In [212…
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import plotly.express as px
%matplotlib inline
```

In [147…
```
#reading the data set
shop=pd.read_csv(r"C:\Users\HP\Downloads\shopping_trends_updated.csv")
shop.shape
```

Out[147… (3900, 18)

In [149…
```
shop.to_excel('shopping_trends_updated.xlsx')
shop.head()
```

Out[149…

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

In [32]:
```
#to find the data types in the data
shop.dtypes
```

Out[32]:
```
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 [36]:
```
#to find the data types in the data

shop.columns
```

Out[36]:
```
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 [43]:
```
#to find the information about the data
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 [45]:
```
shop.isnull().sum()
```

Out[45]:
```
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 [11]:
```
shop.describe()
```

Out[11]:

	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 [47]:
```
print(f"The unique values of the 'Gender' column are:{shop['Gender'].unique()}")
print()
```

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

In [13]:
```
shop.describe(include="object")
```

Out[13]:

	Gender	Item Purchased	Category	Location	Size	Color	Season	Subscription Status	Shipping Type	Discount Applied	Promo Code Used	Payment Method	Frequency of Purchases
count	3900	3900	3900	3900	3900	3900	3900	3900	3900	3900	3900	3900	3900
unique	2	25	4	50	4	25	4	2	6	2	2	6	7
top	Male	Blouse	Clothing	Montana	M	Olive	Spring	No	Free Shipping	No	No	PayPal	Every 3 Months
freq	2652	171	1737	96	1755	177	999	2847	675	2223	2223	677	584

In [49]:
```
print(f"The unique values of the 'Category' column are:{shop['Category'].unique()}")
print()
```

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

In [51]:
```
print(f"The unique values of the 'Size' column are:{shop['Size'].unique()}")
print()
```

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

In [53]:
```
print(f"The unique values of the 'Subscription Status' column are:{shop['Subscription Status'].unique()}")
print()
```

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

In [55]:
```
print(f"The unique values of the 'Shipping Type' column are:{shop['Shipping Type'].unique()}")
print()
```

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

In [57]:
```
print(f"The unique values of the 'Discount Applied' column are:{shop['Discount Applied'].unique()}")
print()
```

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

In [59]:
```
print(f"The unique values of the 'Promo Code Used' column are:{shop['Promo Code Used'].unique()}")
print()
```

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

In [61]:
```
print(f"The unique values of the 'Payment Method' column are:{shop['Payment Method'].unique()}")
print()
```

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

In [ ]:
```

```

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 customers

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.

Customer 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 embark on this exciting analysis journey!

In [ ]:

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

In [65]:
```
shop['Age'].value_counts()
```

Out[65]:

	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

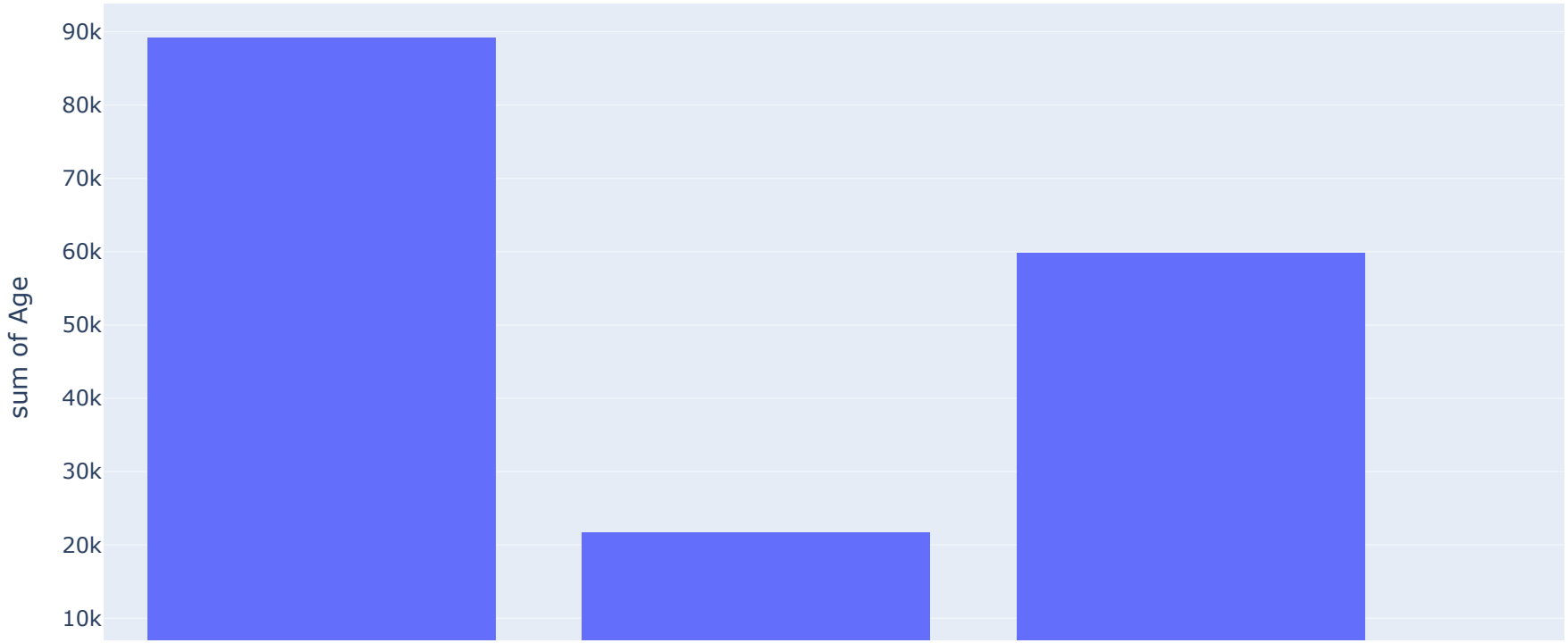
```
21    69
26    69
34    68
48    68
24    68
39    68
70    67
22    66
61    65
60    65
33    63
20    62
67    54
44    51
Name: count, dtype: int64
```

In [67]:
```
shop['Age'].mean()
```

Out[67]:
```
44.06846153846154
```

In [151…
```
shop['Age_Category']=pd.cut(shop['Age'],bins=[0, 15, 18, 30, 50, 70], labels=['child','teen', 'young Adults','Middle-Aged Adul
```

In [153…
```
fig = px.histogram(shop, y='Age' , x='Age_Category')
fig.show()
```



'''Fig 1.Represents how different age groups (categories) are distributed based on age data.'''

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

In [91]:
```
shop.columns
```

Out[91]:
```
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 [155…
```
shop['Category'].unique()
```

Out[155…
```
array(['Clothing', 'Footwear', 'Outerwear', 'Accessories'], dtype=object)
```

