loob98ubx

January 10, 2025

1 Basics import

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

[2]:
```
shop= pd.read_csv('shopping_trends_updated.csv')
```

[3]:
```
shop.head()
```

[3]:

	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
0	1	55	Male	Blouse	Clothing	53	Kentucky	L	Gray	Winter	3.1	Yes	Express	Yes	Yes	14	Venmo	Fortnightly
1	2	19	Male	Sweater	Clothing	64	Maine	L	Maroon	Winter	3.1	Yes	Express	Yes	Yes	2	Cash	Fortnightly
2	3	50	Male	Jeans	Clothing	73	Massachusetts	S	Maroon	Spring	3.1	Yes	Free Shipping	Yes	Yes	23	Credit Card	Weekly
3	4	21	Male	Sandals	Footwear	90	Rhode Island	M	Maroon	Spring	3.5	Yes	Next Day Air	Yes	Yes	49		
4	5	45	Male	Blouse	Clothing	49	Oregon	M	Turquoise	Spring	2.7	Yes	Free Shipping	Yes	Yes	31		

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

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

[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
```

[5]:
```
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
```

[6]:
```
shop.isnull().sum()
```

[6]:
```
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
```

[7]:
```
shop['Gender'].unique()
```

[7]:
```
array(['Male', 'Female'], dtype=object)
```

2 Overall distribution of customer ages

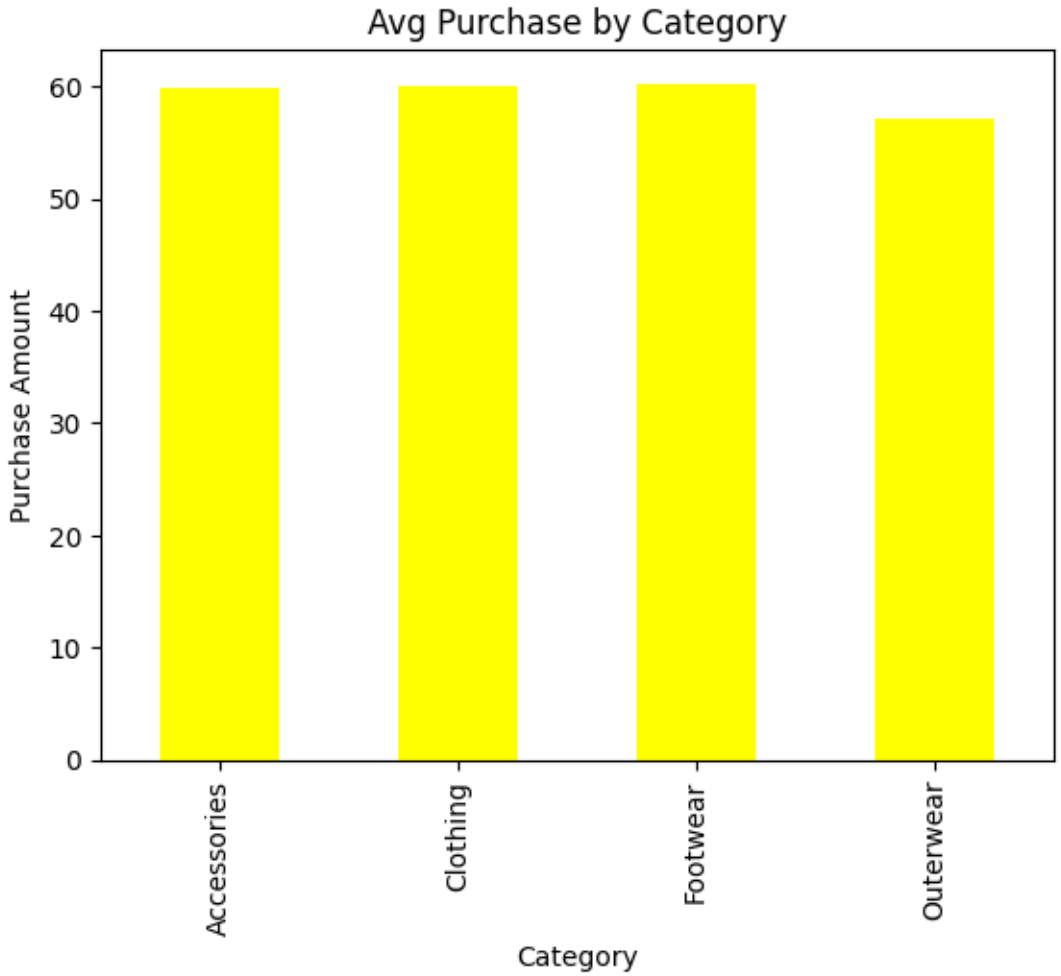
[8]:
```
shop['Age'].value_counts()
```

[8]:
```
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
```

3 Average purchase amount vary across different product categories

[15]:
```
avg_purchase = shop.groupby('Category')['Purchase Amount (USD)'].mean()
avg_purchase.plot(kind='bar', color='yellow')
plt.title('Avg Purchase by Category')
plt.xlabel('Category')
plt.ylabel('Purchase Amount')
plt.show()
```

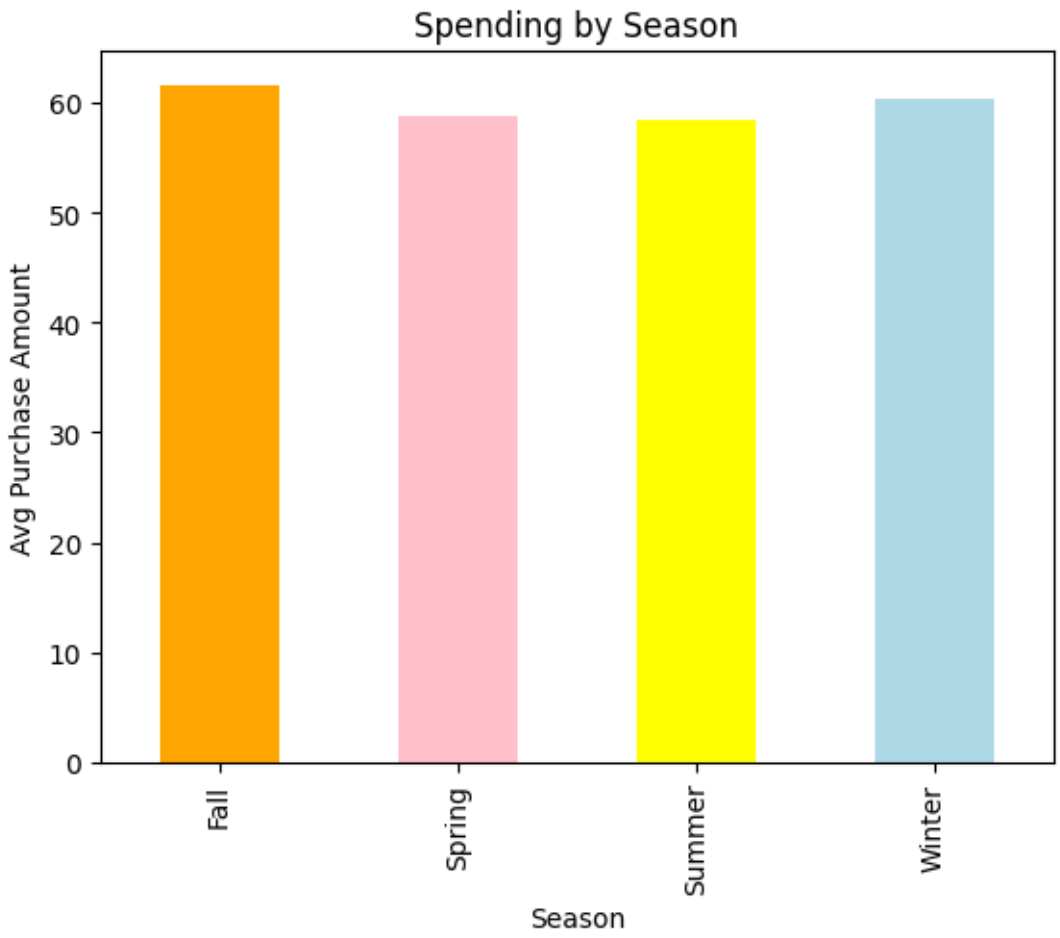


4 Number of Purchases by Gender

[19]:
```
gender_count = shop['Gender'].value_counts()
gender_count.plot(kind='bar', color=['lightblue', 'pink'])
plt.title('Purchases by Gender')
plt.xlabel('Gender')
plt.ylabel('Count')
plt.show()
```

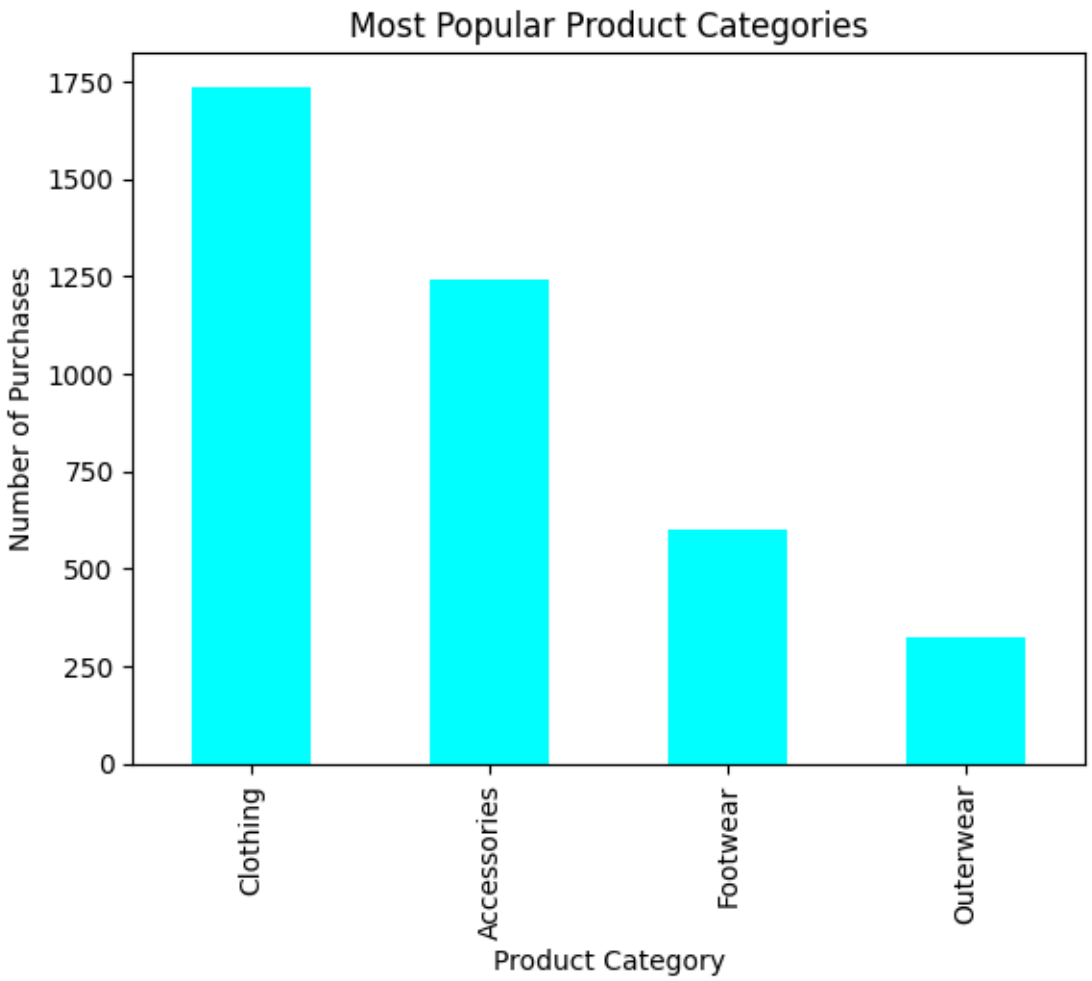
5 Spending by Season

[22]:
```
season_spending = shop.groupby('Season')['Purchase Amount (USD)'].mean()
season_spending.plot(kind='bar', color=['orange','pink','yellow','lightblue'])
plt.title('Spending by Season')
plt.xlabel('Season')
plt.ylabel('Avg Purchase Amount')
plt.show()
```



6 Frequency of Purchases by Age Group

[25]:
```
popular_category = shop['Category'].value_counts()
popular_category.plot(kind='bar', color='cyan')
plt.title('Most Popular Product Categories')
plt.xlabel('Product Category')
plt.ylabel('Number of Purchases')
plt.show()
```



7 Average rating given by customers for each product category?

[26]:
```
avg_rating_by_category = shop.groupby('Category')['Review Rating'].mean()
print(avg_rating_by_category)
```

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

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

[27]:
```
purchase_behavior = shop.groupby('Subscription Status')['Purchase Amount␣
↪(USD)'].mean()
print(purchase_behavior)
```

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

9 8. Which payment method is the most popular among customers?

[28]:
```
popular_payment_method = shop['Payment Method'].value_counts()
print(popular_payment_method)
```

```
Payment Method
PayPal           677
Credit Card      671
Cash             670
Debit Card       636
Venmo            634
Bank Transfer    612
Name: count, dtype: int64
```

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

[29]:
```
promo_usage = shop.groupby('Promo Code Used')['Purchase Amount (USD)'].mean()
print(promo_usage)
```

```
Promo Code Used
No     60.130454
Yes    59.279070
Name: Purchase Amount (USD), dtype: float64
```

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

[30]:
```
popular_colors = shop['Color'].value_counts()
print(popular_colors)
```

```
Color
Olive         177
Yellow        174
Silver        173
Teal          172
Green         169
Black         167
Cyan          166
Violet        166
Gray          159
Maroon        158
Orange        154
Charcoal      153
Pink          153
Magenta       152
Blue          152
Purple        151
Peach         149
Red           148
Beige         147
Indigo        147
Lavender      147
Turquoise     145
White         142
Brown         141
Gold          138
Name: count, dtype: int64
```

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

[31]:
```
discount_impact = shop.groupby('Discount Applied')['Purchase Amount (USD)'].
  ↪mean()
print(discount_impact)
```

```
Discount Applied
No     60.130454
Yes    59.279070
Name: Purchase Amount (USD), dtype: float64
```

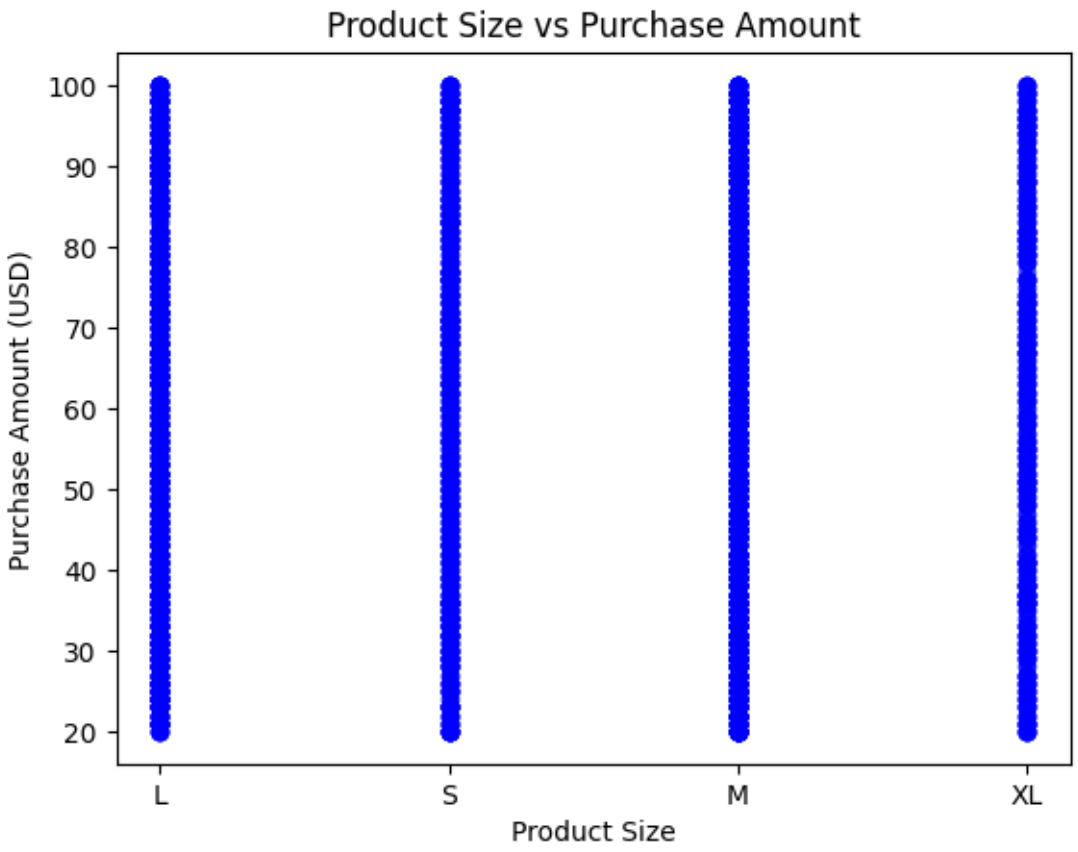
13 Which shipping type is preferred by customers for different product categories?

[32]:
```
preferred_shipping = shop.groupby('Category')['Shipping Type'].agg(lambda x: x.
↪mode()[0])
print(preferred_shipping)
```

```
Category
Accessories     Store Pickup
Clothing            Standard
Footwear       Free Shipping
Outerwear      Free Shipping
Name: Shipping Type, dtype: object
```

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

[43]:
```
plt.scatter(shop['Size'], shop['Purchase Amount (USD)'], color='blue', alpha=0.
↪5)
plt.title('Product Size vs Purchase Amount')
plt.xlabel('Product Size')
plt.ylabel('Purchase Amount (USD)')
plt.show()
```



15 How does the purchase amount differ based on the review ratings given by customers?

[45]:
```
purchase_by_rating = shop.groupby('Review Rating')['Purchase Amount (USD)'].
  ↪mean()
print(purchase_by_rating)
```

```
Review Rating
2.5    62.287879
2.6    59.566038
2.7    59.363636
2.8    57.066176
2.9    56.470588
3.0    60.728395
3.1    58.770701
3.2    61.315789
3.3    59.861842
3.4    59.005495
3.5    58.833333
```

```
3.6    57.322148
3.7    58.974359
3.8    60.873239
3.9    58.926380
4.0    59.237569
4.1    61.959459
4.2    60.853801
4.3    59.673469
4.4    60.525316
4.5    59.489209
4.6    57.683908
4.7    59.283784
4.8    61.881944
4.9    63.885542
5.0    64.352941
Name: Purchase Amount (USD), dtype: float64
```

16 Are there notable differences in purchase behavior between different locations?

[47]:
```
purchase_by_location = shop.groupby('Location')['Purchase Amount (USD)'].mean()
print(purchase_by_location)
```

```
Location
Alabama           59.112360
Alaska            67.597222
Arizona           66.553846
Arkansas          61.113924
California        59.000000
Colorado          56.293333
Connecticut       54.179487
Delaware          55.325581
Florida           55.852941
Georgia           58.797468
Hawaii            57.723077
Idaho             60.075269
Illinois          61.054348
Indiana           58.924051
Iowa              60.884058
Kansas            54.555556
Kentucky          55.721519
Louisiana         57.714286
Maine             56.987013
Maryland          55.755814
Massachusetts     60.888889
Michigan          62.095890
Minnesota         56.556818
```

```
Mississippi       61.037500
Missouri          57.913580
Montana           60.250000
Nebraska          59.448276
Nevada            63.379310
New Hampshire     59.422535
New Jersey        56.746269
New Mexico        61.901235
New York          60.425287
North Carolina    60.794872
North Dakota      62.891566
Ohio              60.376623
Oklahoma          58.346667
Oregon            57.337838
Pennsylvania      66.567568
Rhode Island      61.444444
South Carolina    58.407895
South Dakota      60.514286
Tennessee         61.974026
Texas             61.194805
Utah              62.577465
Vermont           57.176471
Virginia          62.883117
Washington        63.328767
West Virginia     63.876543
Wisconsin         55.946667
Wyoming           60.690141
Name: Purchase Amount (USD), dtype: float64
```