Analysis on Shopping Trends

EDA analysis by Salonee Jadhav

Table of contents

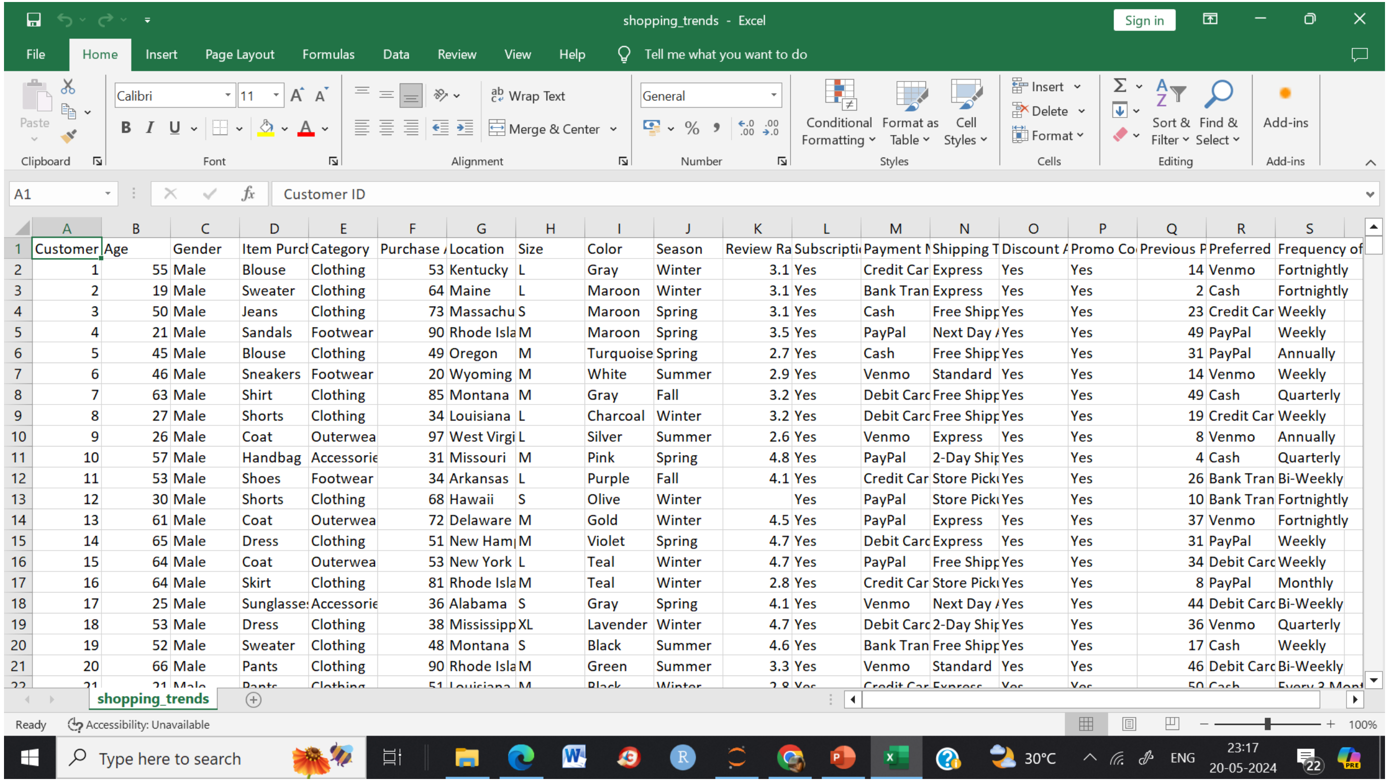
01 Preview of the csv dataset

02 Objective of the analysis

03 Exploratory Data Analysis

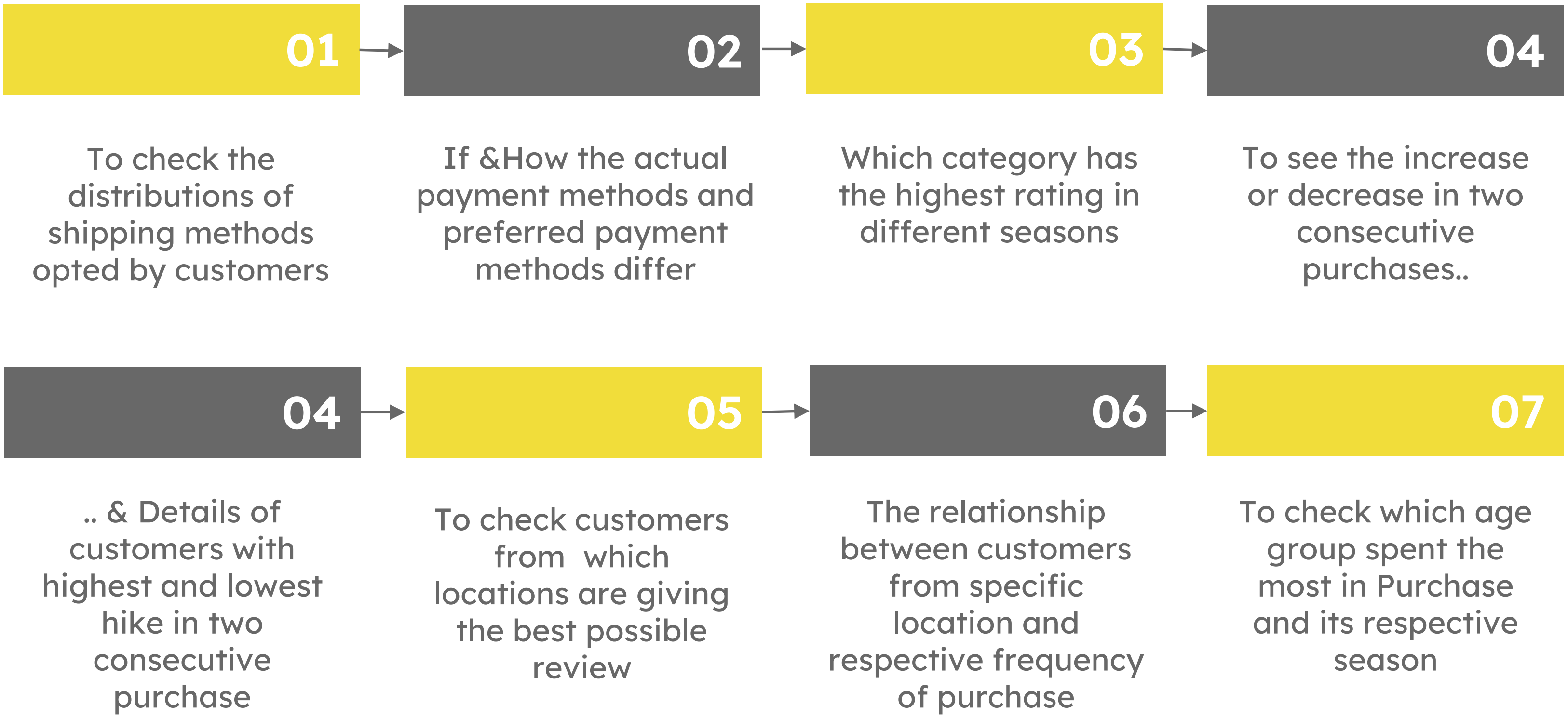
04 Conclusions

Dataset in the question:



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Customer	Age	Gender	Item Purcl	Category	Purchase /	Location	Size	Color	Season	Review Ra	Subscripti	Payment M	Shipping T	Discount /	Promo Co	Previous P	Preferred	Frequency of
2	1	55	Male	Blouse	Clothing	53	Kentucky	L	Gray	Winter	3.1	Yes	Credit Car	Express	Yes	Yes	14	Venmo	Fortnightly
3	2	19	Male	Sweater	Clothing	64	Maine	L	Maroon	Winter	3.1	Yes	Bank Tran	Express	Yes	Yes	2	Cash	Fortnightly
4	3	50	Male	Jeans	Clothing	73	Massachu	S	Maroon	Spring	3.1	Yes	Cash	Free Shipp	Yes	Yes	23	Credit Car	Weekly
5	4	21	Male	Sandals	Footwear	90	Rhode Isla	M	Maroon	Spring	3.5	Yes	PayPal	Next Day /	Yes	Yes	49	PayPal	Weekly
6	5	45	Male	Blouse	Clothing	49	Oregon	M	Turquoise	Spring	2.7	Yes	Cash	Free Shipp	Yes	Yes	31	PayPal	Annually
7	6	46	Male	Sneakers	Footwear	20	Wyoming	M	White	Summer	2.9	Yes	Venmo	Standard	Yes	Yes	14	Venmo	Weekly
8	7	63	Male	Shirt	Clothing	85	Montana	M	Gray	Fall	3.2	Yes	Debit Carc	Free Shipp	Yes	Yes	49	Cash	Quarterly
9	8	27	Male	Shorts	Clothing	34	Louisiana	L	Charcoal	Winter	3.2	Yes	Debit Carc	Free Shipp	Yes	Yes	19	Credit Car	Weekly
10	9	26	Male	Coat	Outerwea	97	West Virgi	L	Silver	Summer	2.6	Yes	Venmo	Express	Yes	Yes	8	Venmo	Annually
11	10	57	Male	Handbag	Accessorie	31	Missouri	M	Pink	Spring	4.8	Yes	PayPal	2-Day Ship	Yes	Yes	4	Cash	Quarterly
12	11	53	Male	Shoes	Footwear	34	Arkansas	L	Purple	Fall	4.1	Yes	Credit Car	Store Picku	Yes	Yes	26	Bank Tran	Bi-Weekly
13	12	30	Male	Shorts	Clothing	68	Hawaii	S	Olive	Winter		Yes	PayPal	Store Picku	Yes	Yes	10	Bank Tran	Fortnightly
14	13	61	Male	Coat	Outerwea	72	Delaware	M	Gold	Winter	4.5	Yes	PayPal	Express	Yes	Yes	37	Venmo	Fortnightly
15	14	65	Male	Dress	Clothing	51	New Ham	M	Violet	Spring	4.7	Yes	Debit Carc	Express	Yes	Yes	31	PayPal	Weekly
16	15	64	Male	Coat	Outerwea	53	New York	L	Teal	Winter	4.7	Yes	PayPal	Free Shipp	Yes	Yes	34	Debit Carc	Weekly
17	16	64	Male	Skirt	Clothing	81	Rhode Isla	M	Teal	Winter	2.8	Yes	Credit Car	Store Picku	Yes	Yes	8	PayPal	Monthly
18	17	25	Male	Sunglasses	Accessorie	36	Alabama	S	Gray	Spring	4.1	Yes	Venmo	Next Day /	Yes	Yes	44	Debit Carc	Bi-Weekly
19	18	53	Male	Dress	Clothing	38	Mississipp	XL	Lavender	Winter	4.7	Yes	Debit Carc	2-Day Ship	Yes	Yes	36	Venmo	Quarterly
20	19	52	Male	Sweater	Clothing	48	Montana	S	Black	Summer	4.6	Yes	Bank Tran	Free Shipp	Yes	Yes	17	Cash	Weekly
21	20	66	Male	Pants	Clothing	90	Rhode Isla	M	Green	Summer	3.3	Yes	Venmo	Standard	Yes	Yes	46	Debit Carc	Bi-Weekly
22	21	21	Male	Pants	Clothing	51	Louisiana	M	Black	Winter	2.8	Yes	Credit Car	Express	Yes	Yes	50	Cash	Every 3 Mont

Objectives:



01

To check the distributions of shipping methods opted by customers

02

If &How the actual payment methods and preferred payment methods differ

03

Which category has the highest rating in different seasons

04

To see the increase or decrease in two consecutive purchases..

04

.. & Details of customers with highest and lowest hike in two consecutive purchase

05

To check customers from which locations are giving the best possible review

06

The relationship between customers from specific location and respective frequency of purchase

07

To check which age group spent the most in Purchase and its respective season

Objectives:

08

To check the avg purchase value by different age groups over all seasons

09

To get the detail information about each category with respect to review ratings

10

To get the overall idea of all the item purchased in all 4 seasons

11

To check the probability of customer aged more and less than 40 years giving more than 4 star rating?

12

To check if the genders are playing important role in opting for subscription?

13

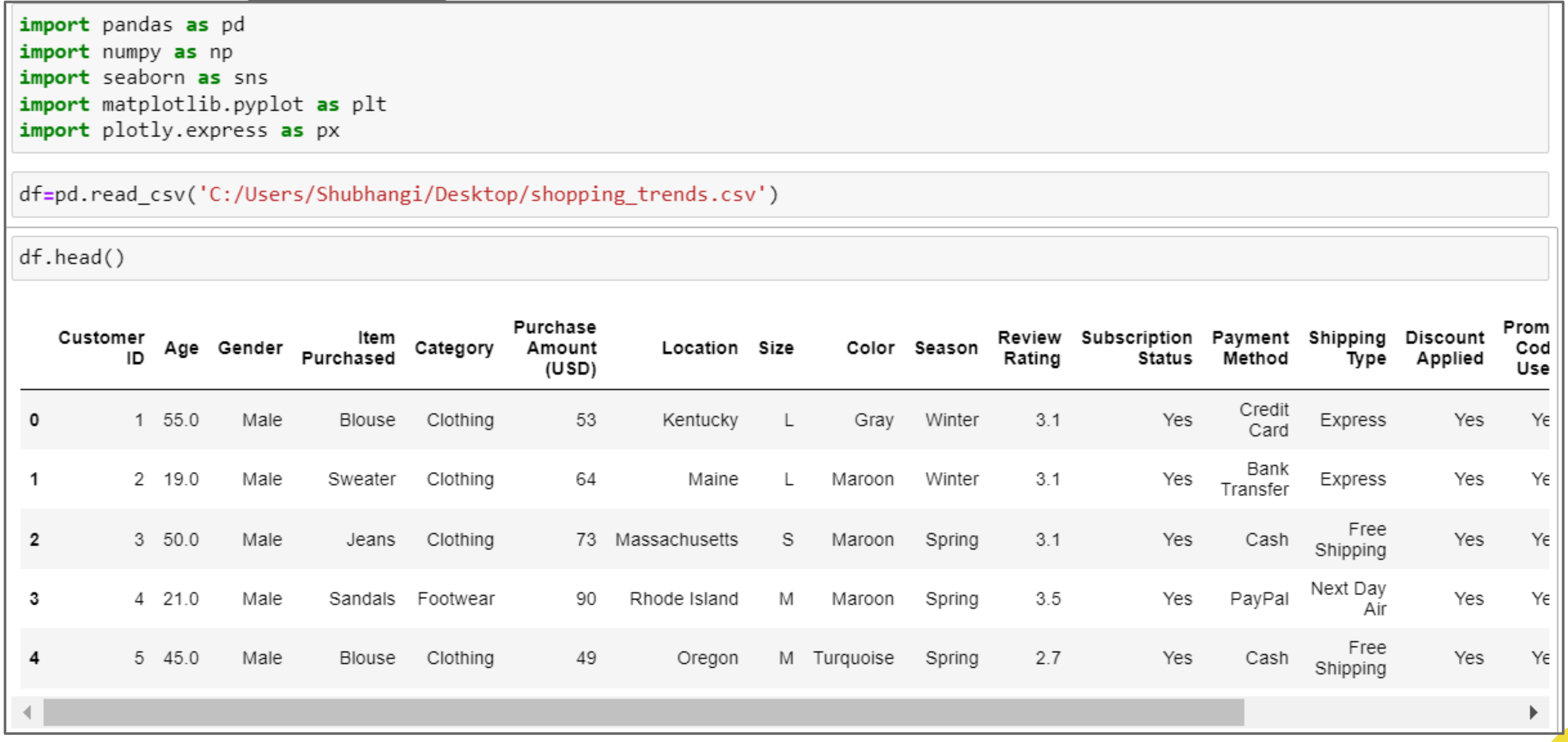
To get idea of the requirement of different sizes of different categories from various locations

14

What is the difference (comparison) in previous purchased amount and present purchase amount?

EDA

Exploratory Data Analysis

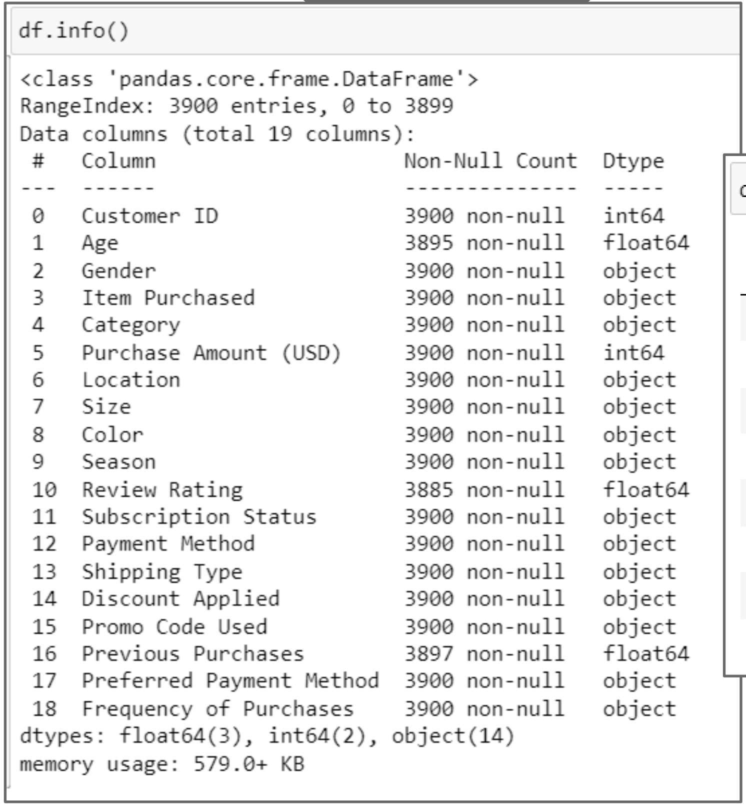


```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import plotly.express as px
```

```
df=pd.read_csv('C:/Users/Shubhangi/Desktop/shopping_trends.csv')
```

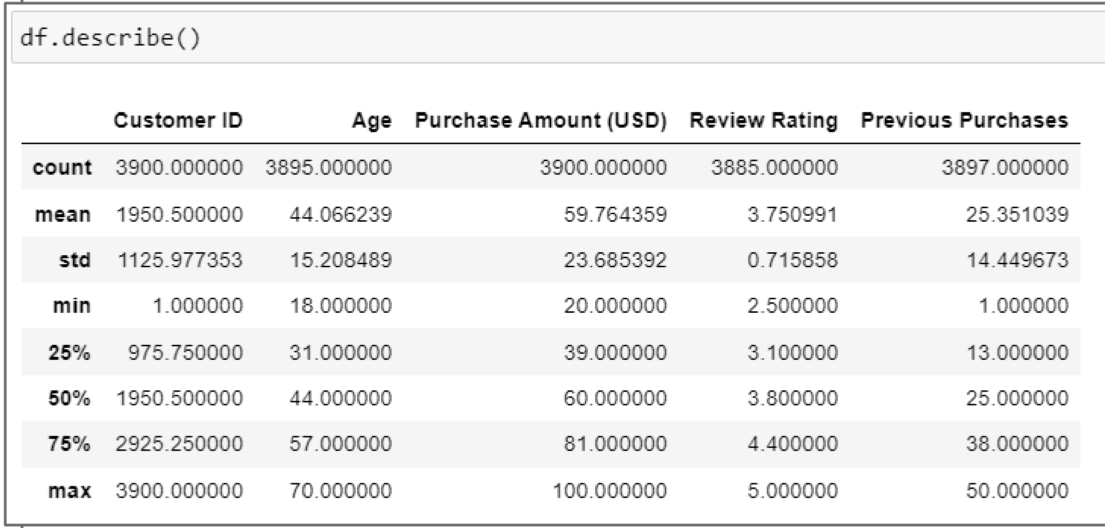
```
df.head()
```

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size	Color	Season	Review Rating	Subscription Status	Payment Method	Shipping Type	Discount Applied	Prom Cod Use
0	1	55.0	Male	Blouse	Clothing	53	Kentucky	L	Gray	Winter	3.1	Yes	Credit Card	Express	Yes	Ye
1	2	19.0	Male	Sweater	Clothing	64	Maine	L	Maroon	Winter	3.1	Yes	Bank Transfer	Express	Yes	Ye
2	3	50.0	Male	Jeans	Clothing	73	Massachusetts	S	Maroon	Spring	3.1	Yes	Cash	Free Shipping	Yes	Ye
3	4	21.0	Male	Sandals	Footwear	90	Rhode Island	M	Maroon	Spring	3.5	Yes	PayPal	Next Day Air	Yes	Ye
4	5	45.0	Male	Blouse	Clothing	49	Oregon	M	Turquoise	Spring	2.7	Yes	Cash	Free Shipping	Yes	Ye



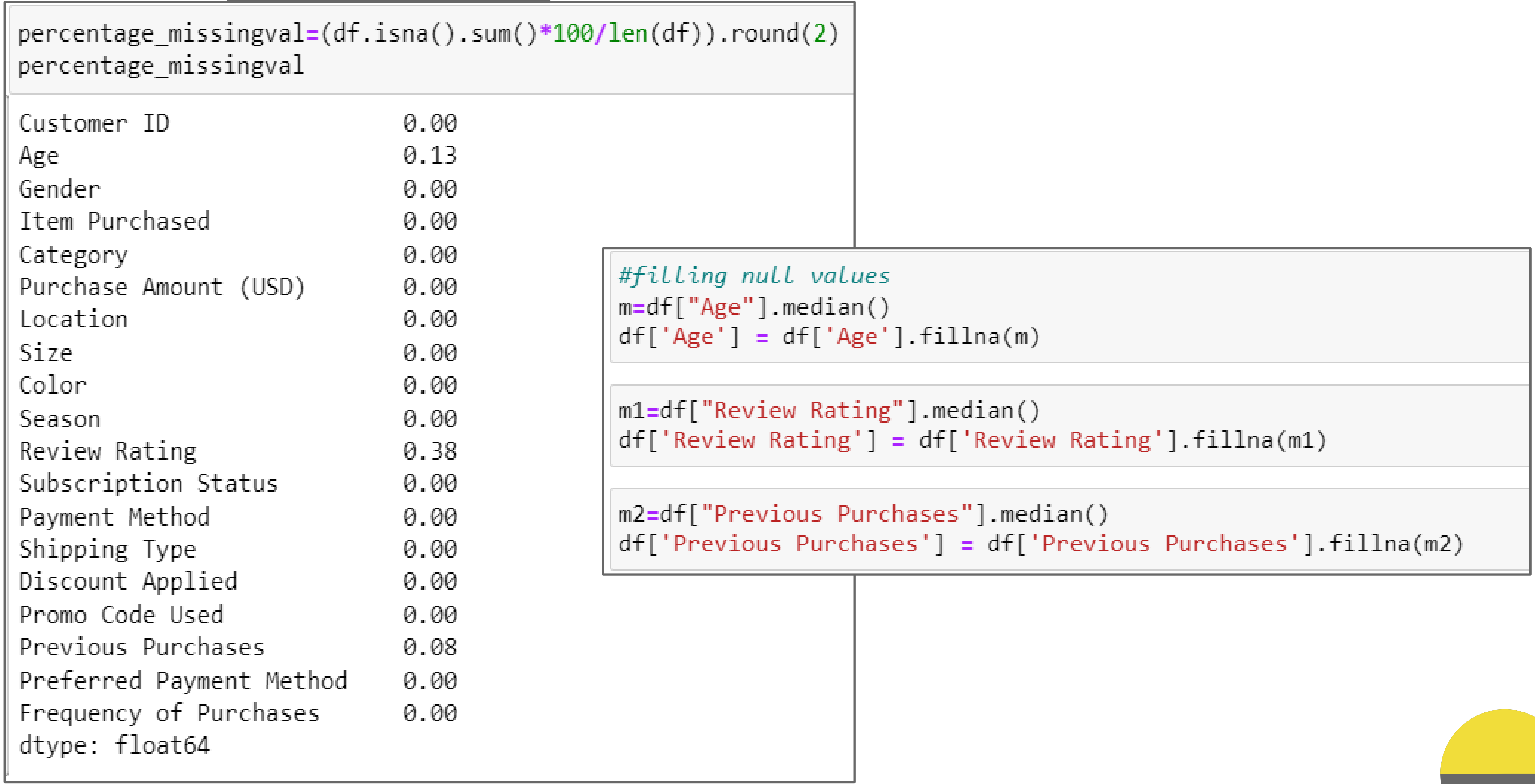
```
df.info()
```

```
<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                     3895 non-null   float64
 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           3885 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      3897 non-null   float64
 17  Preferred Payment Method  3900 non-null   object 
 18  Frequency of Purchases  3900 non-null   object 
dtypes: float64(3), int64(2), object(14)
memory usage: 579.0+ KB
```



```
df.describe()
```

	Customer ID	Age	Purchase Amount (USD)	Review Rating	Previous Purchases
count	3900.000000	3895.000000	3900.000000	3885.000000	3897.000000
mean	1950.500000	44.066239	59.764359	3.750991	25.351039
std	1125.977353	15.208489	23.685392	0.715858	14.449673
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	3.800000	25.000000
75%	2925.250000	57.000000	81.000000	4.400000	38.000000
max	3900.000000	70.000000	100.000000	5.000000	50.000000



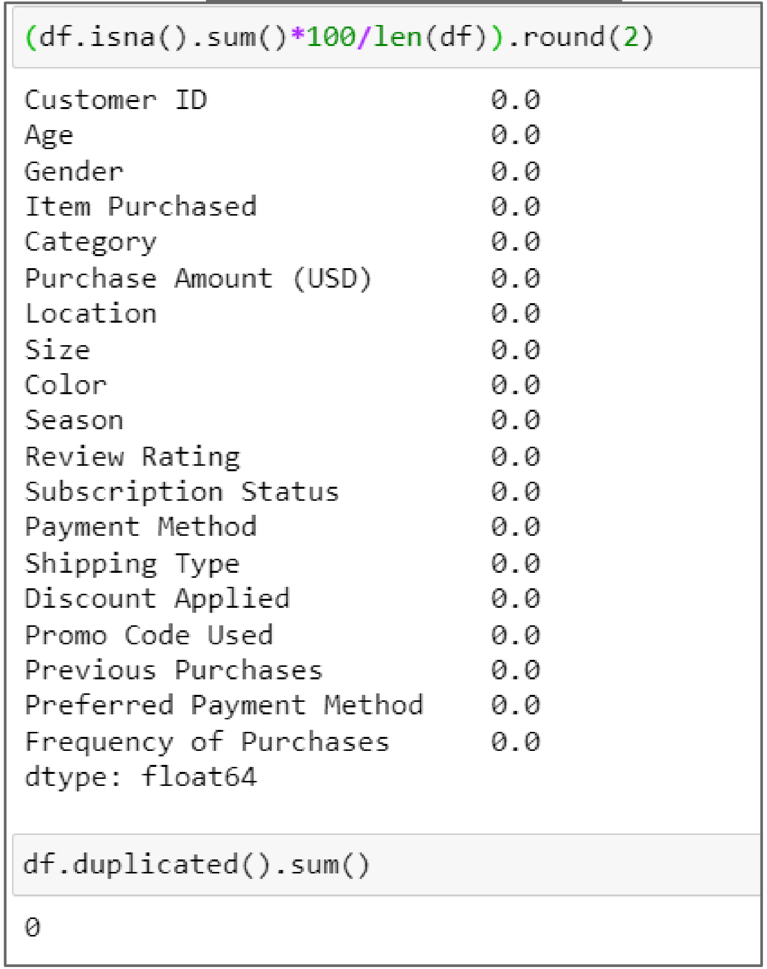
```
percentage_missingval=(df.isna().sum()*100/len(df)).round(2)
percentage_missingval
```

```
Customer ID                 0.00
Age                         0.13
Gender                      0.00
Item Purchased              0.00
Category                    0.00
Purchase Amount (USD)       0.00
Location                    0.00
Size                        0.00
Color                       0.00
Season                      0.00
Review Rating               0.38
Subscription Status         0.00
Payment Method              0.00
Shipping Type               0.00
Discount Applied            0.00
Promo Code Used             0.00
Previous Purchases          0.08
Preferred Payment Method    0.00
Frequency of Purchases      0.00
dtype: float64
```

```
#filling null values
m=df["Age"].median()
df['Age'] = df['Age'].fillna(m)
```

```
m1=df["Review Rating"].median()
df['Review Rating'] = df['Review Rating'].fillna(m1)
```

```
m2=df["Previous Purchases"].median()
df['Previous Purchases'] = df['Previous Purchases'].fillna(m2)
```

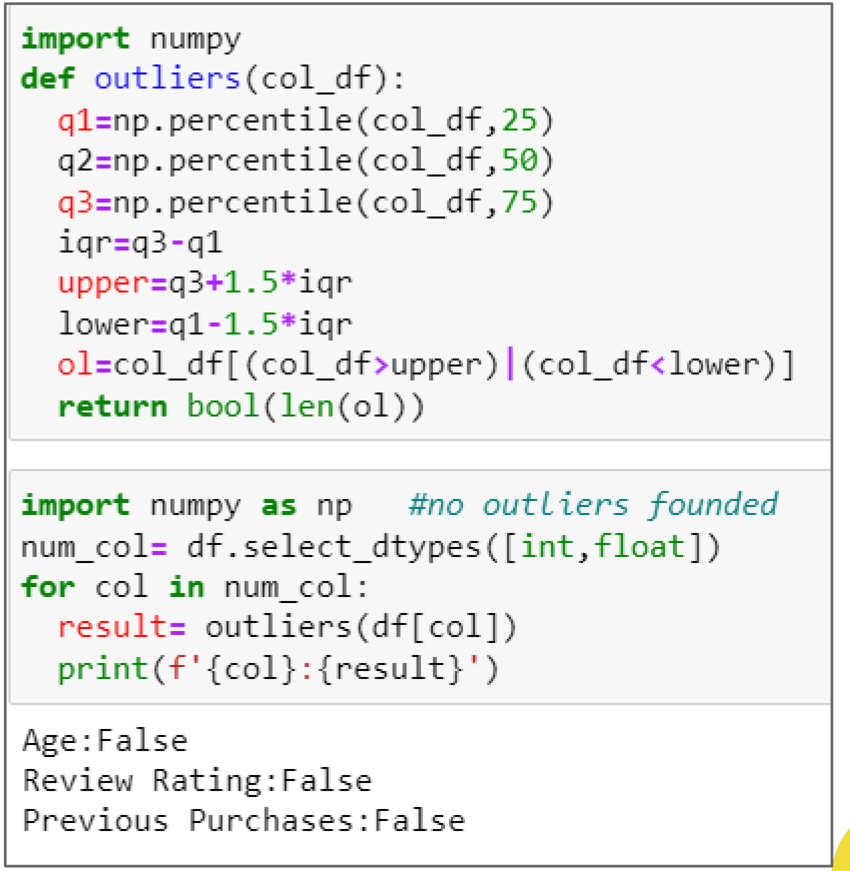


```
(df.isna().sum()*100/len(df)).round(2)
```

```
Customer ID                 0.0
Age                         0.0
Gender                      0.0
Item Purchased              0.0
Category                    0.0
Purchase Amount (USD)       0.0
Location                    0.0
Size                        0.0
Color                       0.0
Season                      0.0
Review Rating               0.0
Subscription Status         0.0
Payment Method              0.0
Shipping Type               0.0
Discount Applied            0.0
Promo Code Used             0.0
Previous Purchases          0.0
Preferred Payment Method    0.0
Frequency of Purchases      0.0
dtype: float64
```

```
df.duplicated().sum()
```

```
0
```

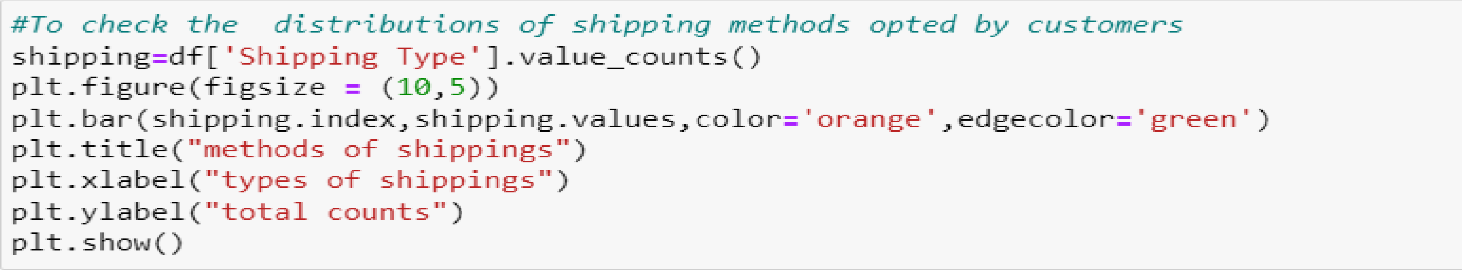


```
import numpy
def outliers(col_df):
  q1=np.percentile(col_df,25)
  q2=np.percentile(col_df,50)
  q3=np.percentile(col_df,75)
  iqr=q3-q1
  upper=q3+1.5*iqr
  lower=q1-1.5*iqr
  ol=col_df[(col_df>upper)|(col_df<lower)]
  return bool(len(ol))
```

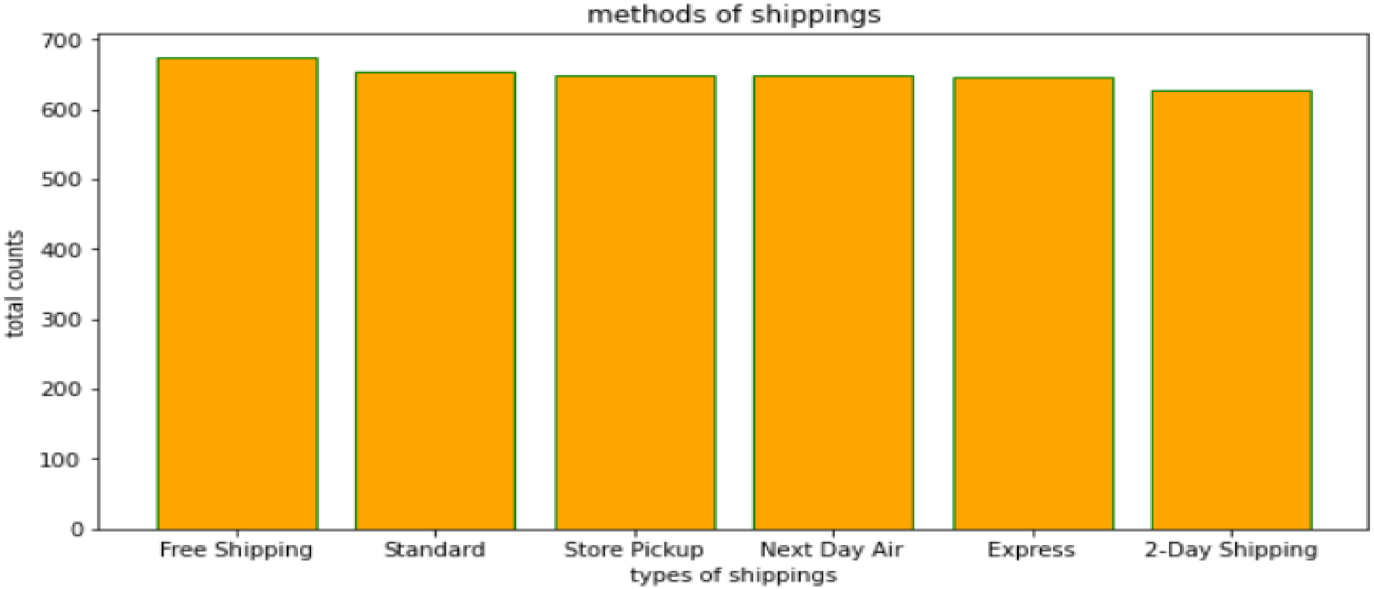
```
import numpy as np   #no outliers founded
num_col= df.select_dtypes([int,float])
for col in num_col:
  result= outliers(df[col])
  print(f'{col}:{result}')
```

```
Age:False
Review Rating:False
Previous Purchases:False
```

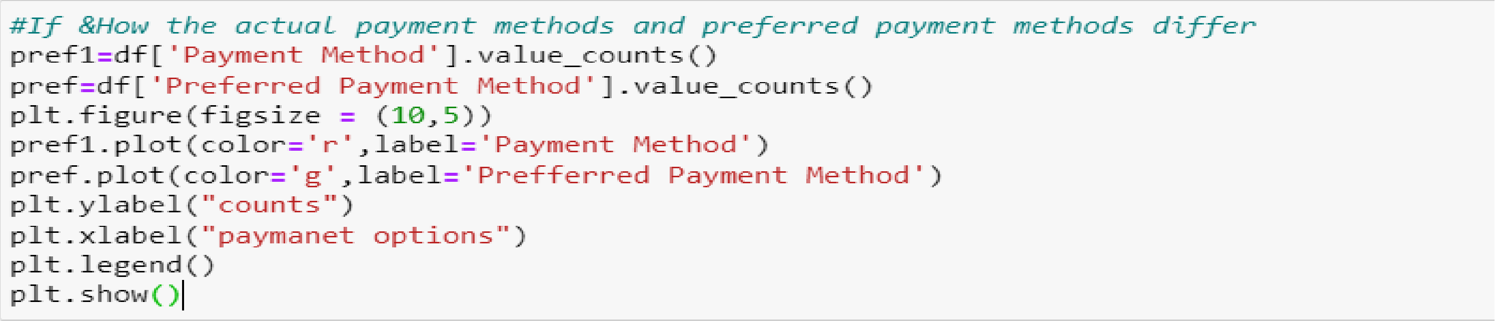
1. To check the distributions of shipping methods opted by customers



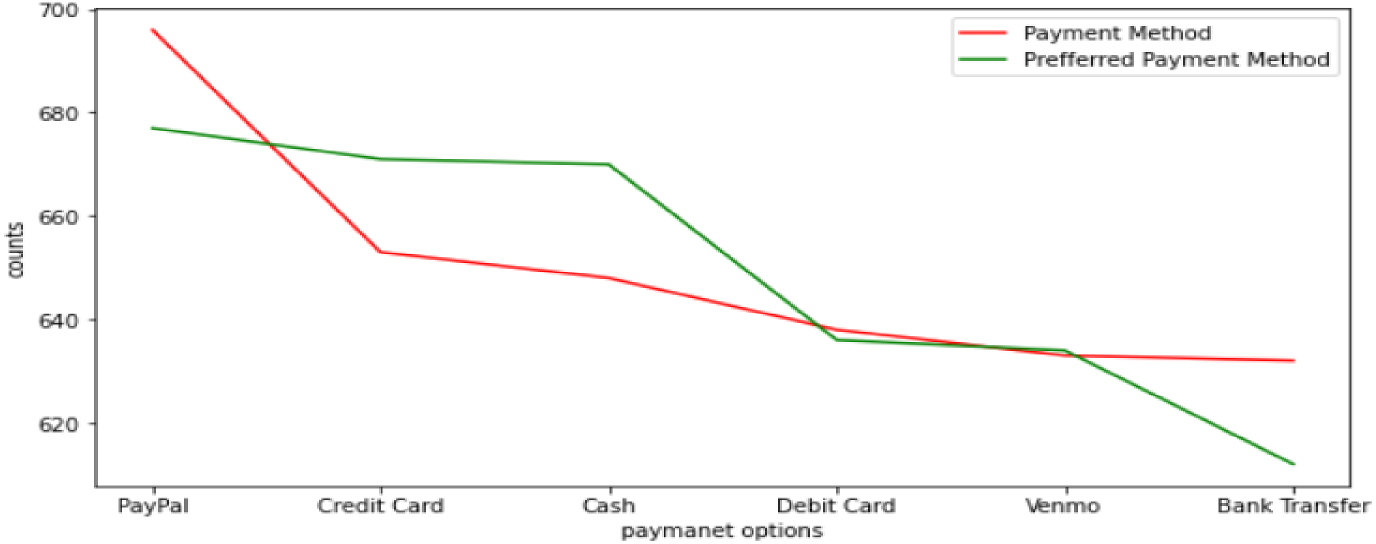
```
#To check the  distributions of shipping methods opted by customers
shipping=df['Shipping Type'].value_counts()
plt.figure(figsize = (10,5))
plt.bar(shipping.index,shipping.values,color='orange',edgecolor='green')
plt.title("methods of shippings")
plt.xlabel("types of shippings")
plt.ylabel("total counts")
plt.show()
```



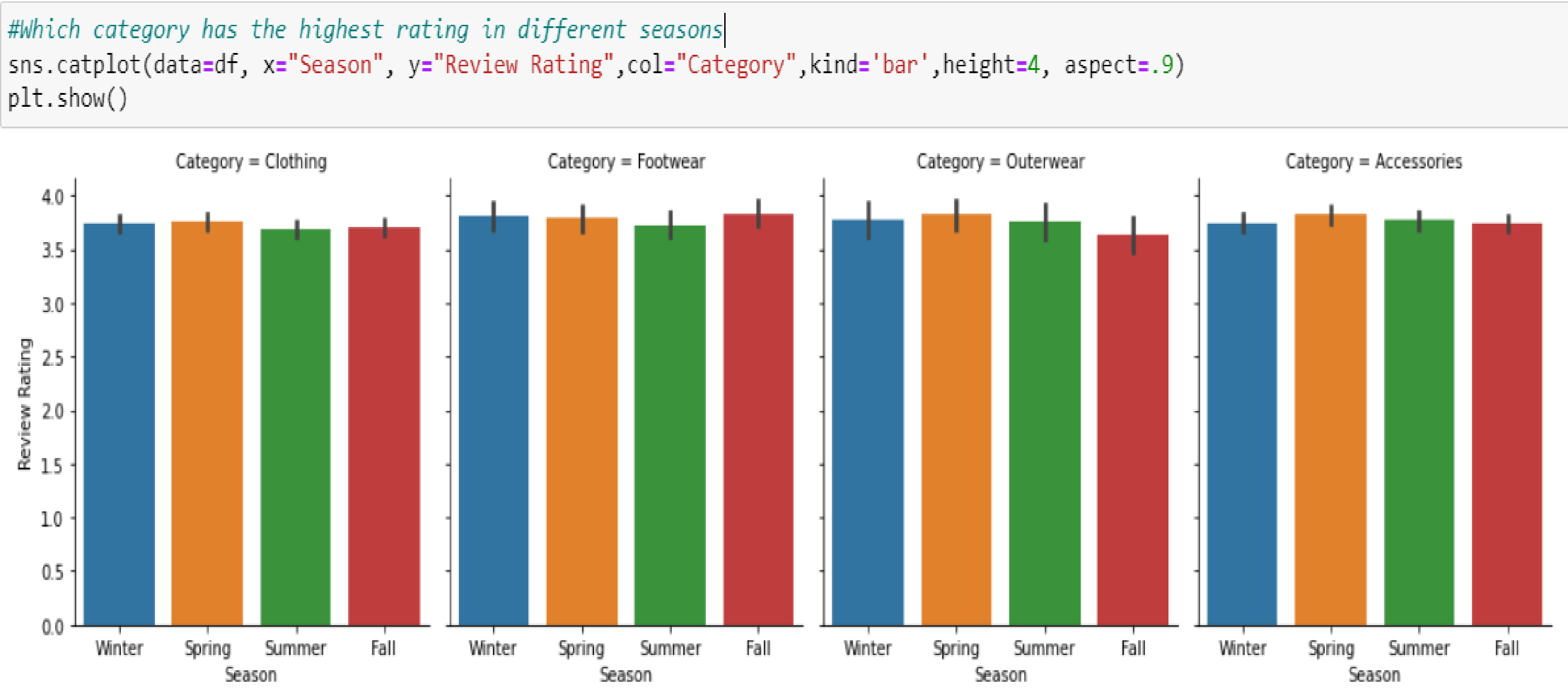
2. If &How the actual payment methods and preferred payment methods differ



```
#If &How the actual payment methods and preferred payment methods differ
pref1=df['Payment Method'].value_counts()
pref=df['Preferred Payment Method'].value_counts()
plt.figure(figsize = (10,5))
pref1.plot(color='r',label='Payment Method')
pref.plot(color='g',label='Prefferred Payment Method')
plt.ylabel("counts")
plt.xlabel("paymanet options")
plt.legend()
plt.show()
```

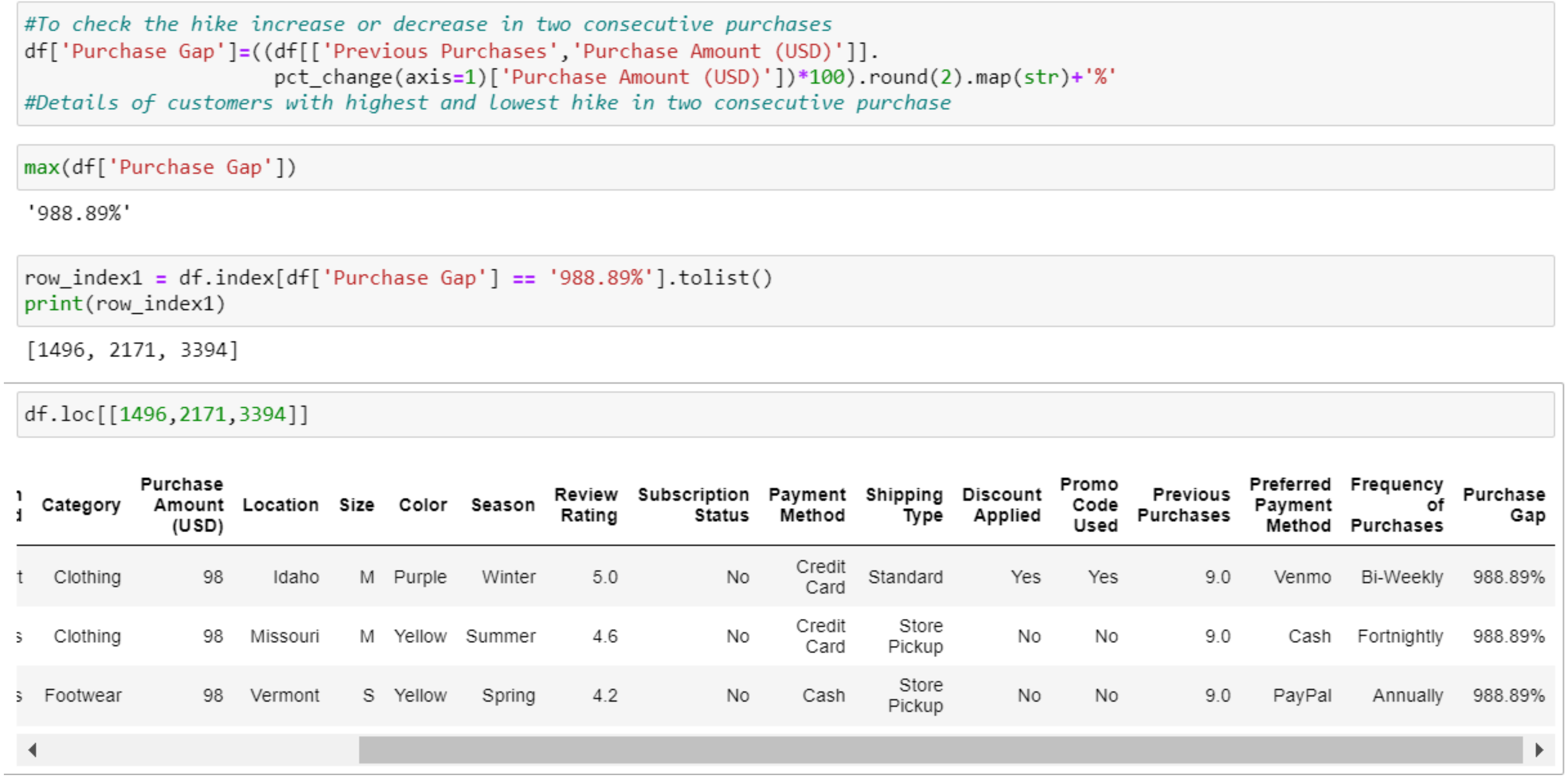


3. Which category has the highest rating in different seasons?



```
#Which category has the highest rating in different seasons
sns.catplot(data=df, x="Season", y="Review Rating",col="Category",kind='bar',height=4, aspect=.9)
plt.show()
```

4. To check the hike increase or decrease in two consecutive purchases & Details of customers with highest and lowest hike in two consecutive purchase



```
#To check the hike increase or decrease in two consecutive purchases
df['Purchase Gap']=((df[['Previous Purchases','Purchase Amount (USD)']].
                    pct_change(axis=1)['Purchase Amount (USD)'])*100).round(2).map(str)+'%'
#Details of customers with highest and lowest hike in two consecutive purchase
```

```
max(df['Purchase Gap'])
```

```
'988.89%'
```

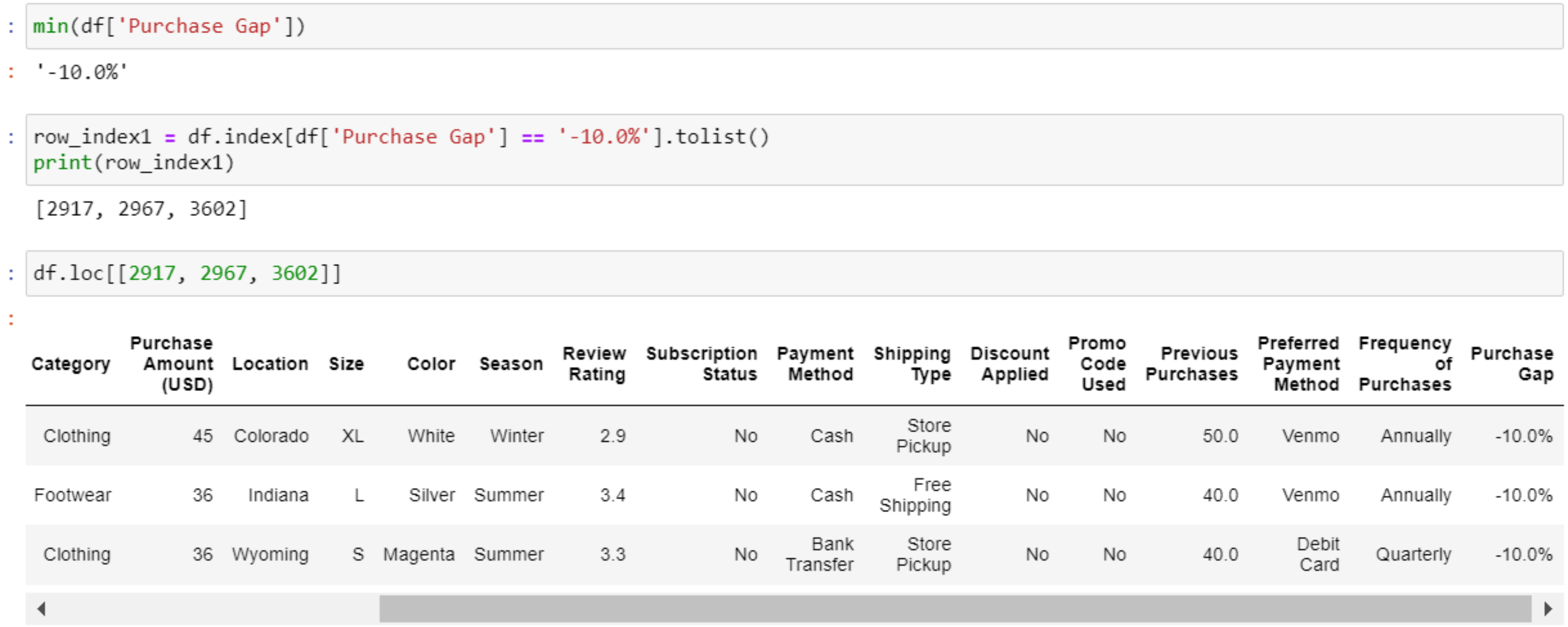
```
row_index1 = df.index[df['Purchase Gap'] == '988.89%'].tolist()
print(row_index1)
```

```
[1496, 2171, 3394]
```

```
df.loc[[1496,2171,3394]]
```

n d	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	Purchase Gap
t	Clothing	98	Idaho	M	Purple	Winter	5.0	No	Credit Card	Standard	Yes	Yes	9.0	Venmo	Bi-Weekly	988.89%
s	Clothing	98	Missouri	M	Yellow	Summer	4.6	No	Credit Card	Store Pickup	No	No	9.0	Cash	Fortnightly	988.89%
s	Footwear	98	Vermont	S	Yellow	Spring	4.2	No	Cash	Store Pickup	No	No	9.0	PayPal	Annually	988.89%

4. To check the hike increase or decrease in two consecutive purchases & Details of customers with highest and lowest hike in two consecutive purchase



```
: min(df['Purchase Gap'])
```

```
: '-10.0%'
```

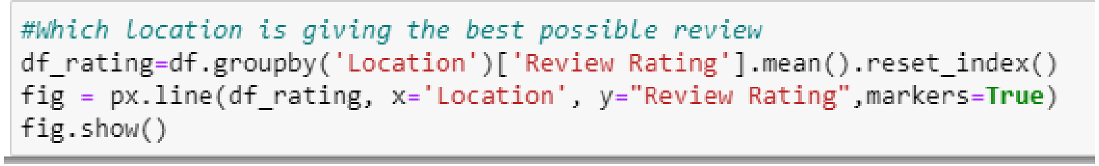
```
: row_index1 = df.index[df['Purchase Gap'] == '-10.0%'].tolist()
  print(row_index1)
```

```
[2917, 2967, 3602]
```

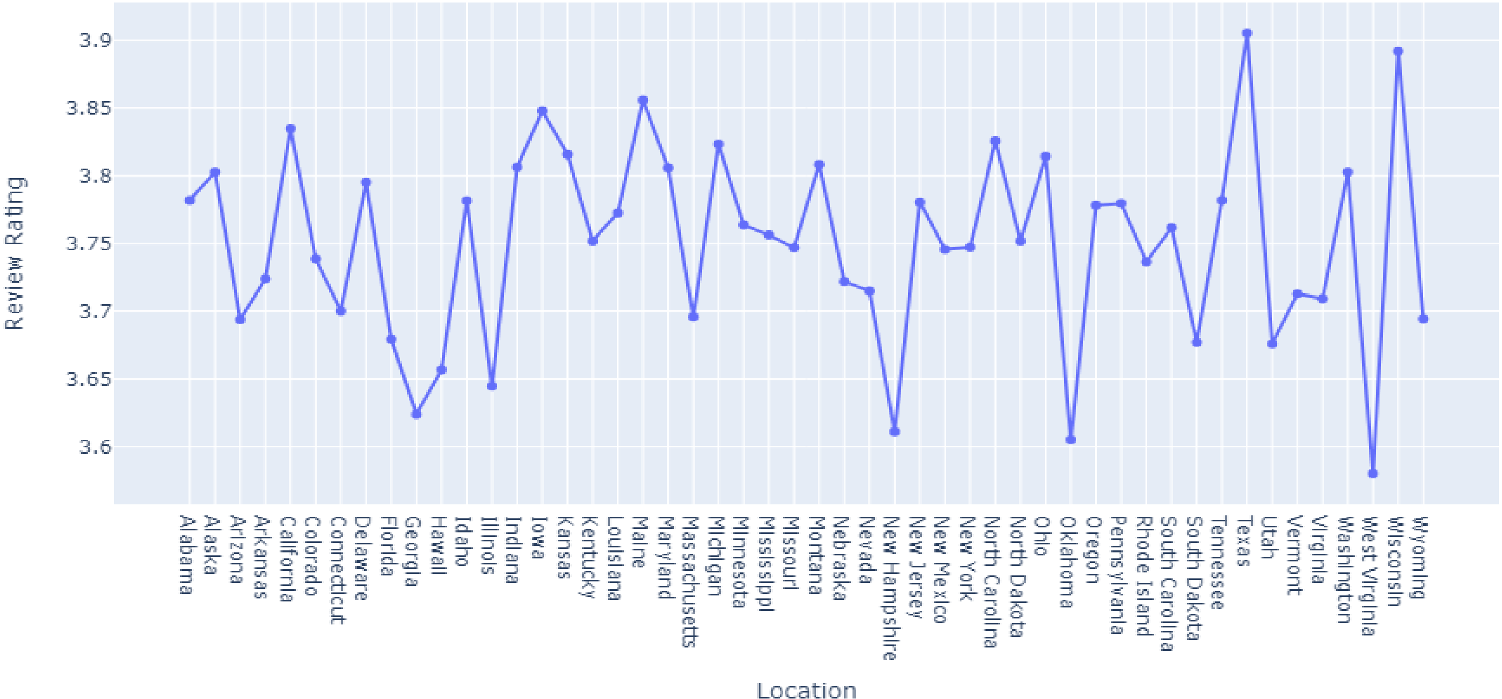
```
: df.loc[[2917, 2967, 3602]]
```

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	Purchase Gap
Clothing	45	Colorado	XL	White	Winter	2.9	No	Cash	Store Pickup	No	No	50.0	Venmo	Annually	-10.0%
Footwear	36	Indiana	L	Silver	Summer	3.4	No	Cash	Free Shipping	No	No	40.0	Venmo	Annually	-10.0%
Clothing	36	Wyoming	S	Magenta	Summer	3.3	No	Bank Transfer	Store Pickup	No	No	40.0	Debit Card	Quarterly	-10.0%

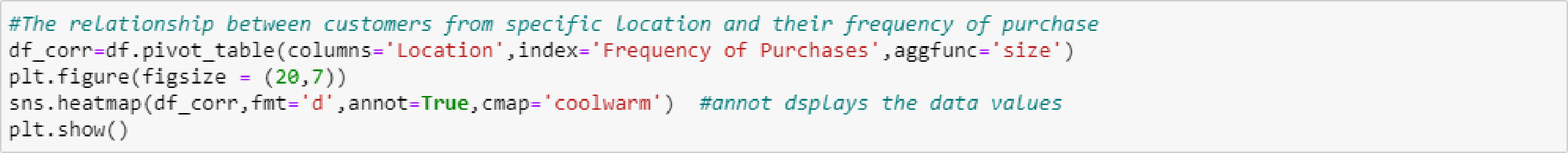
5. To check customers from which locations are giving the best possible review



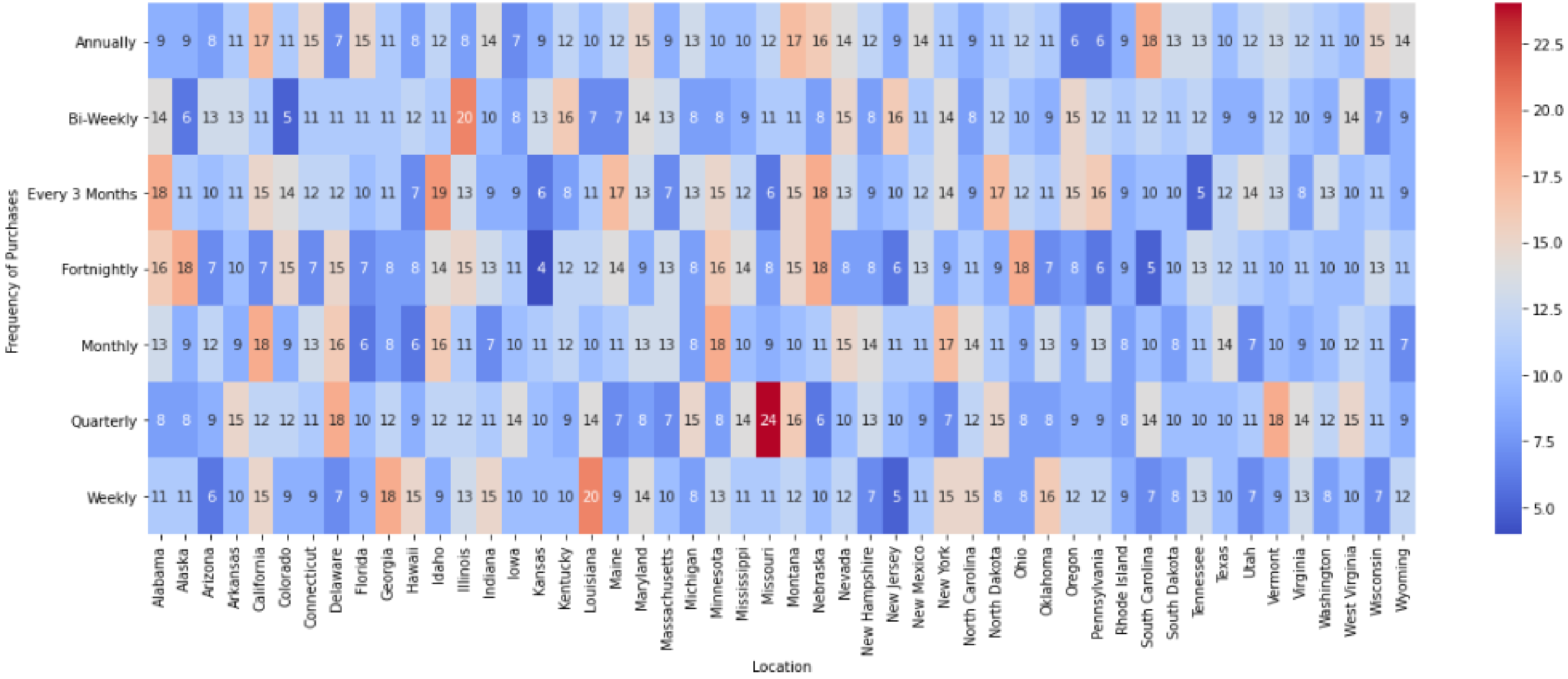
```
#Which location is giving the best possible review
df_rating=df.groupby('Location')['Review Rating'].mean().reset_index()
fig = px.line(df_rating, x='Location', y="Review Rating",markers=True)
fig.show()
```



6. The relationship between customers from specific location and respective frequency of purchase



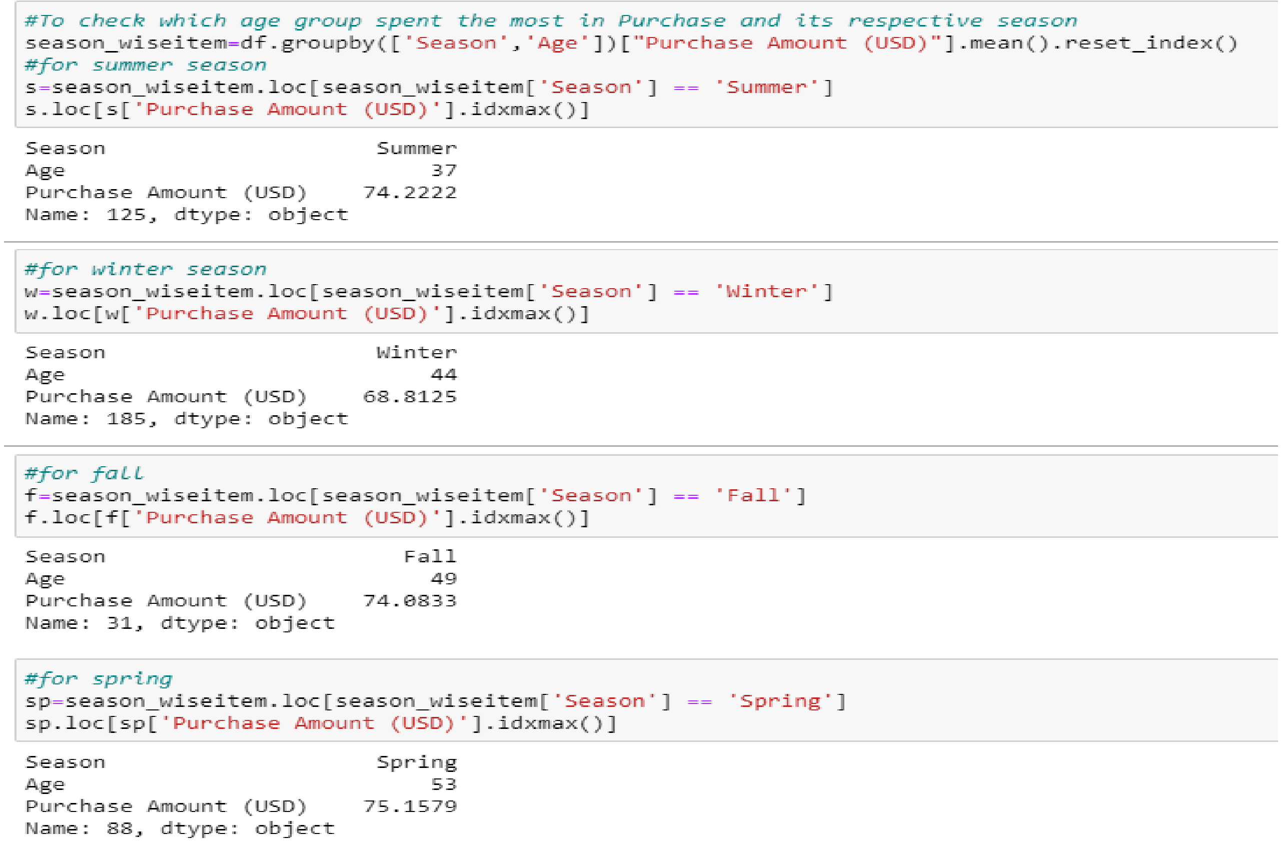
```
#The relationship between customers from specific location and their frequency of purchase
df_corr=df.pivot_table(columns='Location',index='Frequency of Purchases',aggfunc='size')
plt.figure(figsize = (20,7))
sns.heatmap(df_corr,fmt='d',annot=True,cmap='coolwarm')  #annot dsplays the data values
plt.show()
```



Frequency of Purchases	Alabama	Alaska	Arizona	Arkansas	California	Colorado	Connecticut	Delaware	Florida	Georgia	Hawaii	Idaho	Illinois	Indiana	Iowa	Kansas	Kentucky	Louisiana	Maine	Maryland	Massachusetts	Michigan	Minnesota	Mississippi	Missouri
Annually	9	9	8	11	17	11	15	7	15	11	8	12	8	14	7	9	12	10	12	15	9	13	10	10	12
Bi-Weekly	14	6	13	13	11	5	11	11	11	11	12	11	20	10	8	13	16	7	7	14	13	8	8	9	11
Every 3 Months	18	11	10	11	15	14	12	12	10	11	7	19	13	9	9	6	8	11	17	13	7	13	15	12	6
Fortnightly	16	18	7	10	7	15	7	15	7	8	8	14	15	13	11	4	12	12	14	9	13	8	16	14	8
Monthly	13	9	12	9	18	9	13	16	6	8	6	16	11	7	10	11	12	10	11	13	13	8	18	10	9
Quarterly	8	8	9	15	12	12	11	18	10	12	9	12	12	11	14	10	9	14	7	8	7	15	8	14	24
Weekly	11	11	6	10	15	9	9	7	9	18	15	9	13	15	10	10	10	20	9	14	10	8	13	11	11

Frequency of Purchases	Montana	Nebraska	Nevada	New Hampshire	New Jersey	New Mexico	New York	North Carolina	North Dakota	Ohio	Oklahoma	Oregon	Pennsylvania	Rhode Island	South Carolina	South Dakota	Tennessee	Texas	Utah	Vermont	Virginia	Washington	West Virginia	Wisconsin	Wyoming
Annually	17	16	14	12	9	14	11	9	11	12	11	6	6	9	18	13	13	10	12	13	12	11	10	15	14
Bi-Weekly	11	8	15	8	16	11	14	8	12	10	9	15	12	11	12	11	12	9	9	12	10	9	14	7	9
Every 3 Months	15	18	13	9	10	12	14	9	17	12	11	15	16	9	10	10	5	12	14	13	8	13	10	11	9
Fortnightly	15	18	8	8	6	13	9	11	9	18	7	8	6	9	5	10	13	12	11	10	11	10	10	13	11
Monthly	10	11	15	14	11	11	17	14	11	9	13	9	13	8	10	8	11	14	7	10	9	10	12	11	7
Quarterly	16	6	10	13	10	9	7	12	15	8	8	9	9	8	14	10	10	10	11	18	14	12	15	11	9
Weekly	12	10	12	7	5	11	15	15	8	8	16	12	12	9	7	8	13	10	7	9	13	8	10	7	12

7. To check which age group spent the most in Purchase and its respective season



```
#To check which age group spent the most in Purchase and its respective season
season_wiseitem=df.groupby(['Season','Age'])["Purchase Amount (USD)"].mean().reset_index()
#for summer season
s=season_wiseitem.loc[season_wiseitem['Season'] == 'Summer']
s.loc[s['Purchase Amount (USD)'].idxmax()]
```

```
Season                    Summer
Age                           37
Purchase Amount (USD)    74.2222
Name: 125, dtype: object
```

```
#for winter season
w=season_wiseitem.loc[season_wiseitem['Season'] == 'Winter']
w.loc[w['Purchase Amount (USD)'].idxmax()]
```

```
Season                    Winter
Age                           44
Purchase Amount (USD)    68.8125
Name: 185, dtype: object
```

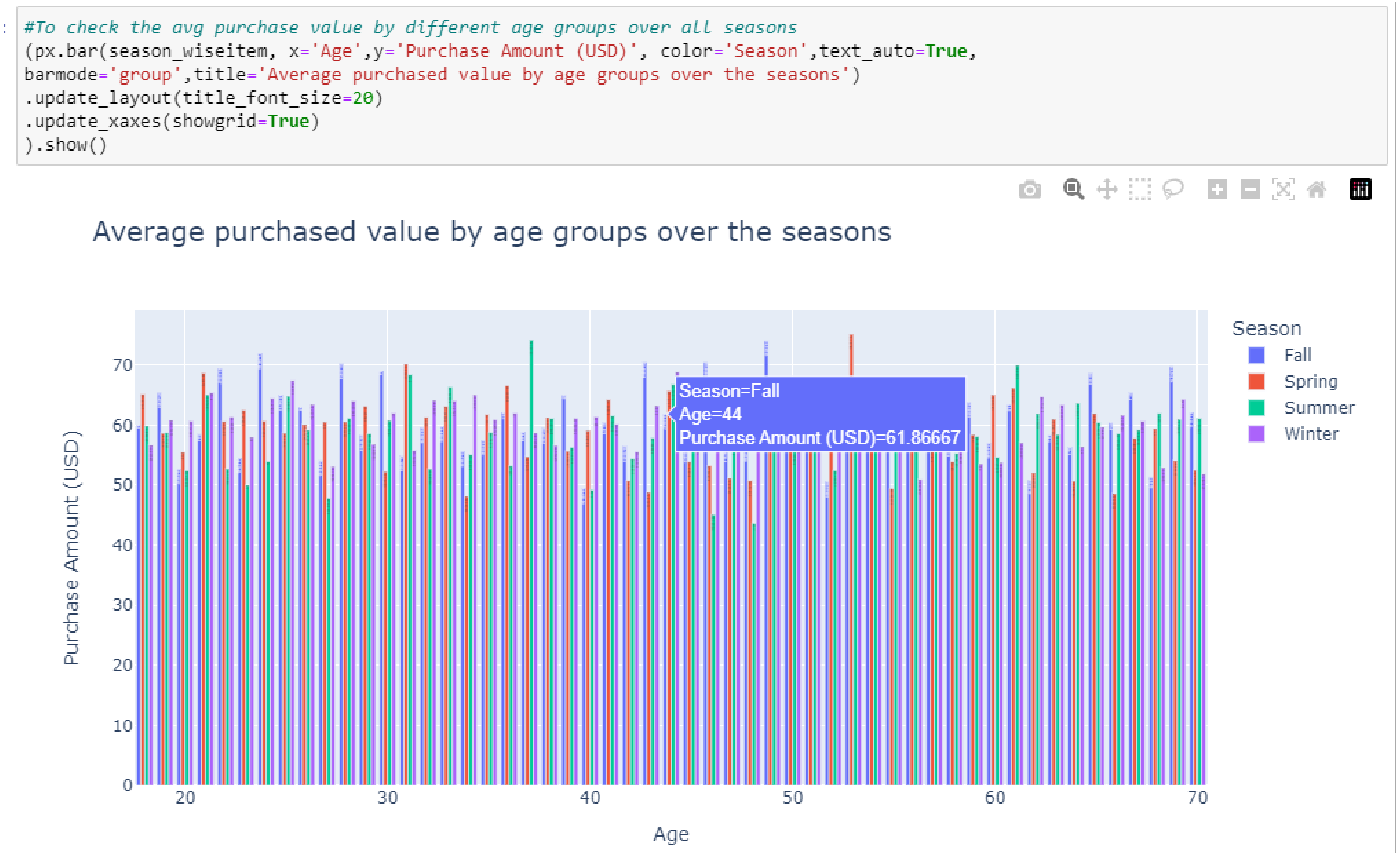
```
#for fall
f=season_wiseitem.loc[season_wiseitem['Season'] == 'Fall']
f.loc[f['Purchase Amount (USD)'].idxmax()]
```

```
Season                      Fall
Age                           49
Purchase Amount (USD)    74.0833
Name: 31, dtype: object
```

```
#for spring
sp=season_wiseitem.loc[season_wiseitem['Season'] == 'Spring']
sp.loc[sp['Purchase Amount (USD)'].idxmax()]
```

```
Season                    Spring
Age                           53
Purchase Amount (USD)    75.1579
Name: 88, dtype: object
```

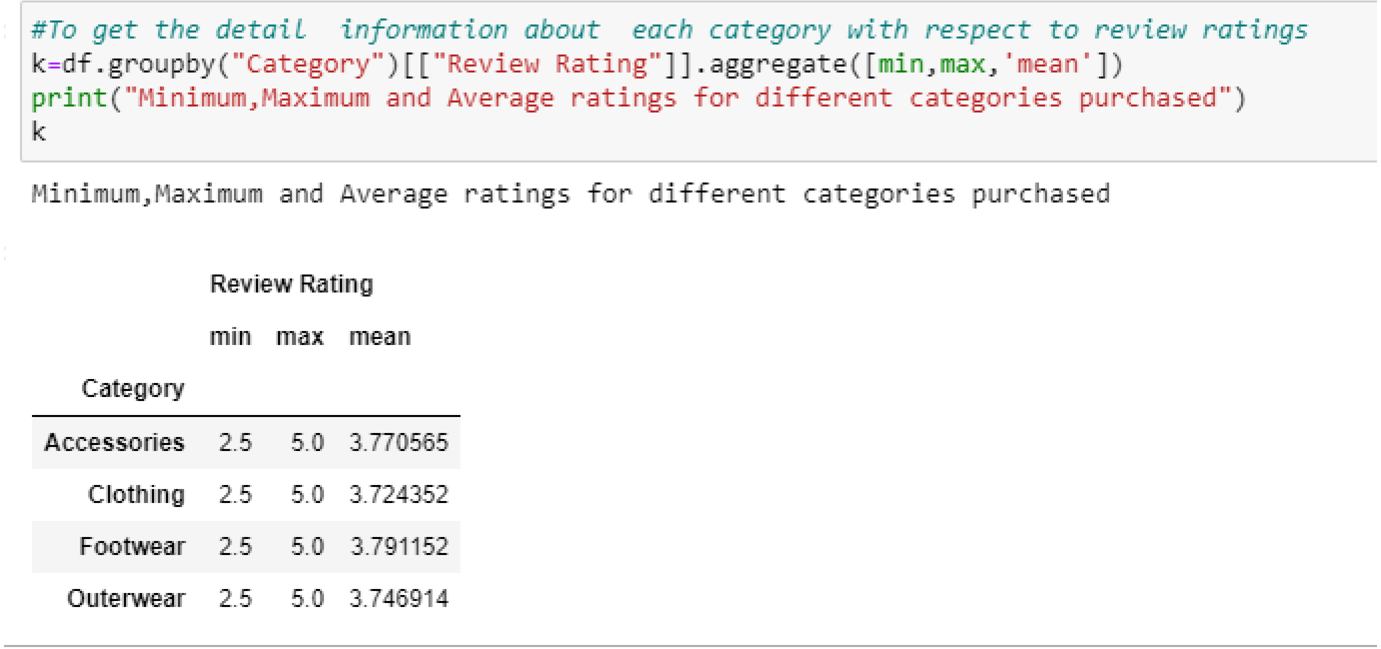
8. To check the avg purchase value by different age groups over all seasons



```
#To check the avg purchase value by different age groups over all seasons
(px.bar(season_wiseitem, x='Age',y='Purchase Amount (USD)', color='Season',text_auto=True,
barmode='group',title='Average purchased value by age groups over the seasons')
.update_layout(title_font_size=20)
.update_xaxes(showgrid=True)
).show()
```

Average purchased value by age groups over the seasons

9. To get the detail information about each category with respect to review ratings

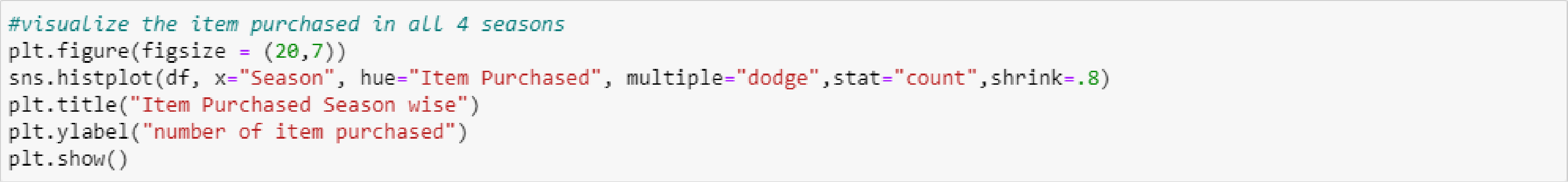


```python
#To get the detail  information about  each category with respect to review ratings
k=df.groupby("Category")[["Review Rating"]].aggregate([min,max,'mean'])
print("Minimum,Maximum and Average ratings for different categories purchased")
k
```

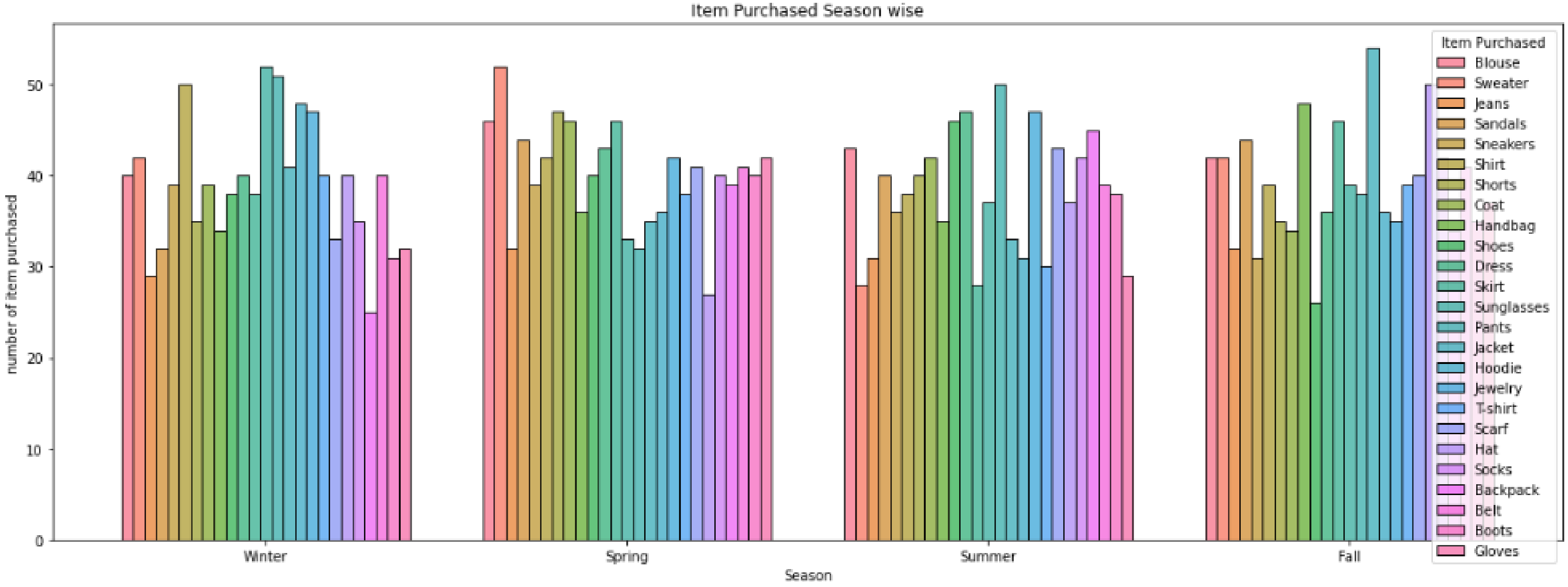
```
Minimum,Maximum and Average ratings for different categories purchased
```

	Review Rating		
	min	max	mean
Category			
Accessories	2.5	5.0	3.770565
Clothing	2.5	5.0	3.724352
Footwear	2.5	5.0	3.791152
Outerwear	2.5	5.0	3.746914

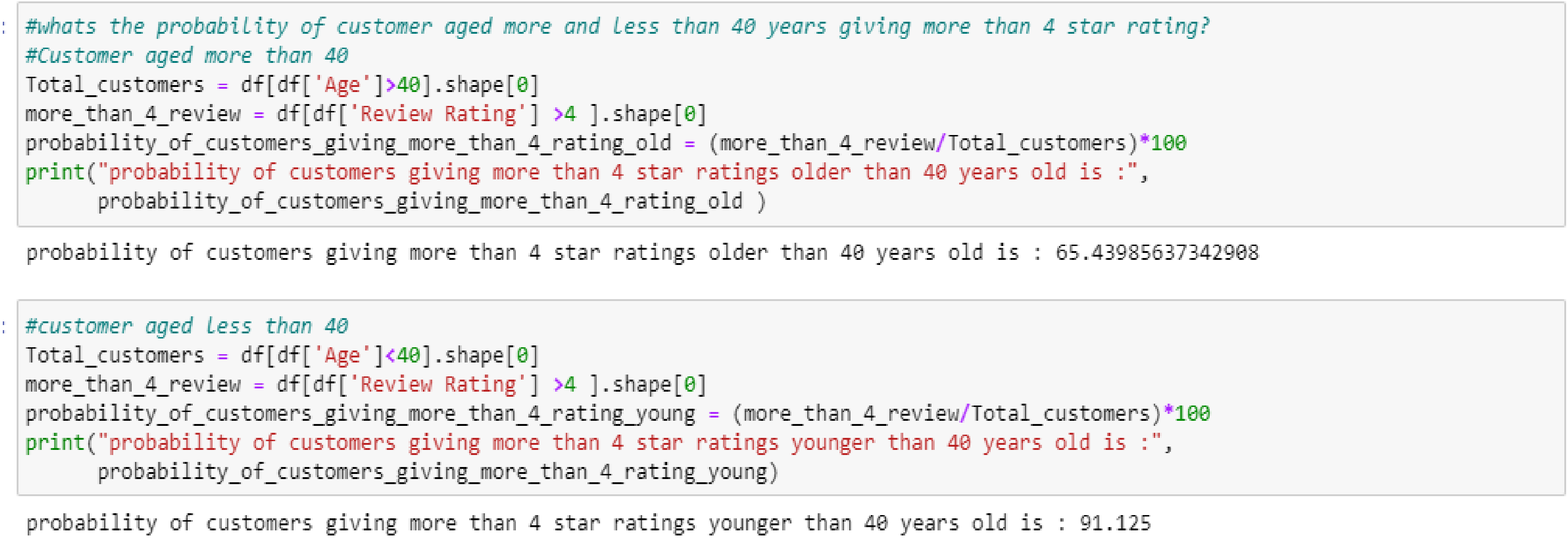
10. To get the overall idea of all the item purchased in all 4 seasons



```
#visualize the item purchased in all 4 seasons
plt.figure(figsize = (20,7))
sns.histplot(df, x="Season", hue="Item Purchased", multiple="dodge",stat="count",shrink=.8)
plt.title("Item Purchased Season wise")
plt.ylabel("number of item purchased")
plt.show()
```



11.To check the probability of customer aged more and less than 40 years giving more than 4 star rating?



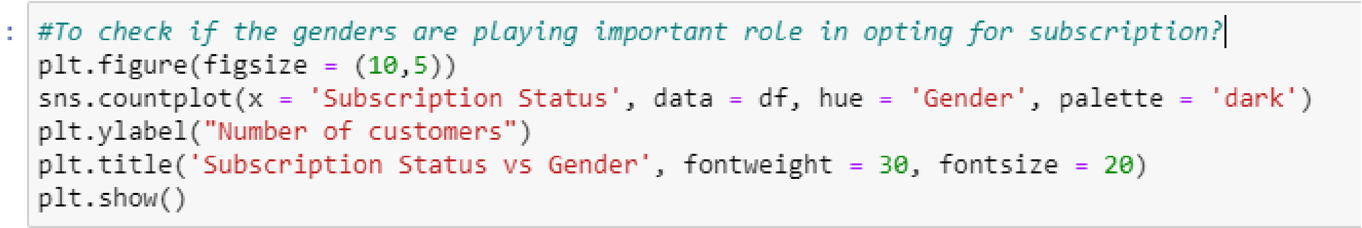
```
#whats the probability of customer aged more and less than 40 years giving more than 4 star rating?
#Customer aged more than 40
Total_customers = df[df['Age']>40].shape[0]
more_than_4_review = df[df['Review Rating'] >4 ].shape[0]
probability_of_customers_giving_more_than_4_rating_old = (more_than_4_review/Total_customers)*100
print("probability of customers giving more than 4 star ratings older than 40 years old is :",
      probability_of_customers_giving_more_than_4_rating_old )
```

```
probability of customers giving more than 4 star ratings older than 40 years old is : 65.43985637342908
```

