my-first-project

October 31, 2023

1 My first project: Shopping Trends

[ ]:
```
#!pip install jovian --upgrade --quiet
import jovian
jovian.commit()
```

```
<IPython.core.display.Javascript object>
```

2 1. Defining configuration

Data details: Consumer Behavior and Shopping Habits Dataset

[1]:
```
# Importing required library
import os
from urllib.request import urlretrieve
import pandas as pd
import numpy as np
import matplotlib
import matplotlib.pyplot as plt
import seaborn as sns

# Configuring styles
sns.set_style("darkgrid")
matplotlib.rcParams['font.size'] = 14
matplotlib.rcParams['figure.figsize'] = (9, 5)
matplotlib.rcParams['figure.facecolor'] = '#00000000'
```

3 2. Downloading or uploading csv data

[3]:
```
#Data url
#url = 'C:\Users\Sunil kumar\Downloads\shopping_trends.csv'
#Downloading file
#urlretrieve(url, 'shopping_trends.csv')

!pip install opendatasets
import opendatasets as od
```

```
od.download('https://www.kaggle.com/datasets/zeesolver/
 ↪consumer-behavior-and-shopping-habits-dataset')
```

```
Collecting opendatasets
  Downloading opendatasets-0.1.22-py3-none-any.whl (15 kB)
Requirement already satisfied: click in /opt/conda/lib/python3.9/site-packages
(from opendatasets) (8.0.3)
Collecting kaggle
  Downloading kaggle-1.5.16.tar.gz (83 kB)
     |                  | 83 kB 5.3 MB/s
  Preparing metadata (setup.py) … done
Requirement already satisfied: tqdm in /opt/conda/lib/python3.9/site-
packages (from opendatasets) (4.62.3)
Requirement already satisfied: six>=1.10 in /opt/conda/lib/python3.9/site-
packages (from kaggle->opendatasets) (1.16.0)
Requirement already satisfied: certifi in /opt/conda/lib/python3.9/site-packages
(from kaggle->opendatasets) (2021.10.8)
Requirement already satisfied: python-dateutil in /opt/conda/lib/python3.9/site-
packages (from kaggle->opendatasets) (2.8.2)
Requirement already satisfied: requests in /opt/conda/lib/python3.9/site-
packages (from kaggle->opendatasets) (2.26.0)
Collecting python-slugify
  Downloading python_slugify-8.0.1-py2.py3-none-any.whl (9.7 kB)
Requirement already satisfied: urllib3 in /opt/conda/lib/python3.9/site-packages
(from kaggle->opendatasets) (1.26.7)
Requirement already satisfied: bleach in /opt/conda/lib/python3.9/site-packages
(from kaggle->opendatasets) (4.1.0)
Requirement already satisfied: packaging in /opt/conda/lib/python3.9/site-
packages (from bleach->kaggle->opendatasets) (21.2)
Requirement already satisfied: webencodings in /opt/conda/lib/python3.9/site-
packages (from bleach->kaggle->opendatasets) (0.5.1)
Collecting text-unidecode>=1.3
  Downloading text_unidecode-1.3-py2.py3-none-any.whl (78 kB)
     |                  | 78 kB 19.6 MB/s
Requirement already satisfied: charset-normalizer~=2.0.0 in
/opt/conda/lib/python3.9/site-packages (from requests->kaggle->opendatasets)
(2.0.0)
Requirement already satisfied: idna<4,>=2.5 in /opt/conda/lib/python3.9/site-
packages (from requests->kaggle->opendatasets) (3.1)
Requirement already satisfied: pyparsing<3,>=2.0.2 in
/opt/conda/lib/python3.9/site-packages (from
packaging->bleach->kaggle->opendatasets) (2.4.7)
Building wheels for collected packages: kaggle
  Building wheel for kaggle (setup.py) … done
  Created wheel for kaggle: filename=kaggle-1.5.16-py3-none-any.whl
size=110701
sha256=2c0dac8432c83417d7cc75bd153a23f43974d23d4da69bf9ce345cce67e84aaf
```

```
  Stored in directory: /home/jovyan/.cache/pip/wheels/d2/ed/a5/da3a0cfb13373d1ac
e41cafa4f2467d858c55c52473ba72799
Successfully built kaggle
Installing collected packages: text-unidecode, python-slugify, kaggle,
opendatasets
Successfully installed kaggle-1.5.16 opendatasets-0.1.22 python-slugify-8.0.1
text-unidecode-1.3
Please provide your Kaggle credentials to download this dataset. Learn more:
http://bit.ly/kaggle-creds
Your Kaggle username: kumarchaitanya1208
Your Kaggle Key: ·····
Downloading consumer-behavior-and-shopping-habits-dataset.zip to ./consumer-
behavior-and-shopping-habits-dataset

100%|      | 146k/146k [00:00<00:00, 197kB/s]
```

[4]:
```
os.listdir('./consumer-behavior-and-shopping-habits-dataset') #Check the␣
↪downloaded or uploaded file name
```

[4]:
```
['shopping_behavior_updated.csv', 'shopping_trends.csv']
```

4 3. Reading and understanding the data

- Load the dataset into a data frame using Pandas
- Explore the number of rows & columns, ranges of values etc.

[5]:
```
shopping_df = pd.read_csv('consumer-behavior-and-shopping-habits-dataset/
↪shopping_trends.csv') #defining data frame
shopping_df #displaying data
```

[5]:

	Customer ID	Age	Gender	Item Purchased	Category	
0	1	55	Male	Blouse	Clothing	
1	2	19	Male	Sweater	Clothing	
2	3	50	Male	Jeans	Clothing	
3	4	21	Male	Sandals	Footwear	
4	5	45	Male	Blouse	Clothing	
…	…	…	…	…	…	
3895	3896	40	Female	Hoodie	Clothing	
3896	3897	52	Female	Backpack	Accessories	
3897	3898	46	Female	Belt	Accessories	
3898	3899	44	Female	Shoes	Footwear	
3899	3900	52	Female	Handbag	Accessories	

	Purchase Amount (USD)	Location	Size	Color	Season	\

```
0                       53       Kentucky    L       Gray  Winter
1                       64          Maine    L     Maroon  Winter
2                       73  Massachusetts    S     Maroon  Spring
3                       90   Rhode Island    M     Maroon  Spring
4                       49         Oregon    M  Turquoise  Spring
…                      …            …  …       …     …
3895                    28       Virginia    L  Turquoise  Summer
3896                    49           Iowa    L      White  Spring
3897                    33     New Jersey    L      Green  Spring
3898                    77      Minnesota    S      Brown  Summer
3899                    81     California    M      Beige  Spring

      Review Rating Subscription Status Payment Method   Shipping Type  \
0               3.1                  Yes    Credit Card         Express
1               3.1                  Yes  Bank Transfer         Express
2               3.1                  Yes           Cash   Free Shipping
3               3.5                  Yes         PayPal    Next Day Air
4               2.7                  Yes           Cash   Free Shipping
…              …                    …             …             …
3895            4.2                   No           Cash  2-Day Shipping
3896            4.5                   No         PayPal    Store Pickup
3897            2.9                   No    Credit Card        Standard
3898            3.8                   No         PayPal         Express
3899            3.1                   No  Bank Transfer    Store Pickup

     Discount Applied Promo Code Used  Previous Purchases  \
0                 Yes             Yes                  14
1                 Yes             Yes                   2
2                 Yes             Yes                  23
3                 Yes             Yes                  49
4                 Yes             Yes                  31
…                …               …                    …
3895               No              No                  32
3896               No              No                  41
3897               No              No                  24
3898               No              No                  24
3899               No              No                  33

     Preferred Payment Method Frequency of Purchases
0                       Venmo            Fortnightly
1                        Cash            Fortnightly
2                 Credit Card                 Weekly
3                      PayPal                 Weekly
4                      PayPal                Annually
…                         …                    …
3895                    Venmo                 Weekly
3896            Bank Transfer              Bi-Weekly
```

```
3897                    Venmo              Quarterly
3898                    Venmo                 Weekly
3899                    Venmo              Quarterly

[3900 rows x 19 columns]
```

[7]:
```
shopping_df.info() #Checking data type and data set details
```

```
<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
```

[8]:
```
shopping_df.describe() #checking numerical values entryies
```

[8]:

	Customer ID	Age	Purchase Amount (USD)	Review Rating	\
count	3900.000000	3900.000000	3900.000000	3900.000000	
mean	1950.500000	44.068462	59.764359	3.749949	
std	1125.977353	15.207589	23.685392	0.716223	
min	1.000000	18.000000	20.000000	2.500000	
25%	975.750000	31.000000	39.000000	3.100000	
50%	1950.500000	44.000000	60.000000	3.700000	
75%	2925.250000	57.000000	81.000000	4.400000	
max	3900.000000	70.000000	100.000000	5.000000	

```
       Previous Purchases
count         3900.000000
mean            25.351538
std             14.447125
min              1.000000
25%             13.000000
50%             25.000000
75%             38.000000
max             50.000000
```

