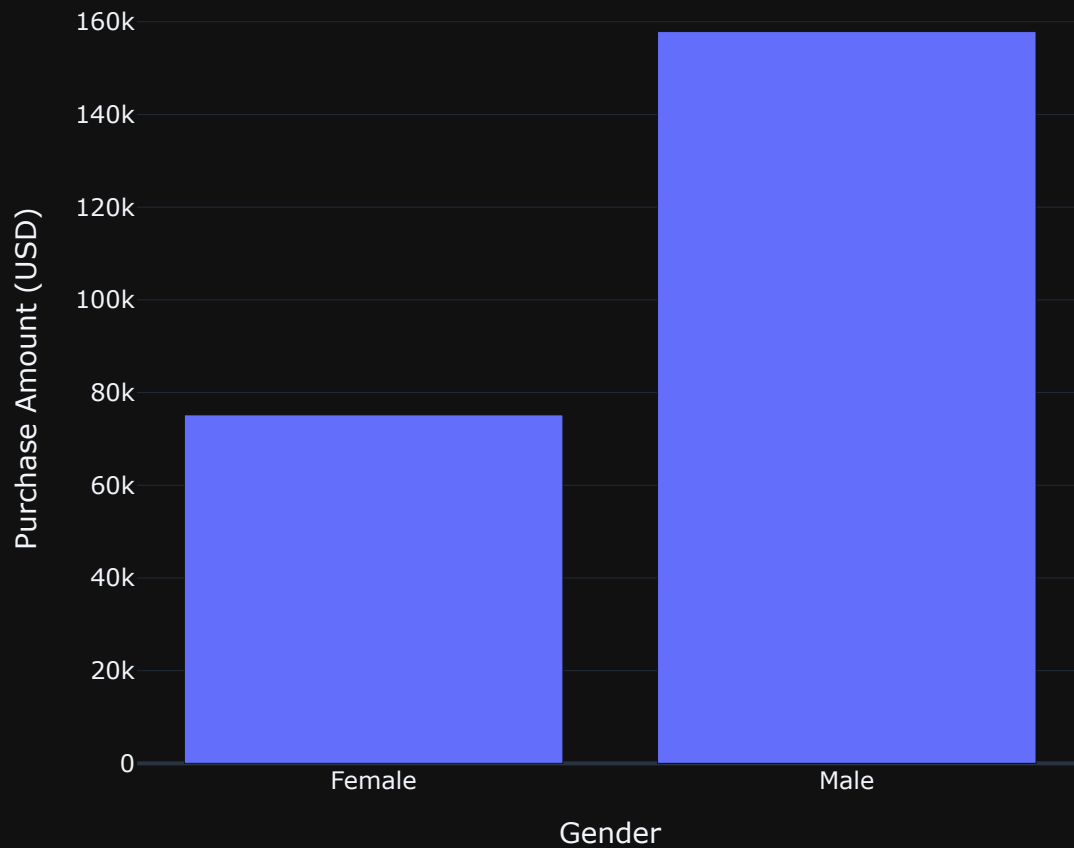
```
In [64]: fig = px.bar(shop_group, y = 'Age', x = 'Category')  
fig.show()
```



## 18 How does the average purchase amount differ between male and female customers?

```
In [65]: shop_group = shop.groupby('Gender')['Purchase Amount (USD)'].sum().reset_index()
```

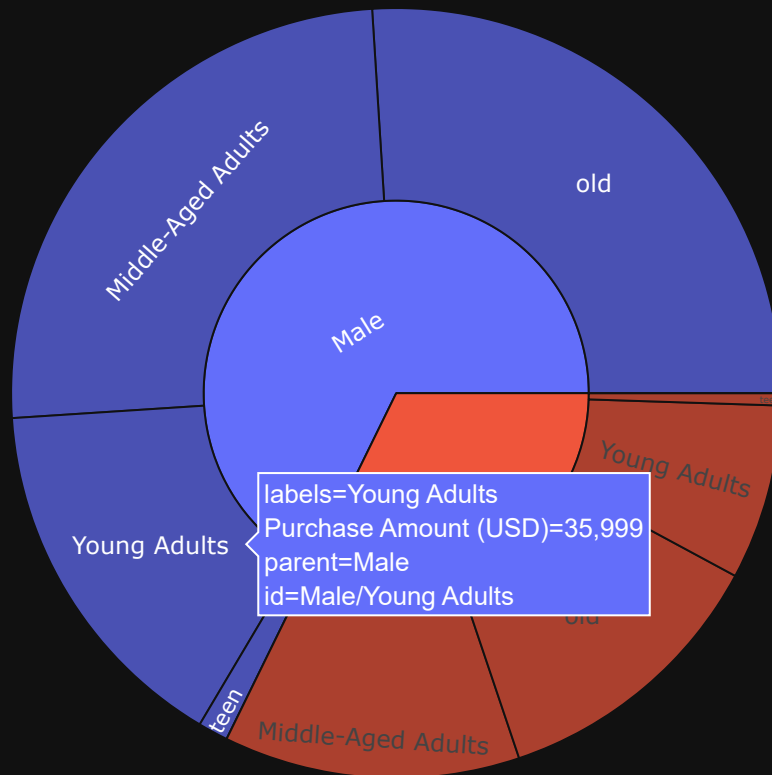
```
In [66]: fig = px.bar(shop_group , x = 'Gender' , y = 'Purchase Amount (USD)')  
fig.show()
```



```
In [67]: px.sunburst(data_frame= shop , path = ['Gender' , 'Age_category'] , values='Purchase Amount (U
```

C:\Users\sharv\AppData\Local\Programs\Python\Python310\lib\site-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 current behavior or observed=True to adopt the future default and silence this warning.



EOF