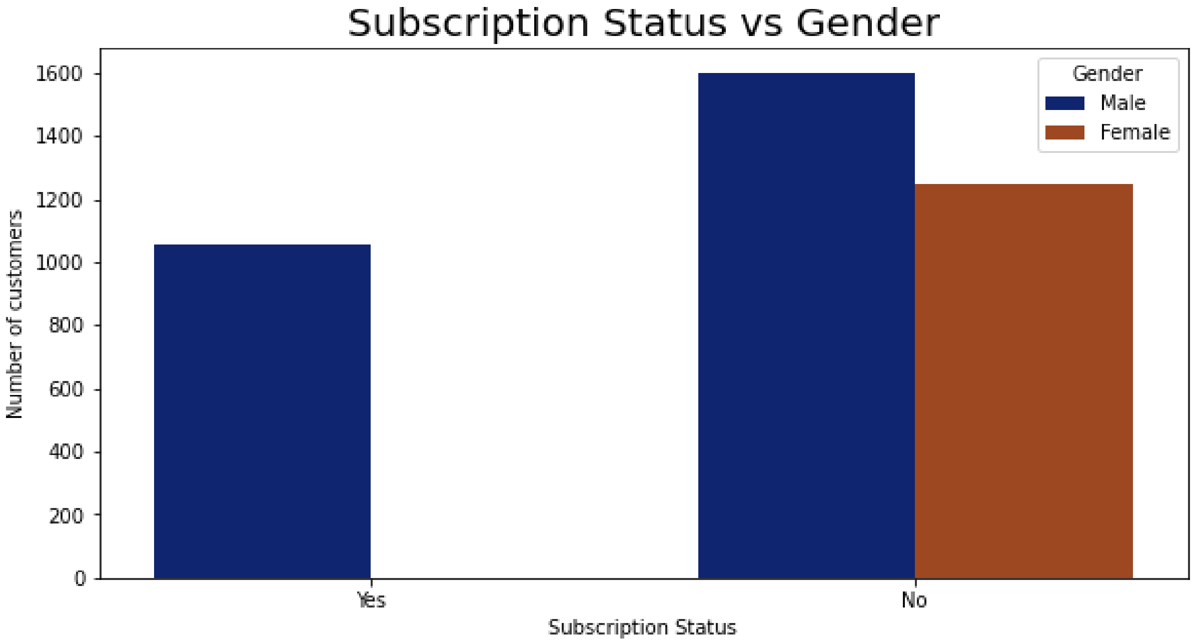
```
#customer aged less than 40
Total_customers = df[df['Age']<40].shape[0]
more_than_4_review = df[df['Review Rating'] >4 ].shape[0]
probability_of_customers_giving_more_than_4_rating_young = (more_than_4_review/Total_customers)*100
print("probability of customers giving more than 4 star ratings younger than 40 years old is :",
      probability_of_customers_giving_more_than_4_rating_young)
```

```
probability of customers giving more than 4 star ratings younger than 40 years old is : 91.125
```

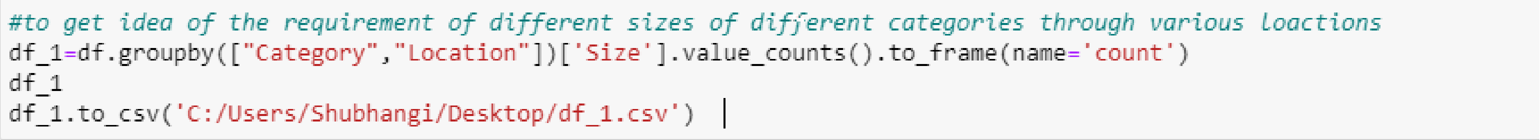
12.To check if the genders are playing important role in opting for subscription?



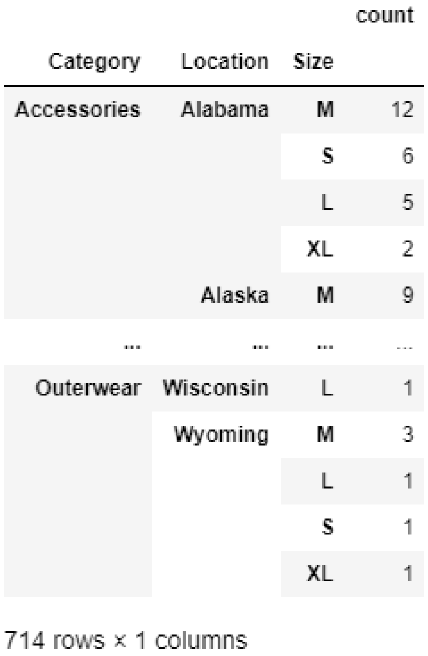
```
#To check if the genders are playing important role in opting for subscription?
plt.figure(figsize = (10,5))
sns.countplot(x = 'Subscription Status', data = df, hue = 'Gender', palette = 'dark')
plt.ylabel("Number of customers")
plt.title('Subscription Status vs Gender', fontweight = 30, fontsize = 20)
plt.show()
```



13. To get idea of the requirement of different sizes of different categories from various locations

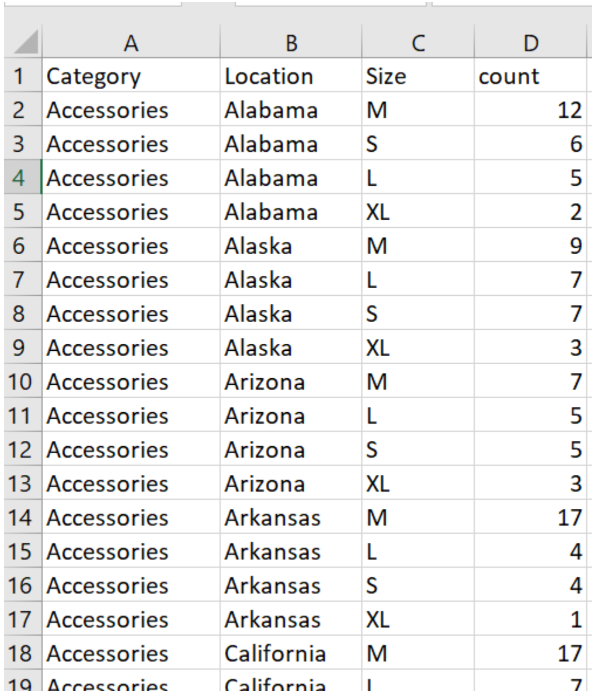


```
#to get idea of the requirement of different sizes of different categories through various loactions
df_1=df.groupby(["Category","Location"])['Size'].value_counts().to_frame(name='count')
df_1
df_1.to_csv('C:/Users/Shubhangi/Desktop/df_1.csv')
```



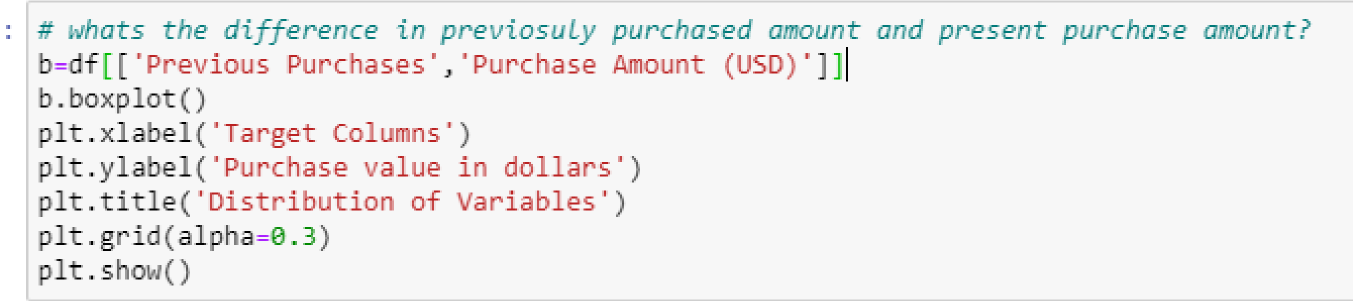
			count
Category	Location	Size	
Accessories	Alabama	M	12
		S	6
		L	5
		XL	2
	Alaska	M	9
...	...	...	...
Outerwear	Wisconsin	L	1
	Wyoming	M	3
		L	1
		S	1
		XL	1

714 rows × 1 columns

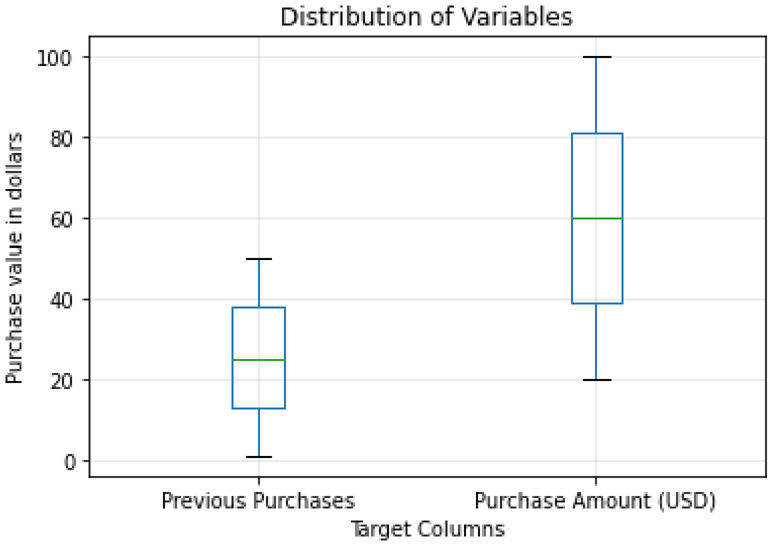


	A	B	C	D
1	Category	Location	Size	count
2	Accessories	Alabama	M	12
3	Accessories	Alabama	S	6
4	Accessories	Alabama	L	5
5	Accessories	Alabama	XL	2
6	Accessories	Alaska	M	9
7	Accessories	Alaska	L	7
8	Accessories	Alaska	S	7
9	Accessories	Alaska	XL	3
10	Accessories	Arizona	M	7
11	Accessories	Arizona	L	5
12	Accessories	Arizona	S	5
13	Accessories	Arizona	XL	3
14	Accessories	Arkansas	M	17
15	Accessories	Arkansas	L	4
16	Accessories	Arkansas	S	4
17	Accessories	Arkansas	XL	1
18	Accessories	California	M	17
19	Accessories	California	L	7

14. What is the difference (comparison) in previously purchased amount and present purchase amount?



```
# whats the difference in previosuly purchased amount and present purchase amount?
b=df[['Previous Purchases','Purchase Amount (USD)']]
b.boxplot()
plt.xlabel('Target Columns')
plt.ylabel('Purchase value in dollars')
plt.title('Distribution of Variables')
plt.grid(alpha=0.3)
plt.show()
```



Conclusions:

01 The most frequently used delivery method is free shipping.

02 PayPal is most preferred as well as most used payment method where as bank transfer is least preferred as well as used payment method among the customers.

03 The clothing category is performing well and getting best reviews in spring season where as footwear category is doing well and getting best reviews in fall season. The outerwear category as well as the accessory category too are getting best reviews in spring season.

04 There is maximum hike between two consecutive purchases of the following customers(customers id): 1496, 2171, 3394 . All the customers gave rating above 4.2 star. Customer id 1496 gave 5 star ,used a promo code resulting in a discount. The mostly Used payment method among the 3 customer is credit card.

Conclusions:

04

There is decrease in between two consecutive purchases of the following customers(customers id): 2917, 2967, 3602. All the customers gave rating below 3.4 star. Customer id 2917 gave 2.9 star. The mostly used Payment method among 3 customer is cash.

05

The customers from Texas are giving 3.90 star rating on average followed by customers from Wisconsin giving 3.89 ratings. The lowest ratings are from customers coming from West Virginia with 3.5 rating.

06

There’s a very strong correlation between Customers from Missouri and the frequency of purchase which is quarterly. Hence Customers from Missouri tends to buy products every 3 months followed by customers from Illinois who tends to buy products bi weekly.

07

Following are the seasons and the respective age groups with maximum amount spent in shopping:
Summer-Age37-74.22$
Winter-Age44-68.81$
Fall-Age49-74.08$
Spring-Age53-75.15$

Conclusions:

08 To get a better understanding of all age groups spending habit through out the seasons. Example for age group 44 In fall season average spending limit is 61.86$

09 The minimum and the maximum ratings for every individual category i.e; Clothing, Footwear, Outerwear, Accessory are 2.5 and 5 respectively, with mean rating 3.72, 3.79, 3.74 , 3.77 respectively.

10 In winter most purchased item is Sunglass.
In spring most purchased item is Sweater.
In summer most purchased item is Pant.
In fall most purchased item is Jacket

11 The customers older than 40 years are less likely to give more than 4 stars with probability 65.43 than the customers younger than 40 years are more likely to give greater than 4 stars with probability 91.125

Conclusions:

12 Males are opting for subscription way more than females.

13 To get idea of the requirement of different sizes of different categories from various locations.
Example: Customers from Hawaii requires L size garment more than S size OR Customers from Washington requires M size more than size L and S.

14 For previously purchased amount the range of purchase is 1$ to 50$ with mean of 25$ and with no exception of extreme expensive or cheap purchase. As for Current purchased amount the range of purchase is 20$ to 100$ with mean of 60$ and with no exception of extreme expensive or cheap purchase.

Methods used in analysis(EDA)

Data Importing

Data preprocessing

Data visualization

Univariate and Bivariate Data Analysis.

Thanks!

HAPPY ANALYSIS !