In [112…
```
import pandas as pd

df = pd.read_csv(r"C:\Users\Dell\customer_shopping_behavior.csv")
df
```

Out[112…

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location
0	1	55	Male	Blouse	Clothing	53	Kentucky
1	2	19	Male	Sweater	Clothing	64	Maine
2	3	50	Male	Jeans	Clothing	73	Massachusetts
3	4	21	Male	Sandals	Footwear	90	Rhode Island
4	5	45	Male	Blouse	Clothing	49	Oregon
...	...	...	...	...	...	...	...
3895	3896	40	Female	Hoodie	Clothing	28	Virginia
3896	3897	52	Female	Backpack	Accessories	49	Iowa
3897	3898	46	Female	Belt	Accessories	33	New Jersey
3898	3899	44	Female	Shoes	Footwear	77	Minnesota
3899	3900	52	Female	Handbag	Accessories	81	California

3900 rows × 18 columns

In [113…
```
df.columns
```

Out[113…
```
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 [114…
```
df.info()
df.isnull().sum()
```

```
<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
```

Out[114…

```
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 [115…

```
# Review rating fill null values with median

df['Review Rating']=df.groupby('Category')['Review Rating'].transform(lambda x
df

df.isnull().sum()
```

Out[115…
```
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 [116…
```
#  Change column name lower case and replace space with (_)
df.columns=df.columns.str.lower()
df.columns
df.columns=df.columns.str.replace(' ','_')
df.columns
```

Out[116…
```
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 [117…
```
# Change a coulumn name
df=df.rename(columns={'purchase_amount_(usd)':'purchase_amount'})
df.columns
```

Out[117…
```
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 [118…
```
# Need a coulumn age_group based on age
labels=['Young adult','Adult','Middel_Aged','Senior']
df['age_group']=pd.qcut(df['age'], q=4, labels = labels)
df.columns
df.head(10)
```

Out[118…

	customer_id	age	gender	item_purchased	category	purchase_amount	
0	1	55	Male	Blouse	Clothing	53	
1	2	19	Male	Sweater	Clothing	64	
2	3	50	Male	Jeans	Clothing	73	Ma
3	4	21	Male	Sandals	Footwear	90	F
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	V
9	10	57	Male	Handbag	Accessories	31	

In [119…

```
# frequency_of_purchases need to create column as numeric values for easy to c
df['frequency_of_purchases'].unique()

frequency_mapping={'Fortnightly':14, 'Weekly': 7, 'Annually': 365, 'Quarterly'
       'Monthly': 30, 'Every 3 Months': 90}
df['purchases_frequency_days']=df['frequency_of_purchases'].map(frequency_mapp
df.columns
df[['purchases_frequency_days','frequency_of_purchases']].head(5)
df
```

Out[119…

	customer_id	age	gender	item_purchased	category	purchase_amount
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
...	...	...	...	...	...	...
3895	3896	40	Female	Hoodie	Clothing	28
3896	3897	52	Female	Backpack	Accessories	49
3897	3898	46	Female	Belt	Accessories	33
3898	3899	44	Female	Shoes	Footwear	77
3899	3900	52	Female	Handbag	Accessories	81

3900 rows × 20 columns

In [126…
```
df.columns
df.shape
```

Out[126… (3900, 19)

In [127…
```
pip install mysql-connector-python sqlalchemy pymysql
```

```
Collecting mysql-connector-pythonNote: you may need to restart the kernel to us
e updated packages.

  Downloading mysql_connector_python-9.5.0-cp313-cp313-win_amd64.whl.metadata
(7.7 kB)
Requirement already satisfied: sqlalchemy in c:\users\dell\anaconda3\lib\site-p
ackages (2.0.39)
Collecting pymysql
  Downloading pymysql-1.1.2-py3-none-any.whl.metadata (4.3 kB)
Requirement already satisfied: greenlet!=0.4.17 in c:\users\dell\anaconda3\lib\
site-packages (from sqlalchemy) (3.1.1)
Requirement already satisfied: typing-extensions>=4.6.0 in c:\users\dell\anacon
da3\lib\site-packages (from sqlalchemy) (4.12.2)
Downloading mysql_connector_python-9.5.0-cp313-cp313-win_amd64.whl (16.5 MB)
   ---------------------------------------- 0.0/16.5 MB ? eta -:--:--
   -- ------------------------------------- 1.0/16.5 MB 8.0 MB/s eta 0:00:02
   --------------- ------------------------ 6.8/16.5 MB 20.2 MB/s eta 0:00:01
   ------------------------------- -------- 13.1/16.5 MB 24.4 MB/s eta 0:00:01
   ---------------------------------------  16.3/16.5 MB 25.2 MB/s eta 0:00:01
   ---------------------------------------- 16.5/16.5 MB 19.7 MB/s eta 0:00:00
Downloading pymysql-1.1.2-py3-none-any.whl (45 kB)
Installing collected packages: pymysql, mysql-connector-python

   ---------------------------------------- 0/2 [pymysql]
   ---------- ----------------------------- 1/2 [mysql-connector-python]
   ---------- ----------------------------- 1/2 [mysql-connector-python]
   ---------- ----------------------------- 1/2 [mysql-connector-python]
   ---------- ----------------------------- 1/2 [mysql-connector-python]
   ---------- ----------------------------- 1/2 [mysql-connector-python]
   ---------- ----------------------------- 1/2 [mysql-connector-python]
   ---------- ----------------------------- 1/2 [mysql-connector-python]
   ---------- ----------------------------- 1/2 [mysql-connector-python]
   ---------- ----------------------------- 1/2 [mysql-connector-python]
   ---------- ----------------------------- 1/2 [mysql-connector-python]
   ---------- ----------------------------- 1/2 [mysql-connector-python]
   ---------- ----------------------------- 1/2 [mysql-connector-python]
   ---------- ----------------------------- 1/2 [mysql-connector-python]
   ---------------------------------------- 2/2 [mysql-connector-python]

Successfully installed mysql-connector-python-9.5.0 pymysql-1.1.2
```

In [132…
```
df.to_csv(r"D:\data sets\cleaned_data.csv", index=False)
```

In [ ]: