

```
In [2]: import pandas as pd
import numpy as np

In [3]: a = pd.read_csv("customer_shopping_behavior.csv")

In [4]: a.head(8)

Out[4]:
   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            0    1    55      Male         Blouse  Clothing                53    Kentucky  L      Gray  Winter        3.1             Yes      Express          Yes          Yes          14          Venmo      Fortnightly
1            1    2    19      Male         Sweater  Clothing                64      Maine  L    Maroon  Winter        3.1             Yes      Express          Yes          Yes           2          Cash      Fortnightly
2            2    3    50      Male          Jeans  Clothing                73  Massachusetts  S    Maroon   Spring        3.1             Yes  Free Shipping          Yes          Yes          23      Credit Card        Weekly
3            3    4    21      Male         Sandals  Footwear                90  Rhode Island  M    Maroon   Spring        3.5             Yes  Next Day Air          Yes          Yes          49          PayPal        Weekly
4            4    5    45      Male         Blouse  Clothing                49      Oregon  M  Turquoise   Spring        2.7             Yes  Free Shipping          Yes          Yes          31          PayPal        Annually
5            5    6    46      Male         Sneakers  Footwear                20      Wyoming  M      White  Summer        2.9             Yes   Standard          Yes          Yes          14          Venmo        Weekly
6            6    7    63      Male          Shirt  Clothing                85      Montana  M      Gray    Fall        3.2             Yes  Free Shipping          Yes          Yes          49          Cash      Quarterly
7            7    8    27      Male          Shorts  Clothing                34      Louisiana  L  Charcoal   Winter        3.2             Yes  Free Shipping          Yes          Yes          19      Credit Card        Weekly

In [5]: a.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         3863 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 [7]: a.describe()

Out[7]:
   Customer ID      Age  Purchase Amount (USD)  Review Rating  Previous Purchases
count  3900.000000  3900.000000          3900.000000  3863.000000          3900.000000
mean    1950.500000    44.068462          59.764359    3.750065          25.351538
std     1125.977353    15.207589          23.685392    0.716983          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    3.800000          25.000000
75%     2925.250000    57.000000          81.000000    4.400000          38.000000
max     3900.000000    70.000000          100.000000    5.000000          50.000000

In [8]: # For all dataset includeing Number and String both
a.describe(include = "all")

Out[8]:
   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
count  3900.000000  3900.000000    3900          3900          3900          3900.000000    3900    3900    3900    3900    3863.000000          3900          3900          3900          3900          3900.000000          3900          3900
unique      NaN      NaN      NaN           2           4           NaN          50      4      25      4           NaN           2           6           2           2           NaN           6           7
top         NaN      NaN      Male         Blouse  Clothing           NaN    Montana  M  Olive   Spring          NaN          No  Free Shipping          No          No          NaN          PayPal        Every 3 Months
freq         NaN      NaN      2652          171          1737           NaN    96  1755    177    999          NaN          2847          675          2223          2223          NaN          677          584
mean    1950.500000    44.068462      NaN          NaN          NaN          59.764359      NaN    NaN    NaN    NaN    3.750065          NaN          NaN          NaN          NaN          25.351538          NaN          NaN
std     1125.977353    15.207589      NaN          NaN          NaN          23.685392      NaN    NaN    NaN    NaN    0.716983          NaN          NaN          NaN          NaN          14.447125          NaN          NaN
min       1.000000    18.000000      NaN          NaN          NaN          20.000000      NaN    NaN    NaN    NaN    2.500000          NaN          NaN          NaN          NaN          1.000000          NaN          NaN
25%      975.750000    31.000000      NaN          NaN          NaN          39.000000      NaN    NaN    NaN    NaN    3.100000          NaN          NaN          NaN          NaN          13.000000          NaN          NaN
50%     1950.500000    44.000000      NaN          NaN          NaN          60.000000      NaN    NaN    NaN    NaN    3.800000          NaN          NaN          NaN          NaN          25.000000          NaN          NaN
75%     2925.250000    57.000000      NaN          NaN          NaN          81.000000      NaN    NaN    NaN    NaN    4.400000          NaN          NaN          NaN          NaN          38.000000          NaN          NaN
max     3900.000000    70.000000      NaN          NaN          NaN          100.000000      NaN    NaN    NaN    NaN    5.000000          NaN          NaN          NaN          NaN          50.000000          NaN          NaN

In [9]: a.isnull().sum()

Out[9]:
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   37
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 [17]: a["Review Rating"] = a.groupby("Category")["Review Rating"].transform(lambda x: x.fillna(x.median()))

In [19]: a.isnull().sum()

Out[19]:
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 [21]: a.columns = a.columns.str.lower()
a.columns = a.columns.str.replace(" ", "_")

In [22]: a.columns

Out[22]:
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 [24]: a = a.rename(columns={ "purchase_amount_usd": "purchase_amount"})

In [25]: a.columns

Out[25]:
Index(['customer_id', 'age', 'gender', 'item_purchased', 'category',
      'purchase_amount', '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 [30]: # Create a group wise age_group
labels = ["Young Adult", "Adult", "Middle-aged", "Senior"]
a["age_group"] = pd.qcut(a["age"], q = 4, labels = labels)

In [31]: a[["age", "age_group"]].head(7)

Out[31]:
   age  age_group
0    55  Middle-aged
1    19  Young Adult
2    50  Middle-aged
3    21  Young Adult
4    45  Middle-aged
5    46  Middle-aged
6    63    Senior

In [43]: # Create a column for Purchase_frequency_days

frequency_mapping = {
    "Fortnightly": 14,
    "Weekly": 7,
    "Monthly": 30,
    "Quarterly": 90,
    "Bi-Weekly": 14,
    "Annually": 365,
    "Every 3 months": 90
}
a["purchase_frequency_days"] = a["frequency_of_purchases"].map(frequency_mapping)

In [44]: a[["purchase_frequency_days", "frequency_of_purchases"]].head(7)

Out[44]:
   purchase_frequency_days  frequency_of_purchases
0                14.0      Fortnightly
1                14.0      Fortnightly
2                 7.0       Weekly
3                 7.0       Weekly
4               365.0      Annually
5                 7.0       Weekly
6                 90.0      Quarterly

In [45]: a[["discount_applied", "promo_code_used"]].head(10)

Out[45]:
   discount_applied  promo_code_used
0             Yes             Yes
1             Yes             Yes
2             Yes             Yes
3             Yes             Yes
4             Yes             Yes
5             Yes             Yes
6             Yes             Yes
7             Yes             Yes
8             Yes             Yes
9             Yes             Yes

In [50]: [a["discount_applied"] == a["promo_code_used"]].all()

Out[50]:
np.True_

In [51]: a.columns

Out[51]:
Index(['customer_id', 'age', 'gender', 'item_purchased', 'category',
      'purchase_amount', 'location', 'size', 'color', 'season',
      'review_rating', 'subscription_status', 'shipping_type',
      'discount_applied', 'promo_code_used', 'previous_purchases',
      'payment_method', 'frequency_of_purchases', 'age_group',
      'purchase_frequency_days'],
      dtype='object')

In [52]: a = a.drop("promo_code_used", axis = 1)

In [53]: a.columns

Out[53]:
Index(['customer_id', 'age', 'gender', 'item_purchased', 'category',
      'purchase_amount', 'location', 'size', 'color', 'season',
      'review_rating', 'subscription_status', 'shipping_type',
      'discount_applied', 'previous_purchases', 'payment_method',
      'frequency_of_purchases', 'age_group', 'purchase_frequency_days'],
      dtype='object')

In [56]: pip install psycopg2-binary sqlalchemy

Collecting psycopg2-binary
  Downloading psycopg2_binary-2.9.11-cp313-cp313-win_amd64.whl.metadata (5.1 kB)
Requirement already satisfied: sqlalchemy in c:\users\kishan gupta\anaconda3\lib\site-packages (2.0.39)
Requirement already satisfied: greenlet==0.4.17 in c:\users\kishan gupta\anaconda3\lib\site-packages (from sqlalchemy) (3.1.1)
Requirement already satisfied: typing-extensions>=4.6.0 in c:\users\kishan gupta\anaconda3\lib\site-packages (from sqlalchemy) (4.12.2)
Downloading psycopg2_binary-2.9.11-cp313-cp313-win_amd64.whl (2.7 MB)
----- 0.0/2.7 MB ? eta -:--:--
----- 1.0/2.7 MB 7.9 MB/s eta 0:00:01
----- 2.6/2.7 MB 8.1 MB/s eta 0:00:01
----- 2.7/2.7 MB 7.2 MB/s eta 0:00:00
Installing collected packages: psycopg2-binary
Successfully installed psycopg2-binary-2.9.11
Note: you may need to restart the kernel to use updated packages.

In [58]: from sqlalchemy import create_engine

# Step 1: Connect to PostgreSQL
# Replace placeholders with your actual details
username = "postgres" # default user
password = "root123" # the password you set during installation
host = "localhost" # if running locally
port = "5432" # default PostgreSQL port
database = "customer_behavior" # the database you created in pgAdmin

# Create the connection engine
engine = create_engine("psycopg2://(username):(password)@host:(port)/(database)")

# Step 2: Load DataFrame into PostgreSQL
table_name = "customer" # choose any table name
a.to_sql(table_name, engine, if_exists="replace", index=False)

# Print confirmation
print(f"Data successfully loaded into table '{table_name}' in database '{database}'.")

Data successfully loaded into table 'customer' in database 'customer_behavior'.

In [ ]:
```