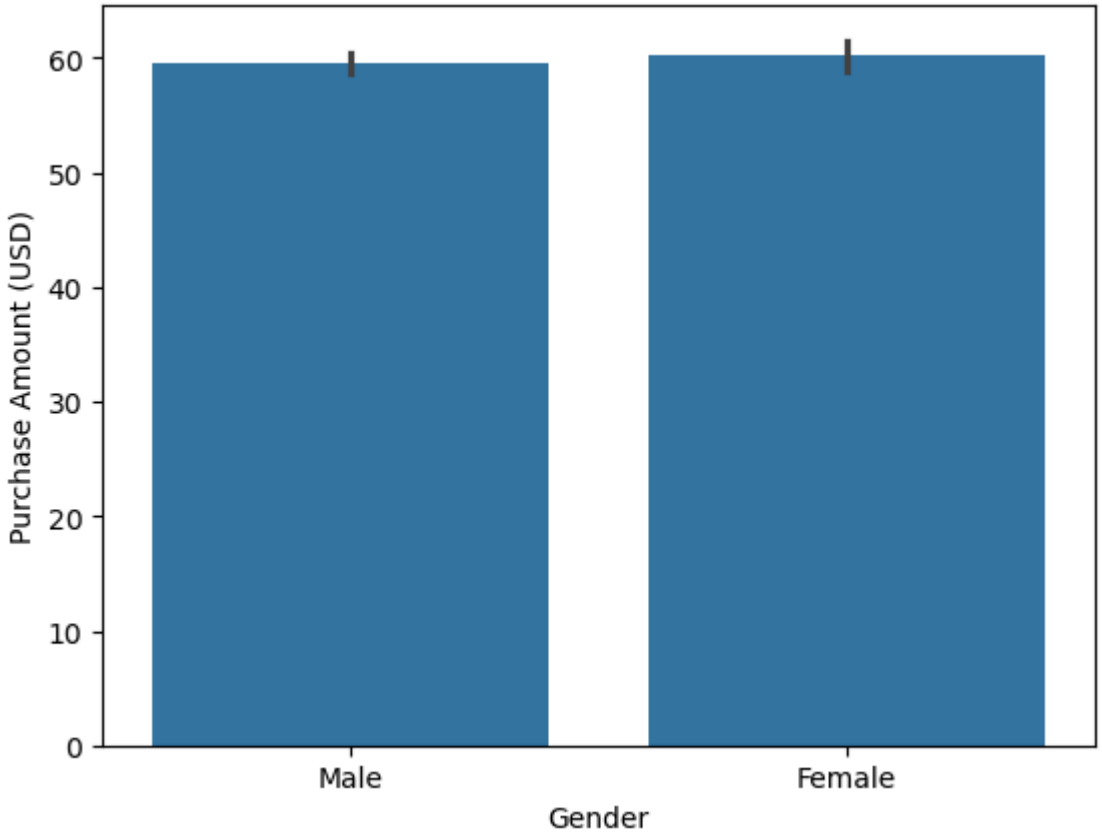
In [157…
```
shop.groupby('Category')['Purchase Amount (USD)'].mean()
```

Out[157…
```
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 [159…
```
sns.barplot(shop,x='Gender',y='Purchase Amount (USD)')
```

Out[159…
```
<Axes: xlabel='Gender', ylabel='Purchase Amount (USD)'>
```



''' Fig2:Represents the highest no.of purchases based on gender'''

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

In [161…

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

Out[161…

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 [165…

```
fig=px.histogram(shop,x='Item Purchased',color='Category')
fig.show()
```

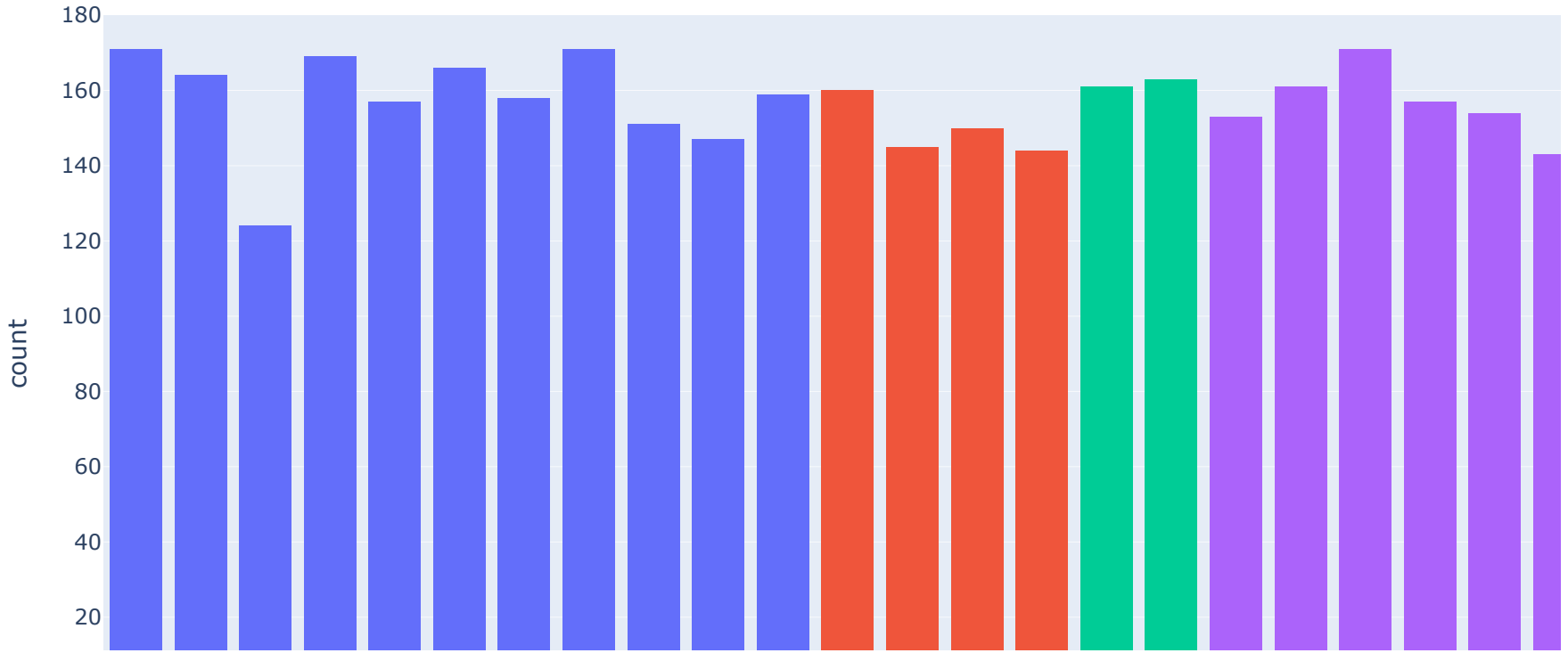


Fig3.Represents the distribution of items purchased, categorized by different groups.

In [ ]:

5. Are there any specific seasons or months where customer spending is significantly higher?

In [115…
```
shop['Season'].unique()
```

Out[115…
```
array(['Winter', 'Spring', 'Summer', 'Fall'], dtype=object)
```

In [21]:
```
shop['Season'].value_counts()
```

Out[21]:
```
Season
Spring    999
Fall      975
Winter    971
Summer    955
Name: count, dtype: int64
```

In [23]:
```
fig=px.histogram(shop,x='Season' ,range_y=[800,1200])
fig.show()
```

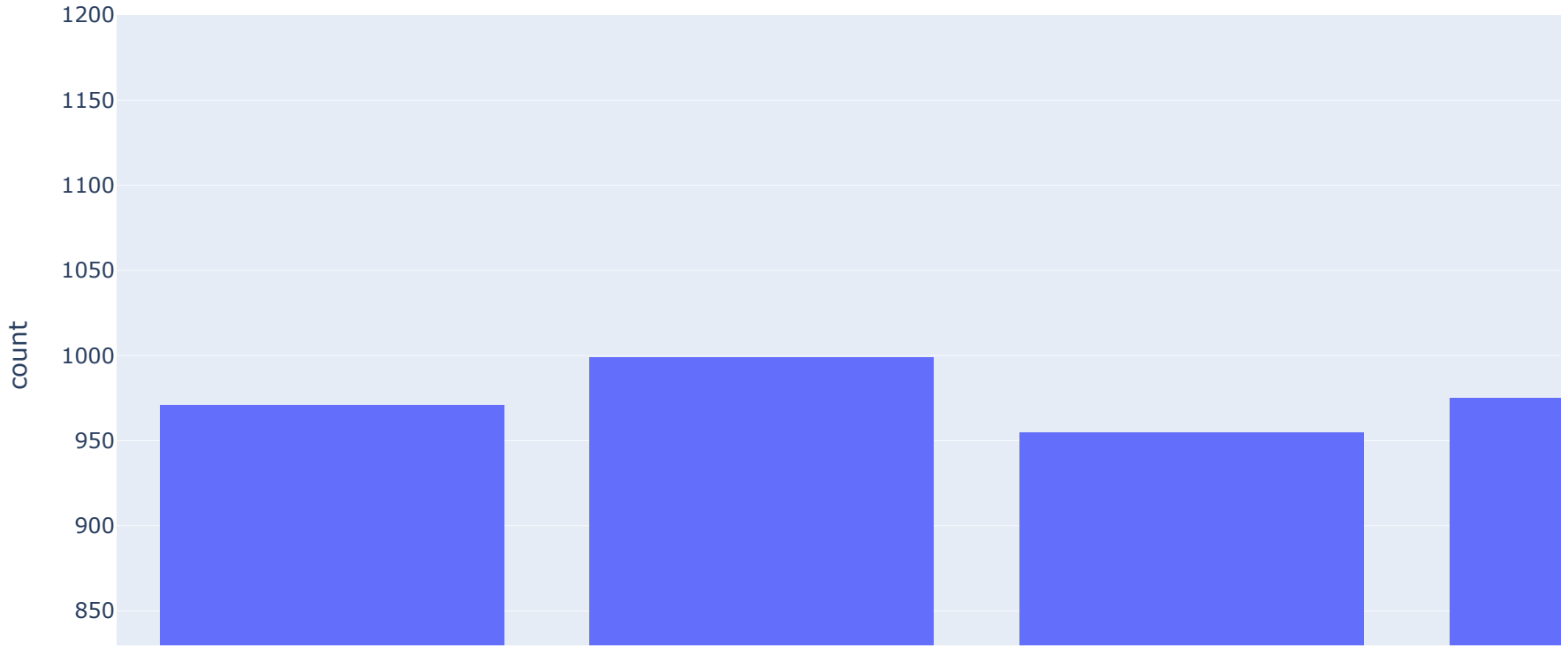


Fig4.Represents possibly sales are distributed across different seasons

In [ ]:

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

In [27]:
```
shop.groupby('Category')['Review Rating'].mean()
```

Out[27]:
```
Category
Accessories    3.768629
Clothing       3.723143
Footwear       3.790651
Outerwear      3.746914
Name: Review Rating, dtype: float64
```

In [43]:
```
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
```

In [48]:
```
fig = px.bar(shop_groupby ,x= 'Category' , y = 'Review Rating' )
fig.show()
```

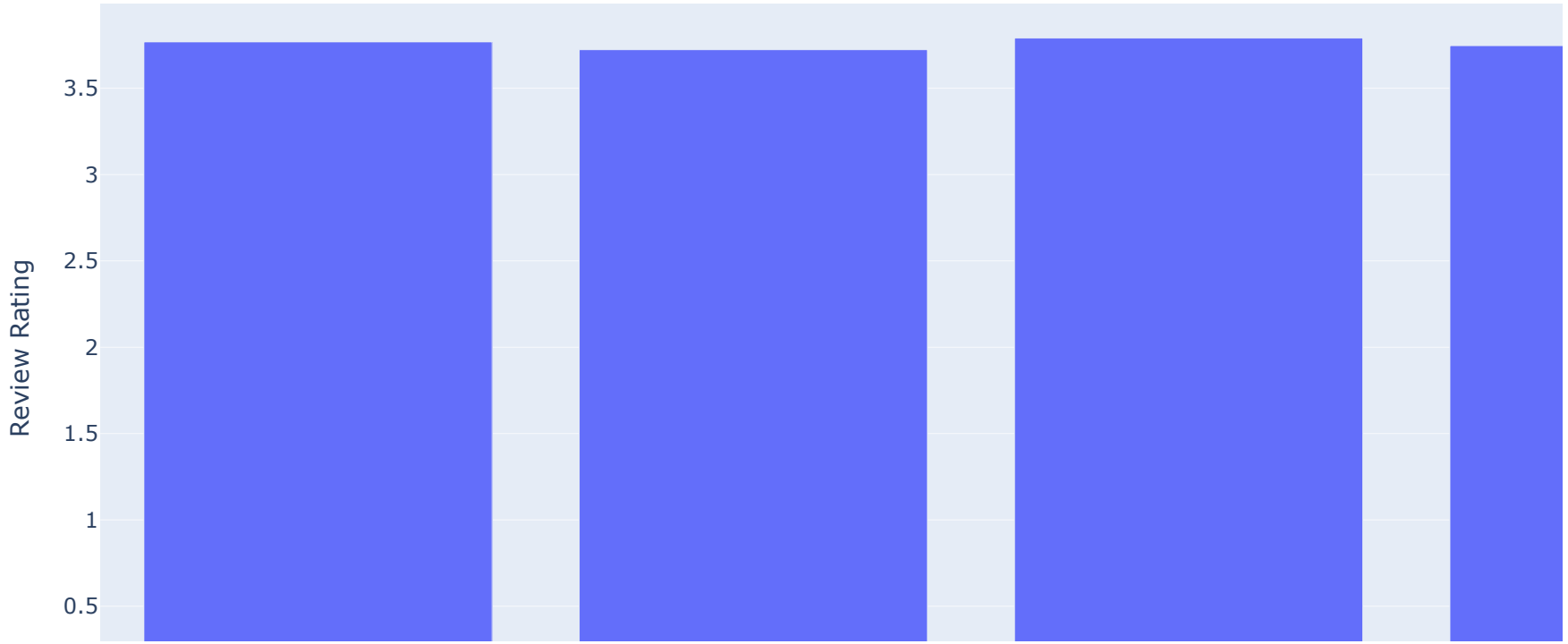


Fig5 Represents average rating given by customers for each product category

In [ ]:

7 Are there any notable differences in purchase behavior between subscribed and non-subscribed customers?

In [61]:
```
shop.columns
```

Out[61]:
```
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 [63]:
```
shop['Subscription Status'].value_counts()
```

Out[63]:
```
Subscription Status
No     2847
Yes    1053
Name: count, dtype: int64
```

In [71]:
```
sns.barplot(shop  , x = 'Subscription Status' , y = 'Purchase Amount (USD)')
```

Out[71]:
```
<Axes: xlabel='Subscription Status', ylabel='Purchase Amount (USD)'>
```

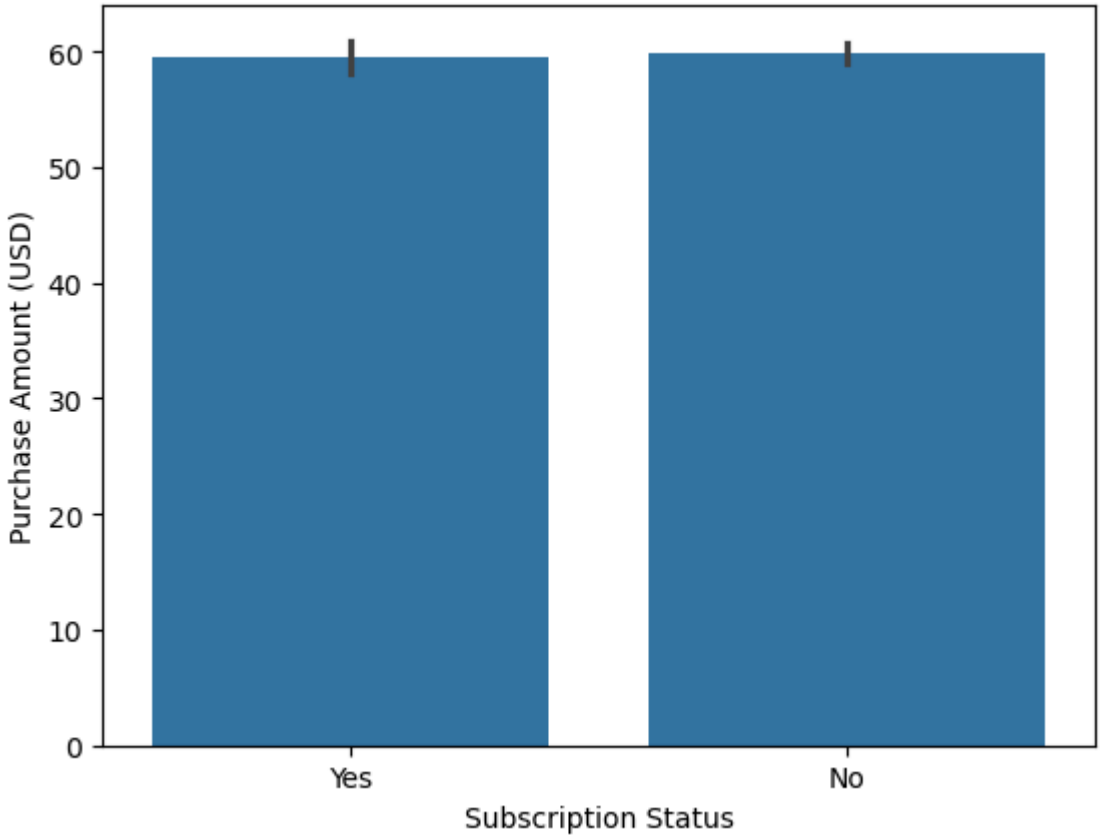


Fig6:Represents how purchase amounts (in USD) vary based on different subscription statuses

In [75]:
```
shop.groupby('Subscription Status')['Purchase Amount (USD)'].mean()
```

Out[75]:
```
Subscription Status
No     59.865121
Yes    59.491928
Name: Purchase Amount (USD), dtype: float64
```

In [ ]:
```

```

8 Which payment method is the most popular among customers?

In [80]:
```
shop.groupby('Payment Method')['Purchase Amount (USD)'].mean().sort_values(ascending= True)
```

Out[80]:
```
Payment Method
Venmo            58.949527
PayPal           59.245199
Cash             59.704478
Bank Transfer    59.712418
Credit Card      60.074516
Debit Card       60.915094
Name: Purchase Amount (USD), dtype: float64
```

In [82]:
```
sns.barplot(shop ,x='Payment Method' , y = 'Purchase Amount (USD)')
plt.show()
```

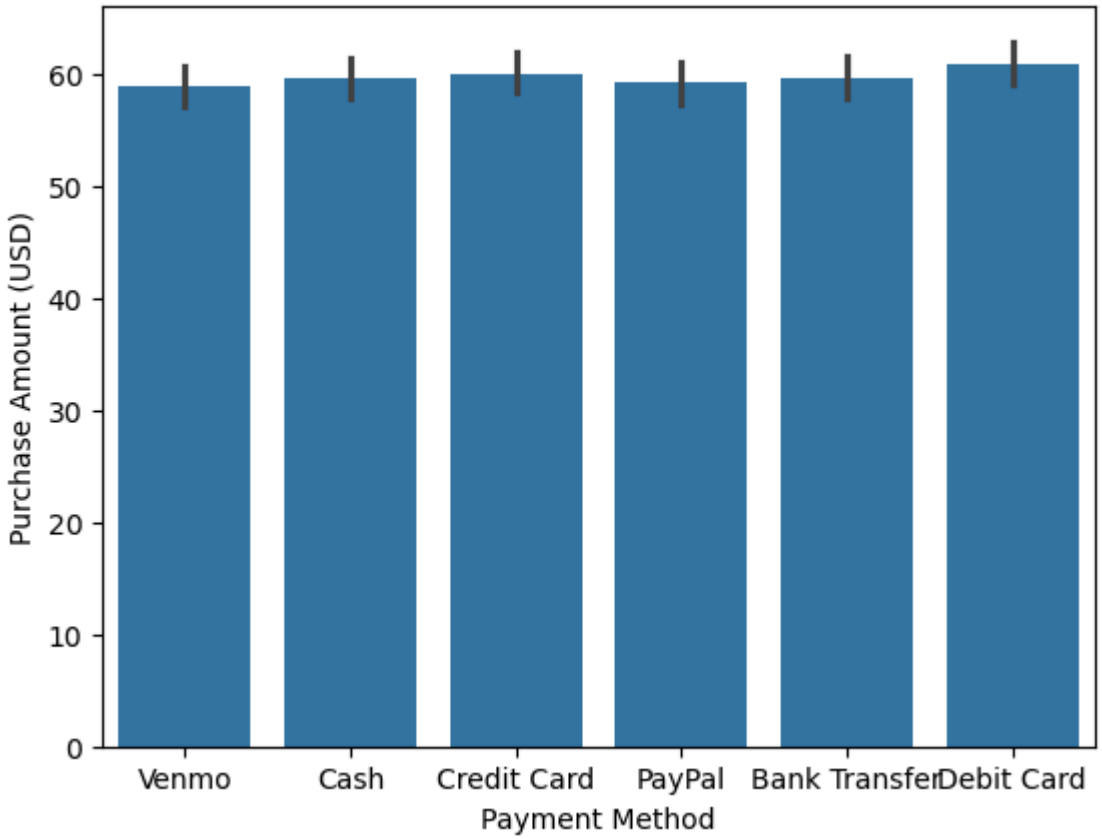


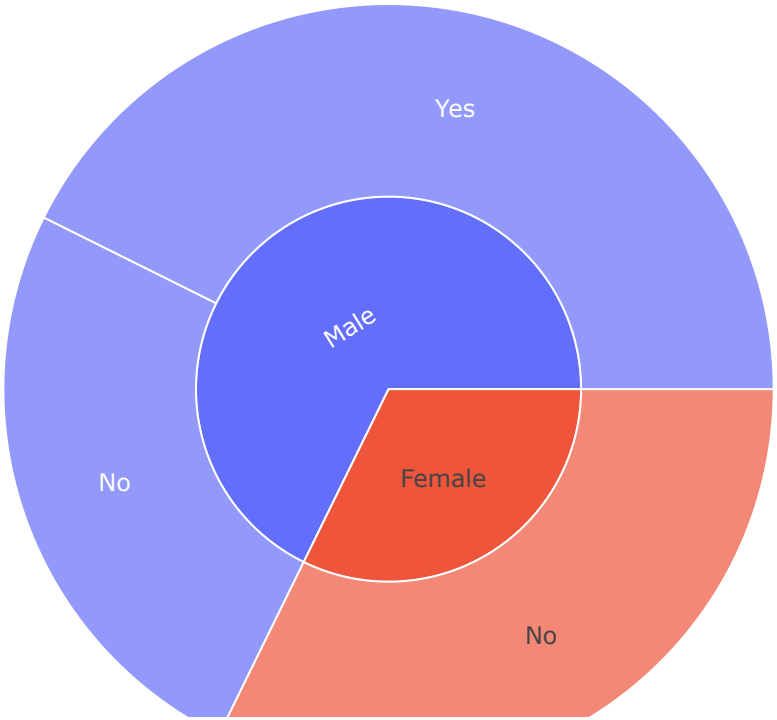
Fig7 Represents Most Popular Payment Method among Customers

In [ ]:

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

In [95]:
```
shop_groupby=shop.groupby('Promo Code Used')['Purchase Amount (USD)'].sum().reset_index()
```

In [93]:
```
fig = px.sunburst(shop , path=['Gender' , 'Promo Code Used'] , values='Purchase Amount (USD)')
fig.show()
```



In [ ]:

In [97]:
```
fig  =  px.bar(shop_groupby , x= 'Promo Code Used' , y = 'Purchase Amount (USD)')
fig.show()
```

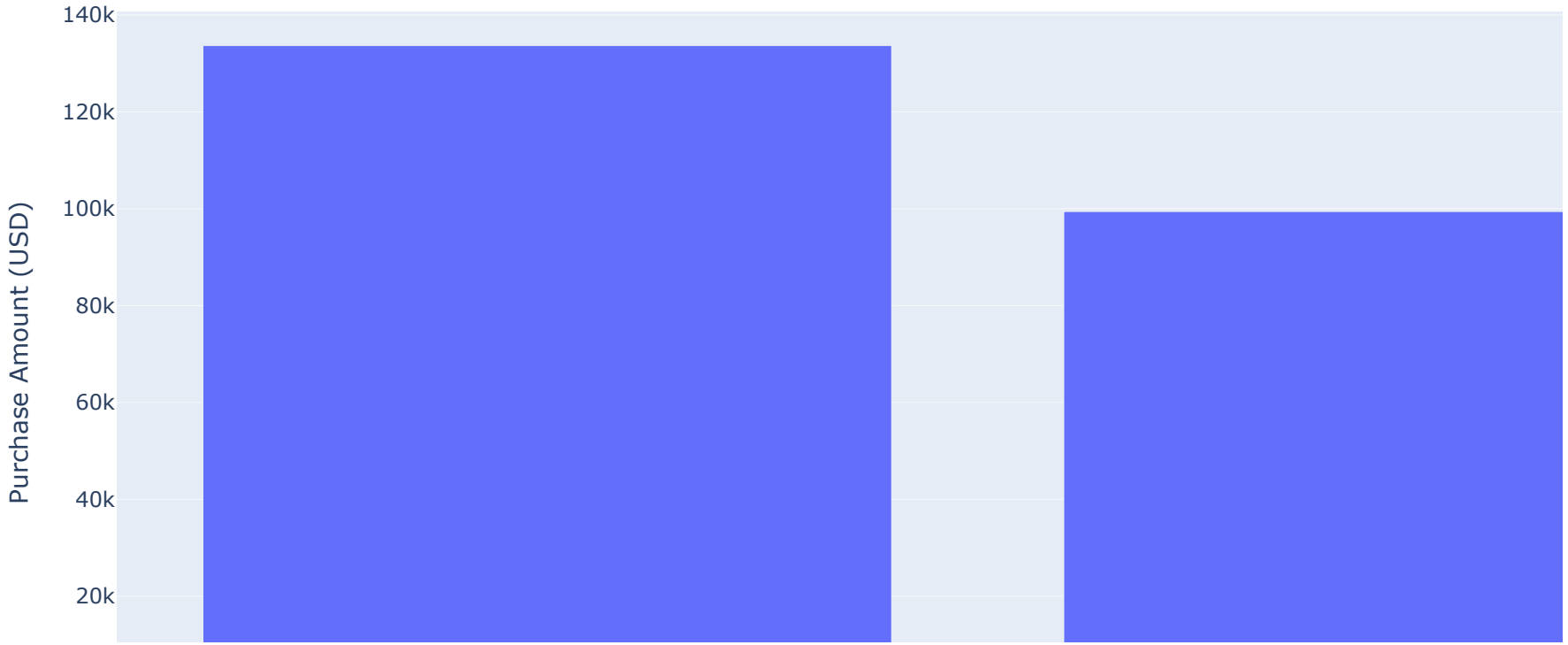


Fig8:Represents who use promo codes tend to spend more than those who don't.

In [ ]:

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

In [45]:
```
shop['Age_Category'].unique()
```

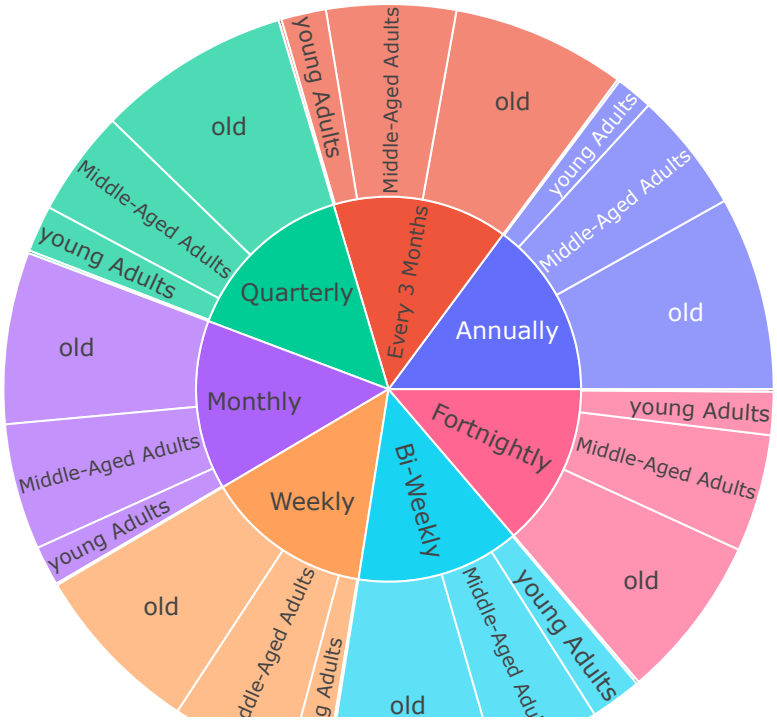
Out[45]:
```
['old', 'young Adults', 'Middle-Aged Adults', 'teen']
Categories (5, object): ['child' < 'teen' < 'young Adults' < 'Middle-Aged Adults' < 'old']
```

In [47]:
```
shop_group = shop.groupby('Frequency of Purchases')['Age'].sum()
```

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

```
C:\Users\HP\anaconda3\Lib\site-packages\plotly\express\_core.py:1706: FutureWarning:

The default of observed=False is deprecated and will be changed to True in a future version of pandas. Pass observed=False to r
etain 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 [58]:
```
shop_group = shop.groupby('Size')['Purchase Amount (USD)'].sum().reset_index()
```

In [60]:
```
fig  = px.bar(shop_group , x = 'Size' , y ='Purchase Amount (USD)'  )
fig.show()
```

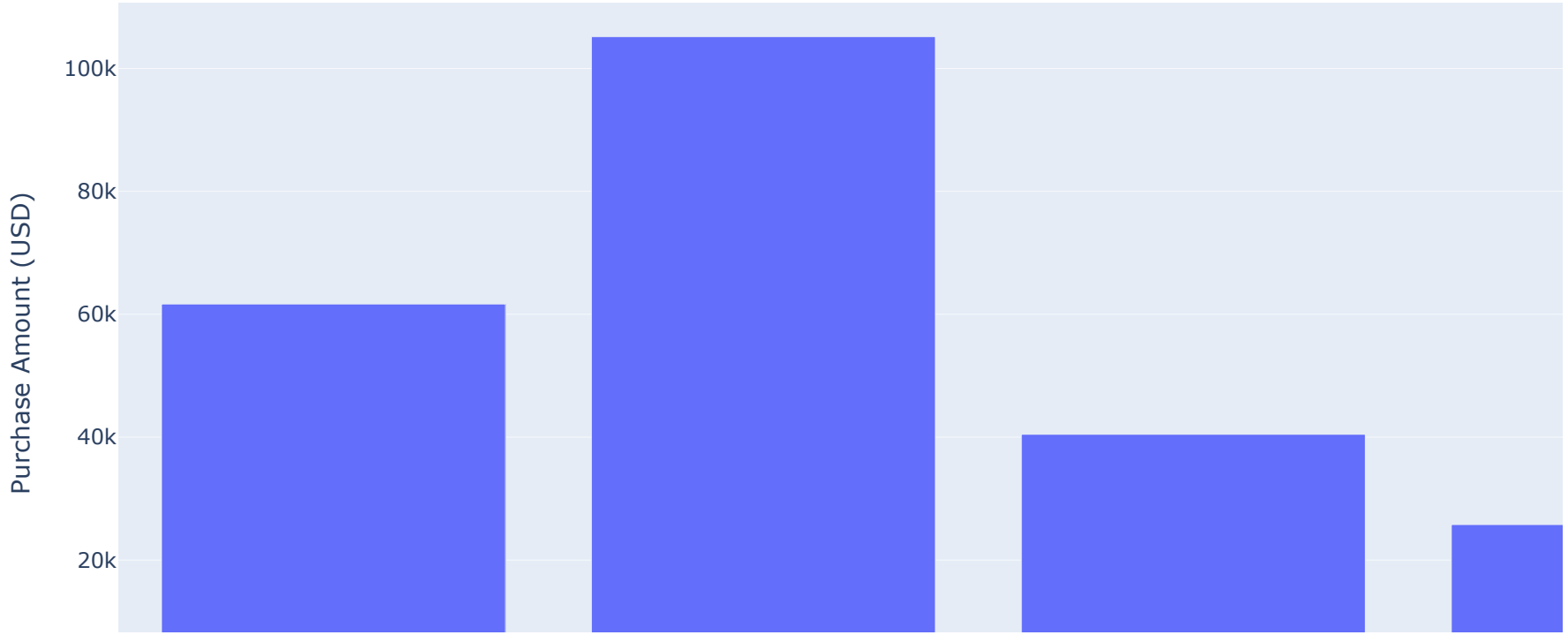


Fig9:Represents correlations between the size of the product and the purchase amount

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

In [65]:
```
shop.groupby('Category')['Shipping Type'].value_counts().sort_values(ascending= False)
```

Out[65]:
```
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: count, dtype: int64
```

In [ ]:
```

```

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

In [70]:
```
shop_group = shop.groupby('Discount Applied')['Purchase Amount (USD)'].sum().reset_index()
```

In [72]:
```
px.histogram(shop_group , x = 'Discount Applied' , y = 'Purchase Amount (USD)')
```

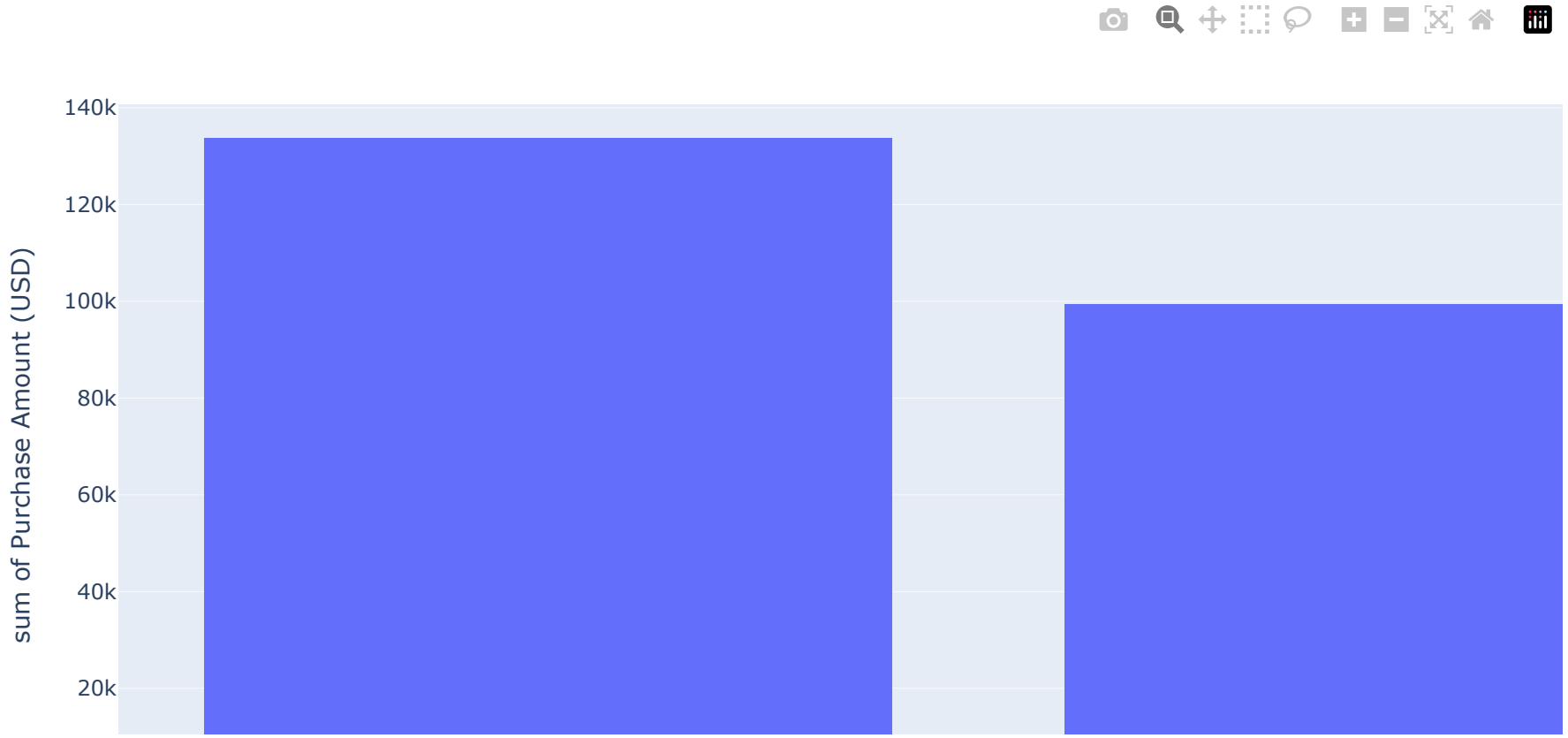


Fig10:Represents the presence of a discount affect the purchase decision of customers?

In [ ]:

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

In [76]:
```
shop['Color'].value_counts().nlargest(5)
```

Out[76]:
```
Color
Olive     177
Yellow    174
Silver    173
Teal      172
Green     169
Name: count, dtype: int64
```

In [78]:
```
px.histogram(shop , x = 'Color')
```

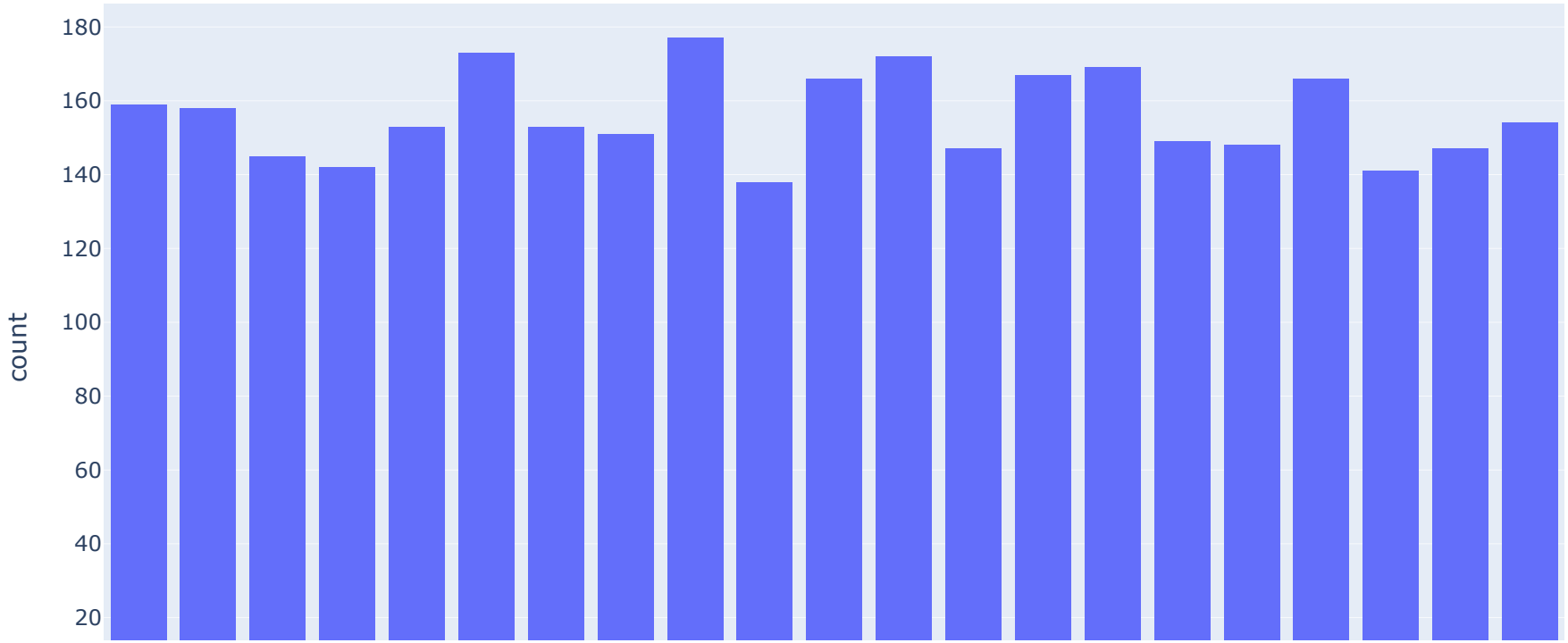


Fig11:Represents specific colors that are more popular among customers

In [ ]:

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

In [83]:
```
shop['Previous Purchases'].mean()
```

Out[83]:
```
25.35153846153846
```

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

In [90]:
```
shop.groupby('Location')['Purchase Amount (USD)'].mean().sort_values(ascending = True)
```

Out[90]:

Location	
Connecticut	54.179487
Kansas	54.555556
Delaware	55.325581
Kentucky	55.721519
Maryland	55.755814
Florida	55.852941
Wisconsin	55.946667
Colorado	56.293333
Minnesota	56.556818
New Jersey	56.746269
Maine	56.987013
Vermont	57.176471
Oregon	57.337838
Louisiana	57.714286
Hawaii	57.723077
Missouri	57.913580
Oklahoma	58.346667
South Carolina	58.407895
Georgia	58.797468
Indiana	58.924051
California	59.000000
Alabama	59.112360
New Hampshire	59.422535
Nebraska	59.448276
Idaho	60.075269
Montana	60.250000
Ohio	60.376623
New York	60.425287
South Dakota	60.514286
Wyoming	60.690141
North Carolina	60.794872
Iowa	60.884058
Massachusetts	60.888889
Mississippi	61.037500
Illinois	61.054348
Arkansas	61.113924
Texas	61.194805
Rhode Island	61.444444
New Mexico	61.901235

```
Tennessee         61.974026
Michigan          62.095890
Utah              62.577465
Virginia          62.883117
North Dakota      62.891566
Washington        63.328767
Nevada            63.379310
West Virginia     63.876543
Arizona           66.553846
Pennsylvania      66.567568
Alaska            67.597222
Name: Purchase Amount (USD), dtype: float64
```

In [96]:
```
fig = px.bar(shop, x = 'Location' , y = 'Purchase Amount (USD)')
fig.show()
```

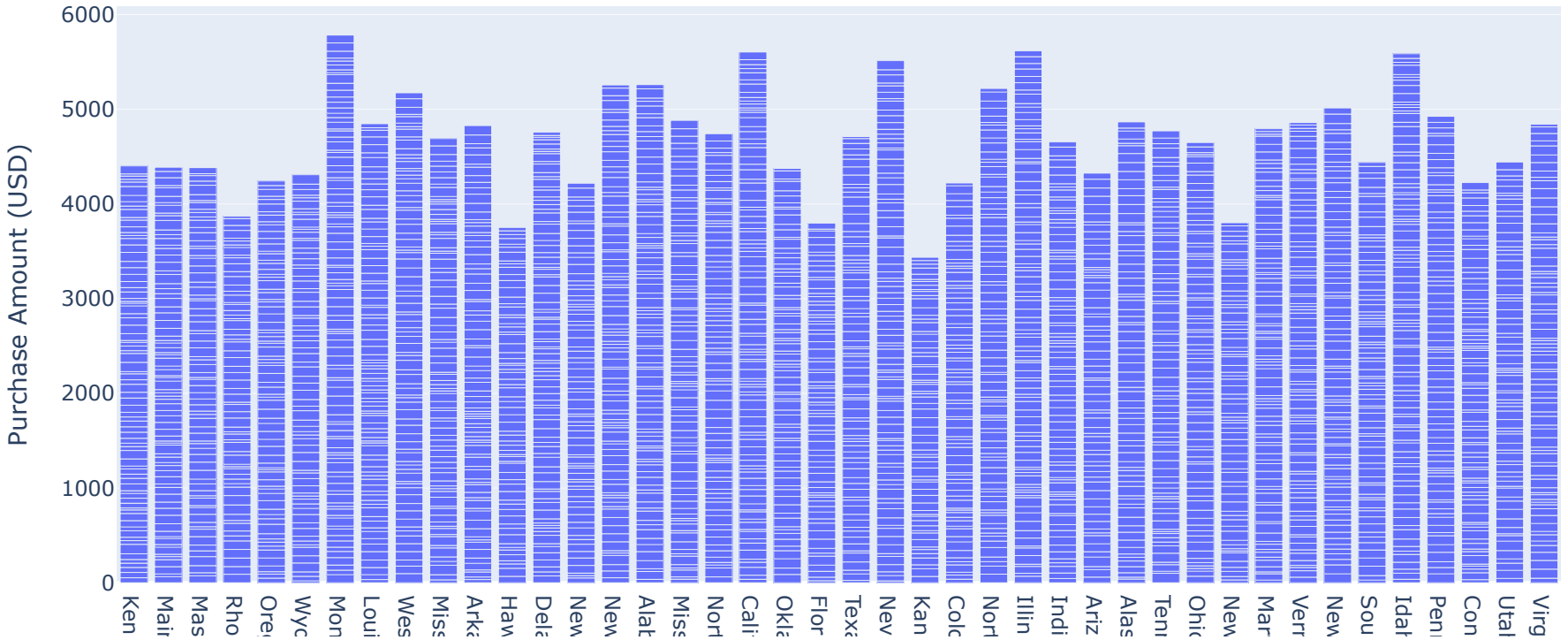


Fig12:Represents there any noticeable differences in purchase behavior between different locations

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

In [175…
```
shop_group = shop.groupby('Category')['Age'].mean().reset_index()
```

In [177…
```
fig = px.bar(shop_group ,y = 'Age' , x= 'Category')
fig.show()
```

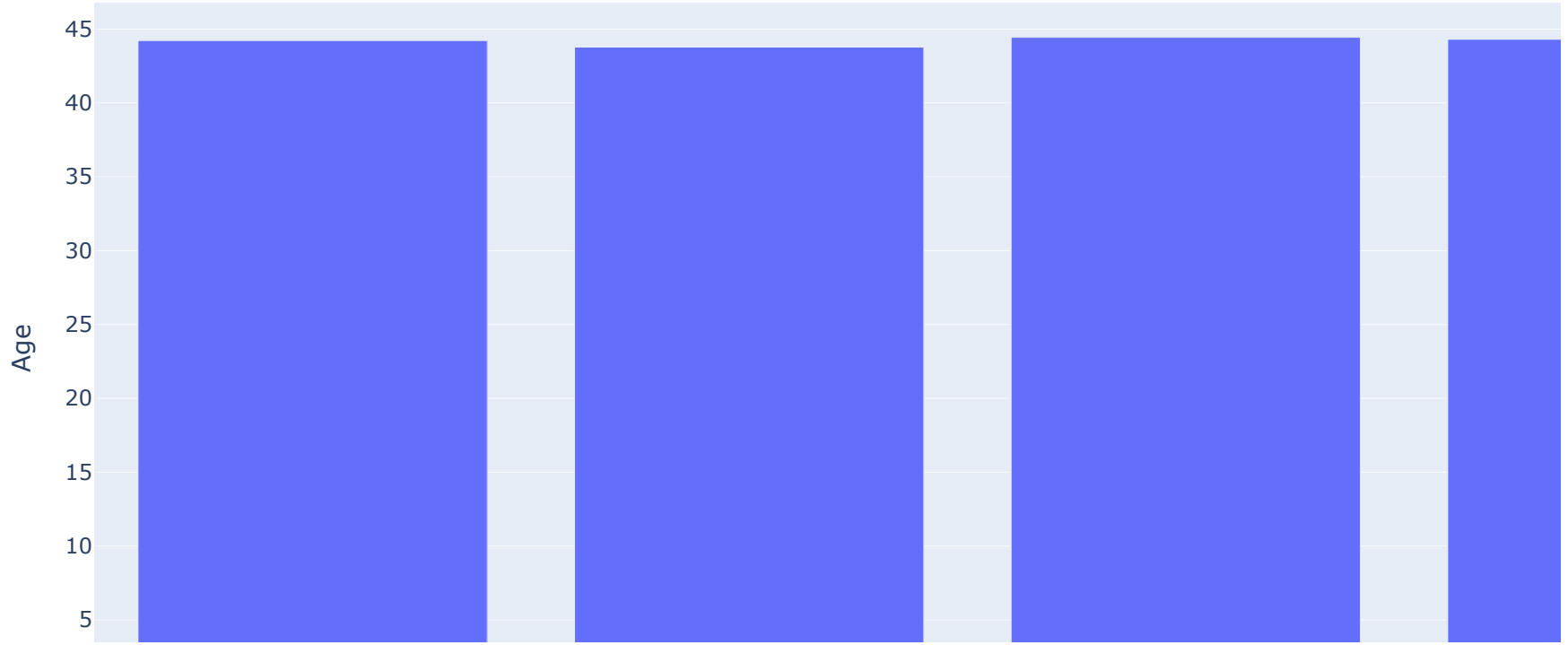


Fig13:Represents a relationship between customer age and the category of products they purchase

In [ ]:

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

In [189…
```
shop_group = shop.groupby('Gender')['Purchase Amount (USD)'].sum().reset_index()
```

In [199…
```
fig = px.bar(shop_group ,y = 'Purchase Amount (USD)' , x= 'Gender')
fig.show()
```

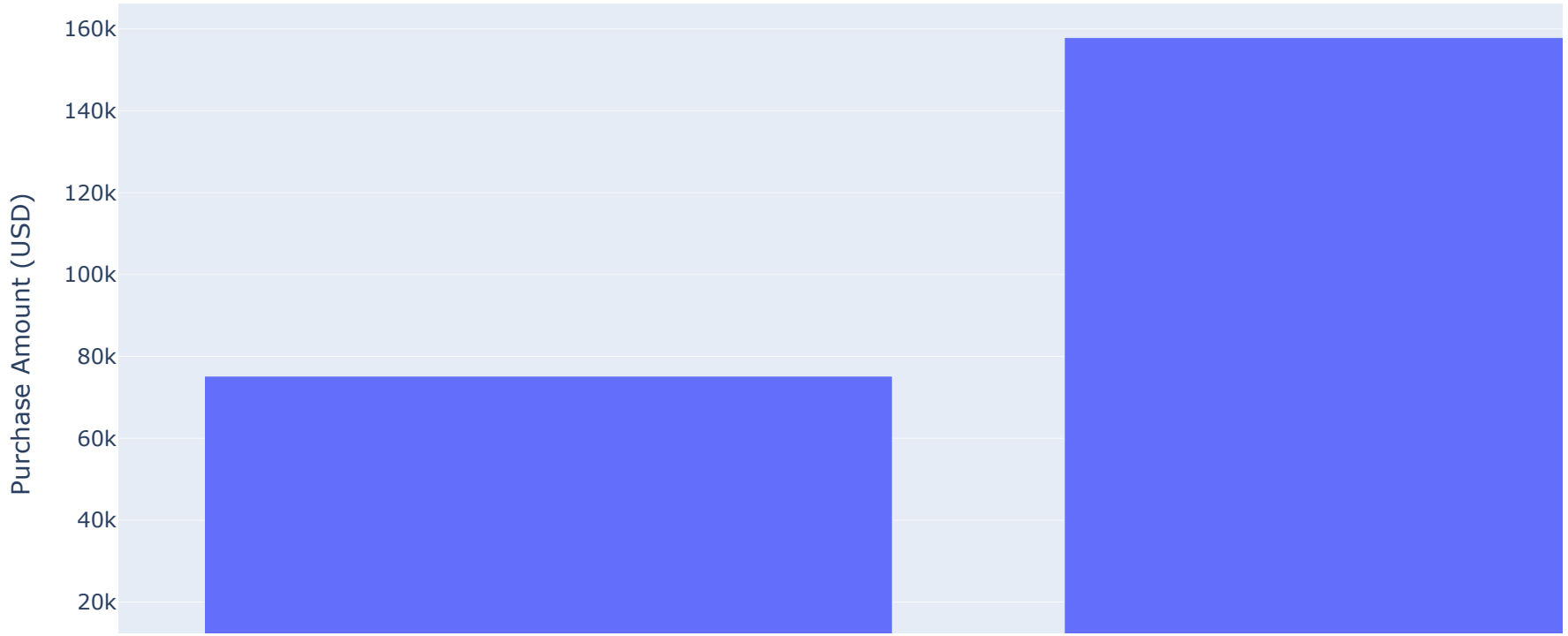


Fig14:Represents the average purchase amount differ between male and female customers

In [ ]:

In [ ]:

In [ ]: