SHOPPING TRENDS ANALYSIS

Project by- Simranjit kaur

Hello! Welcome to my project where I dig deeper into the sales data of a store selling clothing, footwear, outerwear and fashion accessories across 50 different locations. In this project, we harness the capabilities of powerful Python libraries, particularly Pandas for data manipulation and organization while visualizing insightful narratives using visualization libraries, prominently Matplotlib, Seaborn and Plotly.

By doing so we not only gain a comprehensive understanding of the shopping behaviour of customers but also provide a solid foundation for informed decision-making to enhance overall customer experiences and optimize operational strategies.

So let's begin our analytical journey to decipher key trends by answering relevant business questions such as:-

1. WHAT IS THE PROPORTION OF MALE AND FEMALE CUSTOMERS VISITING THE STORE?
2. WHAT IS THE AGE DISTRIBUTION OF THE CUSTOMERS VISITING THE STORE?
3. MAXIMUM NUMBER OF CUSTOMERS VISITING THE STORE BELONG TO WHICH AGE GROUP?
4. RANK THE TYPES OF PRODUCTS SOLD BASED ON THE NUMBER OF ITEMS SOLD FOR EACH TYPE. (GENDER WISE)
5. WHAT ARE THE SALES FOR EACH TYPE OF PRODUCT SOLD?
6. WHICH ARE THE MOST POPULAR SIZES AMONG EACH TYPE OF ITEM SOLD?
7. WHICH ARE THE MOST POPULAR ITEMS AMONG THE CUSTOMERS?
8. WHICH ARE THE MOST POPULAR COLOURS AMONG EACH TYPE OF CUSTOMER? (GENDER WISE)
9. PLOT THE CUSTOMER SATISFACTION TREND W.R.T EACH TYPE OF ITEM SOLD.
10. WHICH ARE THE BEST 10 PERFORMING STORES?
11. WHICH ARE THE 10 WORST-PERFORMING STORES?
12. WHAT PROPORTION OF CUSTOMERS ARE SUBSCRIBERS? (GENDER WISE)
13. WHICH PAYMENT METHOD IS PREFERRED BY THE CUSTOMERS?
14. PLOT THE SHOPPING FREQUENCY TRENDS BY BOTH GENDERS.
15. HOW DO SEASONAL DIFFERENCES AFFECT THE PREFERRED SHIPPING TYPE BY CUSTOMERS?

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

[2]:
```
#opening the file
df=pd.read_csv("C:/Users/lenovo/OneDrive/Desktop/simran/projects/
 ↪shopping_trends_updated.csv")
df.head(5)
```

[2]:

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	\
0	1	55	Male	Blouse	Clothing	53	
1	2	19	Male	Sweater	Clothing	64	
2	3	50	Male	Jeans	Clothing	73	
3	4	21	Male	Sandals	Footwear	90	
4	5	45	Male	Blouse	Clothing	49	

	Location	Size	Color	Season	Review Rating	Subscription Status	\
0	Kentucky	L	Gray	Winter	3.1	Yes	
1	Maine	L	Maroon	Winter	3.1	Yes	
2	Massachusetts	S	Maroon	Spring	3.1	Yes	
3	Rhode Island	M	Maroon	Spring	3.5	Yes	
4	Oregon	M	Turquoise	Spring	2.7	Yes	

	Shipping Type	Discount Applied	Promo Code Used	Previous Purchases	\
0	Express	Yes	Yes	14	
1	Express	Yes	Yes	2	
2	Free Shipping	Yes	Yes	23	
3	Next Day Air	Yes	Yes	49	
4	Free Shipping	Yes	Yes	31	

	Payment Method	Frequency of Purchases
0	Venmo	Fortnightly
1	Cash	Fortnightly
2	Credit Card	Weekly

```
3          PayPal                 Weekly
4          PayPal                Annually
```

[3]:
```
df.shape
```

[3]:
```
(3900, 18)
```

[4]:
```
df.info()
```

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

[5]:
```
df.columns
```

[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')
```

[6]:
```
df.set_index('Customer ID',inplace=True)
```

[7]:
```
for col in df.columns:
    if df[col].dtype=='object':
        print(col)
        print(df[col].unique())
        print(df[col].nunique())
        print('**********************************')
```

```
Gender
['Male' 'Female']
2
**********
Item Purchased
['Blouse' 'Sweater' 'Jeans' 'Sandals' 'Sneakers' 'Shirt' 'Shorts' 'Coat'
 'Handbag' 'Shoes' 'Dress' 'Skirt' 'Sunglasses' 'Pants' 'Jacket' 'Hoodie'
 'Jewelry' 'T-shirt' 'Scarf' 'Hat' 'Socks' 'Backpack' 'Belt' 'Boots'
 'Gloves']
25
**********
Category
['Clothing' 'Footwear' 'Outerwear' 'Accessories']
4
**********
Location
['Kentucky' 'Maine' 'Massachusetts' 'Rhode Island' 'Oregon' 'Wyoming'
 'Montana' 'Louisiana' 'West Virginia' 'Missouri' 'Arkansas' 'Hawaii'
 'Delaware' 'New Hampshire' 'New York' 'Alabama' 'Mississippi'
 'North Carolina' 'California' 'Oklahoma' 'Florida' 'Texas' 'Nevada'
 'Kansas' 'Colorado' 'North Dakota' 'Illinois' 'Indiana' 'Arizona'
 'Alaska' 'Tennessee' 'Ohio' 'New Jersey' 'Maryland' 'Vermont'
 'New Mexico' 'South Carolina' 'Idaho' 'Pennsylvania' 'Connecticut' 'Utah'
 'Virginia' 'Georgia' 'Nebraska' 'Iowa' 'South Dakota' 'Minnesota'
 'Washington' 'Wisconsin' 'Michigan']
50
**********
Size
['L' 'S' 'M' 'XL']
4
**********
Color
['Gray' 'Maroon' 'Turquoise' 'White' 'Charcoal' 'Silver' 'Pink' 'Purple'
 'Olive' 'Gold' 'Violet' 'Teal' 'Lavender' 'Black' 'Green' 'Peach' 'Red'
 'Cyan' 'Brown' 'Beige' 'Orange' 'Indigo' 'Yellow' 'Magenta' 'Blue']
25
**********
Season
['Winter' 'Spring' 'Summer' 'Fall']
4
```

```
**********
Subscription Status
['Yes' 'No']
2
**********
Shipping Type
['Express' 'Free Shipping' 'Next Day Air' 'Standard' '2-Day Shipping'
 'Store Pickup']
6
**********
Discount Applied
['Yes' 'No']
2
**********
Promo Code Used
['Yes' 'No']
2
**********
Payment Method
['Venmo' 'Cash' 'Credit Card' 'PayPal' 'Bank Transfer' 'Debit Card']
6
**********
Frequency of Purchases
['Fortnightly' 'Weekly' 'Annually' 'Quarterly' 'Bi-Weekly' 'Monthly'
 'Every 3 Months']
7
**********
```

[8]:
```
df.describe()
```

[8]:

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

[9]:
```
df.isnull().sum().sum()
```

[9]: 0

[10]:
```
df.duplicated().sum().sum()
```

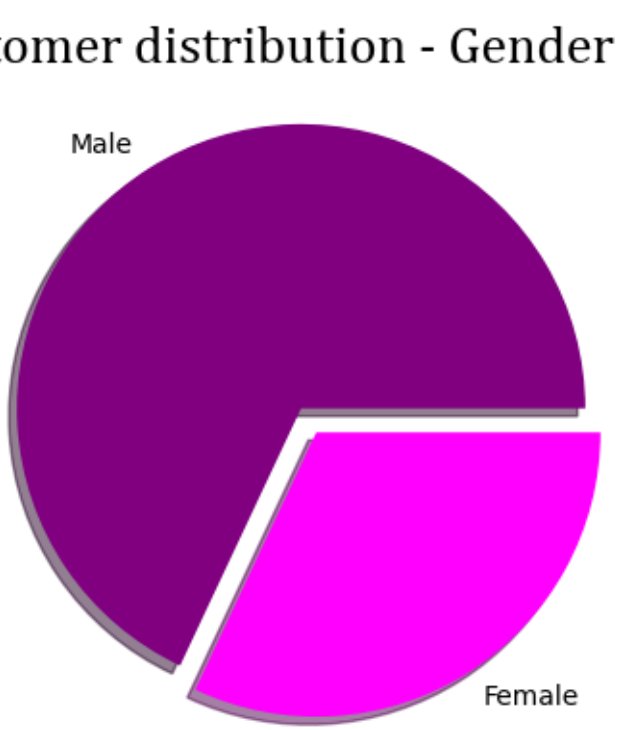
[10]: 0

[11]:
```
df['Gender'].value_counts()
```

[11]:
```
Gender
Male      2652
Female    1248
Name: count, dtype: int64
```

[12]:
```
f={'color':'k', 'size':20,'family':'Cambria'}
plt.pie(x=df['Gender'].value_counts(),data=df,labels=df['Gender'].
  unique(),shadow=True,explode=(0.1,0),colors=('purple','magenta'))
plt.title('Customer distribution - Gender wise',f)
```

[12]:
```
Text(0.5, 1.0, 'Customer distribution - Gender wise')
```



[13]:
```
df['Age'].describe()
```

[13]:
```
count    3900.000000
mean       44.068462
std        15.207589
min        18.000000
25%        31.000000
50%        44.000000
```

```
75%        57.000000
max        70.000000
Name: Age, dtype: float64
```

[14]:
```
df['Age Group']= pd.cut(df['Age'],
 bins=[0,15,18,30,50,70],labels=['Child','Teenager','Young
 Adult','Middle-aged Adult','Old'])
df['Age Group'].value_counts()
```

[14]:
```
Age Group
Old                   1476
Middle-aged Adult     1475
Young Adult            880
Teenager                69
Child                    0
Name: count, dtype: int64
```

[15]:
```
df.head(5)
```

[15]:

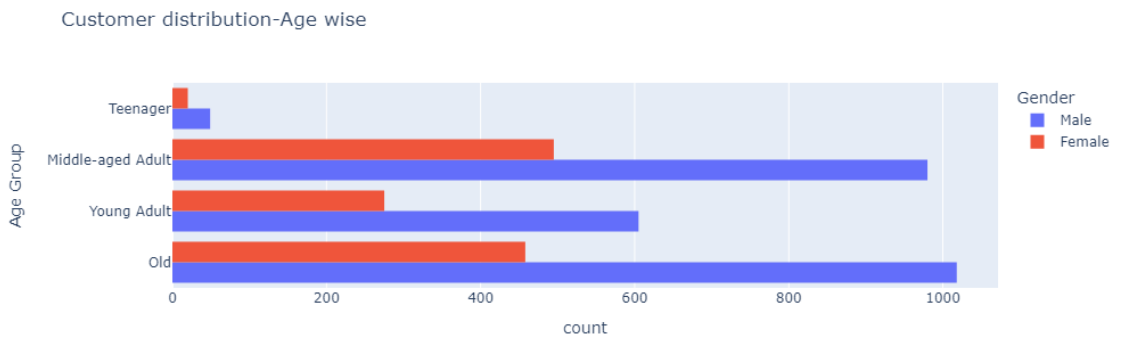
	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	\
0	1	55	Male	Blouse	Clothing	53	
1	2	19	Male	Sweater	Clothing	64	
2	3	50	Male	Jeans	Clothing	73	
3	4	21	Male	Sandals	Footwear	90	
4	5	45	Male	Blouse	Clothing	49	

	Location	Size	Color	Season	Review Rating	Subscription Status	\
0	Kentucky	L	Gray	Winter	3.1	Yes	
1	Maine	L	Maroon	Winter	3.1	Yes	
2	Massachusetts	S	Maroon	Spring	3.1	Yes	
3	Rhode Island	M	Maroon	Spring	3.5	Yes	
4	Oregon	M	Turquoise	Spring	2.7	Yes	

	Shipping Type	Discount Applied	Promo Code Used	Previous Purchases	\
0	Express	Yes	Yes	14	
1	Express	Yes	Yes	2	
2	Free Shipping	Yes	Yes	23	
3	Next Day Air	Yes	Yes	49	
4	Free Shipping	Yes	Yes	31	

	Payment Method	Frequency of Purchases	Age Group
0	Venmo	Fortnightly	Old
1	Cash	Fortnightly	Young Adult
2	Credit Card	Weekly	Middle-aged Adult
3	PayPal	Weekly	Young Adult
4	PayPal	Annually	Middle-aged Adult

[16]:
```
fig=px.histogram(df, y='Age Group',title='Customer distribution-Age␣
 ↪wise',color='Gender',barmode='group')
fig.show()
```



[12]:
```
df['Category'].unique()
```

[12]:
```
array(['Clothing', 'Footwear', 'Outerwear', 'Accessories'], dtype=object)
```

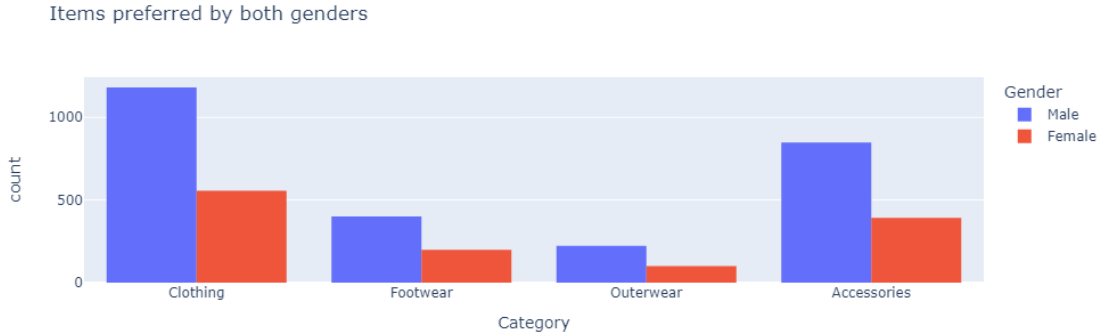
[32]:
```
n=df['Category'].value_counts().sort_values(ascending=False)
n
```

[32]:
```
Category
Clothing       1737
Accessories    1240
Footwear        599
Outerwear       324
Name: count, dtype: int64
```

[34]:
```
df.groupby('Category')['Purchase Amount (USD)'].sum().
 sort_values(ascending=False)
```

[34]:
```
Category
Clothing       104264
Accessories     74200
Footwear        36093
Outerwear       18524
Name: Purchase Amount (USD), dtype: int64
```

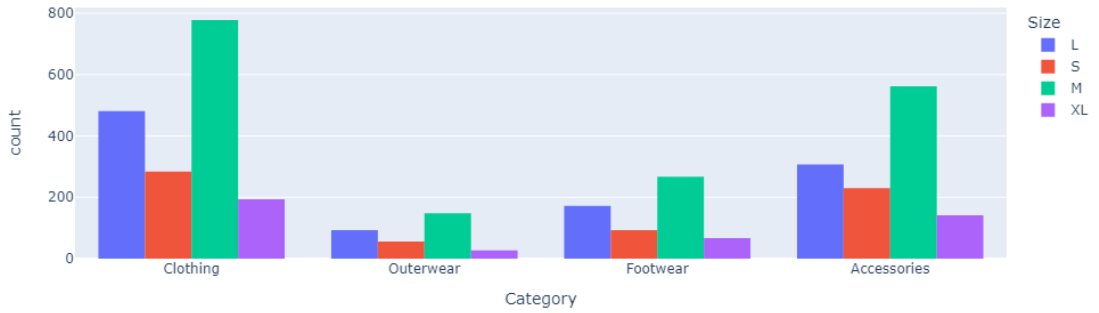
[17]:
```
fig=px.histogram(df, x='Category',color='Gender',barmode='group',title='Items␣
 preferred by both genders')
fig.show()
```



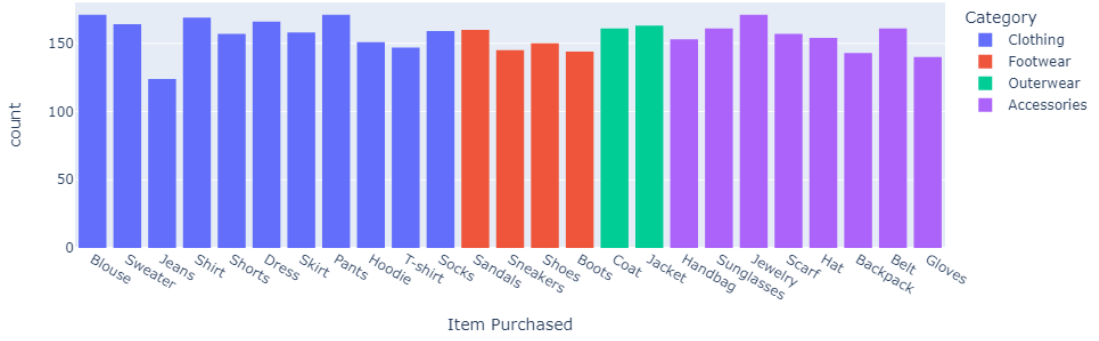
[6]:
```
df.groupby('Size')['Category'].value_counts().sort_values(ascending=False)
```

[6]:
```
Size  Category
M     Clothing       778
      Accessories    562
L     Clothing       481
      Accessories    307
S     Clothing       284
M     Footwear       267
S     Accessories    230
XL    Clothing       194
L     Footwear       172
M     Outerwear      148
XL    Accessories    141
L     Outerwear       93
S     Footwear        93
XL    Footwear        67
S     Outerwear       56
XL    Outerwear       27
Name: count, dtype: int64
```

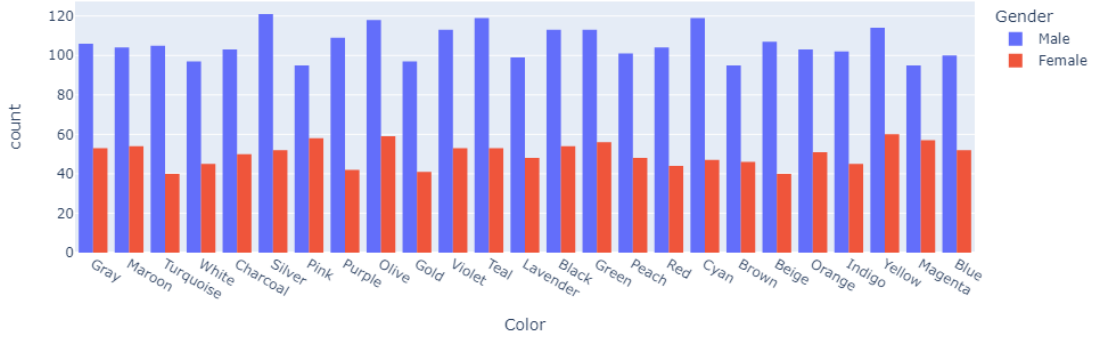
[69]:
```
fig=px.histogram(df, x='Category', color='Size',barmode='group')
fig.show()
```



[61]:
```
fig = px.histogram(df , x = 'Item Purchased' , color = 'Category')
fig.show()
```



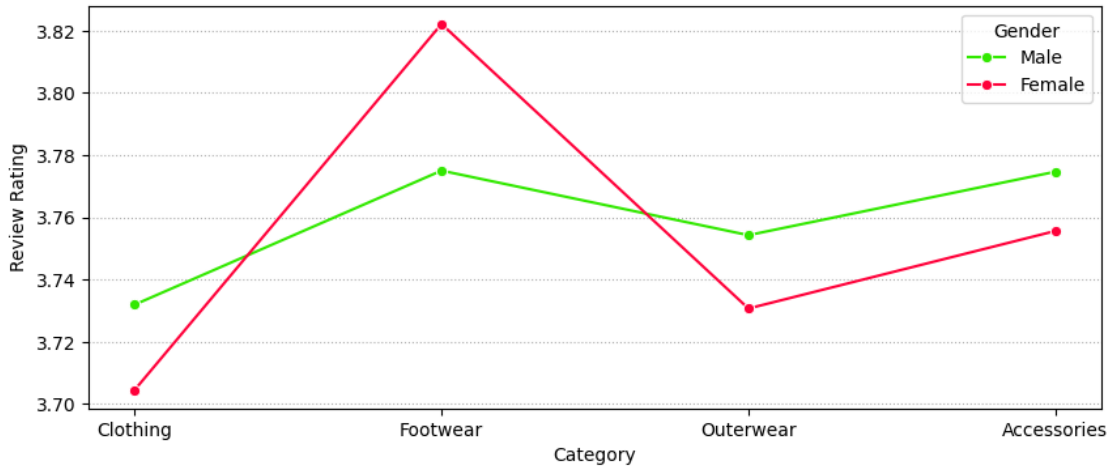
[60]:
```
px.histogram(df, x='Color',color='Gender',barmode='group')
```



[40]:
```
Ratings_given=df.groupby('Category')['Review Rating'].mean().
  sort_values(ascending=False)
Ratings_given
```

[40]:
```
Category
Footwear       3.790651
Accessories    3.768629
Outerwear      3.746914
Clothing       3.723143
Name: Review Rating, dtype: float64
```

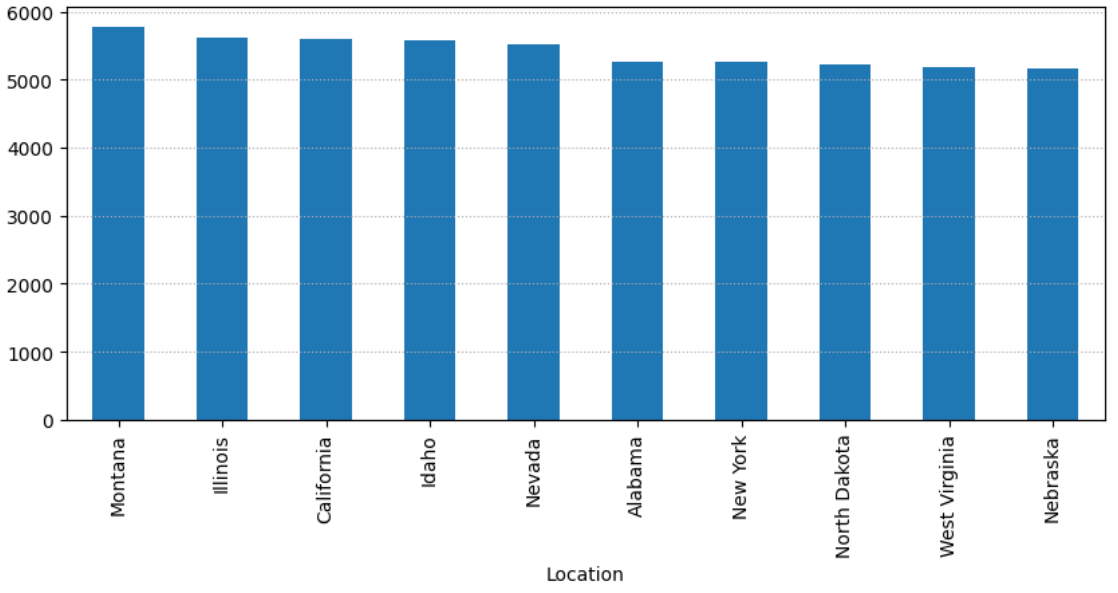
[9]:
```
plt.figure(figsize=(10,4))
sns.lineplot(x='Category',y='Review Rating',data=df, hue='Gender',␣
  palette='prism',marker='o',errorbar=None)
plt.grid(axis='y',ls='dotted')
```



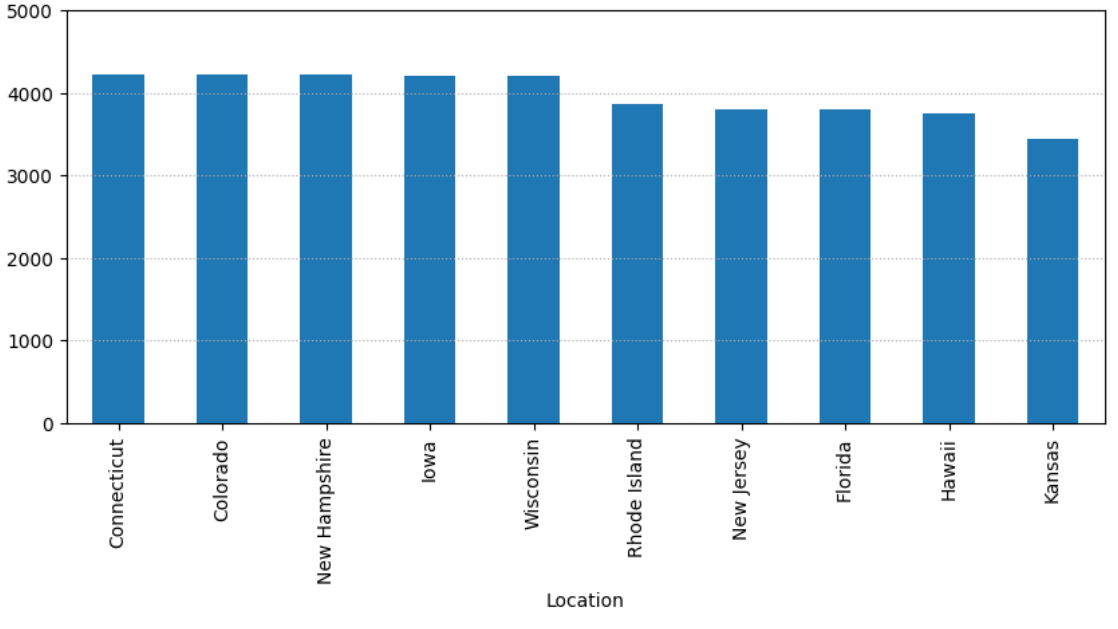
[103]:
```
df['Location'].nunique()
```

[103]:
```
50
```

[51]:
```
plt.figure(figsize=(10,4))
df.groupby('Location')['Purchase Amount (USD)'].sum().nlargest(10).
  plot(kind='bar')
plt.grid(axis='y', ls='dotted')
```



[7]:
```
plt.figure(figsize=(10,4))
df.groupby('Location')['Purchase Amount (USD)'].sum().nsmallest(10).
  sort_values(ascending=False).plot(kind='bar')
plt.yticks(np.arange(0,6000,1000))
plt.grid(axis='y', ls='dotted')
```



[113]:
```
df['Promo Code Used'].value_counts()
```

[113]:
```
Promo Code Used
No      2223
Yes     1677
Name: count, dtype: int64
```

[114]:
```
df['Discount Applied'].value_counts()
```

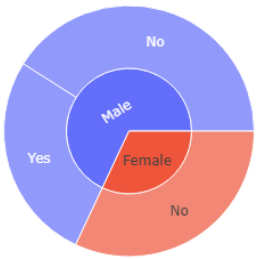
[114]:
```
Discount Applied
No      2223
Yes     1677
Name: count, dtype: int64
```

[38]:
```
df.drop('Promo Code Used', axis=1, inplace=True)
```

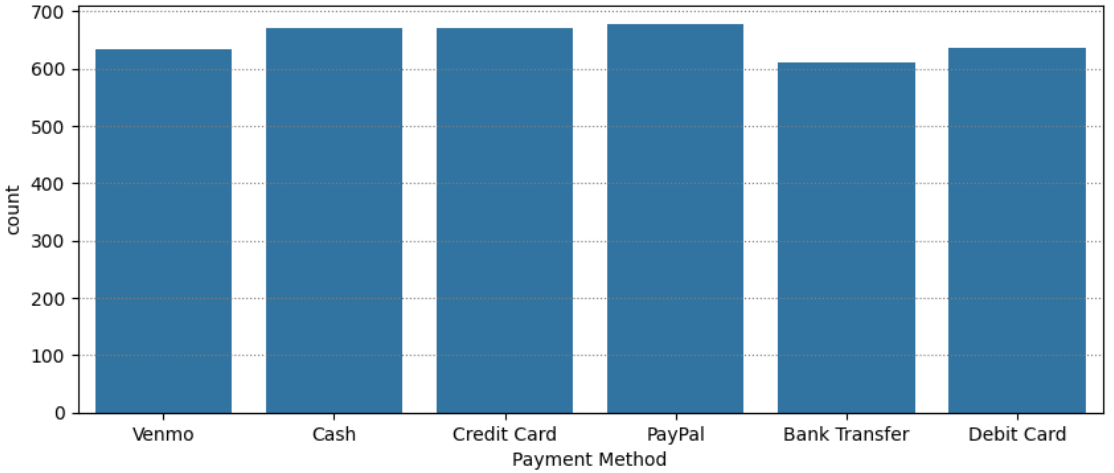
[43]:
```
df['Subscription Status'].value_counts()
```

[43]:
```
Subscription Status
No      2847
Yes     1053
Name: count, dtype: int64
```

[49]:
```
px.sunburst(df, path=['Gender','Subscription Status'])
```



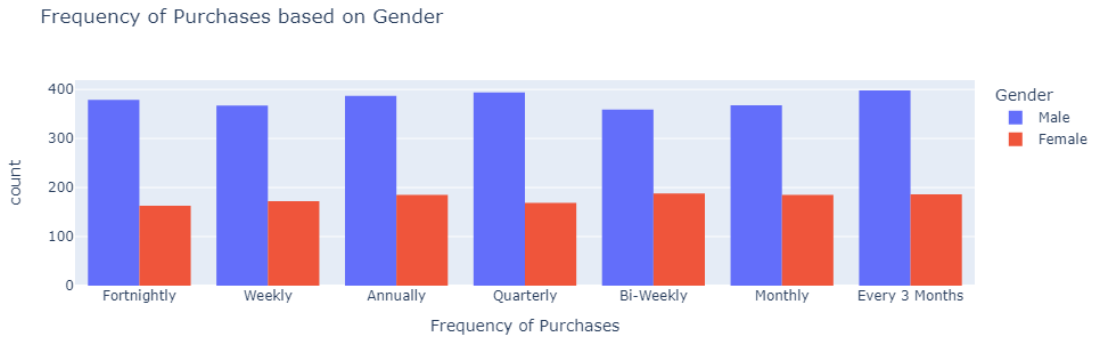
[43]:
```
plt.figure(figsize=(10,4))
sns.countplot(x='Payment Method',data=df)
plt.grid(axis='y', ls='dotted', color='grey')
```



[57]:
```
df.groupby('Gender')['Frequency of Purchases'].value_counts()
```

[57]:
```
Gender  Frequency of Purchases
Female  Bi-Weekly                 188
        Every 3 Months            186
        Annually                  185
        Monthly                   185
        Weekly                    172
        Quarterly                 169
        Fortnightly               163
Male    Every 3 Months            398
        Quarterly                 394
        Annually                  387
        Fortnightly               379
        Monthly                   368
        Weekly                    367
        Bi-Weekly                 359
Name: count, dtype: int64
```

[44]:
```
fig = px.histogram(df , x = 'Frequency of Purchases',color =
  'Gender',barmode='group',title='Frequency of Purchases based on Gender')
fig.show()
```



[13]:
```
df['Shipping Type'].value_counts()
```

[13]:
```
Shipping Type
Free Shipping     675
Standard          654
Store Pickup      650
Next Day Air      648
Express           646
2-Day Shipping    627
Name: count, dtype: int64
```

[81]:
```
df.head(2)
```

[81]:

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
1	55	Male	Blouse	Clothing	53	Kentucky	L	Gray	Winter	3.1	Yes	Express	Yes	Yes	14	Venmo	Fortnightly
2	19	Male	Sweater	Clothing	64	Maine	L	Maroon	Winter	3.1	Yes	Express	Yes	Yes	2	Cash	Fortnightly

[53]:
```
fig=px.histogram(df, x='Shipping Type',color='Season',barmode='group')
fig.show()
```

[ ]:

THANK YOU