[9]:
```
shopping_df.columns #Checking columns names
```

[9]:
```
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')
```

[10]:
```
shopping_df.Category.unique() #Output ares unique value of a column
```

[10]:
```
array(['Clothing', 'Footwear', 'Outerwear', 'Accessories'], dtype=object)
```

[11]:
```
shopping_df.Category.nunique() # Output are count of unique values of a column
```

[11]:
```
4
```

[12]:
```
shopping_df.sample(10)
```

[12]:

	Customer ID	Age	Gender	Item Purchased	Category	\
304	305	40	Male	Dress	Clothing	
2464	2465	47	Male	Jacket	Outerwear	
1910	1911	29	Male	Sandals	Footwear	
3701	3702	46	Female	Sandals	Footwear	
2016	2017	53	Male	Shoes	Footwear	
32	33	36	Male	Jacket	Outerwear	
2681	2682	23	Female	Shoes	Footwear	
750	751	25	Male	T-shirt	Clothing	
3464	3465	36	Female	Coat	Outerwear	
1971	1972	54	Male	Hat	Accessories	

	Purchase Amount (USD)	Location	Size	Color	Season	\
304	84	Oregon	M	Green	Winter	
2464	84	Hawaii	L	Teal	Winter	
1910	38	New York	S	Yellow	Fall	
3701	66	New Mexico	M	Charcoal	Spring	

```
2016                    47          Maine    S    Violet  Winter
32                      67         Kansas    M    Silver  Summer
2681                    23  Massachusetts    L     Olive  Summer
750                     95       Maryland    L    Violet    Fall
3464                    59           Utah    S     Peach  Winter
1971                    29   Pennsylvania   XL    Indigo  Spring

      Review Rating Subscription Status Payment Method  Shipping Type  \
304             3.4                  Yes          Venmo   Next Day Air
2464            2.8                   No    Debit Card        Express
1910            3.8                   No          Venmo        Express
3701            4.5                   No           Cash   Store Pickup
2016            3.4                   No    Credit Card       Standard
32              4.9                  Yes  Bank Transfer  Free Shipping
2681            4.8                   No  Bank Transfer       Standard
750             3.4                  Yes  Bank Transfer  Free Shipping
3464            2.7                   No    Debit Card   Next Day Air
1971            3.4                   No         PayPal   Next Day Air

     Discount Applied Promo Code Used  Previous Purchases  \
304               Yes             Yes                  45
2464               No              No                   4
1910               No              No                  13
3701               No              No                  36
2016               No              No                  20
32                Yes             Yes                  37
2681               No              No                  17
750               Yes             Yes                   5
3464               No              No                  12
1971               No              No                   1

     Preferred Payment Method Frequency of Purchases
304               Credit Card              Bi-Weekly
2464                     Cash              Bi-Weekly
1910              Debit Card              Quarterly
3701                     Cash               Annually
2016              Debit Card            Fortnightly
32                     Venmo               Annually
2681              Debit Card                Monthly
750                    Venmo               Annually
3464                  PayPal               Annually
1971                   Venmo                Monthly
```

[13]:
```
shopping_df.head(10)
```

[13]:
```
   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
5	6	46	Male	Sneakers	Footwear		20
6	7	63	Male	Shirt	Clothing		85
7	8	27	Male	Shorts	Clothing		34
8	9	26	Male	Coat	Outerwear		97
9	10	57	Male	Handbag	Accessories		31

	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	
5	Wyoming	M	White	Summer	2.9	Yes	
6	Montana	M	Gray	Fall	3.2	Yes	
7	Louisiana	L	Charcoal	Winter	3.2	Yes	
8	West Virginia	L	Silver	Summer	2.6	Yes	
9	Missouri	M	Pink	Spring	4.8	Yes	

	Payment Method	Shipping Type	Discount Applied	Promo Code Used	\
0	Credit Card	Express	Yes	Yes	
1	Bank Transfer	Express	Yes	Yes	
2	Cash	Free Shipping	Yes	Yes	
3	PayPal	Next Day Air	Yes	Yes	
4	Cash	Free Shipping	Yes	Yes	
5	Venmo	Standard	Yes	Yes	
6	Debit Card	Free Shipping	Yes	Yes	
7	Debit Card	Free Shipping	Yes	Yes	
8	Venmo	Express	Yes	Yes	
9	PayPal	2-Day Shipping	Yes	Yes	

	Previous Purchases	Preferred Payment Method	Frequency of Purchases
0	14	Venmo	Fortnightly
1	2	Cash	Fortnightly
2	23	Credit Card	Weekly
3	49	PayPal	Weekly
4	31	PayPal	Annually
5	14	Venmo	Weekly
6	49	Cash	Quarterly
7	19	Credit Card	Weekly
8	8	Venmo	Annually
9	4	Cash	Quarterly

[14]:
```
#To change any data type

#Shopping_df['objecttypecolumnname'] = pd.to_numeric(Shopping_df.
↪objecttypecolumnname, errors='coerce')
```

[ ]:
```

```

5 4. Data Preparation and Cleaning

- Handle missing, incorrect and invalid data
- Perform any additional steps (parsing dates, creating additional columns, merging multiple dataset etc.)

[15]:
```
# There is nothing to clean
```

[16]:
```
#top 10 rating
shopping_df.sort_values('Review Rating',ascending = False).head(10)
```

[16]:

	Customer ID	Age	Gender	Item Purchased	Category	\
776	777	49	Male	Shirt	Clothing	
1664	1665	19	Male	Handbag	Accessories	
1277	1278	19	Male	Blouse	Clothing	
2632	2633	24	Male	Scarf	Accessories	
965	966	43	Male	Boots	Footwear	
3554	3555	39	Female	Dress	Clothing	
3820	3821	34	Female	Belt	Accessories	
3243	3244	45	Female	Hoodie	Clothing	
2319	2320	56	Male	Sandals	Footwear	
1301	1302	42	Male	Shirt	Clothing	

	Purchase Amount (USD)	Location	Size	Color	Season	\
776	60	Alabama	M	Maroon	Spring	
1664	53	Minnesota	M	Purple	Spring	
1277	97	Rhode Island	L	Green	Spring	
2632	27	Alaska	M	Black	Fall	
965	55	Delaware	L	Black	Spring	
3554	55	Louisiana	M	Indigo	Spring	
3820	65	New Hampshire	M	Purple	Winter	
3243	99	Idaho	L	Black	Summer	
2319	38	Maine	XL	Turquoise	Winter	
1301	54	Maine	M	Orange	Fall	

	Review Rating	Subscription Status	Payment Method	Shipping Type	\
776	5.0	Yes	PayPal	Standard	
1664	5.0	No	PayPal	Next Day Air	
1277	5.0	No	Debit Card	Express	

```
2632            5.0                  No          Venmo   Next Day Air
965             5.0                 Yes     Debit Card   Store Pickup
3554            5.0                  No  Bank Transfer  Free Shipping
3820            5.0                  No    Credit Card       Standard
3243            5.0                  No          Venmo   Store Pickup
2319            5.0                  No     Debit Card  Free Shipping
1301            5.0                  No          Venmo       Express

     Discount Applied Promo Code Used  Previous Purchases  \
776               Yes             Yes                   8
1664              Yes             Yes                  35
1277              Yes             Yes                   5
2632               No              No                  14
965               Yes             Yes                  10
3554               No              No                   5
3820               No              No                  34
3243               No              No                  11
2319               No              No                  29
1301              Yes             Yes                  21

     Preferred Payment Method Frequency of Purchases
776                     Venmo         Every 3 Months
1664                   PayPal                 Weekly
1277                    Venmo              Quarterly
2632                     Cash               Annually
965                Credit Card               Annually
3554               Credit Card            Fortnightly
3820                Debit Card              Quarterly
3243           Bank Transfer         Every 3 Months
2319                   PayPal                Monthly
1301                Debit Card                 Weekly
```

[ ]:
```
#top 10 purchage
shopping_df.sort_values('Purchase Amount (USD)', ascending = False).head(10)
```

[18]:
```
# Group by item purchage

shopping_ItemPurchased_df = shopping_df.groupby('Item Purchased')[['Purchase␣
 ↪Amount (USD)']].sum()
shopping_ItemPurchased_df
```

[18]:
```
                Purchase Amount (USD)
Item Purchased
Backpack                         8636
Belt                             9635
Blouse                          10410
Boots                            9018
```

```
Coat                    9275
Dress                  10320
Gloves                  8477
Handbag                 8857
Hat                     9375
Hoodie                  8767
Jacket                  9249
Jeans                   7548
Jewelry                10010
Pants                  10090
Sandals                 9200
Scarf                   9561
Shirt                  10332
Shoes                   9240
Shorts                  9433
Skirt                   9402
Sneakers                8635
Socks                   9252
Sunglasses              9649
Sweater                 9462
T-shirt                 9248
```

[ ]:
```
#group by Frequency

shopping_frequency_df = shopping_df.groupby('Frequency of␣
↪Purchases')[['Customer ID']].count()
shopping_frequency_df
```

[15]:
```
shopping_df['Gender'].value_counts()
```

[15]:
```
Male      2652
Female    1248
Name: Gender, dtype: int64
```

[10]:
```
gender_count=shopping_df.Gender.value_counts() #gender wise customer
gender_count
```

[10]:
```
Male      2652
Female    1248
Name: Gender, dtype: int64
```

[23]:
```
shopping_df['Item Purchased'].value_counts() #Item wise sale
```

[23]:
```
Blouse        171
Jewelry       171
Pants         171
Shirt         169
```

```
Dress          166
Sweater        164
Jacket         163
Belt           161
Sunglasses     161
Coat           161
Sandals        160
Socks          159
Skirt          158
Shorts         157
Scarf          157
Hat            154
Handbag        153
Hoodie         151
Shoes          150
T-shirt        147
Sneakers       145
Boots          144
Backpack       143
Gloves         140
Jeans          124
Name: Item Purchased, dtype: int64
```

[30]:
```
shopping_df['Location'].value_counts().head(10)
```

[30]:
```
Montana       96
California    95
Idaho         93
Illinois      92
Alabama       89
Minnesota     88
Nebraska      87
New York      87
Nevada        87
Maryland      86
Name: Location, dtype: int64
```

6 5. Exploratory Analysis and Visualization

- Compute the mean, sum, range and other interesting statistics for numeric columns
- Explore distributions of numeric columns using histograms etc.
- Explore relationship between columns using scatter plots, bar charts etc.
- Make a note of interesting insights from the exploratory analysis

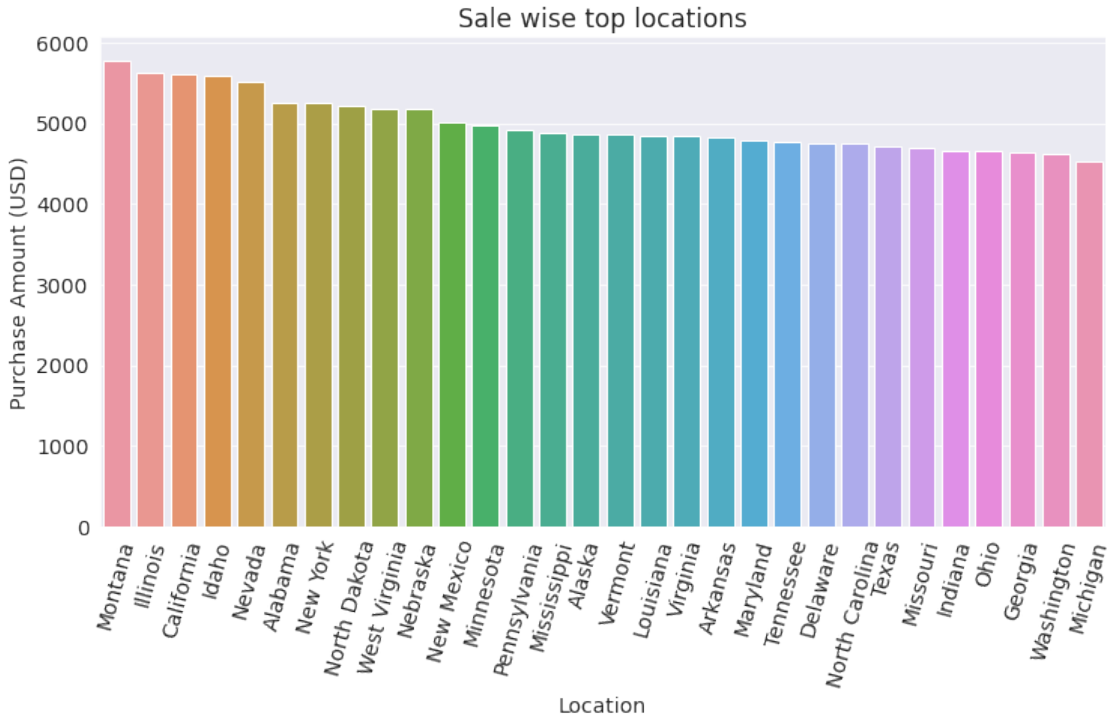
Top location vs sale

[16]:
```
top_location = shopping_df.groupby('Location')[['Purchase Amount (USD)']].sum().
↪sort_values('Purchase Amount (USD)', ascending = False).head(30)
top_location
```

[16]:

Location	Purchase Amount (USD)
Montana	5784
Illinois	5617
California	5605
Idaho	5587
Nevada	5514
Alabama	5261
New York	5257
North Dakota	5220
West Virginia	5174
Nebraska	5172
New Mexico	5014
Minnesota	4977
Pennsylvania	4926
Mississippi	4883
Alaska	4867
Vermont	4860
Louisiana	4848
Virginia	4842
Arkansas	4828
Maryland	4795
Tennessee	4772
Delaware	4758
North Carolina	4742
Texas	4712
Missouri	4691
Indiana	4655
Ohio	4649
Georgia	4645
Washington	4623
Michigan	4533

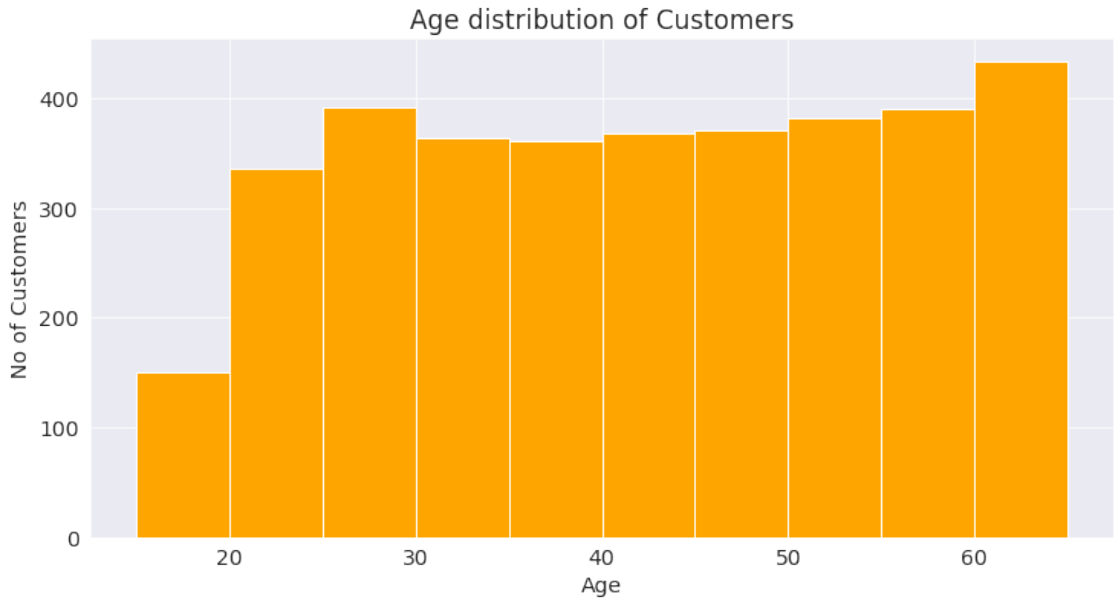
[18]:
```
plt.figure(figsize=(12,6))
plt.xticks(rotation=75)
plt.title('Sale wise top locations')
sns.barplot(x=top_location.index, y= top_location['Purchase Amount (USD)']);
```



Age Distribution of customer

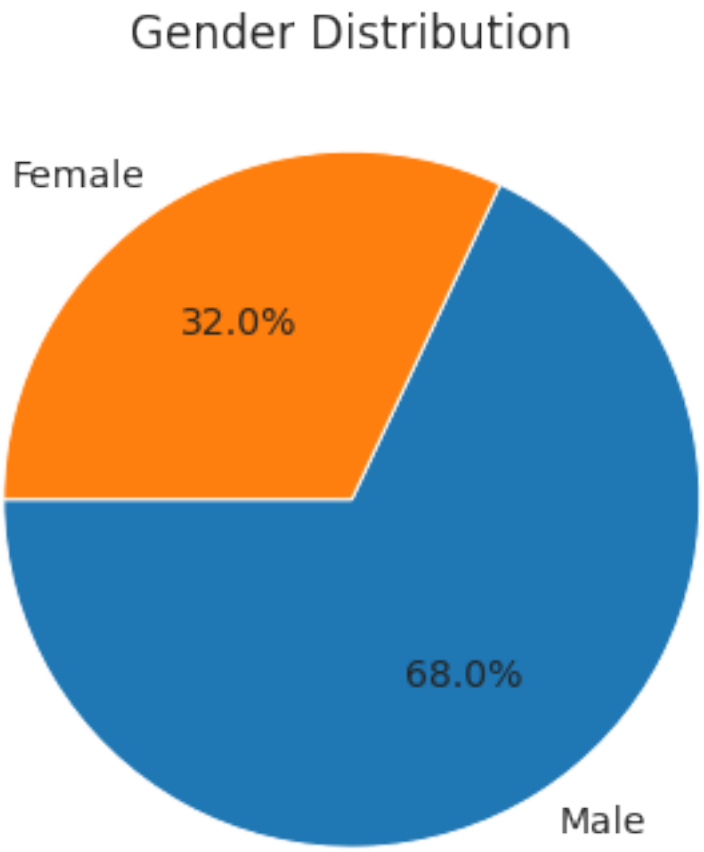
[13]:
```
plt.figure(figsize=(12,6))
plt.title('Age distribution of Customers')
plt.xlabel('Age')
plt.ylabel('No of Customers')

plt.hist(shopping_df.Age, bins=np.arange(15,70,5), color='orange');
```

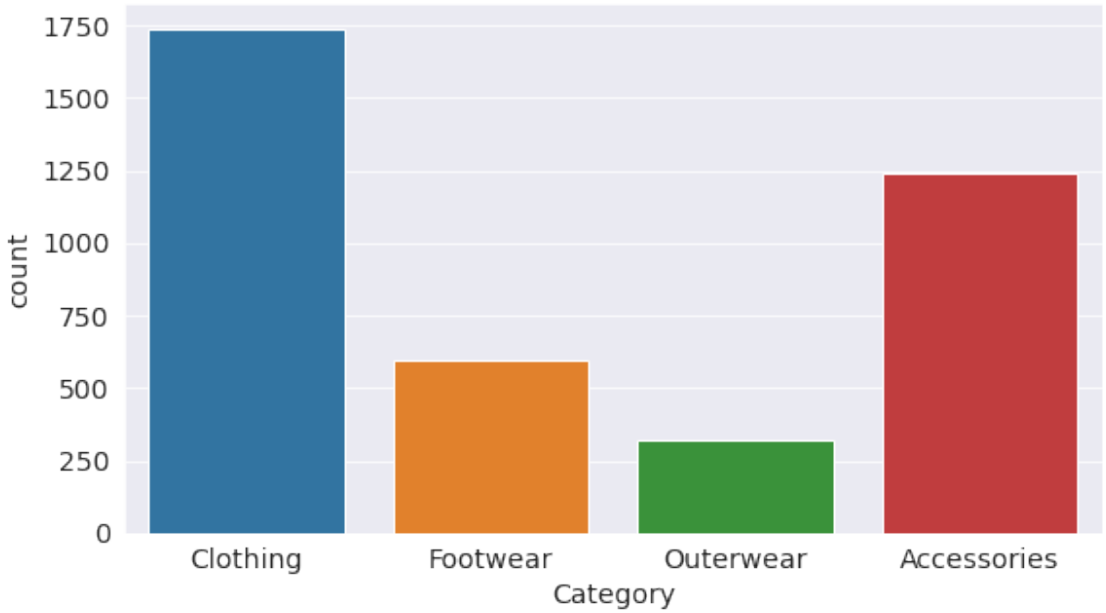


[23]:
```
#Pie Chart of gender distribution

plt.figure(figsize=(12,6))
plt.title('Gender Distribution')
plt.pie(gender_count, labels=gender_count.index, autopct='%1.1f%%',␣
 ↪startangle=180);
```



[25]:
```
sns.countplot(x=shopping_df.Category);
```



[ ]:

7 6. Question and Answer

[ ]:

[ ]:

8 7. Inferences and Conclusion

[ ]:

[ ]: