Identifying Shopping Trends using Data Analysis

📌 **Customer ID** - Unique identifier for each customer.
📌 **Age** - Age of the customer.
📌 **Gender** - Gender of the customer (Male/Female).
📌 **Item Purchased** - The item purchased by the customer.
📌 **Category** - Category of the item purchased.
📌 **Purchase Amount (USD)** - The amount of the purchase in USD.
📌 **Location** - Location where the purchase was made.
📌 **Size** - Size of the purchased item.
📌 **Color** - Color of the purchased item.
📌 **Season** - Season during which the purchase was made.
📌 **Review Rating** - Rating given by the customer for the purchased item.
📌 **Subscription Status** - Indicates if the customer has a subscription (Yes/No).
📌 **Shipping Type** - Type of shipping chosen by the customer.
📌 **Discount Applied** - Indicates if a discount was applied to the purchase (Yes/No).
📌 **Promo Code Used** - Indicates if a promo code was used for the purchase (Yes/No).
📌 **Previous Purchases** - Number of previous purchases made by the customer.
📌 **Payment Method** - Customer's most preferred payment method.
📌 **Frequency of Purchases** - Frequency at which the customer makes purchases (e.g., Weekly, Fortnightly, Monthly).

In [1]:
```
pip install WordCloud
```

```
Defaulting to user installation because normal site-packages is not write
able
Requirement already satisfied: WordCloud in c:\users\aryan\appdata\roamin
g\python\python312\site-packages (1.9.4)
Requirement already satisfied: numpy>=1.6.1 in c:\programdata\anaconda3\l
ib\site-packages (from WordCloud) (1.26.4)
Requirement already satisfied: pillow in c:\programdata\anaconda3\lib\sit
e-packages (from WordCloud) (10.4.0)
Requirement already satisfied: matplotlib in c:\programdata\anaconda3\lib
\site-packages (from WordCloud) (3.9.2)
Requirement already satisfied: contourpy>=1.0.1 in c:\programdata\anacond
a3\lib\site-packages (from matplotlib->WordCloud) (1.2.0)
Requirement already satisfied: cycler>=0.10 in c:\programdata\anaconda3\l
ib\site-packages (from matplotlib->WordCloud) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\programdata\anacon
da3\lib\site-packages (from matplotlib->WordCloud) (4.51.0)
Requirement already satisfied: kiwisolver>=1.3.1 in c:\programdata\anacon
da3\lib\site-packages (from matplotlib->WordCloud) (1.4.4)
Requirement already satisfied: packaging>=20.0 in c:\users\aryan\appdata
\roaming\python\python312\site-packages (from matplotlib->WordCloud) (24.
2)
Requirement already satisfied: pyparsing>=2.3.1 in c:\programdata\anacond
a3\lib\site-packages (from matplotlib->WordCloud) (3.1.2)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\aryan\app
data\roaming\python\python312\site-packages (from matplotlib->WordCloud)
(2.9.0.post0)
Requirement already satisfied: six>=1.5 in c:\users\aryan\appdata\roaming
\python\python312\site-packages (from python-dateutil>=2.7->matplotlib->W
ordCloud) (1.17.0)
Note: you may need to restart the kernel to use updated packages.
```

In [2]:
```
# importing libraries
import numpy as np # Importing the numpy library for array operations and m
import pandas as pd # Use for exploring the data
import seaborn as sns # it has also plot
import matplotlib.pyplot as plt # for some extra plot functions
import plotly.express as px # this library can makes interactive plot
import warnings
warnings.simplefilter(action='ignore', category=FutureWarning)
```

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

In [3]:
```
# reading the data set
shop = pd.read_csv('shopping_trends.csv')
```

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

Out[4]:
```
(3900, 19)
```

In [5]:
```
shop.to_excel('shopping_trends_updated.xlsx')
```

In [6]:
```
shop.head()
```

Out[6]:

	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 [7]:
```
shop.dtypes
```

Out[7]:
```
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
Payment Method             object
Shipping Type              object
Discount Applied           object
Promo Code Used            object
Previous Purchases          int64
Preferred Payment Method   object
Frequency of Purchases     object
dtype: object
```

In [8]:
```
# it shows the names of the columns
shop.columns
```

Out[8]:
```
Index(['Customer ID', 'Age', 'Gender', 'Item Purchased', 'Category',
       'Purchase Amount (USD)', 'Location', 'Size', 'Color', 'Season',
       'Review Rating', 'Subscription Status', 'Payment Method',
       'Shipping Type', 'Discount Applied', 'Promo Code Used',
       'Previous Purchases', 'Preferred Payment Method',
       'Frequency of Purchases'],
      dtype='object')
```

In [9]:
```
shop.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3900 entries, 0 to 3899
Data columns (total 19 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  Payment Method            3900 non-null   object 
 13  Shipping Type             3900 non-null   object 
 14  Discount Applied          3900 non-null   object 
 15  Promo Code Used           3900 non-null   object 
 16  Previous Purchases        3900 non-null   int64  
 17  Preferred Payment Method  3900 non-null   object 
 18  Frequency of Purchases    3900 non-null   object 
dtypes: float64(1), int64(4), object(14)
memory usage: 579.0+ KB
```

In [10]:
```
shop.shape
```

Out[10]:
```
(3900, 19)
```

In [11]:
```
shop.isnull().sum()
```

Out[11]:
```
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
Payment Method              0
Shipping Type               0
Discount Applied            0
Promo Code Used             0
Previous Purchases          0
Preferred Payment Method    0
Frequency of Purchases      0
dtype: int64
```

In [12]:
```
print(f"The unique values of the 'Gender' column are: {shop['Gender'].uniqu
print()# This will print a blank line
print(f"The unique values of the 'Category' column are: {shop['Category'].u
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['S
print()# This will print a blank line
print(f"The unique values of the 'Shipping Type' column are: {shop['Shippir
print()# This will print a blank line
print(f"The unique values of the 'Discount Applied' column are: {shop['Disc
print()# This will print a blank line
print(f"The unique values of the 'Promo Code Used' column are: {shop['Promc
print()# This will print a blank line
print(f"The unique values of the 'Payment Method' column are: {shop['Paymer
```

