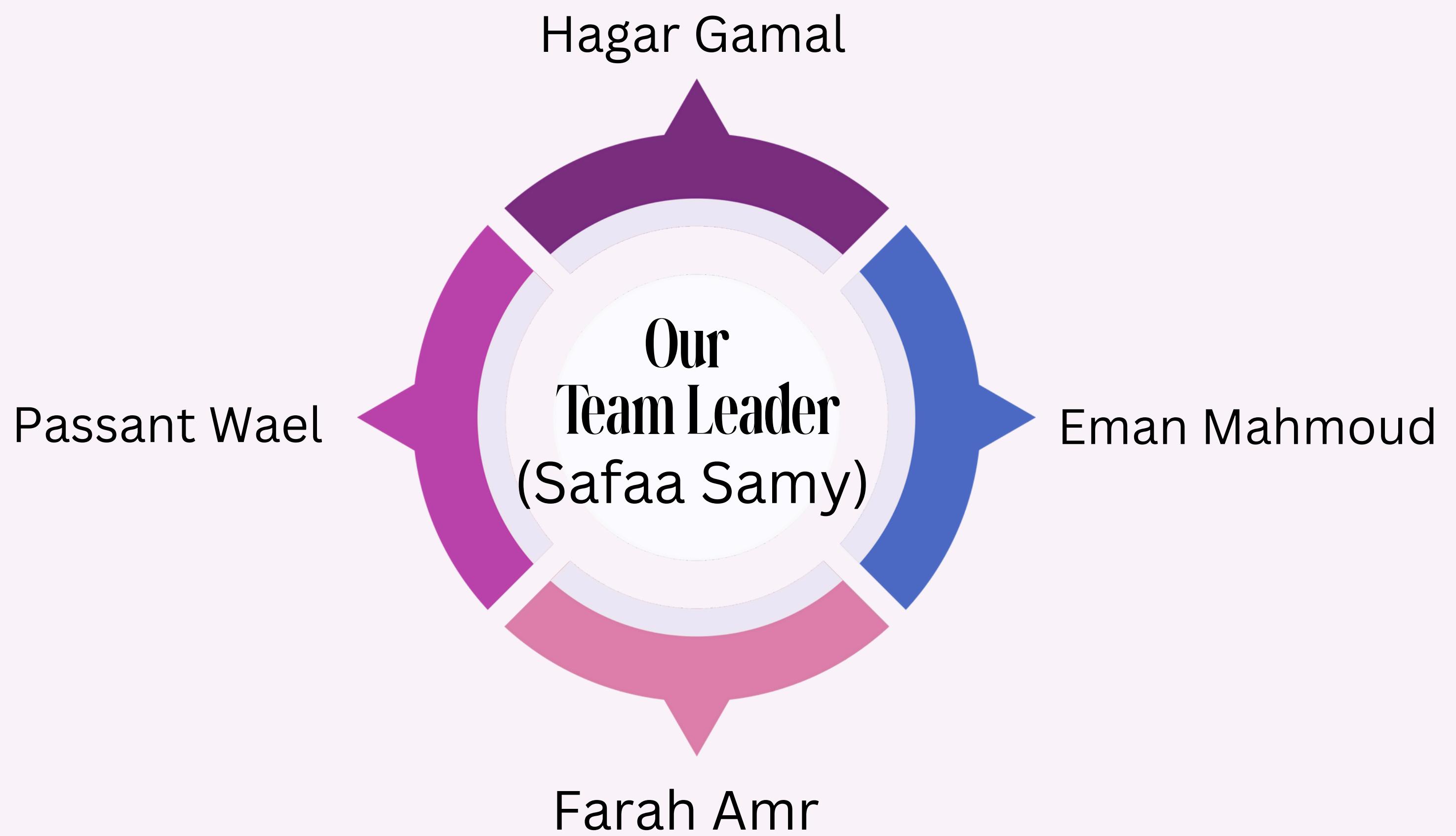




Deal

online store

The Project Team



Data Analysis Project for De al Store

Objective:

To analyze key sales metrics, customer demographics, and supplier performance to optimize business strategies and improve decision-making.

Key Focus Areas:

- Sales Performance: Identifying top products, suppliers, and revenue-generating regions.
- Customer Insights: Understanding age groups and demand trends across categories.
- Operational Efficiency: Evaluating shipping methods and return rates for improvement.

Goal:

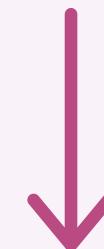
To enhance sales strategies, optimize inventory management, and improve customer satisfaction by leveraging data-driven insights.

Data Analysis Project for De al Store

Our project will be divided into four weeks of challenges.

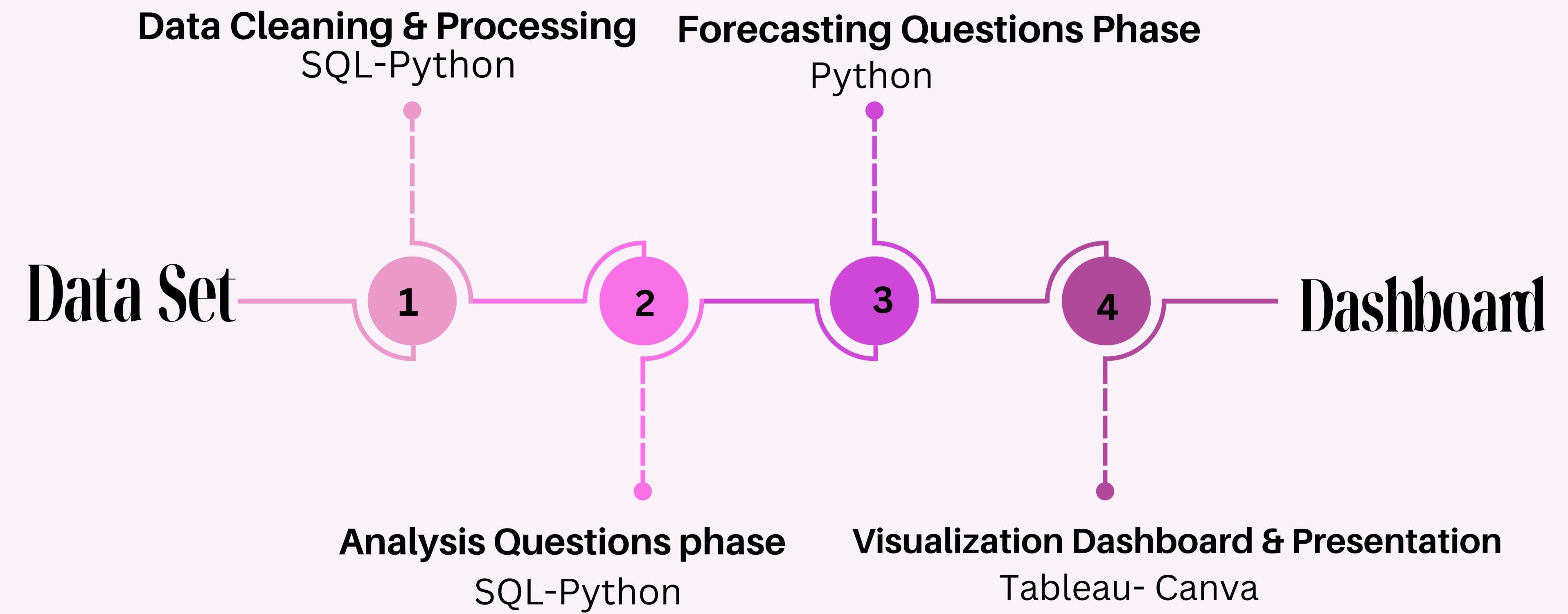


Various tools will be used to obtain many important points for data analysis



(Excel -Python- SQL-Tableau)





WEEK 1

Data Cleaning & Processing



SQL part

- 1- Safaa Samy
- 2- Passant Wael



python part

- 1-Hagar Gamal
- 2-Farah Amr
- 3- Eman Mahmoud

CREATING TABLES (SQL)

1.product:

```
USE E_COMMERCE ;
CREATE TABLE products (
product_id int primary key,
product_name
varchar(40) not null,
price decimal (4,2),
tax_rate int,
discount int,
category varchar(40)
not null,
stock_level int,
return_level decimal (4,2),
seasonality tinyint(1),
popularity_index int
```

2.customer:

```
USE E_COMMERCE ;
CREATE TABLE customers (
customer_id int primary key ,
customer_gender varchar(5)
not null
check(customer_gender int),
('Female','Male')),
customer_location
varchar(50)
not null,
customer_age_group int);
```

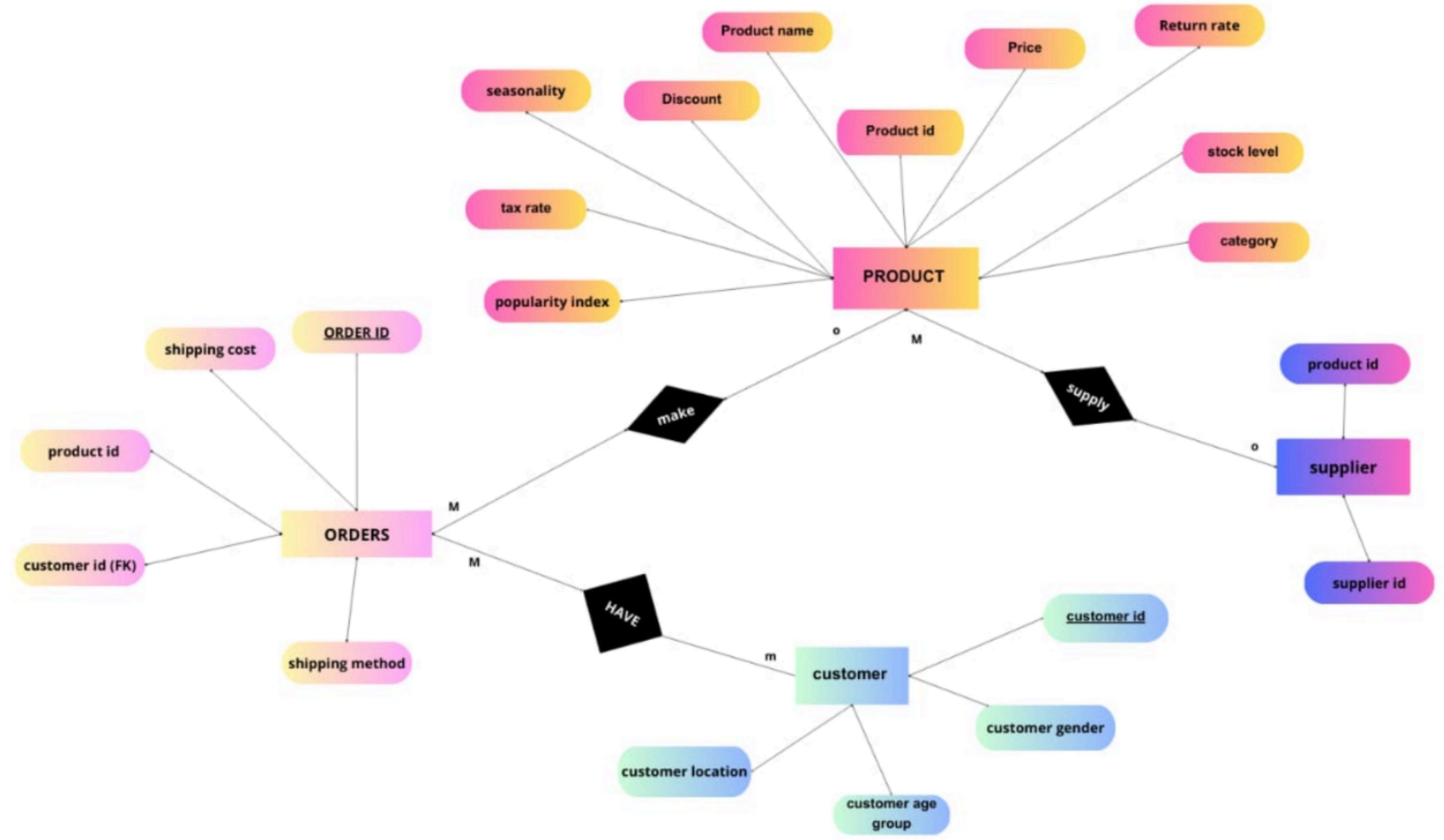
3.create order:

```
USE E_COMMERCE ;
CREATE TABLE ORDERS (
ORDER_ID int primary key
not null,
product_id int ,
customer_id int ,
shipping_cost
decimal(4,2),
shipping_method varchar
(50),
pure_profit decimal(8,2),
foreign key (product_id)
references
products(product_id),
```

4.create supply:

```
USE E_COMMERCE ;
CREATE TABLE supply (
supplier_id int primary
key not null,
product_id int,
foreign key (product_id)
references
products(product_id)
```

ERD Diagram SQL



01 python notebook

creating python notebook
for our data set



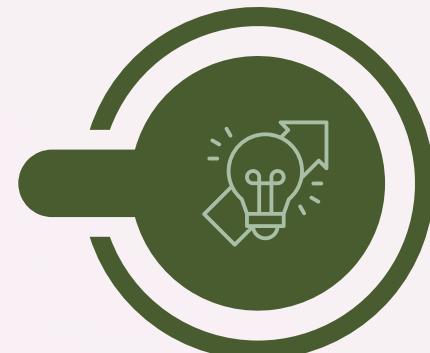
02 Importing libraries ,Dataset

import pandas,matplotlib,
seaborn Dataset



03 Data Exploration

check Data shape ,Head
,tail and info



04 Investigation

check Nulls
duplicates
missing values



Our python notebook Workflow From Exploration to visualization



Data description

check quantiles
IQR
outliers
unique data



Resolution 06

check columns name
renaming of some columns
adding column
count some data values



Testing and Verification 07

check cleaned dataset



Data visualization 08

creating some charts to
visualize the data with a
diffrent charts



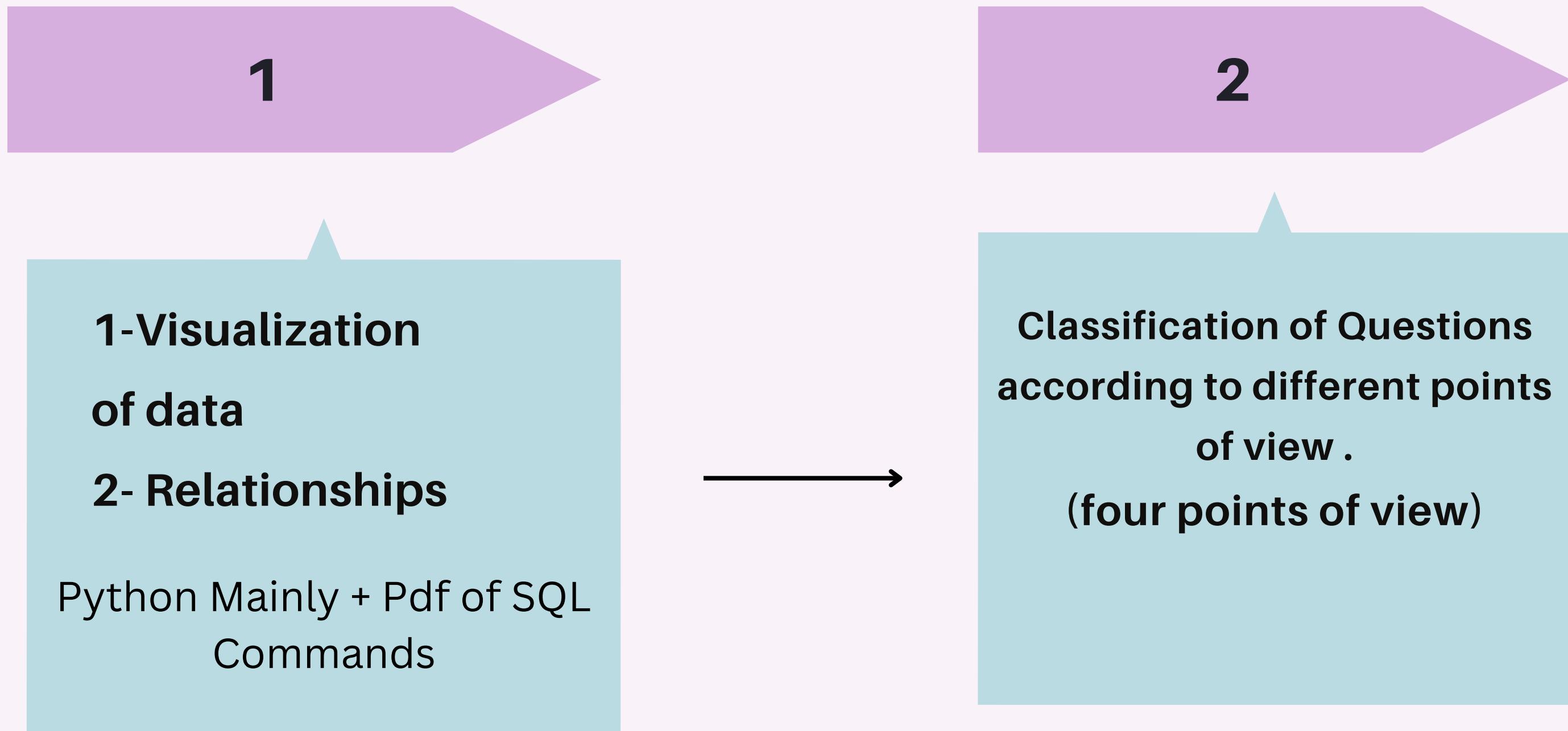
Check nulls by python

```
: ecommerce_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000000 entries, 0 to 999999
Data columns (total 16 columns):
 #   Column           Non-Null Count   Dtype  
 ---  --  
 0   Product ID      1000000 non-null    object 
 1   Product Name    1000000 non-null    object 
 2   Category        1000000 non-null    object 
 3   Price           1000000 non-null    float64
 4   Discount        1000000 non-null    int64  
 5   Tax Rate         1000000 non-null    int64  
 6   Stock Level     1000000 non-null    int64  
 7   Supplier ID     1000000 non-null    object 
 8   Customer Age Group 1000000 non-null    object 
 9   Customer Location 1000000 non-null    object 
 10  Customer Gender 1000000 non-null    object 
 11  Shipping Cost   1000000 non-null    float64
 12  Shipping Method 1000000 non-null    object 
 13  Return Rate     1000000 non-null    float64
 14  Seasonality     1000000 non-null    object 
 15  Popularity Index 1000000 non-null    int64  
dtypes: float64(3), int64(4), object(9)
memory usage: 122.1+ MB
```

week 2

Analysis Questions phase



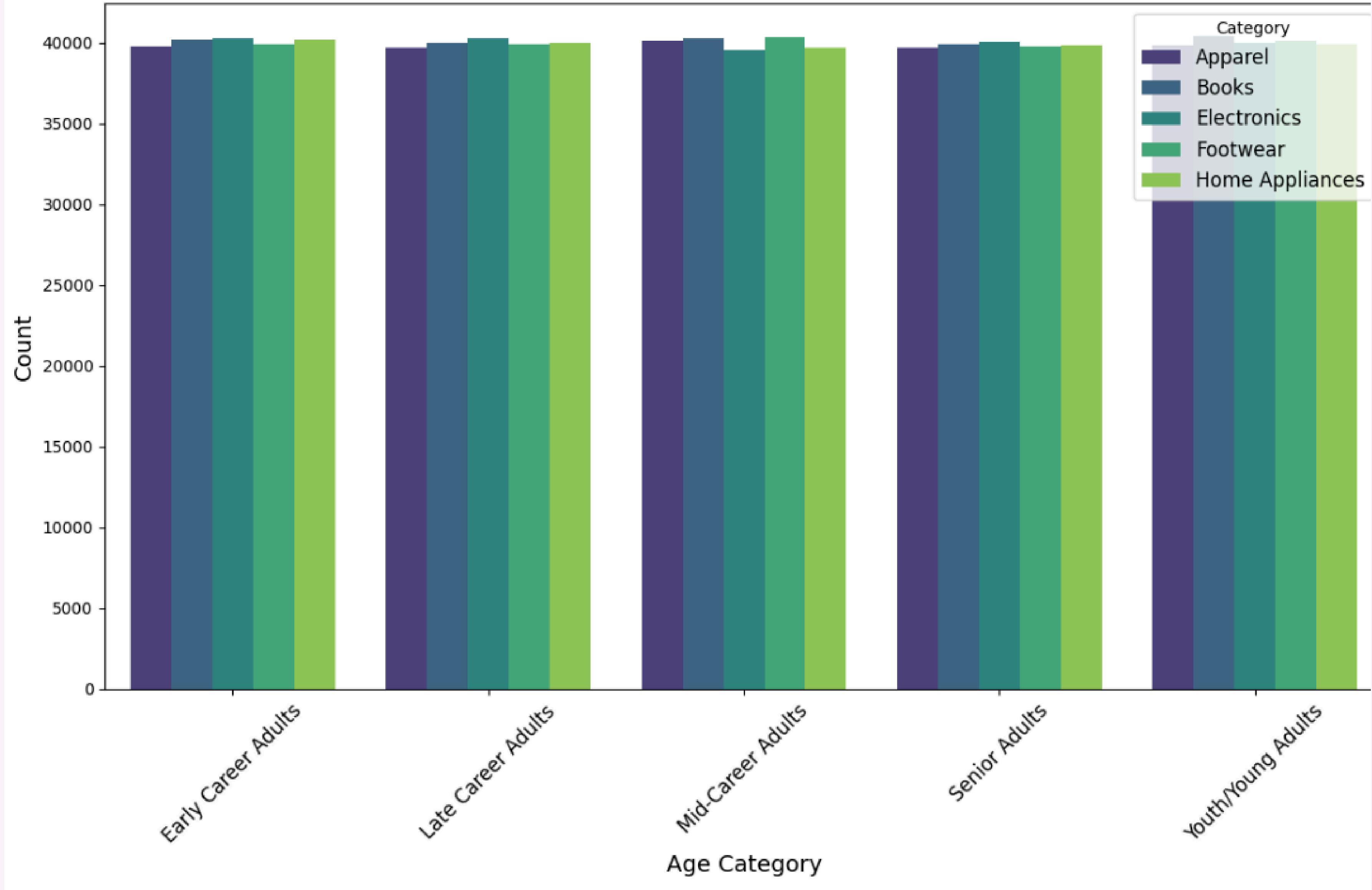


1. How does gender distribution correlate with product purchases?

2. Are customer age groups more likely to purchase a product from a specific product category than others?

3. Top 3 product popularity based on location.

Age Category vs Category Distribution

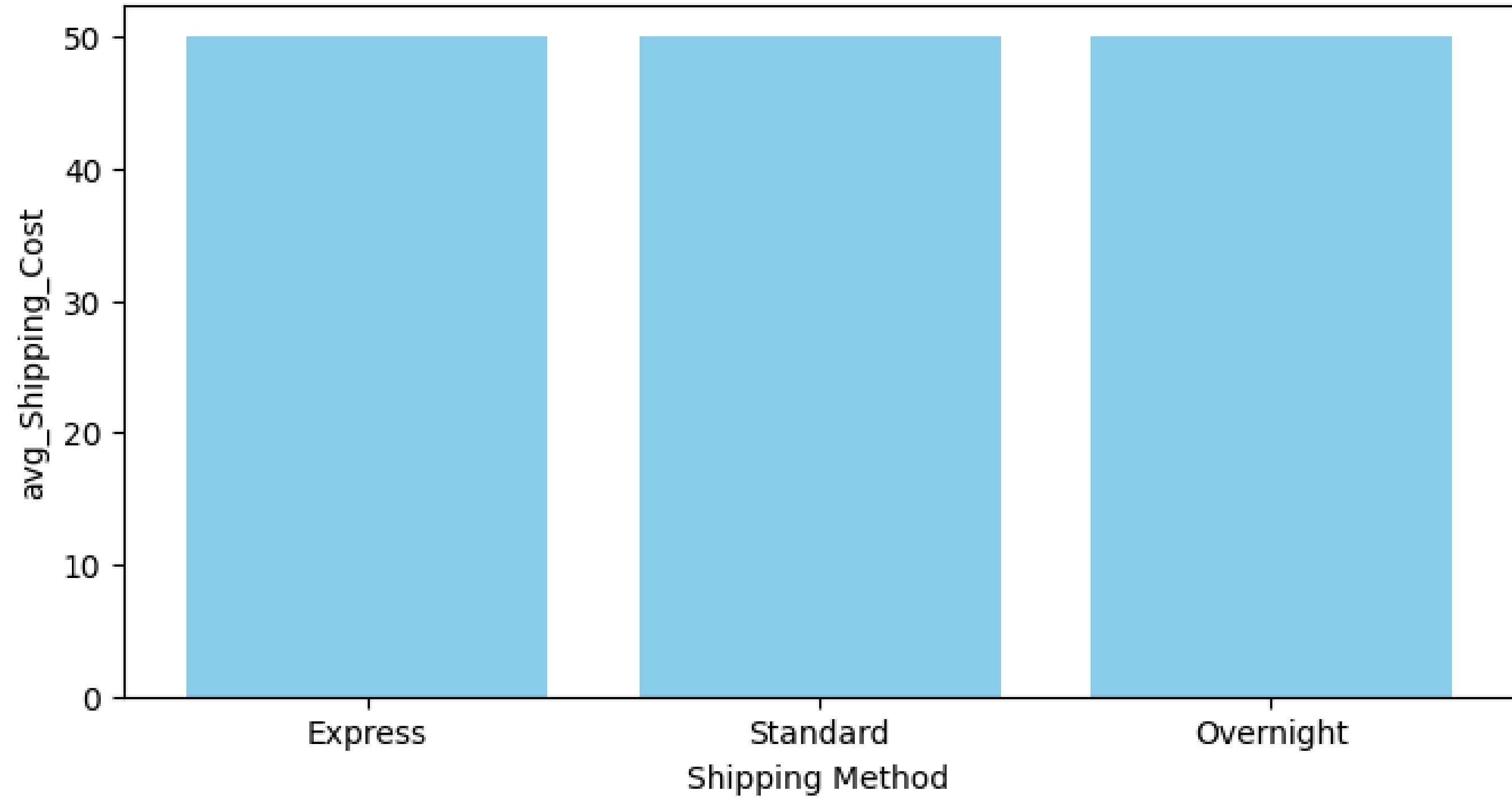


2-Shipping and Logistics

1. What's the most used shipping method for each product category?
2. Are products with higher return rates associated with a specific shipping method?
3. Which regions have the highest 'Shipping Costs'?
4. What is the most cost-effective shipping method based on the average 'Shipping Cost'?

```
<function matplotlib.pyplot.show(close=None, block=None)>
```

Shipping_method_cost_Relation



EXPRESS Method IS The Most Cost-Effective Method

3-Product



- 1. What is the overall distribution of stock levels and the risk of stockouts for items with a Popularity Index 100?**

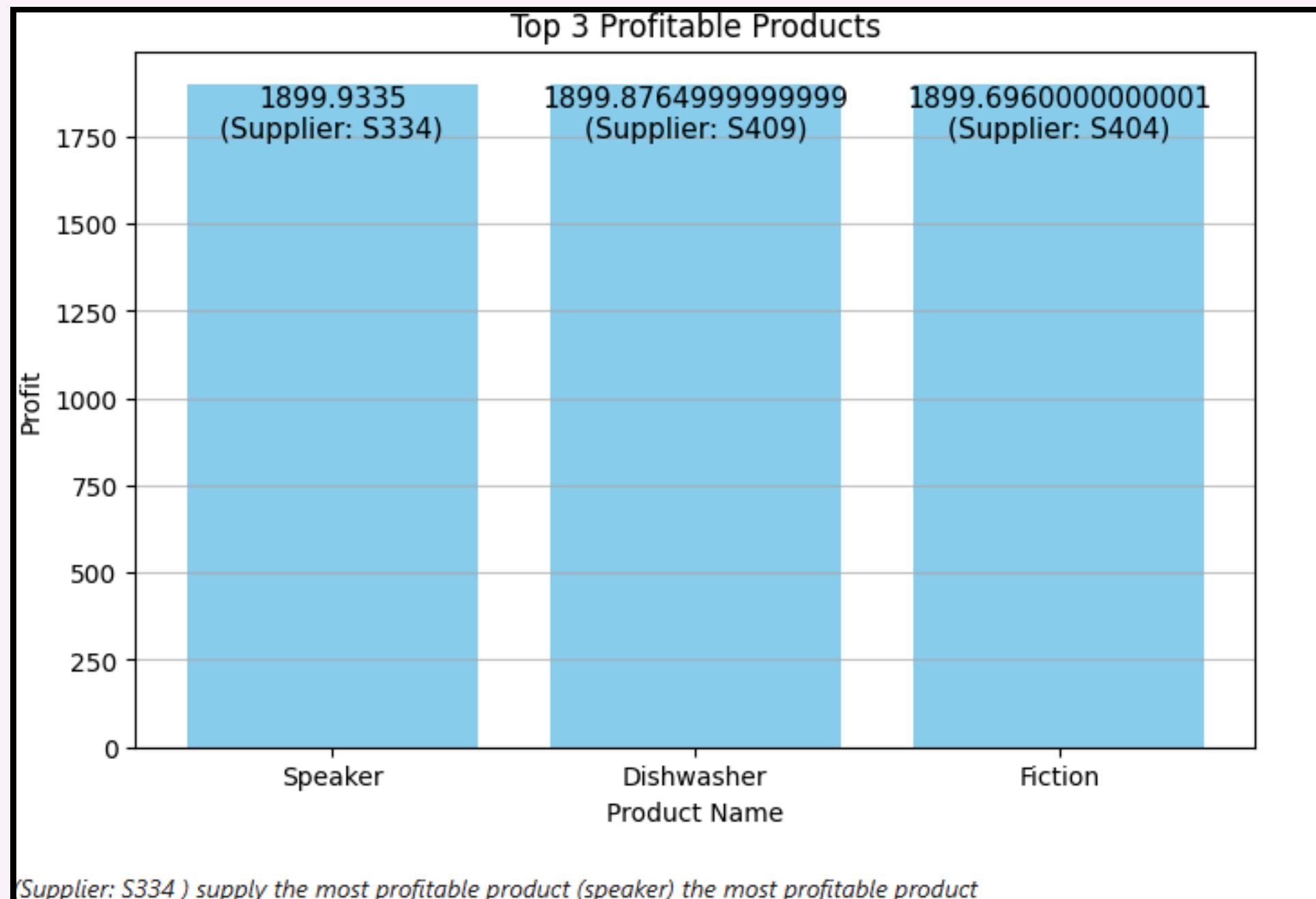
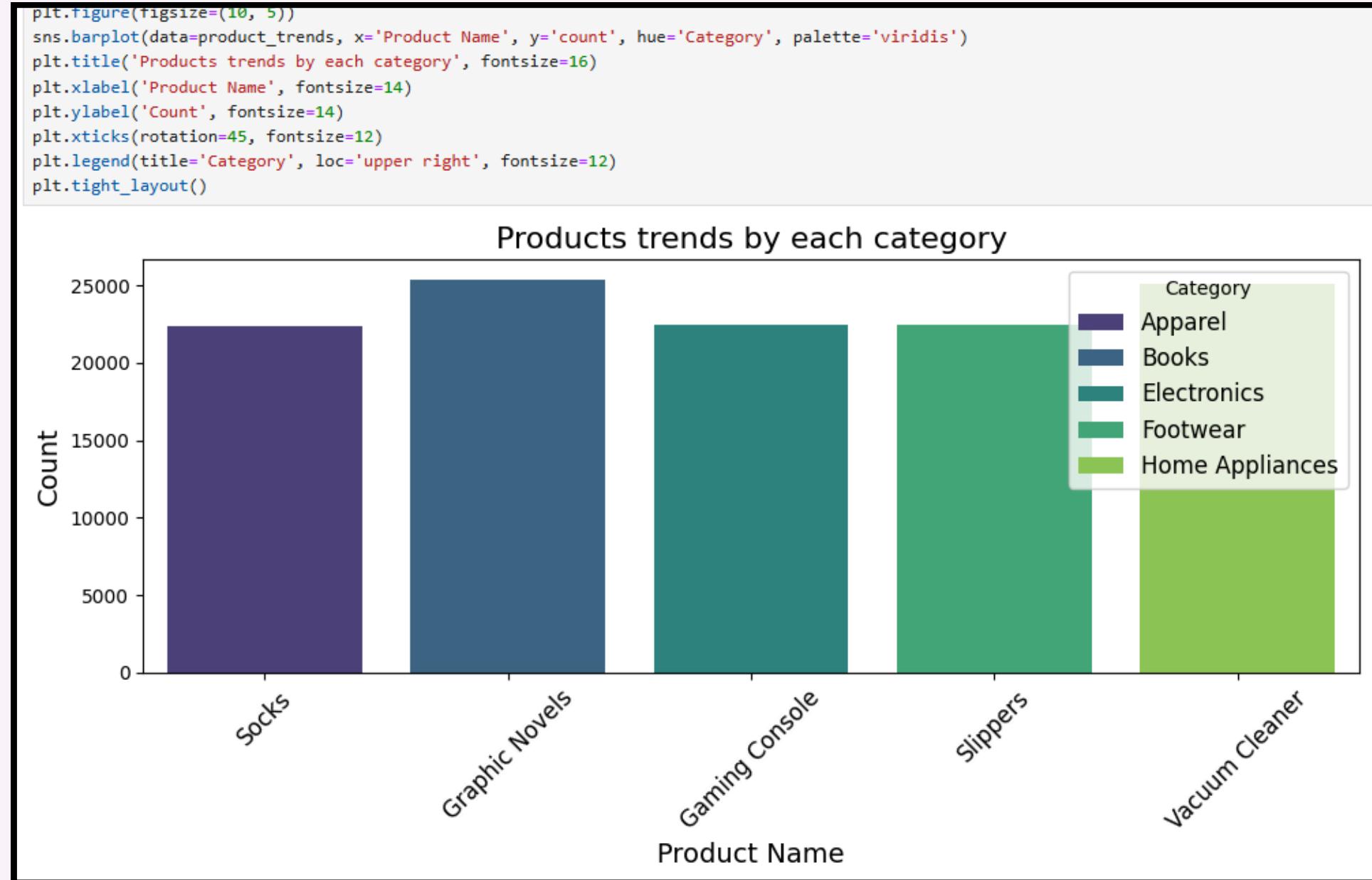
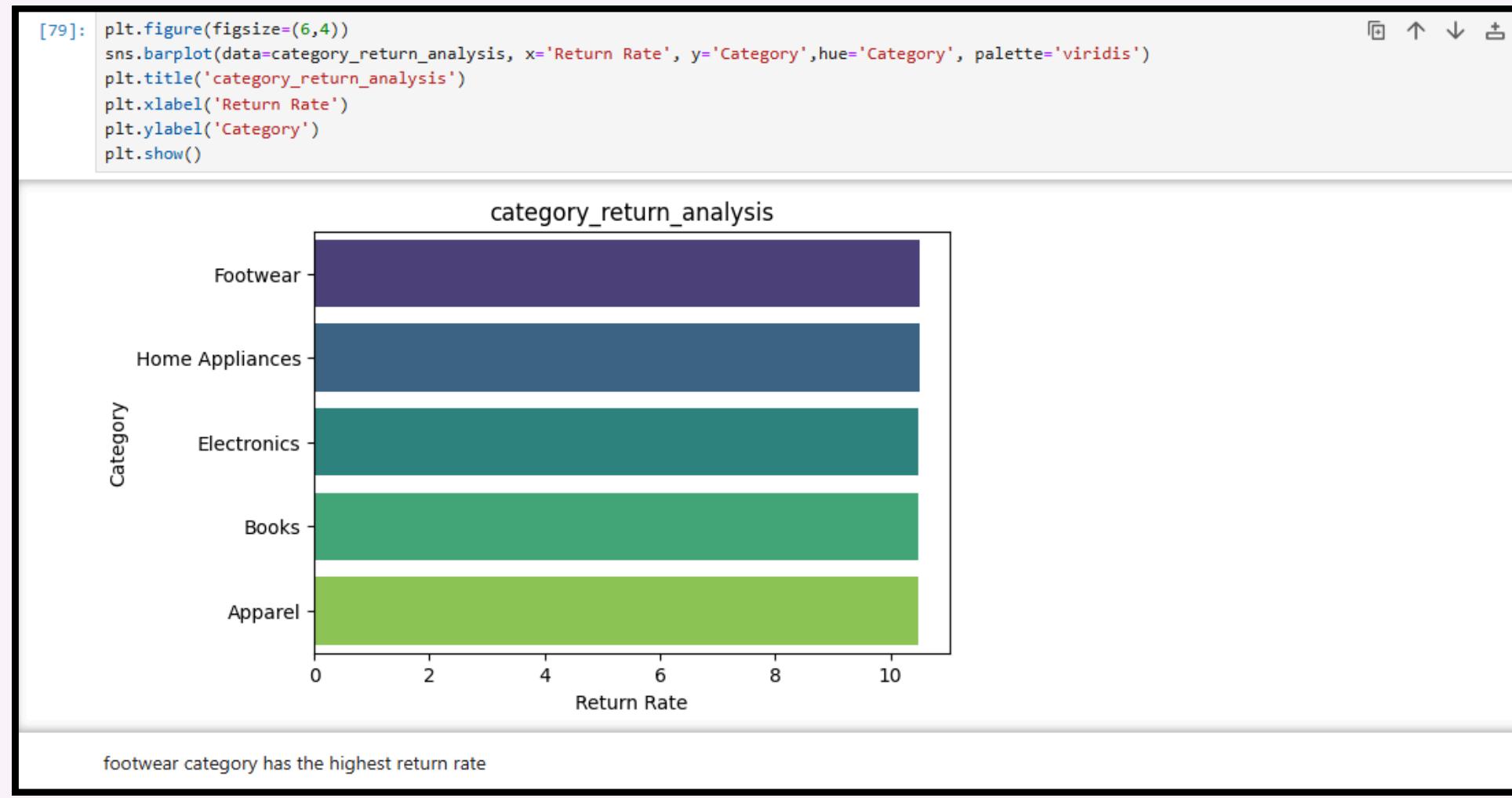
- 2. What are the Top 3 suppliers that supply the top 3 profitable products?**

- 3. What are each City's Trends By Category?**

- 4. Which products perform best overall in sales and popularity?**

- 5. Which categories have the highest return rate?**

- 6. What is the Top product in each category?**



4-Sales and Revenue

1-

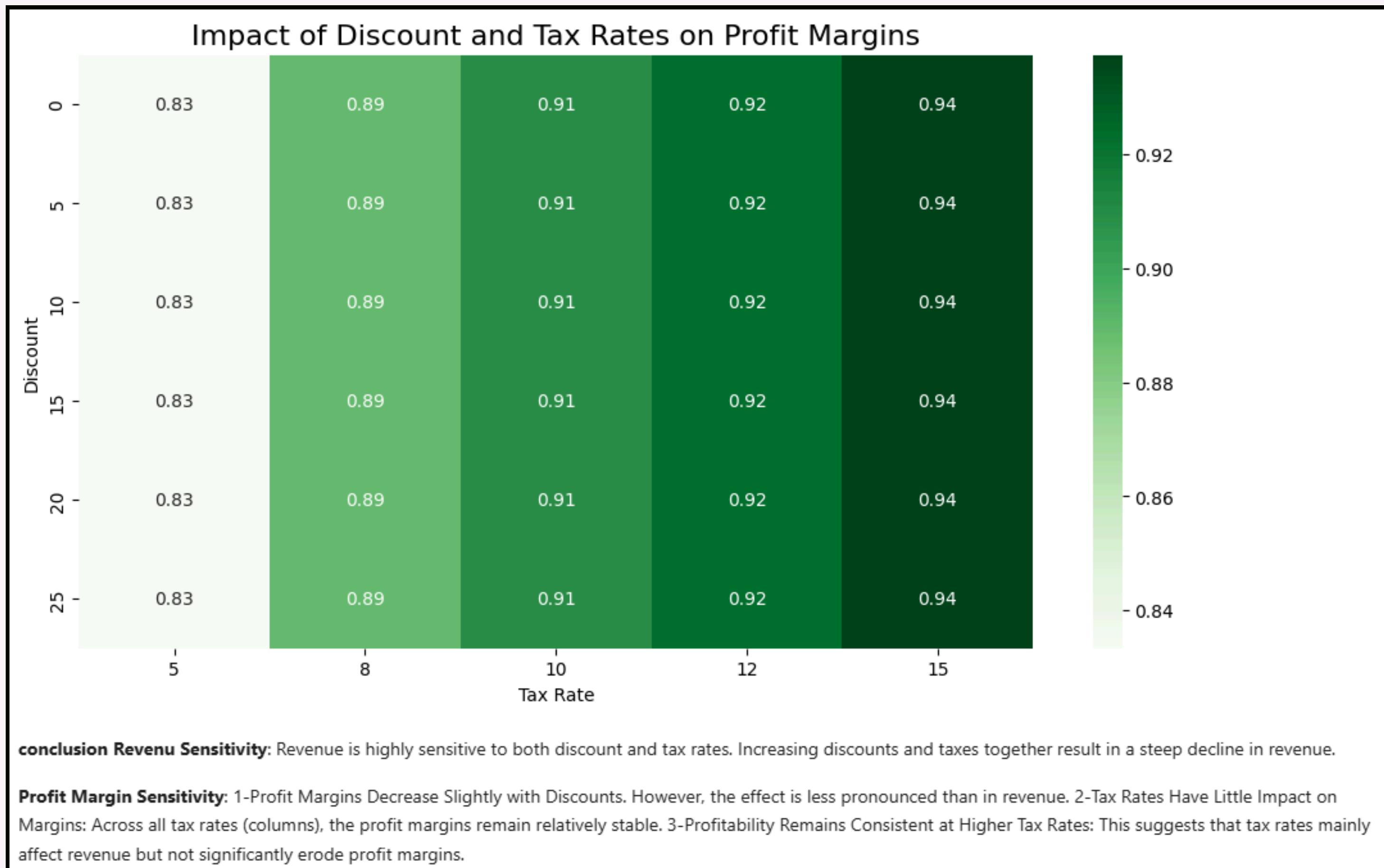
How do discount, and tax rates affect overall revenue and profit margins?

2-

How does the Seasonality impact the sales of specific products?

3-

1. Which customer groups (age, gender) contribute the most to revenue?



2-How does seasonality impact sales and popularity of specific products?

```

def analyze_seasonality_impact(ecommerce_df):
    # Group by seasonality and calculate average popularity and stock levels (as proxy for sales)
    seasonality_summary = ecommerce_df.groupby('Seasonality')[['Popularity Index', 'Stock Level']].mean()
    print("\nSeasonality Impact on Popularity and Sales:")
    print(seasonality_summary)
    return seasonality_summary
if 'Seasonality' in ecommerce_df.columns and 'Popularity Index' in ecommerce_df.columns:
    seasonality_summary = analyze_seasonality_impact(ecommerce_df)

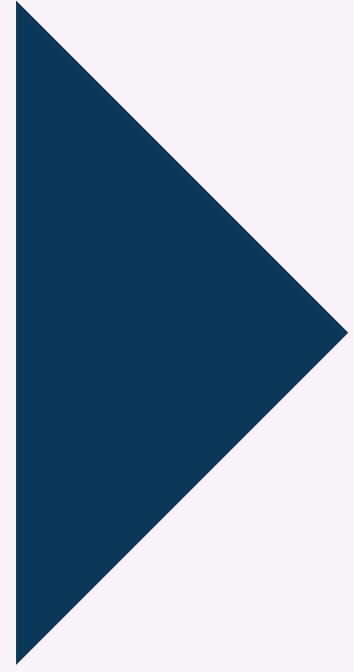
Seasonality Impact on Popularity and Sales:
    Popularity Index  Stock Level
Seasonality
No              49.968993  250.217658
Yes             49.971430  249.839261

Overall Interpretation: Seasonality seems to have a minor impact on both the Popularity Index and Stock Level. Seasonal products have a slightly lower popularity but also a slightly lower stock level, possibly due to demand peaks and inventory adjustments for specific seasons. Non-seasonal products might enjoy more consistent popularity and slightly higher stock levels, suggesting that they are more reliable in terms of steady sales and stock management.

, In conclusion, while seasonality impacts both popularity and stock levels, the differences are relatively small, and other factors might play a larger role in determining the popularity and stock strategies for these products.

```

Week 3



FORECASTING QUESTIONS PHASE



- 1. How does the tax rate impact sales over different customer locations?**
- 2. which regions will require higher shipping costs based on past purchase?**
- 3. How does shipping cost vary across different customer locations over time?**
- 4. Which products are likely to go out of stock soon based on current demand and stock levels?**



**OUR DATASET DOSEN`T CONTAIN ANY DATE OR SERIES TIME SO
We Try To Obtain Questions Depend on Cost ,Stock Level ,Tax rate, and Customer Locations.**

Products Analysis

Category	Customer Age ..	Customer Gend..	Customer Locat..	Shipping Meth..
Home Appli...	25-34	(All)	(All)	Express
Shippin	Return	Tax rate	stock	

336,347

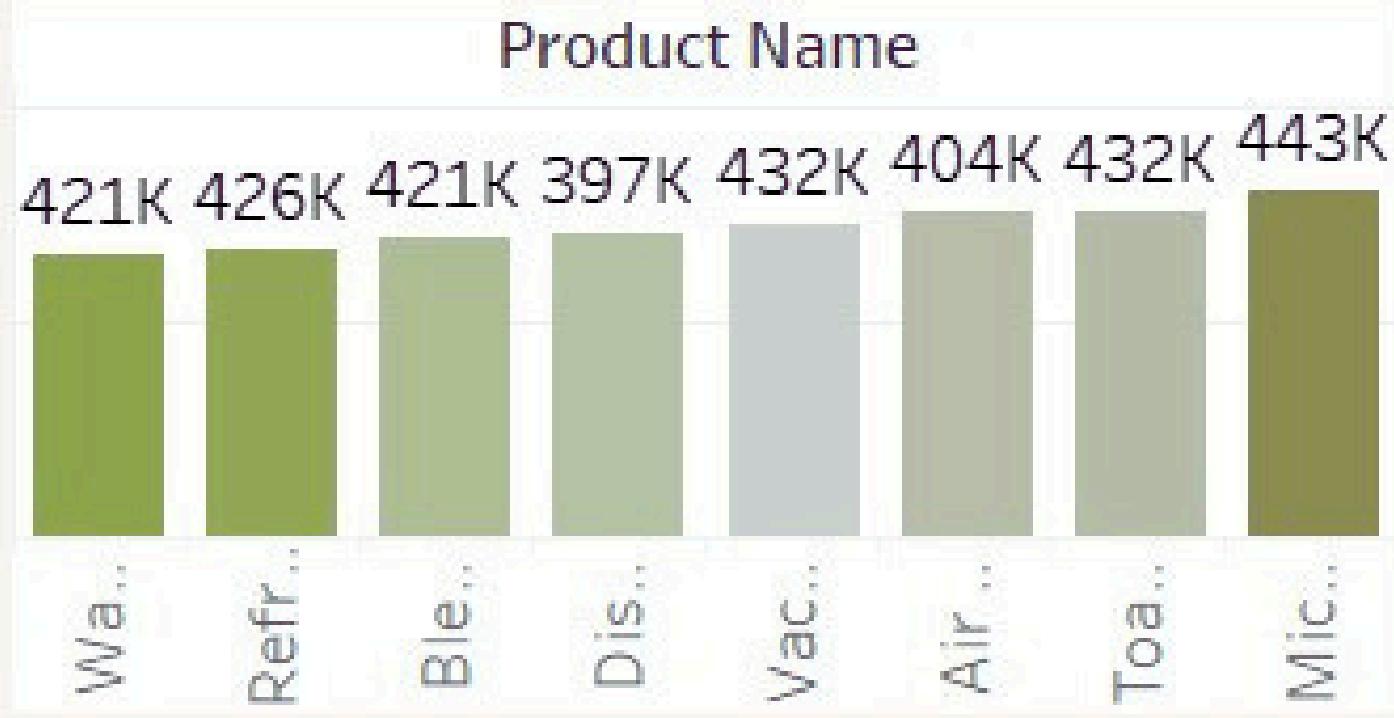
141,402

134,669

3,377,740

Which products are likely to go out of stock soon based on current demand and stock levels?

How does shipping cost vary across different customer locations over time?



Seasonality

165K

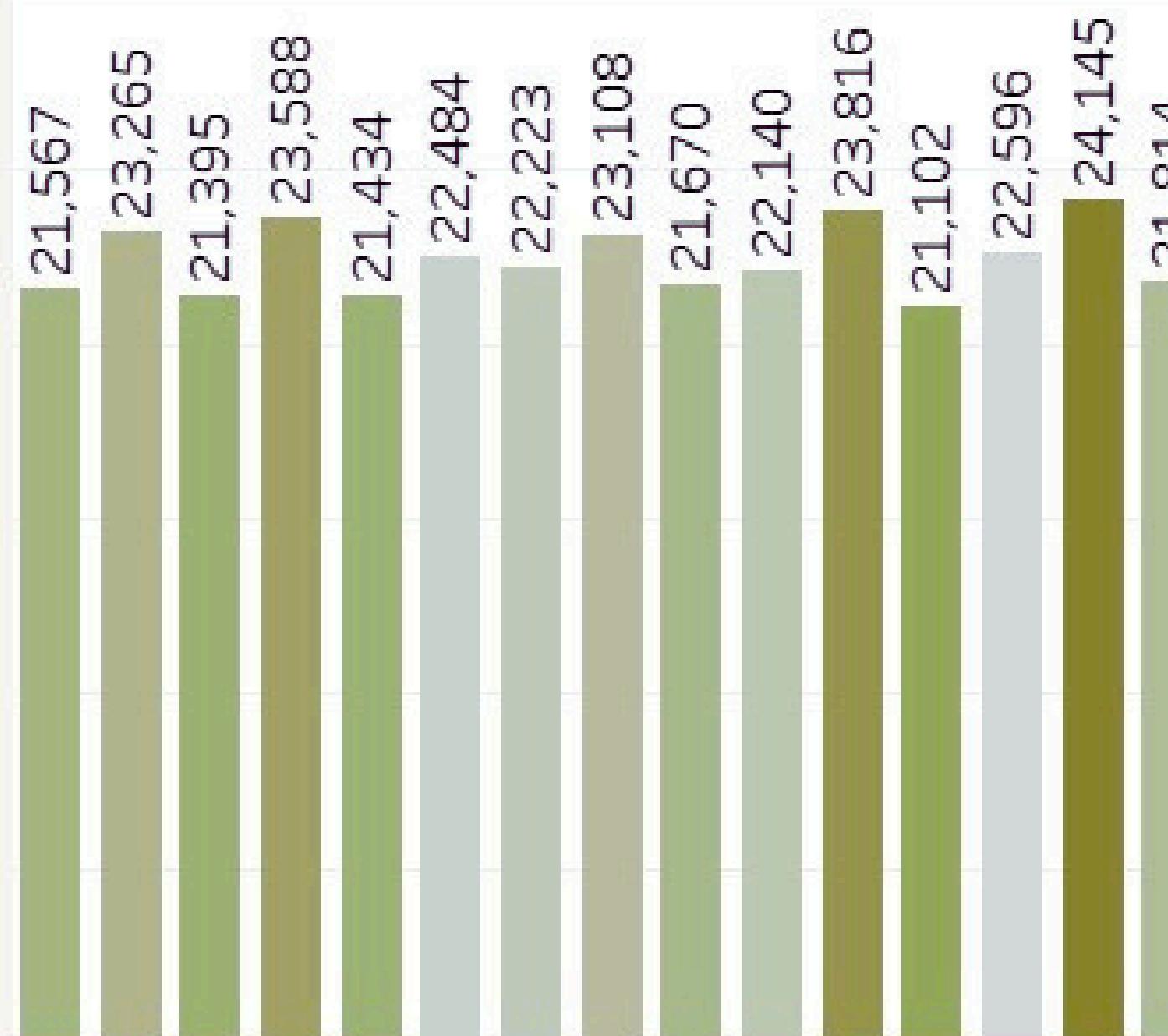
171K

No

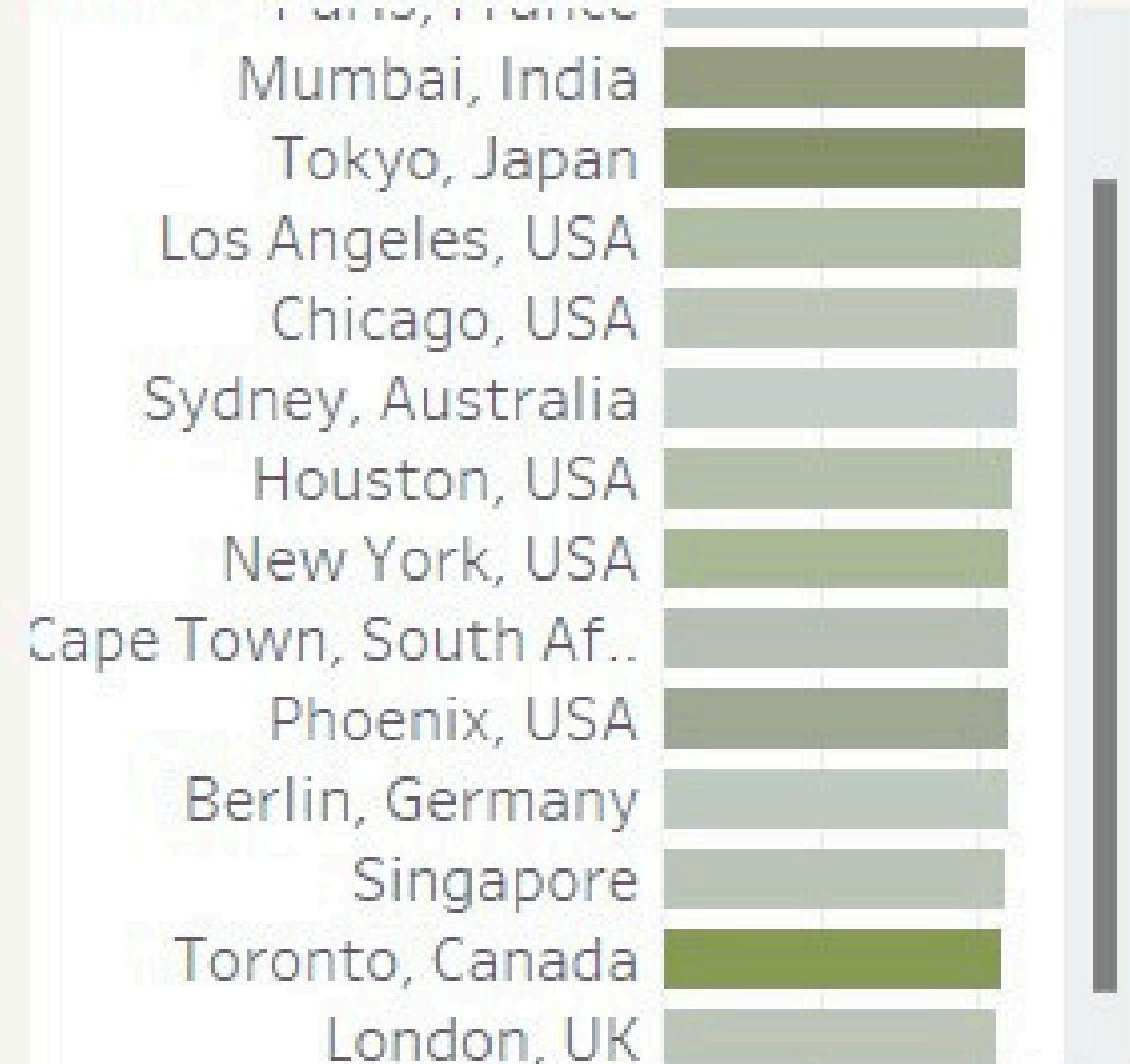
Yes

which regions will require higher shipping costs based on past purchases

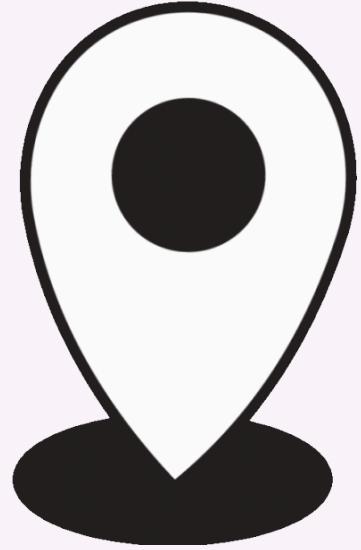
How does tax rate impact sales over different customer locations?



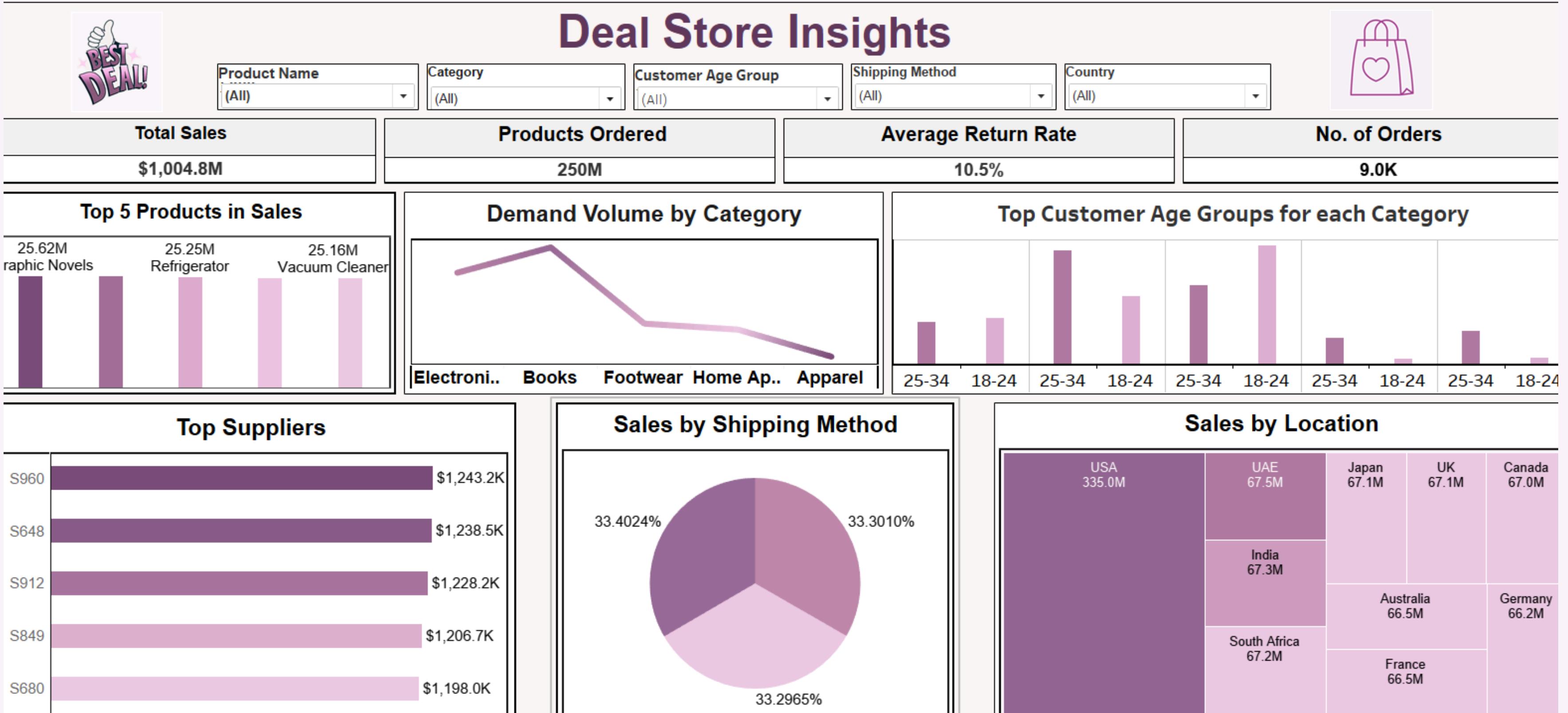
Customer Locati..

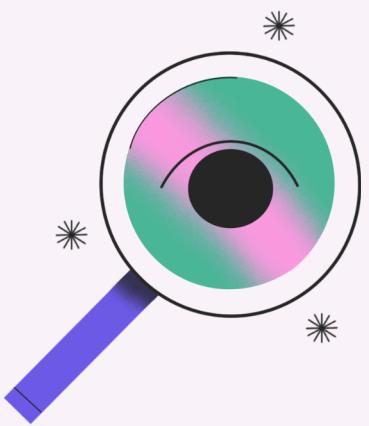


Week 4



Project Dashboard





The most Important Insights



1. Total Sales & Orders:

- The total sales amount to \$1,004.8M.
- 250M products were ordered.
- The total number of orders is 9.0K.



2. The average return rate is 10.5%, indicating a potential issue with product quality or customer satisfaction.



3. The top 3 products in sales are:

- Graphic Novels (25.62M)
- Refrigerator (25.25M)
- Vacuum Cleaner (25.16M)



4. Demand Volume by Category:

- Books have the highest demand, followed by Electronics.
- Apparel has the lowest demand, indicating potential underperformance in this category.



5. Customer Demographics.

- The 25-34 age group is the top customer segment across multiple categories.
- The 18-24 age group also has strong engagement, but less than 25-34.



6. Top Suppliers:

- The leading supplier is S960 with \$1,243.2K in sales.
- Other major suppliers (S648, S912, S849, S680) have similar performance, suggesting a competitive supplier market.



7. Sales by Location:

- The USA leads in sales with \$335M.
- Other top-performing countries include:
(UAE: \$67.5M ,Japan: \$67.1M , UK: \$67.1M,Canada: \$67M)
- This suggests strong international reach, with North America and Asia being key markets.

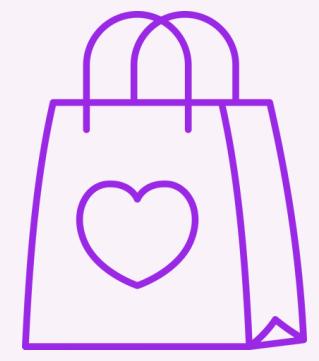


8. Sales by Shipping Method:

Sales are evenly distributed across different shipping methods, with no single method dominating.



Dashboard conclusion



- Books and Electronics drive the highest demand.
- The USA is the most significant market, but international markets contribute significantly.
- Customer demographics show that 25-34-year-olds are the primary buyers.
- Return rates are relatively high at 10.5%, which could indicate quality or satisfaction issues.

**BEST
DEAL!**



Thank
You!