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

The unique values of the 'Category' column are: ['Clothing' 'Footwear' 'O
uterwear' '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 Shi
pping' '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: ['Credit Card' 'Ban
k Transfer' 'Cash' 'PayPal' 'Venmo' 'Debit Card']
```

OBSERVATION:

Upon initial examination of the dataset, it is evident that we have a comprehensive and well-structured dataset with 3900 rows and 18 columns. The data is complete, with no missing values, which allows us to proceed confidently with our analysis.

Let's delve into the columns and their significance in understanding our custome

- **Customer ID:** This column serves as a unique identifier for each customer, enabling us to differentiate between individuals.
- **Age:** The age column provides insights into the age demographics of our customers, helping us understand their preferences and behaviors.
- **Gender:** This column showcases the gender of the customers, enabling us to analyze buying patterns based on gender.
- **Item Purchased:** Here, we can identify the specific products that customers have bought, allowing us to gain an understanding of popular choices.
- **Category:** The category column categorizes the products into different groups such as clothing, footwear, and more, aiding us in analyzing trends within specific product categories.

- **Purchase Amount (USD):** This column reveals the amount customers spent on their purchases, providing insights into their spending habits.
- **Location:** The location column indicates the geographical location of customers, which can help identify regional trends and preferences.
- **Size:** This column denotes the size of the purchased products, assisting in understanding size preferences across different categories.
- **Color:** Here, we can determine the color preferences of customers, aiding in analyzing color trends and their impact on purchasing decisions.
- **Season:** The season column allows us to identify the season during which customers made their purchases, enabling us to explore seasonal shopping trends.
- **Review Rating:** This column showcases the ratings given by customers, providing valuable feedback on product satisfaction and quality.
- **Subscription Status:** This column indicates whether customers have opted for a subscription status, which can help us understand customer loyalty and engagement.
- **Shipping Type:** Here, we can identify the different shipping methods used to deliver products to customers, shedding light on preferred shipping options.
- **Discount Applied:** This column indicates whether a discount was applied to the purchased products, enabling us to analyze the impact of discounts on customer behavior.
- **Promo Code Used:** Here, we can identify whether customers utilized promo codes during their purchases, helping us evaluate the effectiveness of promotional campaigns.
- **Previous Purchases:** This column reveals the number of previous purchases made by customers, aiding in understanding customer loyalty and repeat business.
- **Payment Method:** The payment method column showcases the various methods used by customers to make their purchases, allowing us to analyze preferred payment options.
- **Frequency of Purchases:** This column provides insights into the frequency at which customers make purchases, helping us identify patterns and customer buying habits.

ustomer buying habits. With this rich and diverse dataset, we are well-equipped to explore customer shopping trends, understand their preferences, and uncover valuable insights that can drive informed decision-making and enhance the overall customer experience. Let's

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

In [13]:
```
shop['Age'].value_counts()
```

Out[13]:
```
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: Age, dtype: int64
```

In [14]:
```
shop['Age'].mean()
```

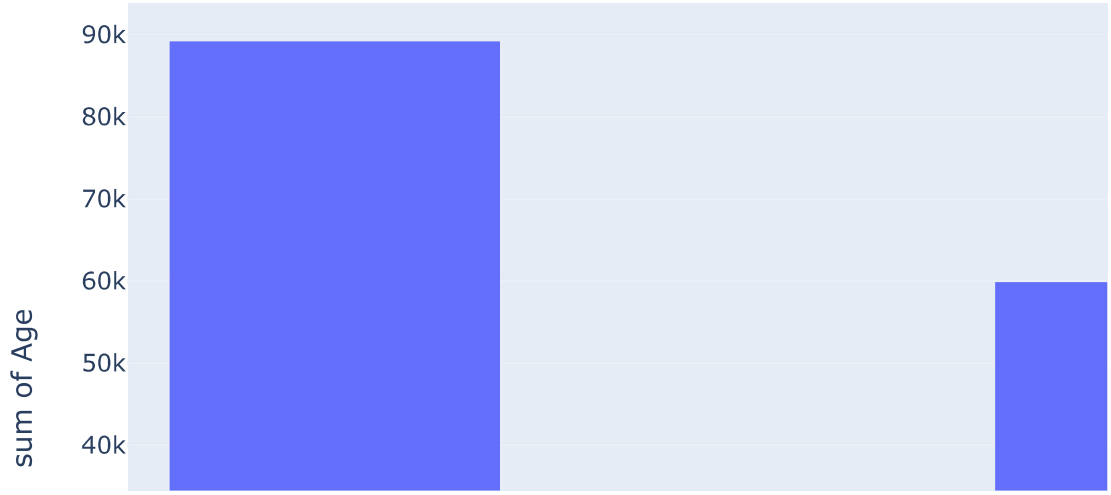
Out[14]:
```
44.06846153846154
```

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

Out[15]:
```
array(['Male', 'Female'], dtype=object)
```

In [16]:
```
shop['Age_category'] = pd.cut(shop['Age'], bins= [0,15, 18 , 30 , 50 , 70]
```

In [17]:
```
fig = px.histogram(shop , y = 'Age' , x = 'Age_category')
fig.show()
```



In [18]:
```
shop.columns
```

Out[18]:
```
Index(['Customer ID', 'Age', 'Gender', 'Item Purchased', 'Category',
       'Purchase Amount (USD)', 'Location', 'Size', 'Color', 'Season',
       'Review Rating', 'Subscription Status', 'Payment Method',
       'Shipping Type', 'Discount Applied', 'Promo Code Used',
       'Previous Purchases', 'Preferred Payment Method',
       'Frequency of Purchases', 'Age_category'],
      dtype='object')
```

In [19]:
```
shop['Category'].unique()
```

Out[19]:
```
array(['Clothing', 'Footwear', 'Outerwear', 'Accessories'], dtype=object)
```

In [20]:
```
shop.groupby('Category')['Purchase Amount (USD)'].mean()
```

Out[20]:
```
Category
Accessories    59.838710
Clothing       60.025331
Footwear       60.255426
Outerwear      57.172840
Name: Purchase Amount (USD), dtype: float64
```

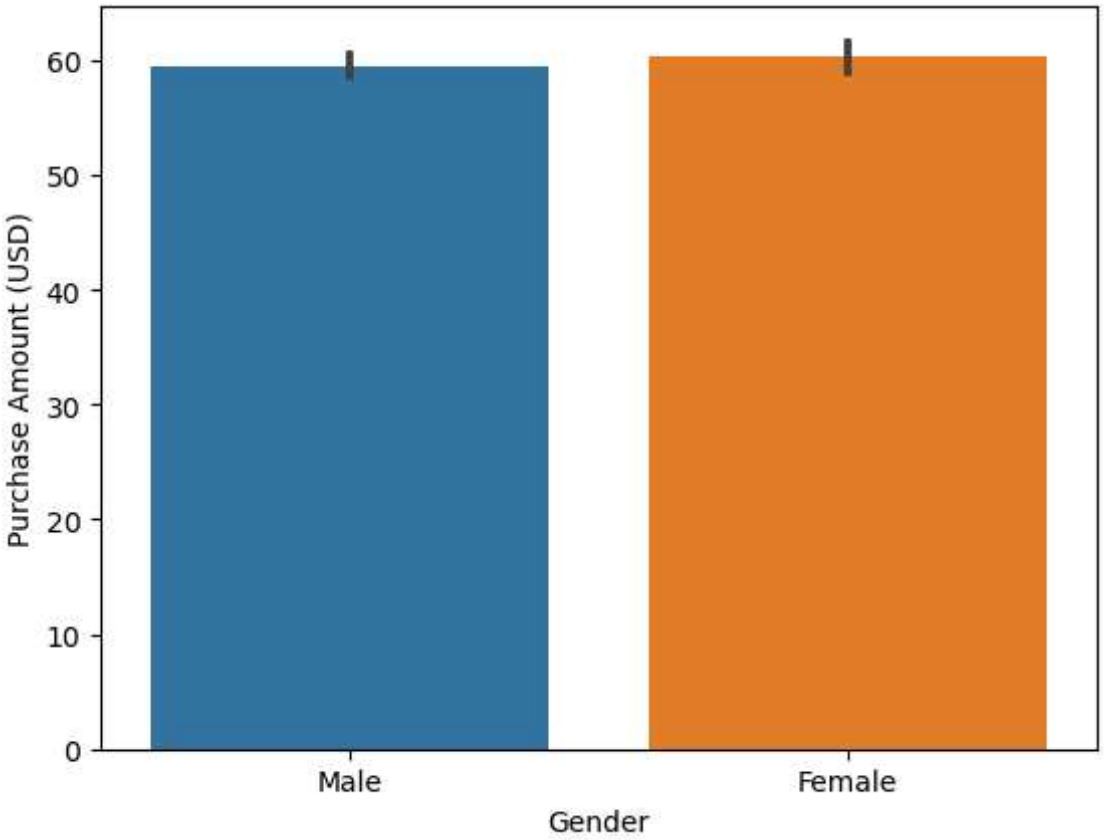
3 Which gender has the highest number of purchases?

In [21]:
```
shop.columns
```

Out[21]:
```
Index(['Customer ID', 'Age', 'Gender', 'Item Purchased', 'Category',
       'Purchase Amount (USD)', 'Location', 'Size', 'Color', 'Season',
       'Review Rating', 'Subscription Status', 'Payment Method',
       'Shipping Type', 'Discount Applied', 'Promo Code Used',
       'Previous Purchases', 'Preferred Payment Method',
       'Frequency of Purchases', 'Age_category'],
      dtype='object')
```

In [22]:
```
sns.barplot(shop , x = 'Gender' , y = 'Purchase Amount (USD)')
```

Out[22]:
```
<Axes: xlabel='Gender', ylabel='Purchase Amount (USD)'>
```



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

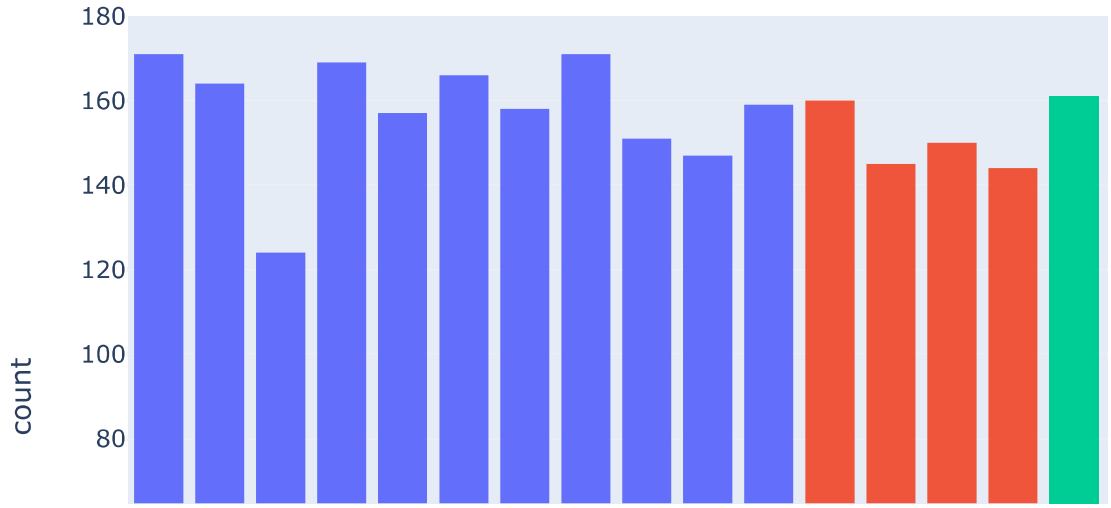
In [23]:
```
shop.columns
```

Out[23]:
```
Index(['Customer ID', 'Age', 'Gender', 'Item Purchased', 'Category',
       'Purchase Amount (USD)', 'Location', 'Size', 'Color', 'Season',
       'Review Rating', 'Subscription Status', 'Payment Method',
       'Shipping Type', 'Discount Applied', 'Promo Code Used',
       'Previous Purchases', 'Preferred Payment Method',
       'Frequency of Purchases', 'Age_category'],
      dtype='object')
```

In [24]:
```
shop.groupby('Category')['Item Purchased'].value_counts()
```

Out[24]:
```
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: Item Purchased, dtype: int64
```

In [25]:
```
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 [26]:
```
shop['Season'].unique()
```

Out[26]:
```
array(['Winter', 'Spring', 'Summer', 'Fall'], dtype=object)
```

In [27]:
```
shop[shop['Season'] == 'Summer'].value_counts().sum()
```

Out[27]:
```
955
```

In [28]:
```
shop[shop['Season'] == 'Winter'].value_counts().sum()
```

Out[28]:
```
971
```

In [29]:
```
shop[shop['Season'] == 'Spring'].value_counts().sum()
```

Out[29]:
```
999
```

In [30]:
```
shop[shop['Season'] == 'Fall'].value_counts().sum()
```

Out[30]: 975

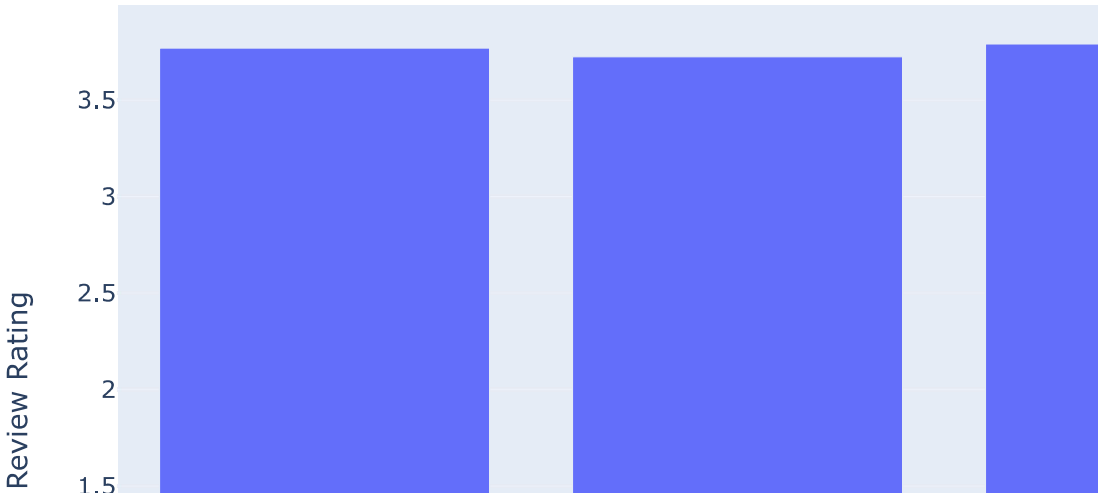
In [31]:
```
fig = px.histogram(shop , x = 'Season' , range_y= [200 , 1500] )

fig.show()
```

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

In [32]:
```
shop_groupby = shop.groupby('Category')['Review Rating'].mean().reset_index
```

In [33]:
```
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 [34]:
```
shop.columns
```

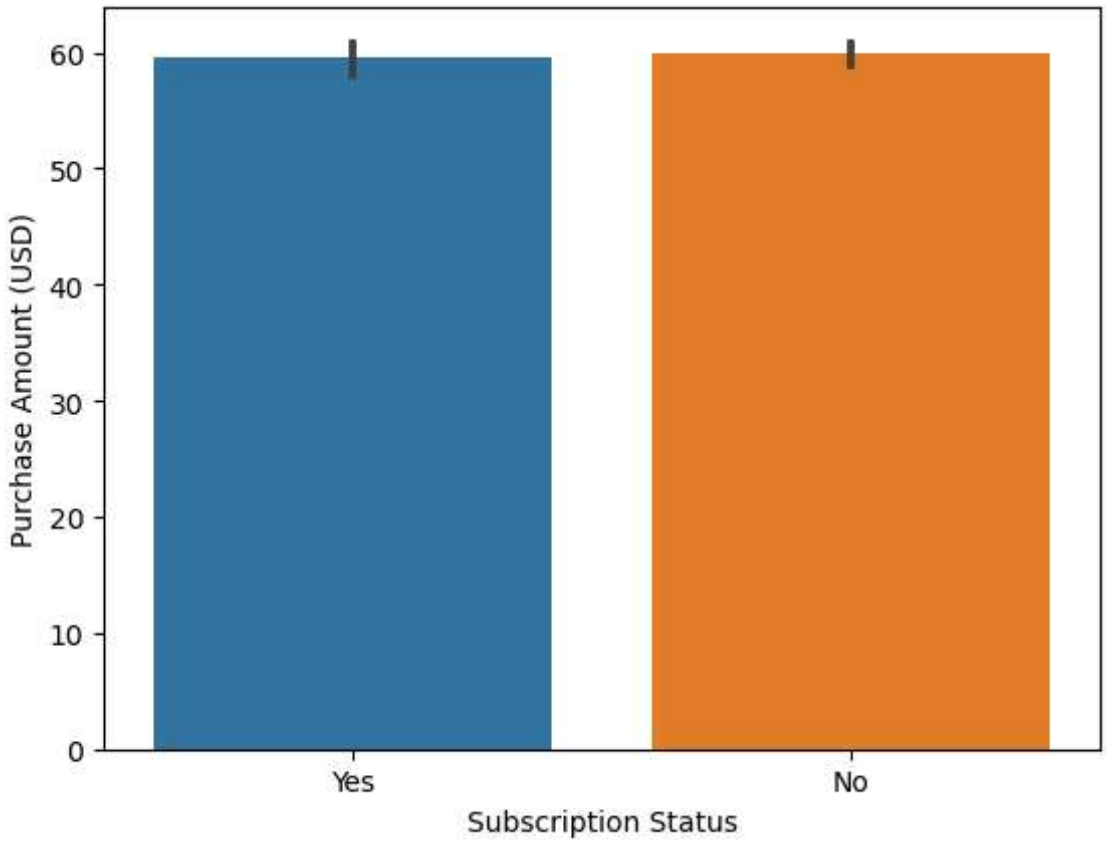
Out[34]:
```
Index(['Customer ID', 'Age', 'Gender', 'Item Purchased', 'Category',
       'Purchase Amount (USD)', 'Location', 'Size', 'Color', 'Season',
       'Review Rating', 'Subscription Status', 'Payment Method',
       'Shipping Type', 'Discount Applied', 'Promo Code Used',
       'Previous Purchases', 'Preferred Payment Method',
       'Frequency of Purchases', 'Age_category'],
      dtype='object')
```

In [35]:
```
shop['Subscription Status'].unique()
```

Out[35]:
```
array(['Yes', 'No'], dtype=object)
```

In [36]:
```
sns.barplot(shop  , x = 'Subscription Status' , y = 'Purchase Amount (USD)'
```

Out[36]:
```
<Axes: xlabel='Subscription Status', ylabel='Purchase Amount (USD)'>
```



In [37]:
```
shop['Purchase Amount (USD)'].sum()
```

Out[37]:
```
233081
```

In [38]:
```
shop.groupby('Subscription Status')['Purchase Amount (USD)'].mean()
```

Out[38]:
```
Subscription Status
No     59.865121
Yes    59.491928
Name: Purchase Amount (USD), dtype: float64
```

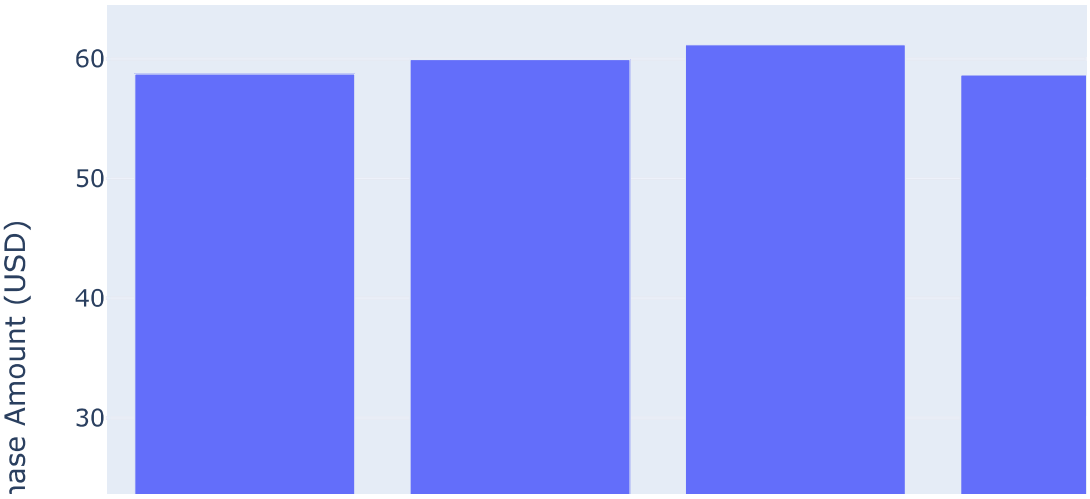
8 Which payment method is the most popular among customers?

In [39]:
```
shop.groupby('Payment Method')['Purchase Amount (USD)'].mean().sort_values(
```

Out[39]:
```
Payment Method
Venmo            61.241960
Credit Card      61.159483
Cash             59.927469
Bank Transfer    58.738924
PayPal           58.697492
Debit Card       58.638231
Name: Purchase Amount (USD), dtype: float64
```

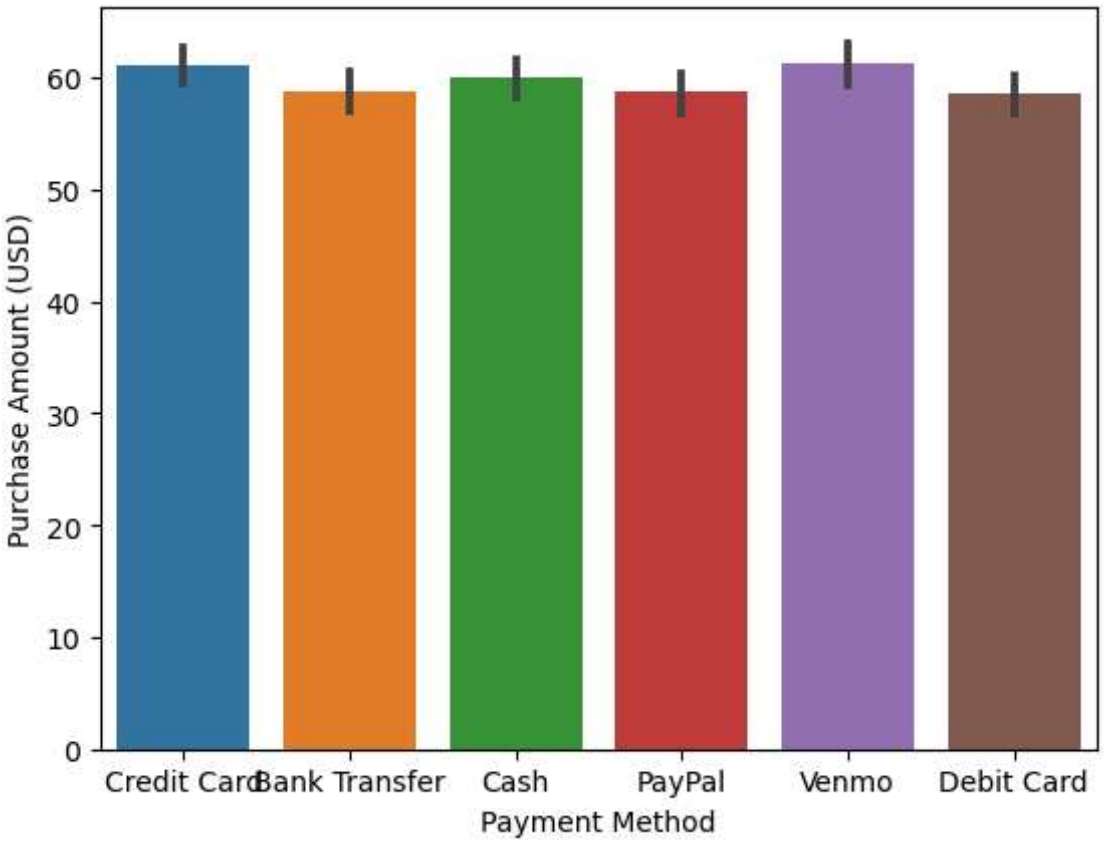
In [40]:
```
shop_groupby = shop.groupby('Payment Method')['Purchase Amount (USD)'].mean
```

In [41]:
```
fig = px.bar(shop_groupby , x = 'Payment Method' , y = 'Purchase Amount (US
fig.show()
```



In [42]:
```
sns.barplot(shop ,x='Payment Method' , y = 'Purchase Amount (USD)')
```

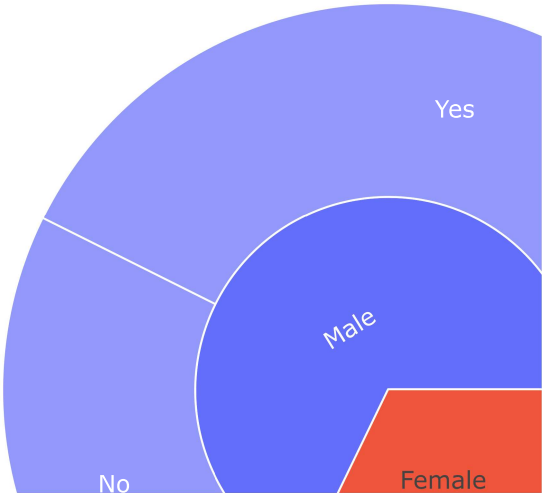
Out[42]:
```
<Axes: xlabel='Payment Method', ylabel='Purchase Amount (USD)'>
```



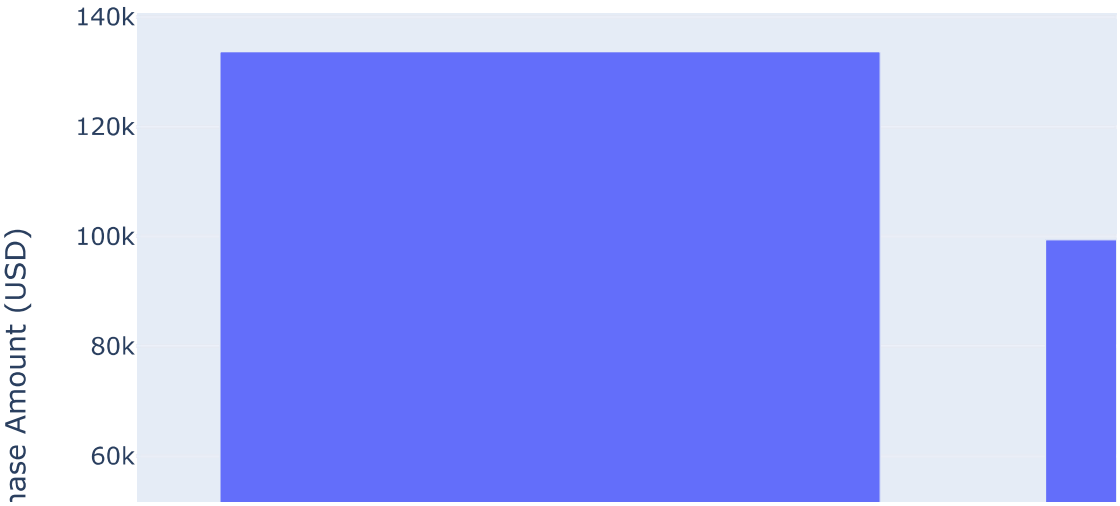
9 Do customers who use promo codes tend to spend more than those who don't?

In [43]:
```
shop_groupby  = shop.groupby('Promo Code Used')['Purchase Amount (USD)'].su
```

In [44]:
```
fig = px.sunburst(shop , path=['Gender' , 'Promo Code Used'] , values='Purc
fig.show()
```



In [45]:
```
fig  =  px.bar(shop_groupby , x= 'Promo Code Used' , y = 'Purchase Amount (
fig.show()
```



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

In [46]:
```
shop[['Age' , 'Age_category']]
```

Out[46]:

	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

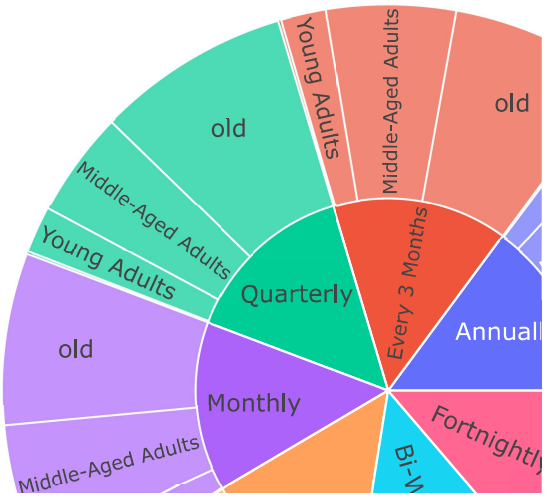
3900 rows × 2 columns

In [47]:
```
shop['Age_category'].unique()
```

Out[47]:
```
['old', 'Young Adults', 'Middle-Aged Adults', 'teen']
Categories (5, object): ['child' < 'teen' < 'Young Adults' < 'Middle-Aged
Adults' < 'old']
```

In [48]:
```
shop_group = shop.groupby('Frequency of Purchases')['Age'].sum()
```

In [49]:
```
px.sunburst(shop , path=['Frequency of Purchases','Age_category'] , values=
```



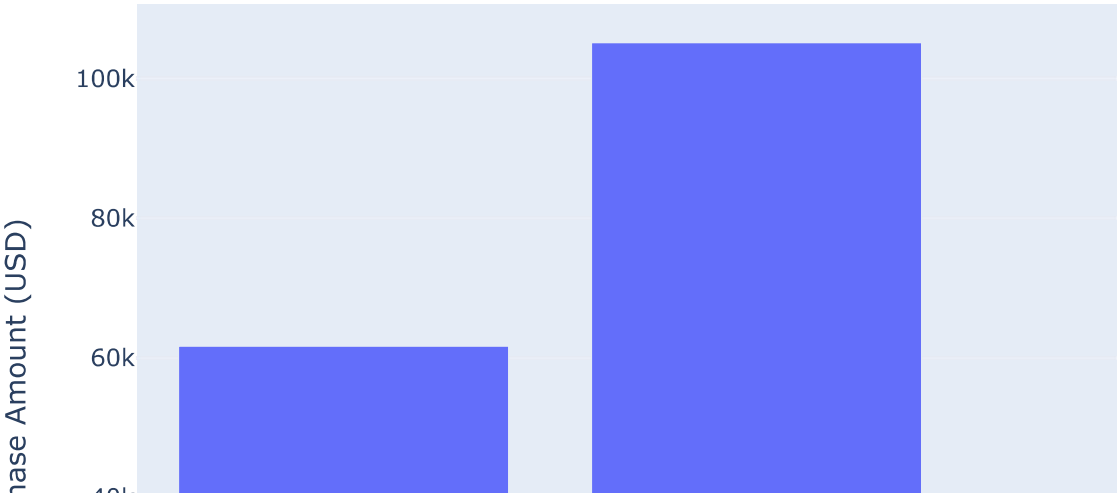
11 Are there any correlations between the size of the product and the purchase amount?

In [50]:
```
shop.columns
```

Out[50]:
```
Index(['Customer ID', 'Age', 'Gender', 'Item Purchased', 'Category',
       'Purchase Amount (USD)', 'Location', 'Size', 'Color', 'Season',
       'Review Rating', 'Subscription Status', 'Payment Method',
       'Shipping Type', 'Discount Applied', 'Promo Code Used',
       'Previous Purchases', 'Preferred Payment Method',
       'Frequency of Purchases', 'Age_category'],
      dtype='object')
```

In [51]:
```
shop_group = shop.groupby('Size')['Purchase Amount (USD)'].sum().reset_inde
```

In [52]:
```
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 [53]:
```
shop.groupby('Category')['Shipping Type'].value_counts().sort_values(ascend
```

Out[53]:
```
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
Name: Shipping Type, dtype: int64
```

In [54]:
```
shop['Shipping_Category'] =shop['Shipping Type'].map({'Express': 0, 'Free S
                                                      'Standard': 3, '2-Da
```

In [55]:
```
shop['Category'].unique()
```

Out[55]:
```
array(['Clothing', 'Footwear', 'Outerwear', 'Accessories'], dtype=object)
```

In [56]:
```
shop['Category_num'] =shop['Category'].map({'Clothing':1, 'Footwear':2, 'Ou
```

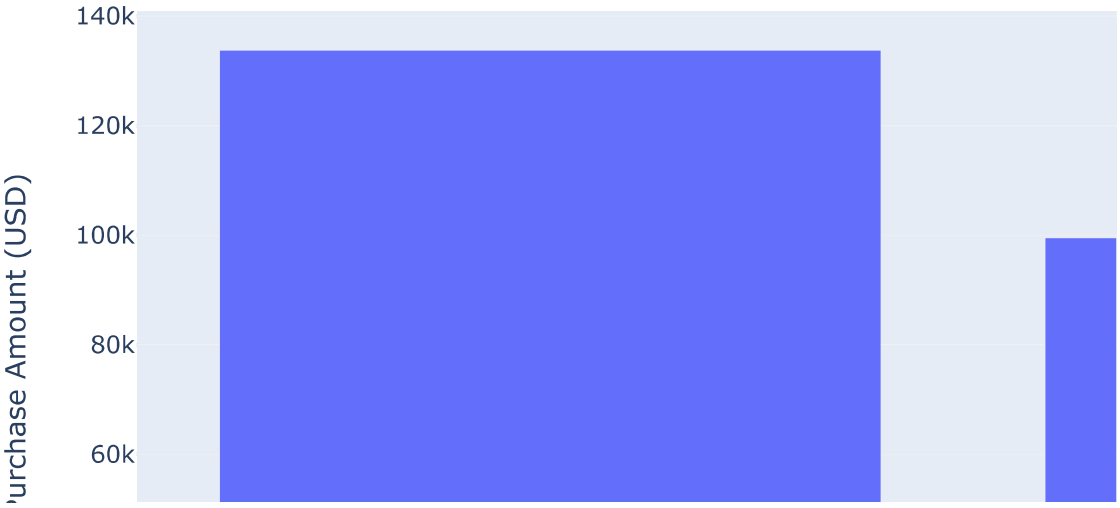
13 How does the presence of a discount affect the purchase decision of customers?

In [57]:
```
shop.columns
```

Out[57]:
```
Index(['Customer ID', 'Age', 'Gender', 'Item Purchased', 'Category',
       'Purchase Amount (USD)', 'Location', 'Size', 'Color', 'Season',
       'Review Rating', 'Subscription Status', 'Payment Method',
       'Shipping Type', 'Discount Applied', 'Promo Code Used',
       'Previous Purchases', 'Preferred Payment Method',
       'Frequency of Purchases', 'Age_category', 'Shipping_Category',
       'Category_num'],
      dtype='object')
```

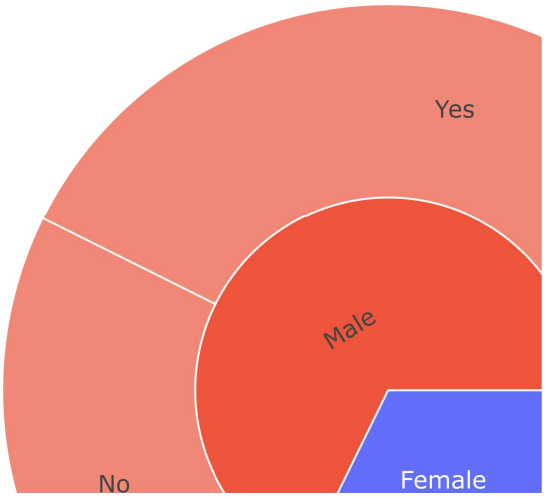
In [58]:
```
shop_group = shop.groupby('Discount Applied')['Purchase Amount (USD)'].sum(
```

In [59]:
```
px.histogram(shop_group , x = 'Discount Applied' , y = 'Purchase Amount (US
```



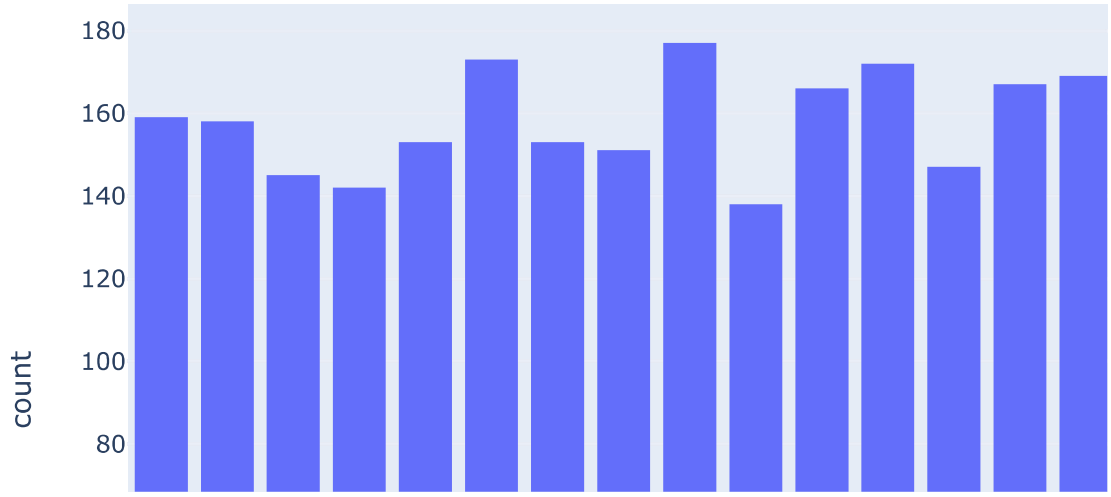
In [60]:
```
fig = px.sunburst(shop , path = ['Gender' , 'Discount Applied'], values='Pu

fig.show()
```



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

In [61]:
```
px.histogram(shop , x = 'Color')
```



In [62]:
```
shop['Color'].value_counts().nlargest(5)
```

Out[62]:
```
Olive     177
Yellow    174
Silver    173
Teal      172
Green     169
Name: Color, dtype: int64
```

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

In [63]:
```
shop['Previous Purchases'].mean()
```

Out[63]:
```
25.35153846153846
```

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

In [64]:
```
shop.groupby('Location')['Purchase Amount (USD)'].mean().sort_values(ascend
```

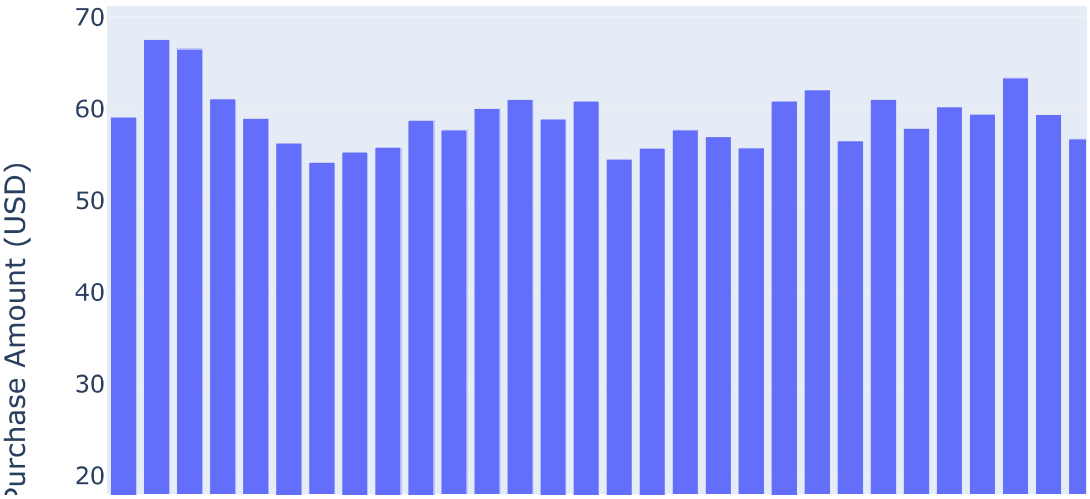
Out[64]:

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 [65]:
```
shop_group = shop.groupby('Location')['Purchase Amount (USD)'].mean().reset
```

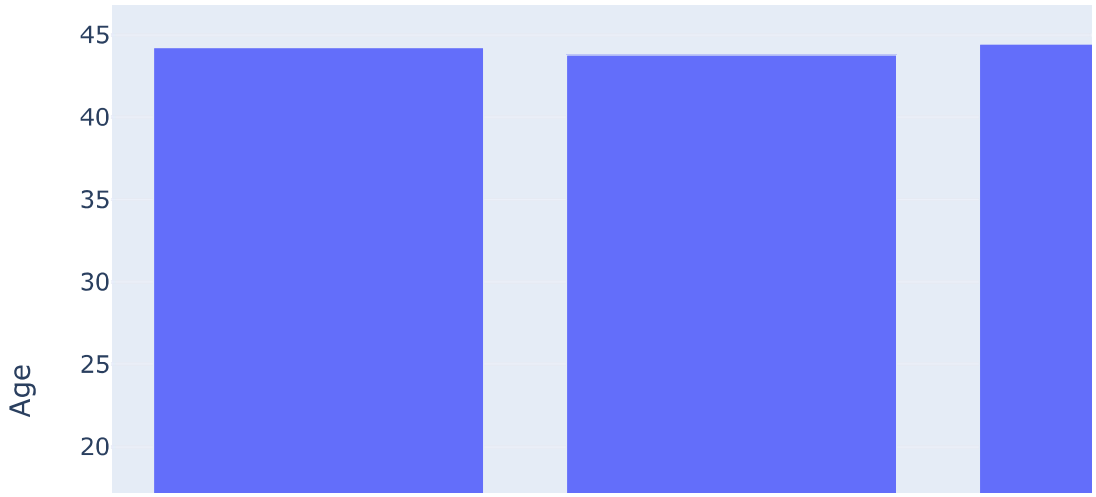
In [66]:
```
fig = px.bar(shop_group, x = 'Location' , y = 'Purchase Amount (USD)')
fig.show()
```



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

In [67]:
```
shop_group = shop.groupby('Category')['Age'].mean().reset_index()
```

In [68]:
```
fig = px.bar(shop_group ,y = 'Age' , x= 'Category')
fig.show()
```

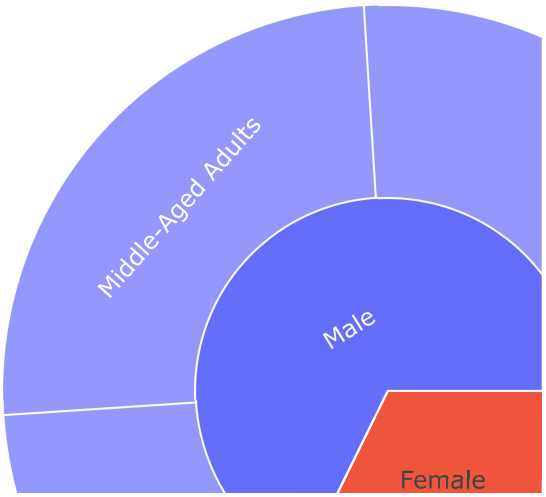


18 How does the average purchase amount differ between male and female customers?

In [69]:
```
shop_group = shop.groupby('Gender')['Purchase Amount (USD)'].sum().reset_ir
```

In [70]:
```
fig = px.bar(shop_group , x = 'Gender' , y = 'Purchase Amount (USD)')
fig.show()
```

In [71]:
```
px.sunburst(data_frame= shop , path = ['Gender' ,'Age_category'] , values='
```



In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: