Collegedunia - Data Analysis Assessment

Ajay Krishnan
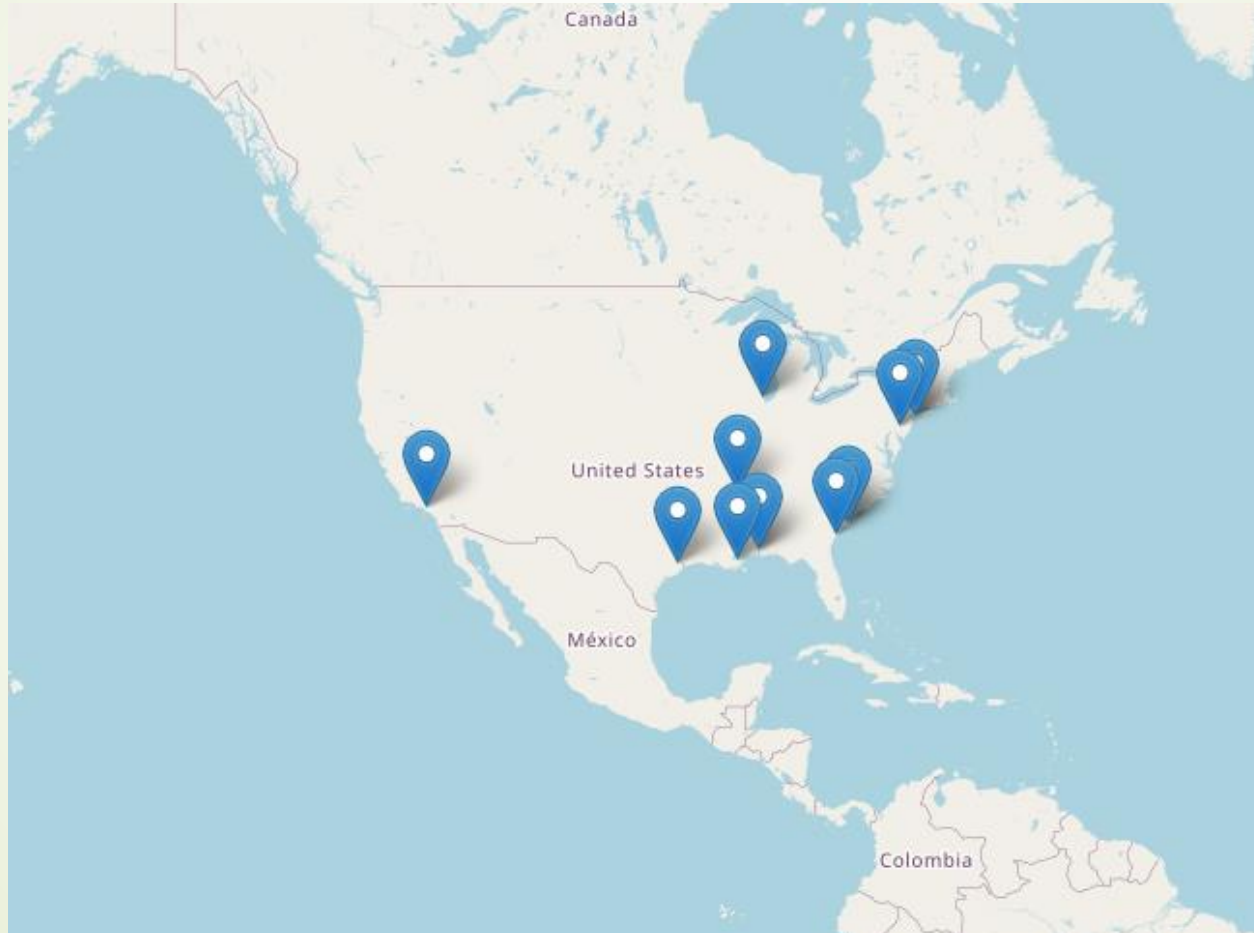
bsajaykrishnanat@gmail.com

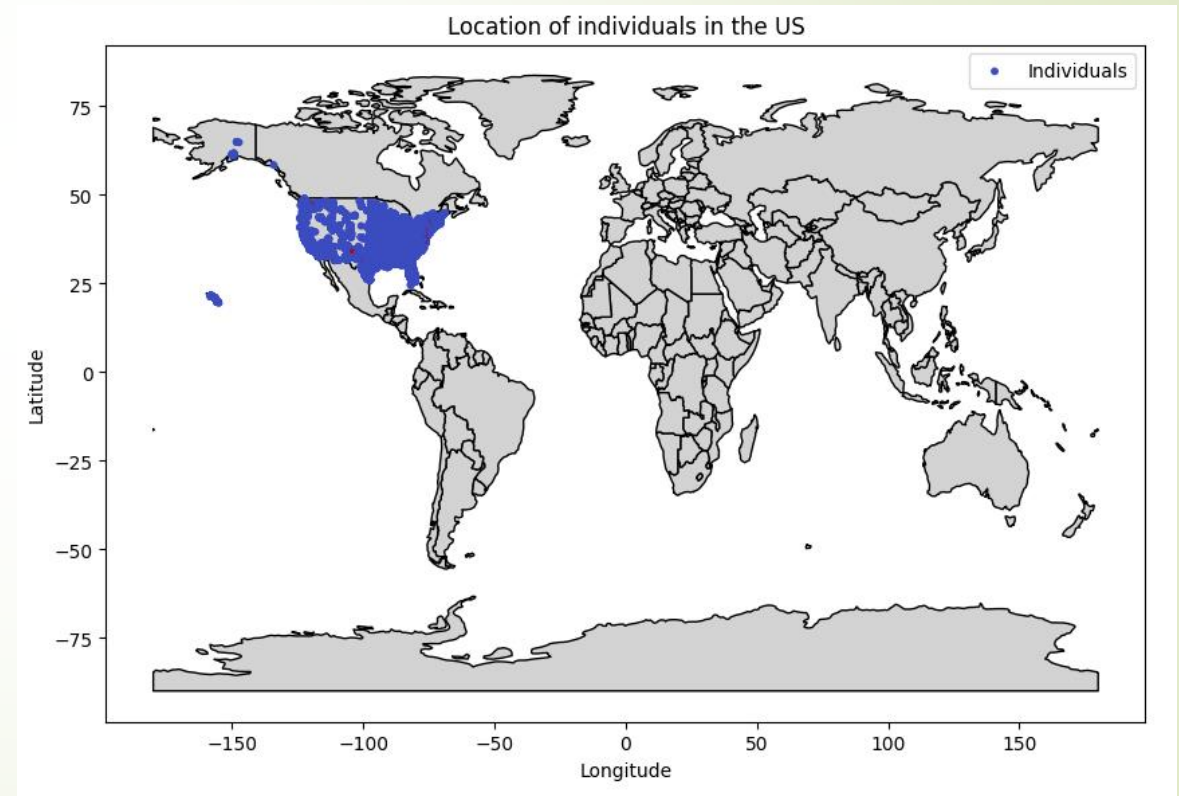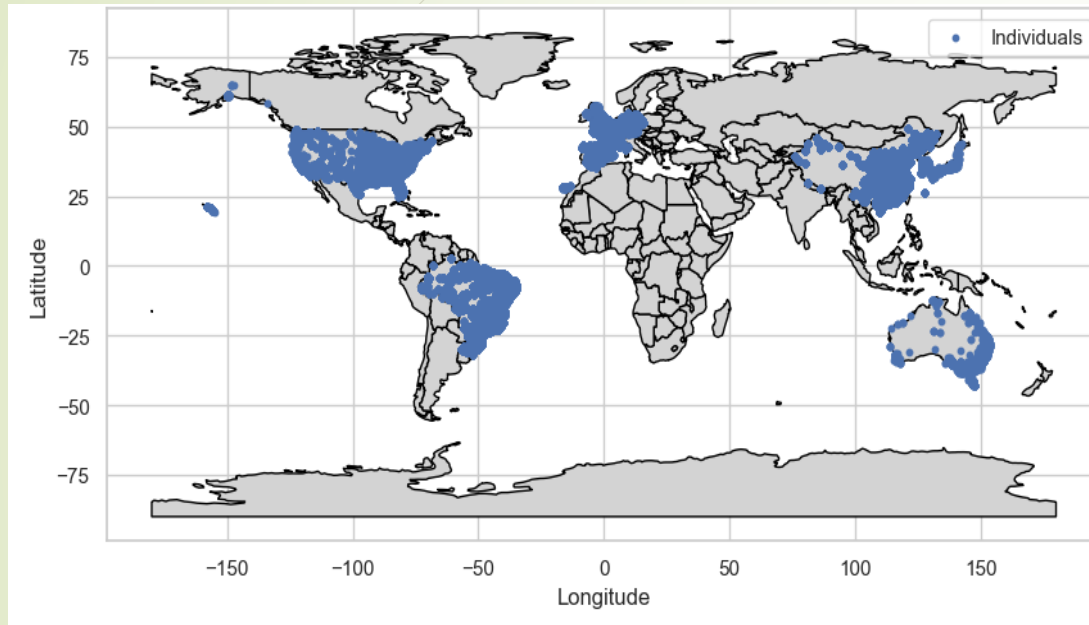LinkedIn :
https://www.linkedin.com/in/ajzkrish/

Geospatial Analysis

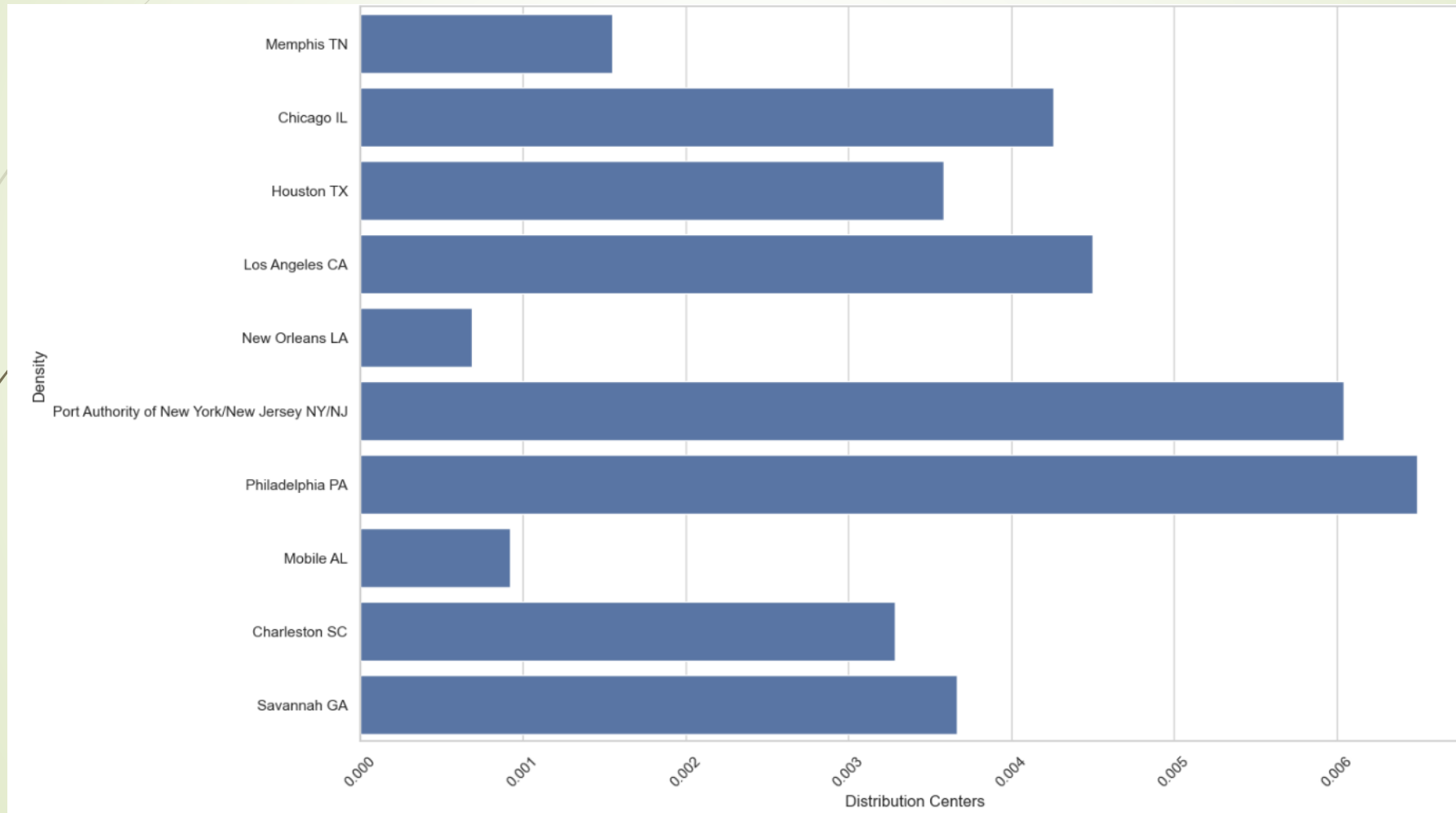# Distribution Centers located in the US
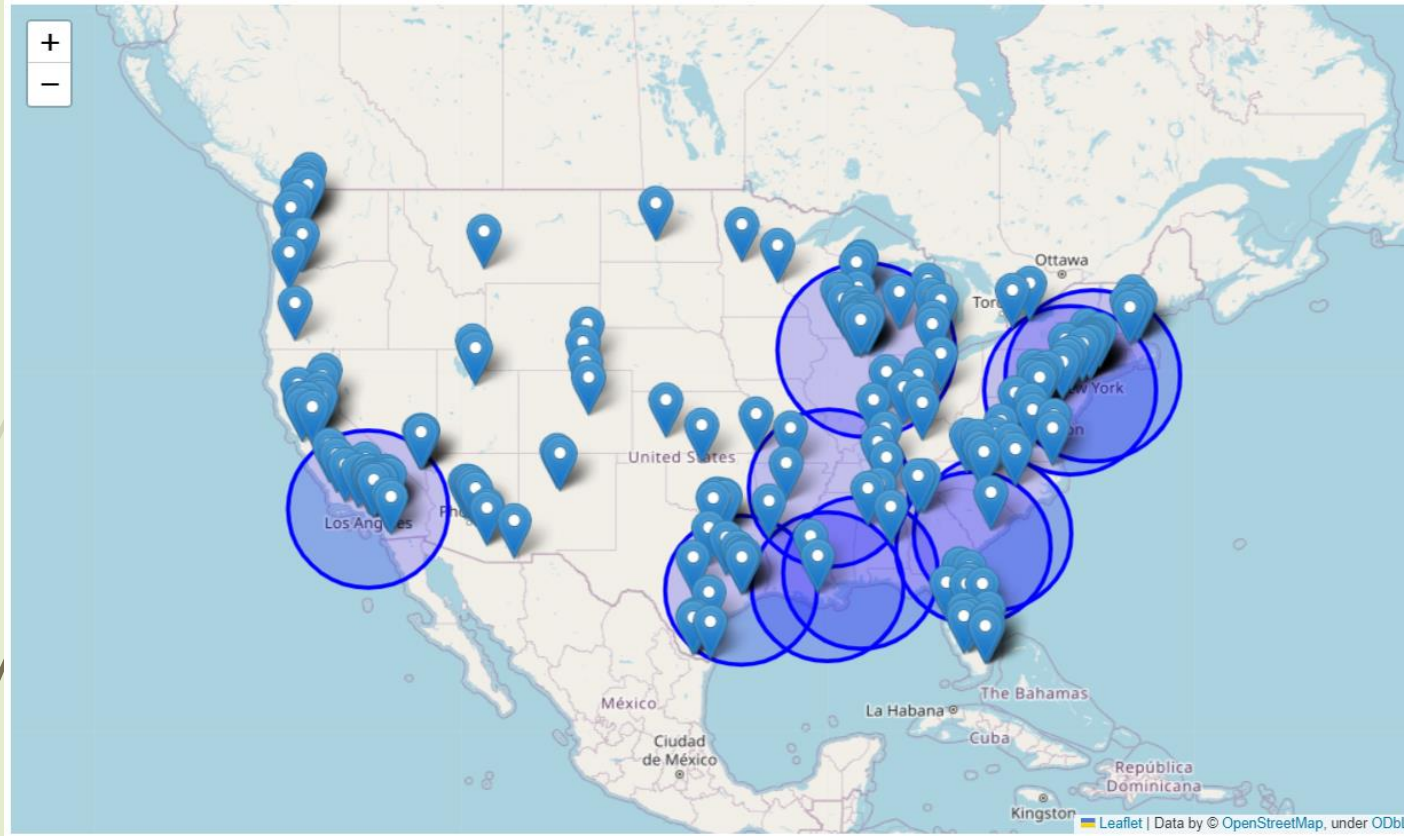
Distribution of users worldwide

Distribution of users in the US

Inference : "Port Authority of New York/New Jersey NY/NJ" and "Philadelphia PA" has the highest user density among all distribution centers

Users and Distribution centers

Inference
- The map clearly indicates that the distribution centers experience high population density in the Los Angeles area and across the eastern United States.

- Implying the need of more distribution centers in remaining parts (north western parts, Hawaii)

User Behaviour Analysis

```
1  events.isnull().sum() * 100/len(events)
```

```
id                   0.000000
user_id             46.286518
sequence_number      0.000000
session_id           0.000000
created_at           0.000000
ip_address           0.000000
city                 0.949028
state                0.000000
postal_code          0.000000
browser              0.000000
traffic_source       0.000000
uri                  0.000000
event_type           0.000000
dtype: float64
```
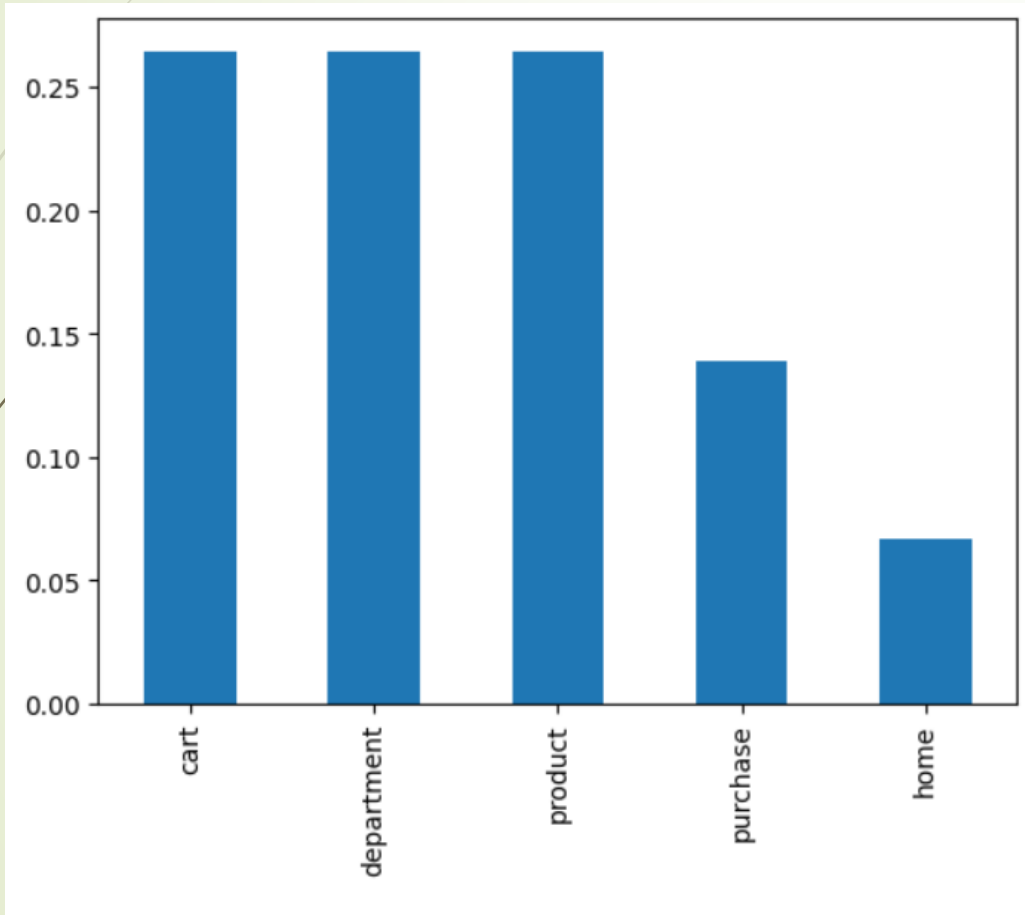
Inference
- Due to 46% of 'user_id' values being null, dropping them is advisable.
- For 'city' column imputation, opting for the mode is simpler and more practical considering computational constraints.
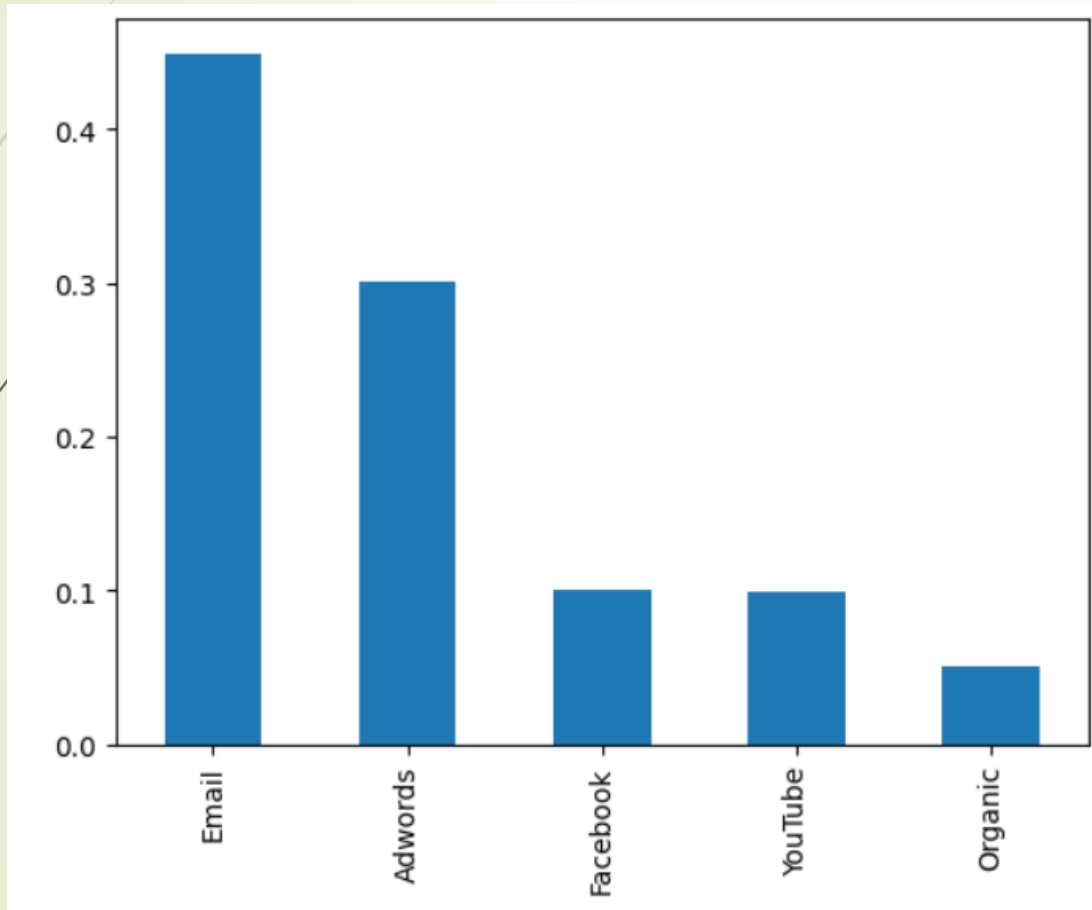
Browser Usage
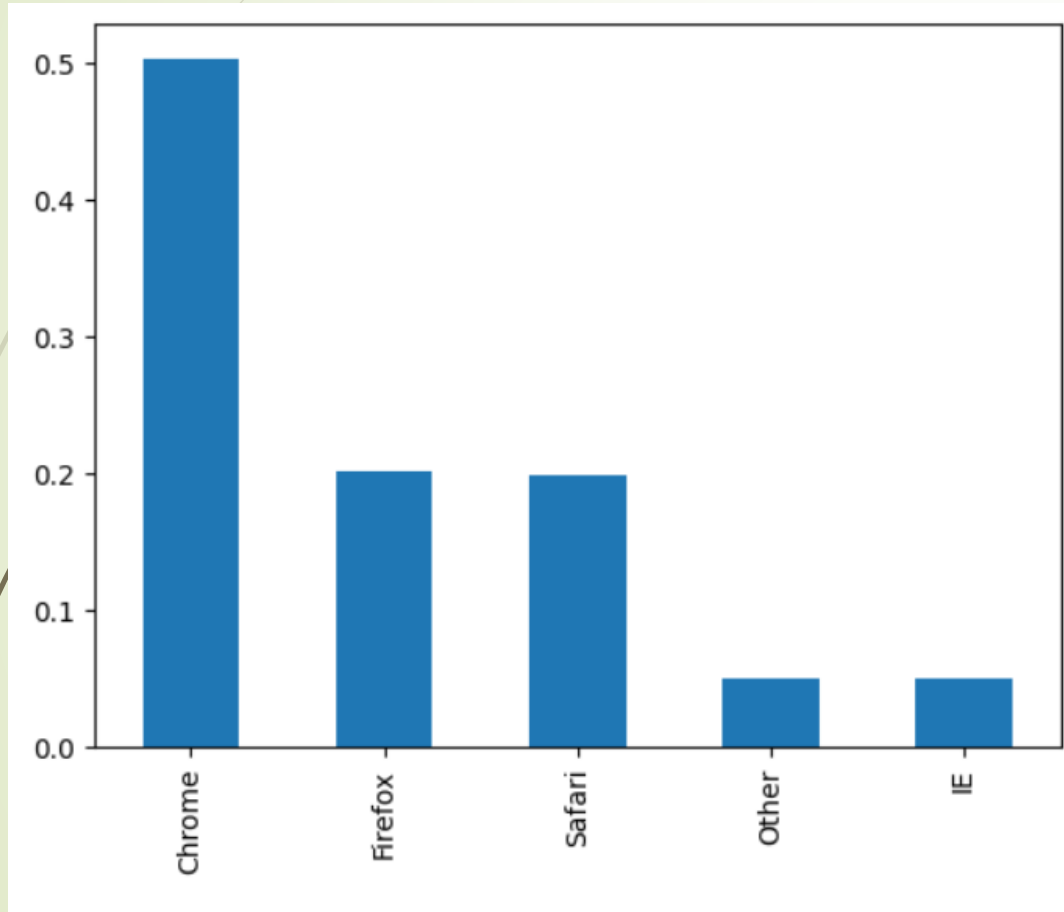
Inference : Chrome is the most used browser (followed by firefox and safari), hence optimising the site for the browser would be fruitful

Event Type

Inference: The purchase percentage is relatively low, suggesting that many individuals are browsing products, adding them to their carts, but not completing the purchase.

Traffic Source

Inference
- High number of traffic is from emails and adwords
- Pushing a little bit more on facebook traffic would be beneficial considering that the number of people using facebook is high.
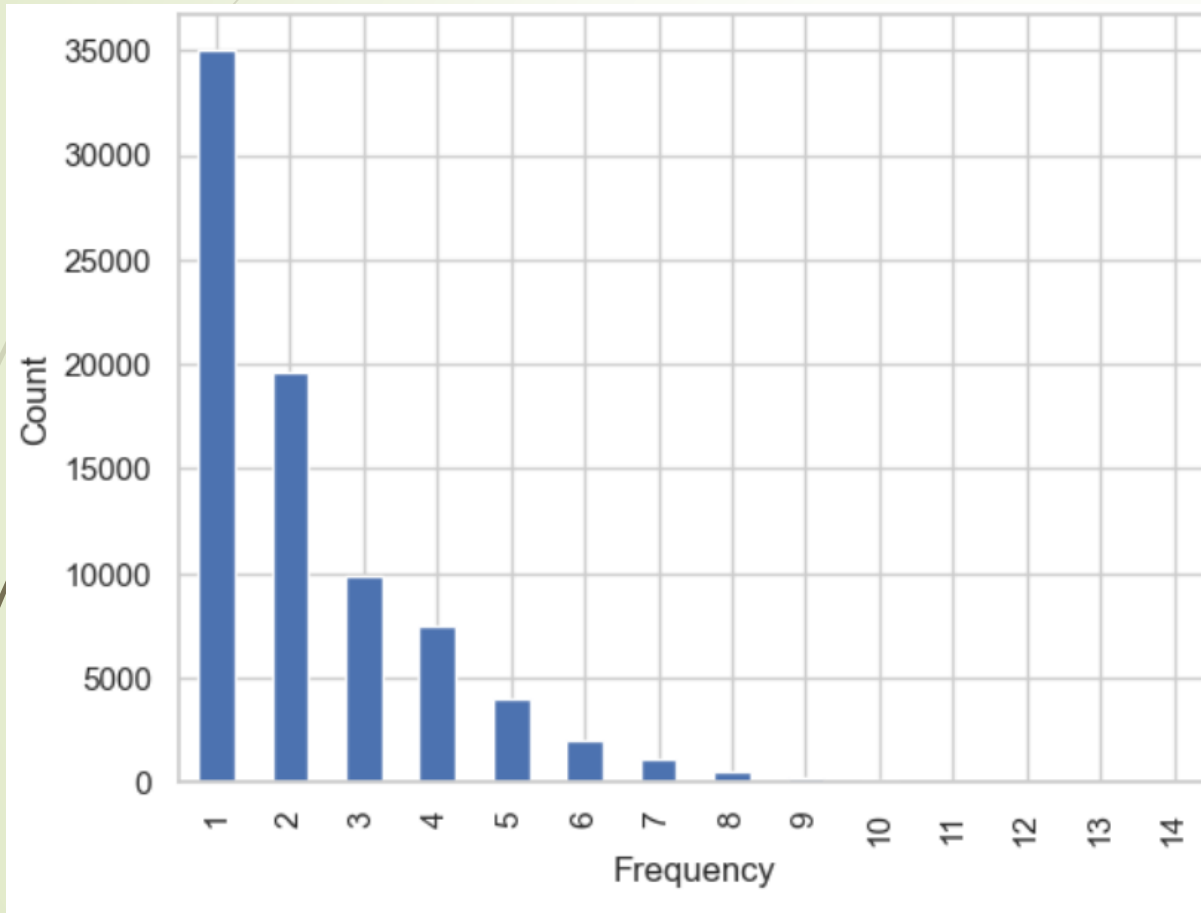
Session Duration
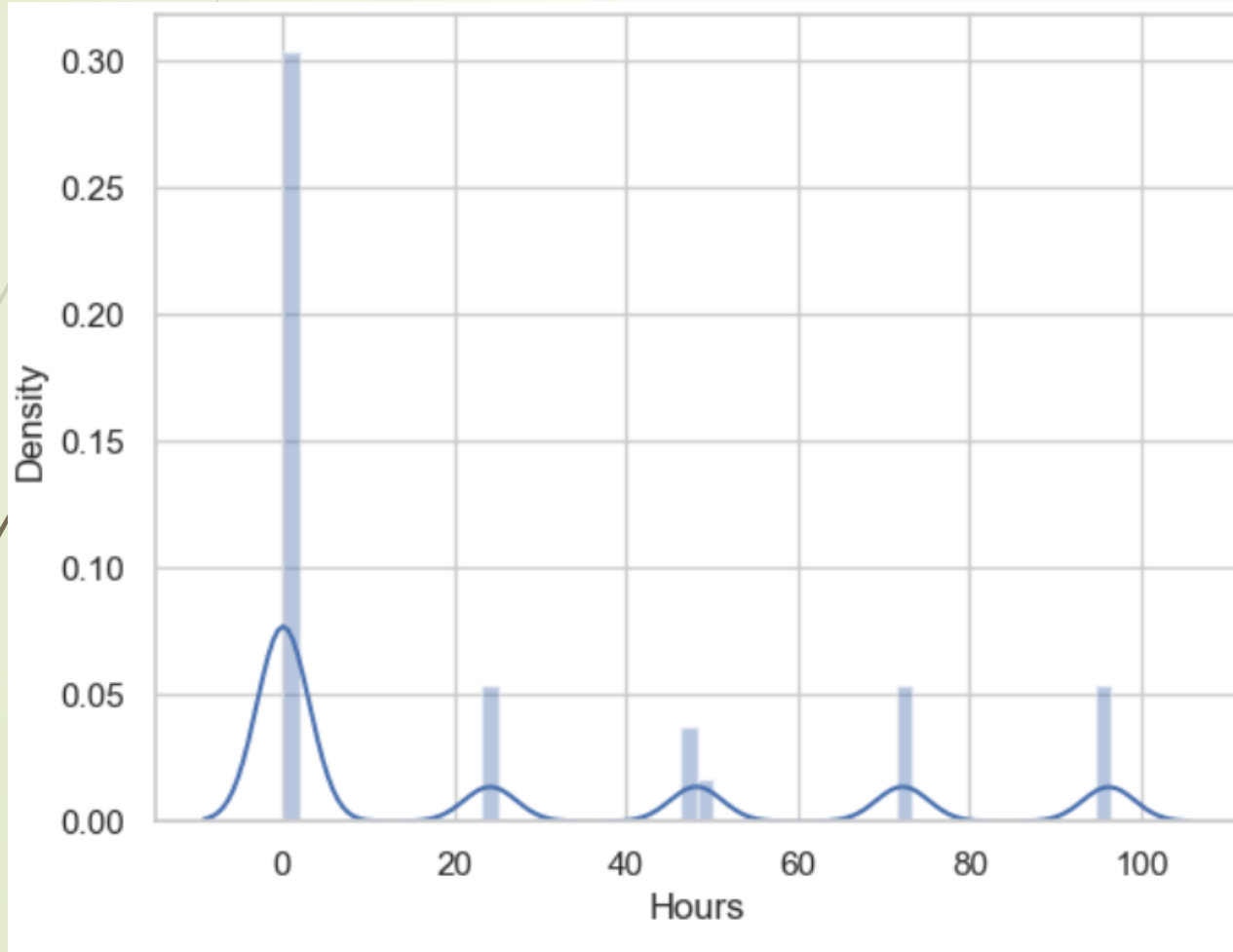
Inference

- Approximately 70% of whole sessions are only less than 25 hours

- Approximately 60% of whole sessions are only less than 0.5 hours

- Approximately 10% of whole sessions are only more than 10 hours

Session Frequency

Inference:
The inference from a high number of users having a session frequency of 1 is that a significant portion of users visit the platform or website only once. This suggests a low level of engagement or interaction beyond the initial visit.

This indicates :
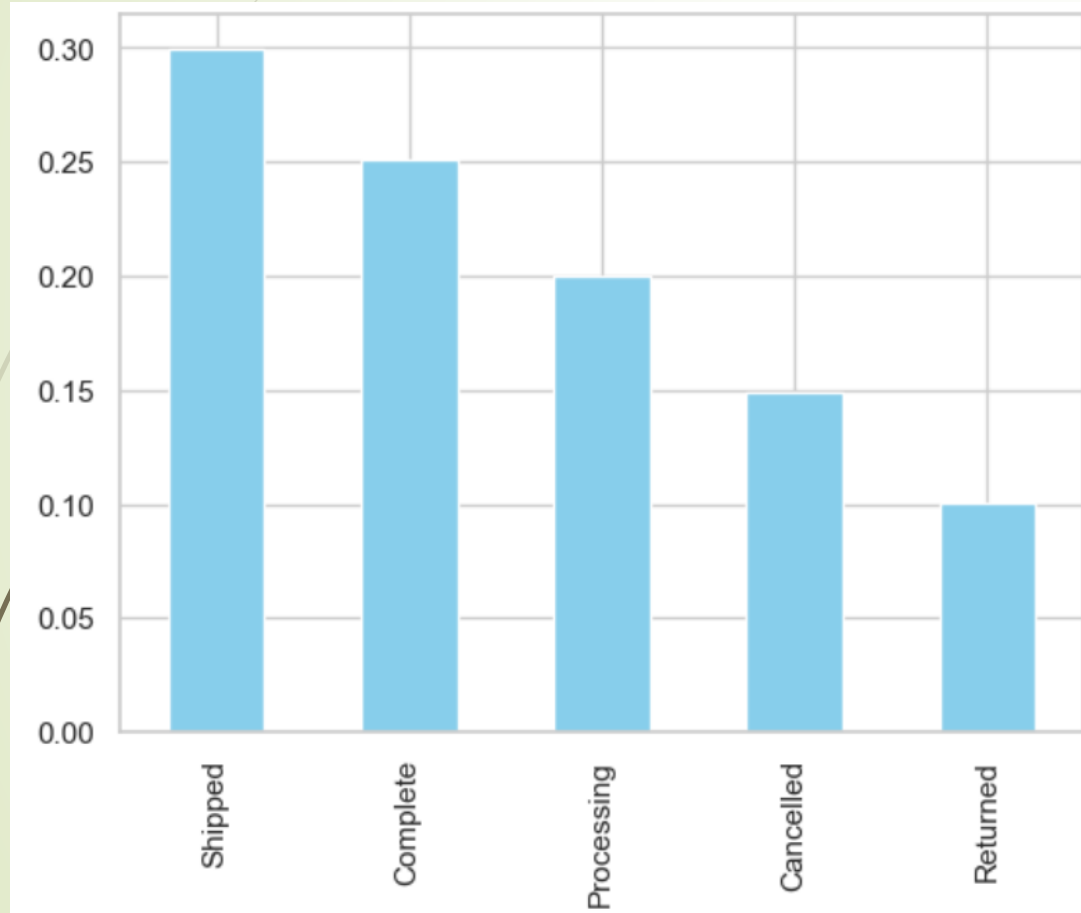Lack of Retention Strategies
User Experience Issues
Users may have visited the platform to make a one-time transaction

Sales and Revenue Analysis

Order Status

Inference
- 55% of orders are either shipped or completed and 10% returned.
- 10% return rate suggests a noteworthy proportion of orders experiencing dissatisfaction or issues significant enough for customers to initiate returns.

Distribution of Sale Prices

Distribution of Sale Prices
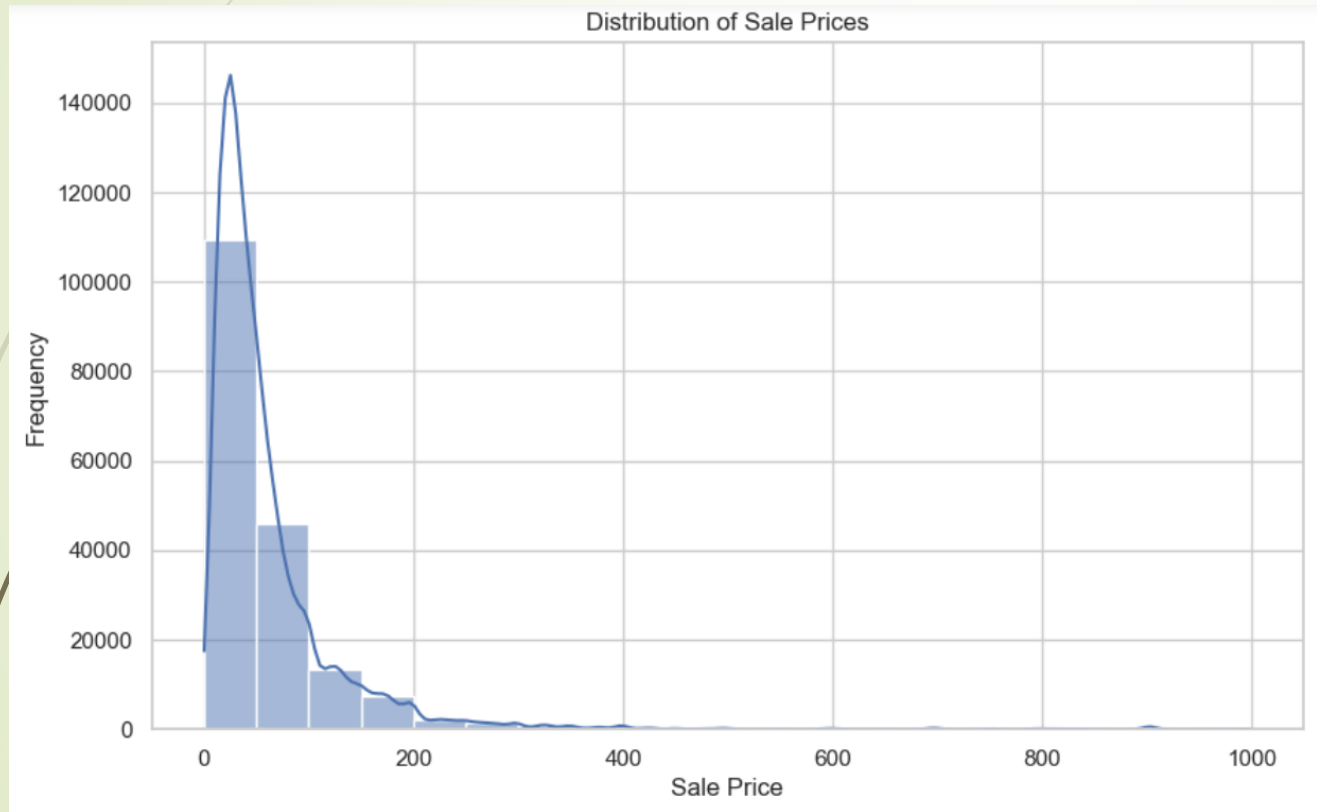
Inference
- Very high proportion of data is priced less than 200.
- This may suggest a market segment or consumer preference for budget-friendly products within the business's product range.

Distribution of Orders Created by Month

Distribution of Orders Created by Month
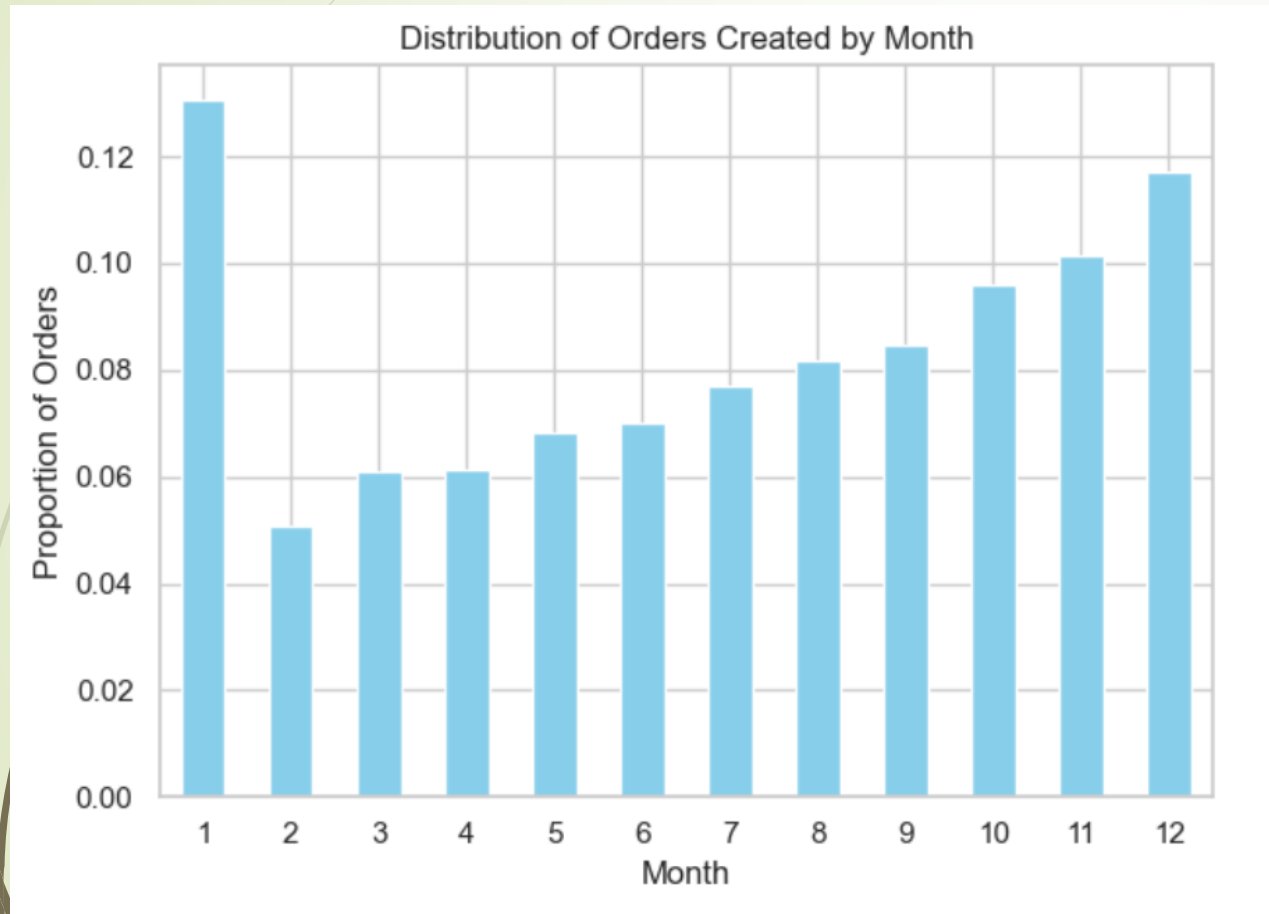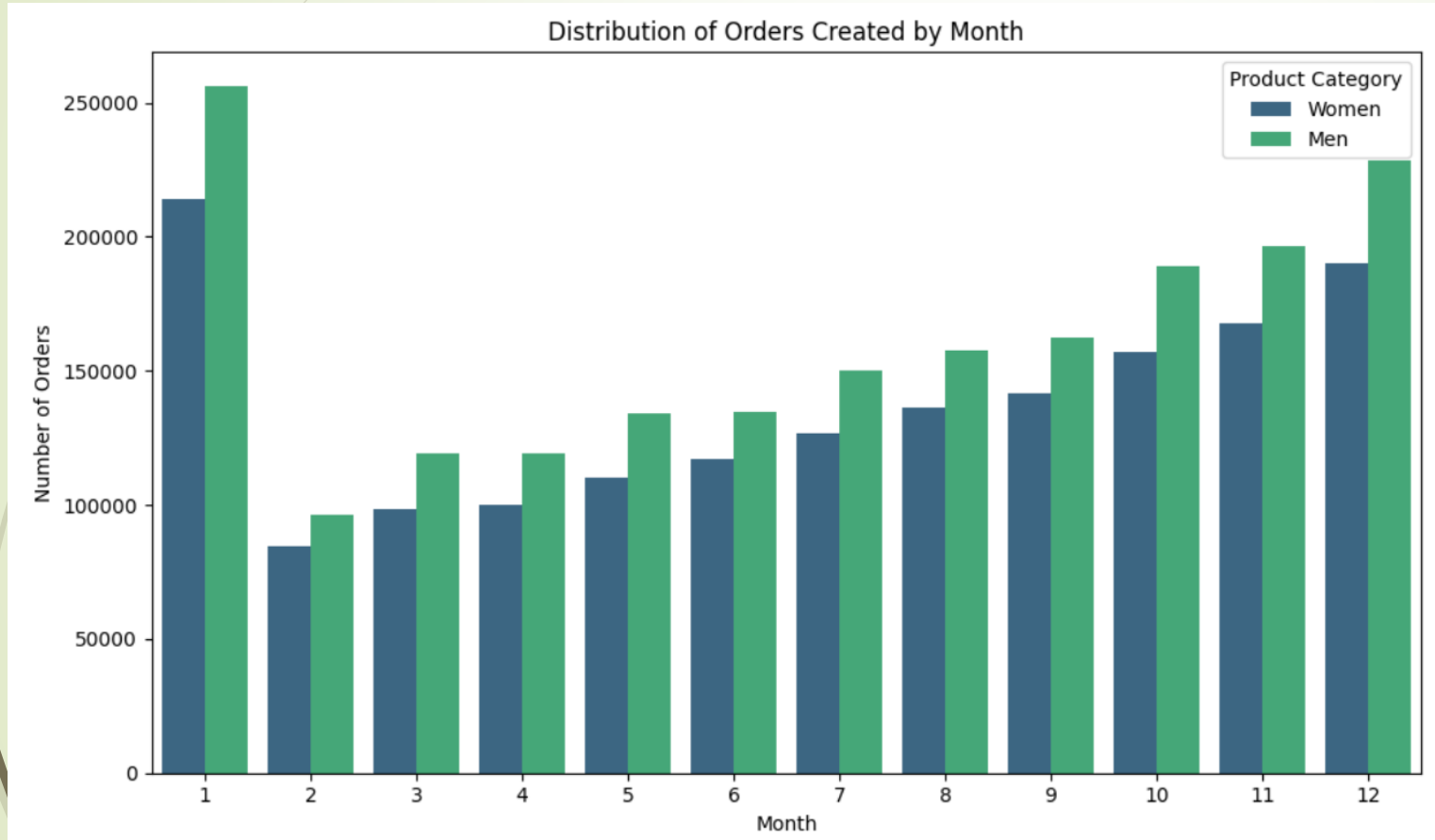
Inference
- Highest number of order are Placed in the winter seasons (nov-jan)
- Understanding and anticipating seasonal fluctuations in demand can be valuable for businesses in planning inventory management, marketing campaigns, and resource allocation to effectively capitalize on peak periods and maximize sales opportunities.

Distribution of Orders Created by Month (Partitioned by gender)

Inference
- In all months, product_department - men have higher number of purchase.
- Businesses can leverage this insight to optimize product offerings, tailor marketing strategies, and enhance customer engagement efforts.

Distribution of Product Categories
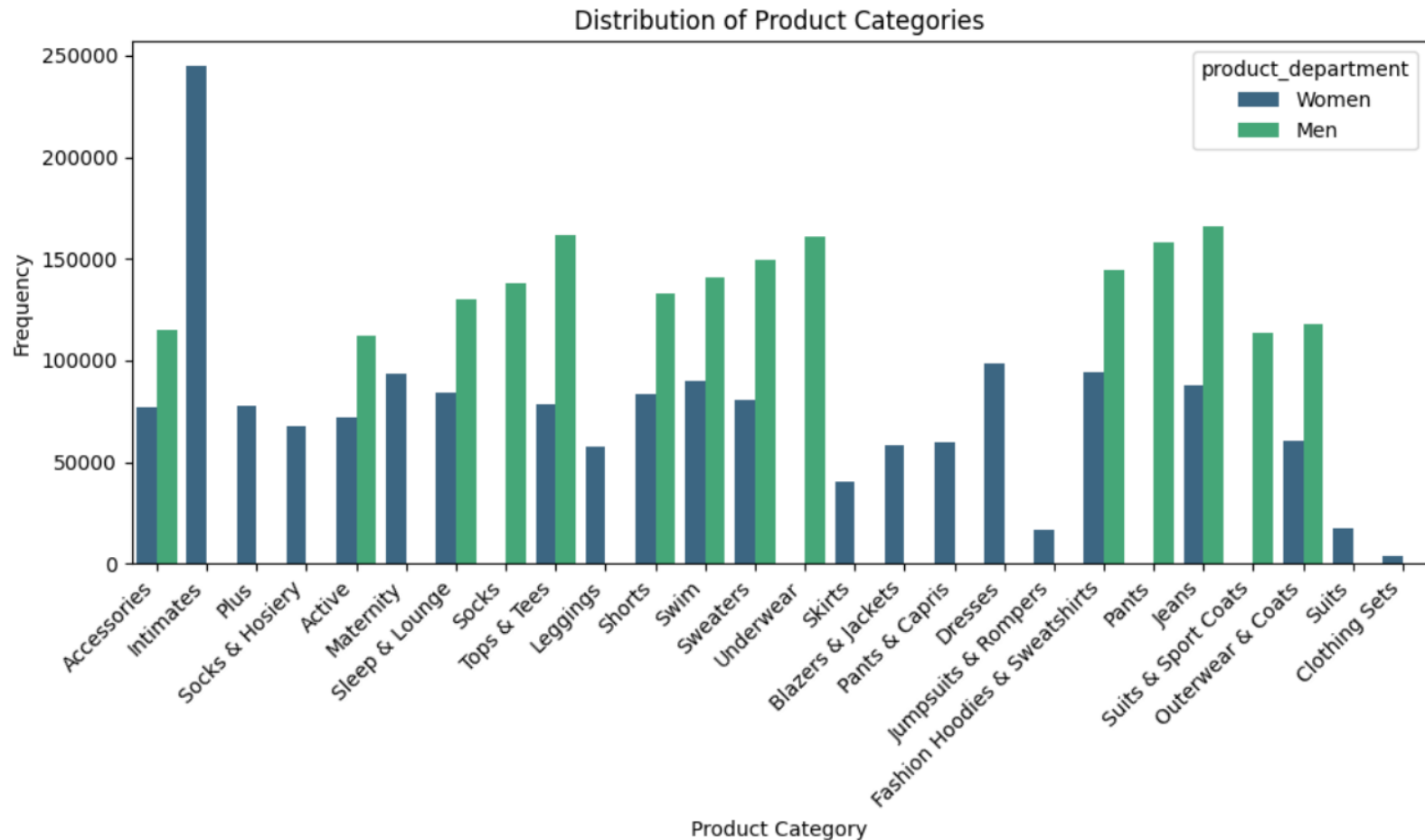
Inference
- There is a notable number of products that are exclusive for women
- The presence of a notable number of products exclusively for women highlights a targeted approach to product segmentation and catering to specific consumer demographics.

Details of revenue from the orders
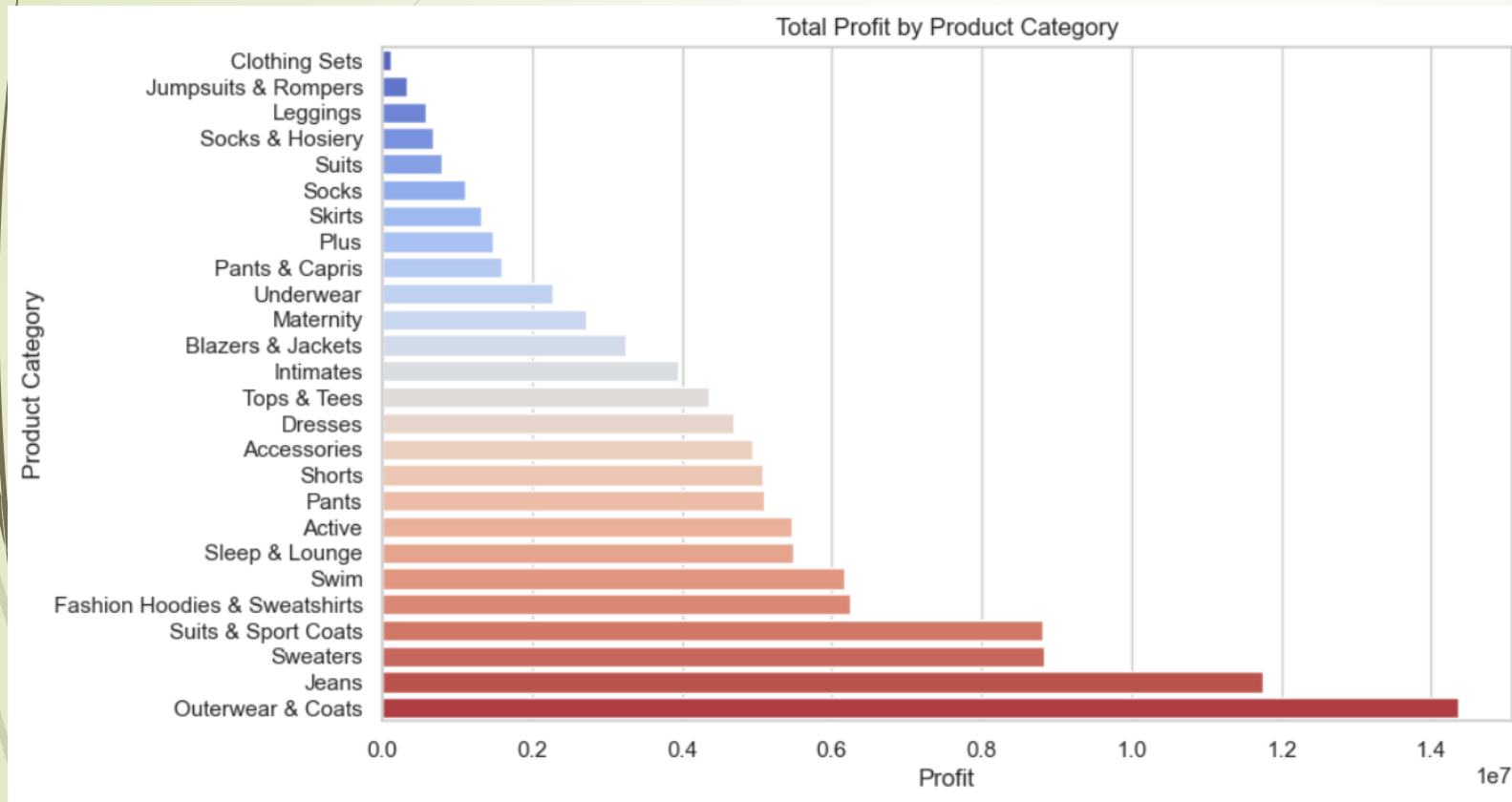
- **Total Revenue made from all the orders is 111559298.7420401**
- **Average revenue per sale 31.109841000841918**

Distribution of profit from each product Categories


Total Profit by Product Category

Inference
- Over coats jeans and sweaters has made highest profit and jumpsuit, clothing sets made lowest profit.
- Businesses can leverage this insight to optimize product offerings, adjust pricing strategies, and allocate resources effectively to maximize profitability and drive overall business performance.

Product Performance Analysis

# Distribution of Product price
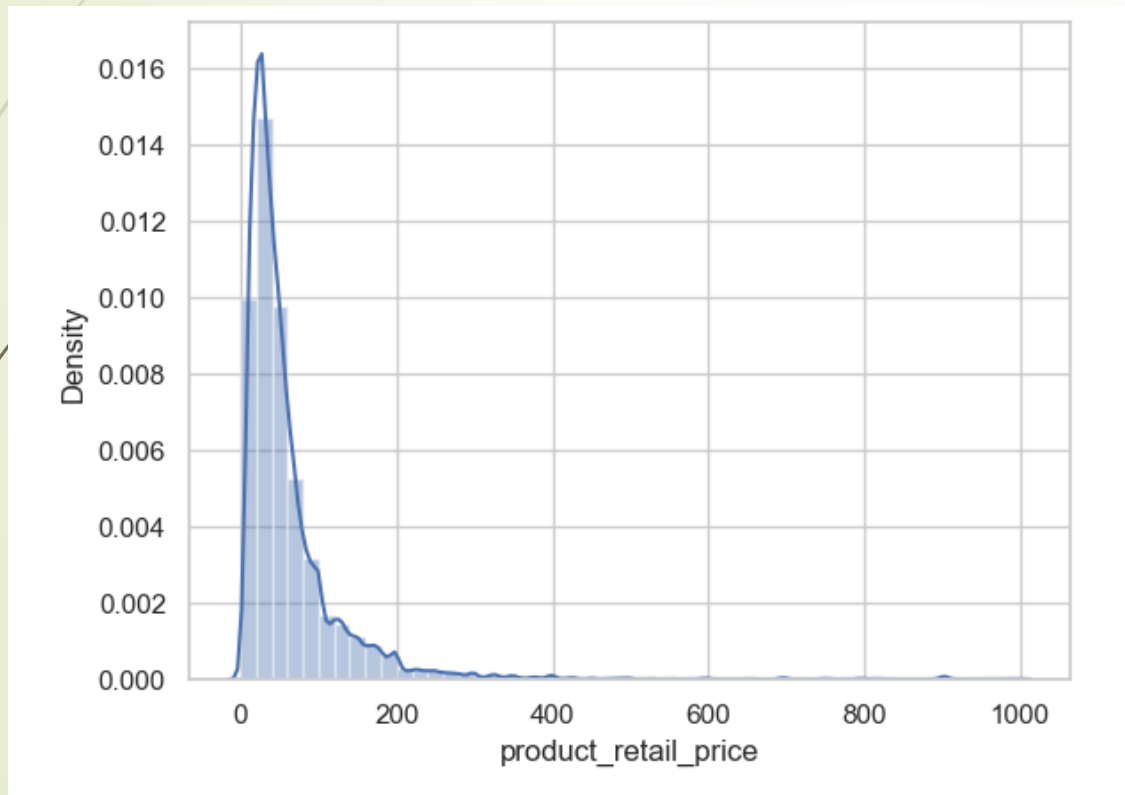
Inference
- A heavily left-skewed distribution of selling prices in the dataset suggests that a majority of products are priced lower than the average or median price.
- This could indicate a pricing strategy focused on affordability, potentially aiming to capture a larger market share by appealing to price-sensitive consumers.

Average Profit by Product Category

Volume of products ordered by Product Category
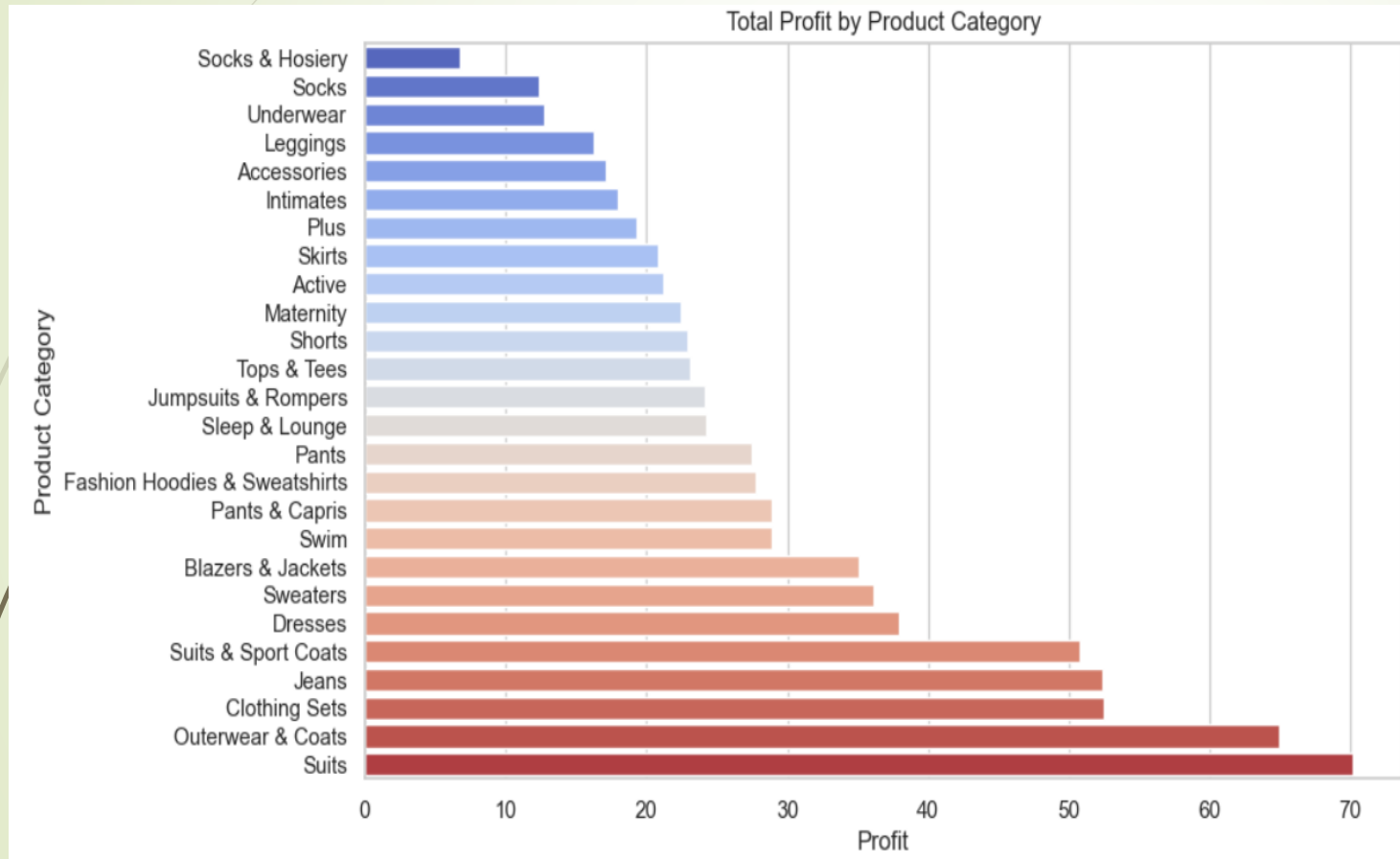
Volume of products ordered by Product Category
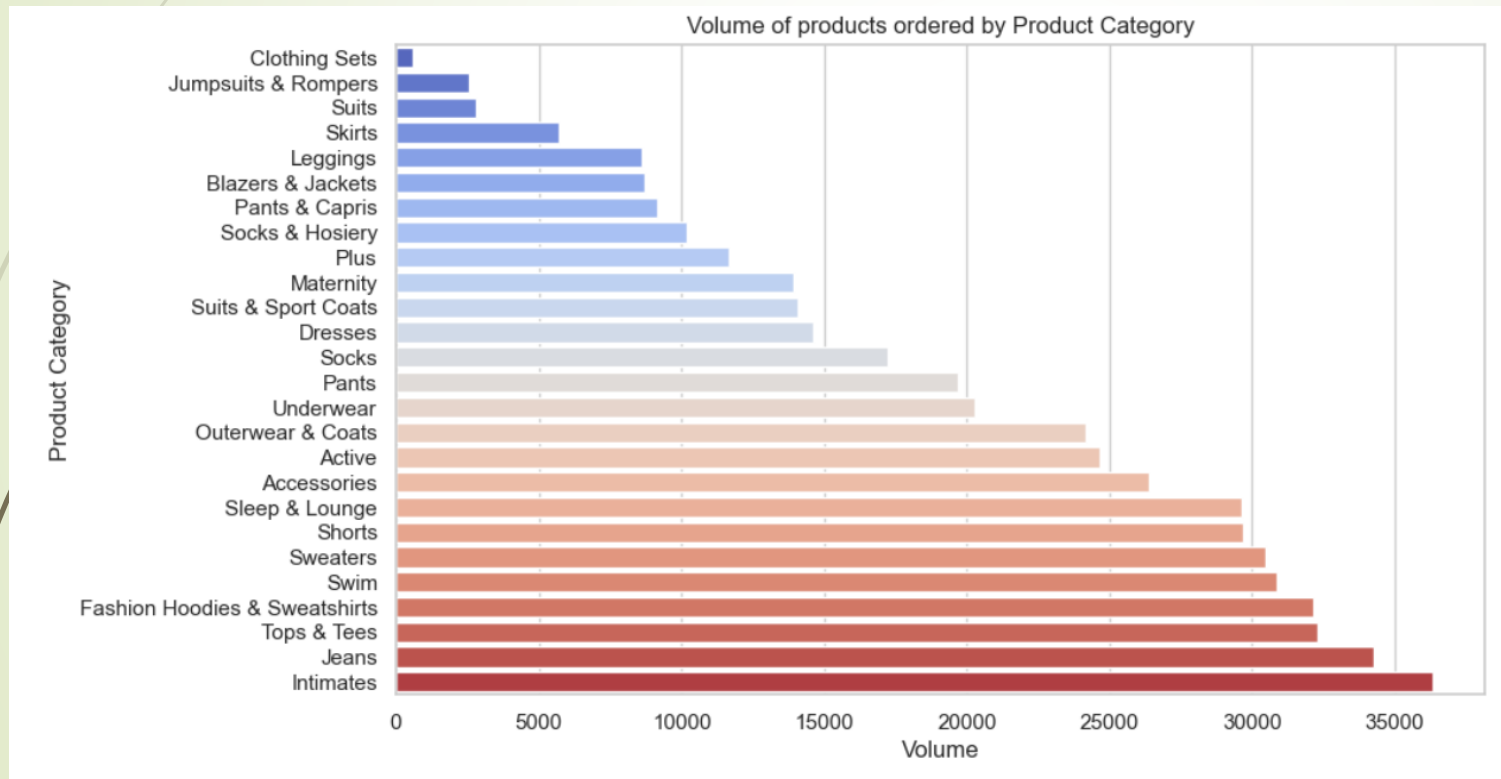
Inference
- Highest volume of products in inventory are from intimate,jeans,tops and tees category
- By focusing on these high-volume product categories, businesses can enhance customer satisfaction, increase repeat purchases, and ultimately, improve overall profitability.

User Demographics Analysis

Distribution of users based on gender

Distribution of users based on gender

F

50.2%

49.8%

M
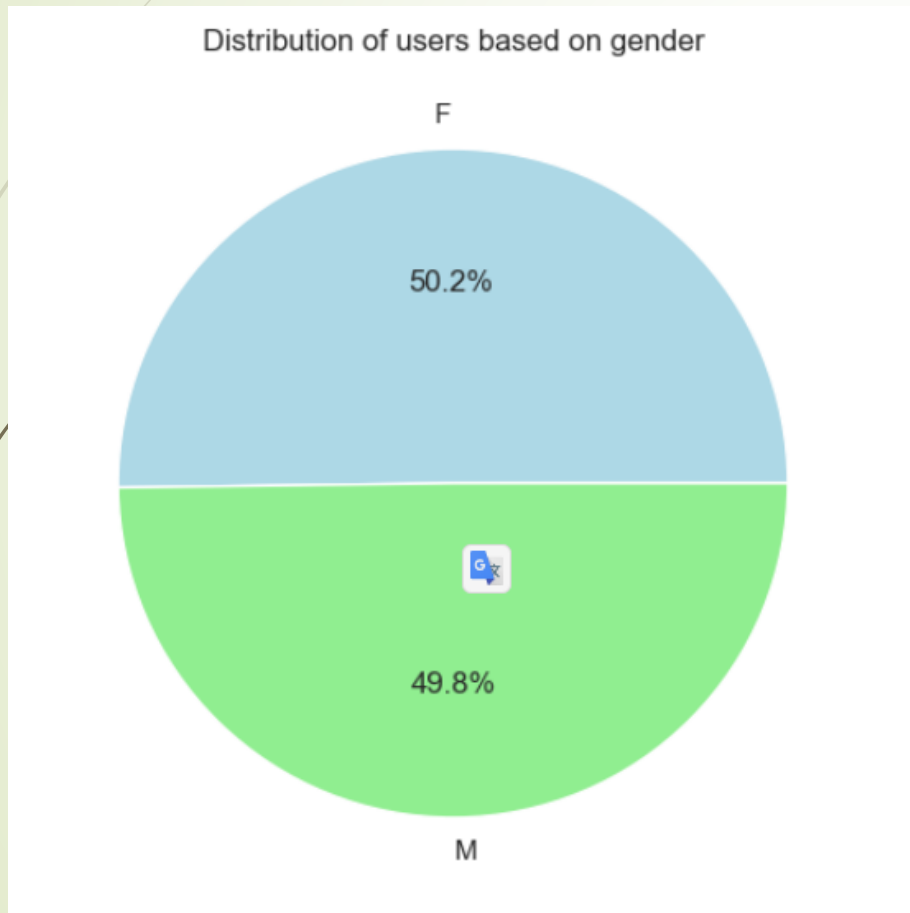
Inference
- The male-female ration is approximately 50-50.
- This suggest the active engagement from both genders.

Distribution of users based on country

Distrubution of users based on country

Inference
- Low user numbers in Austria, Spain, Colombia, and Poland indicate potential for market expansion in these regions.
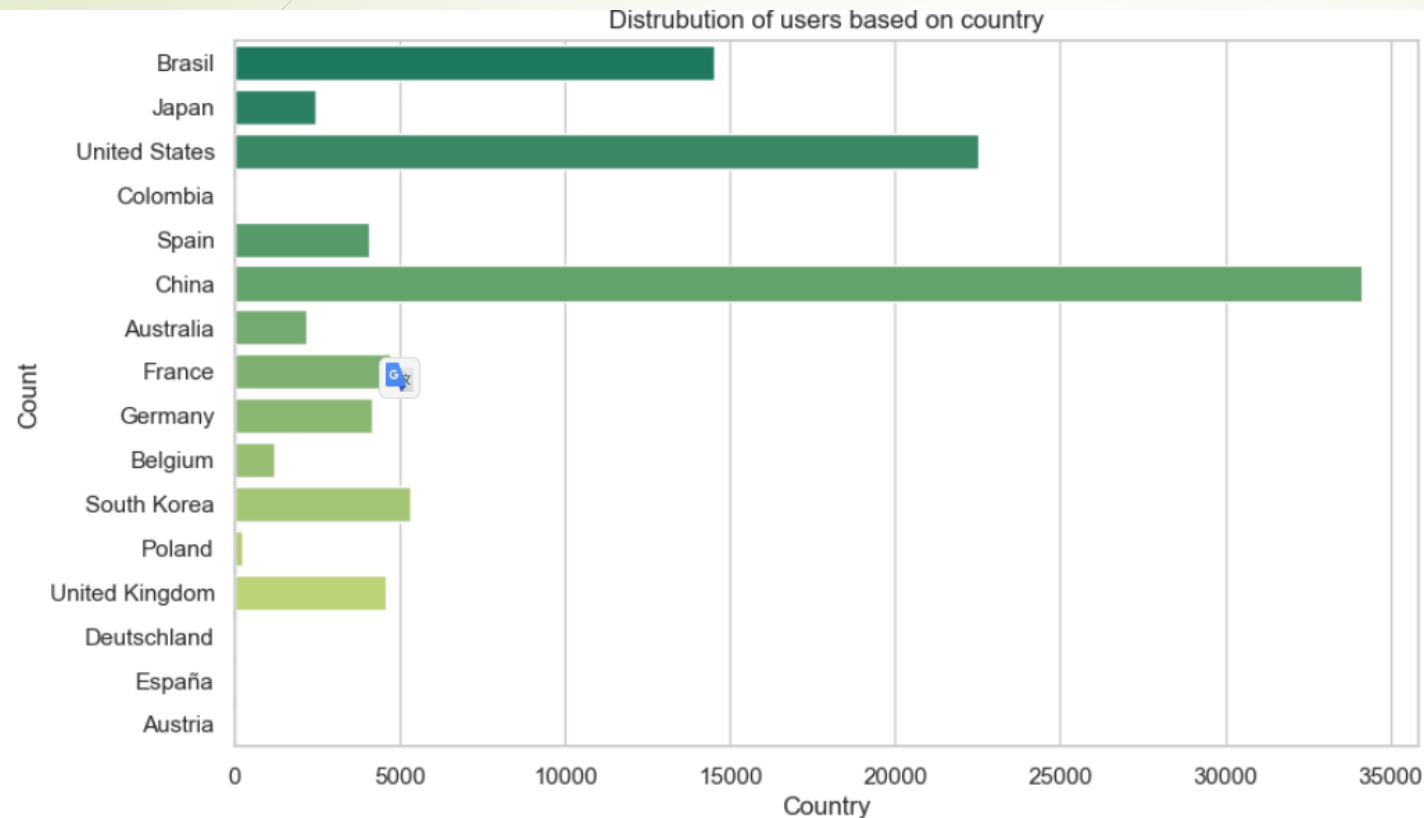
Inference : Highest number of users are from china

Distribution of traffic source in each country

Inference

- In most countries main contributor to traffic_source is 'search'

Distribution of gender in each country

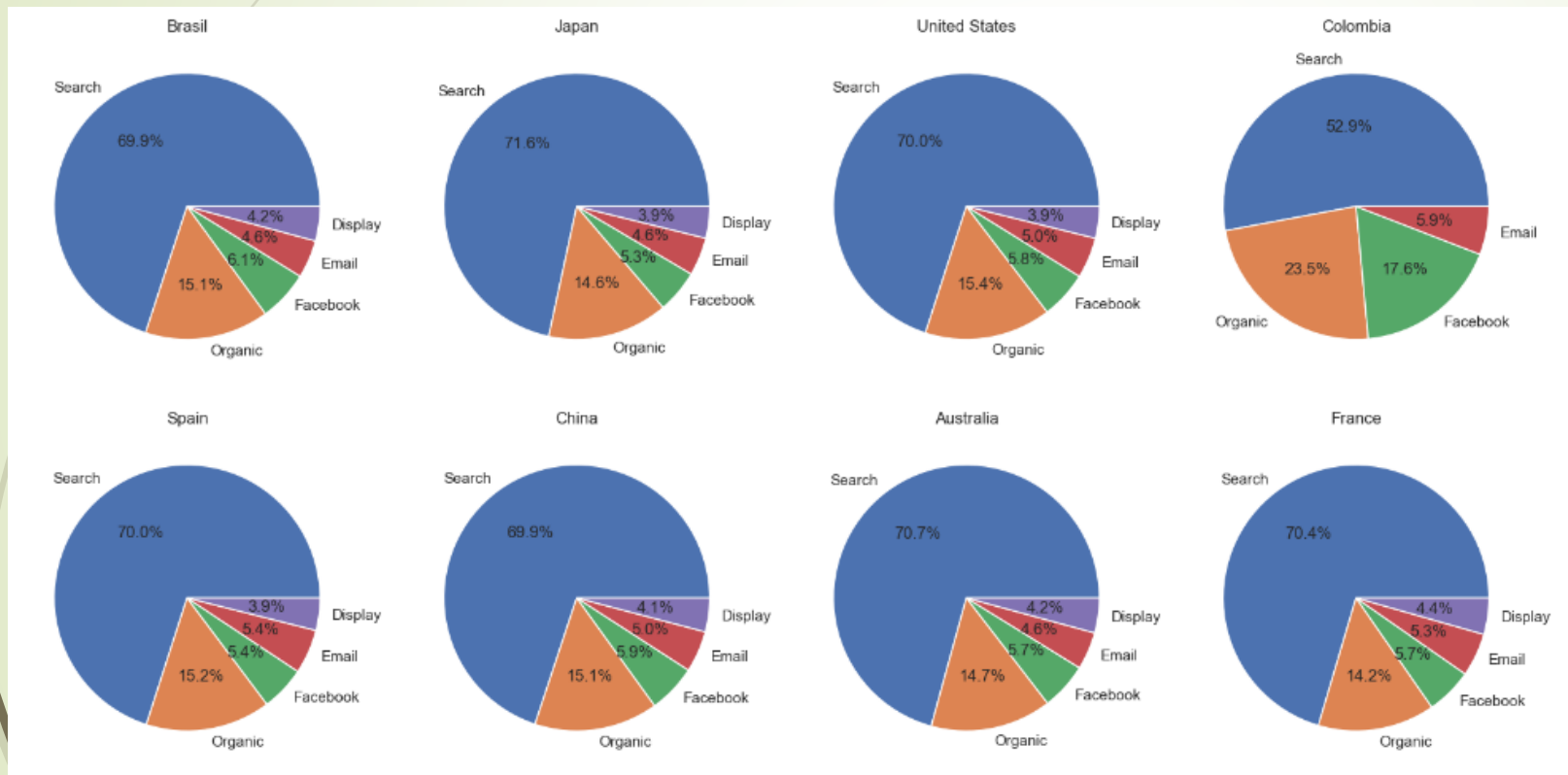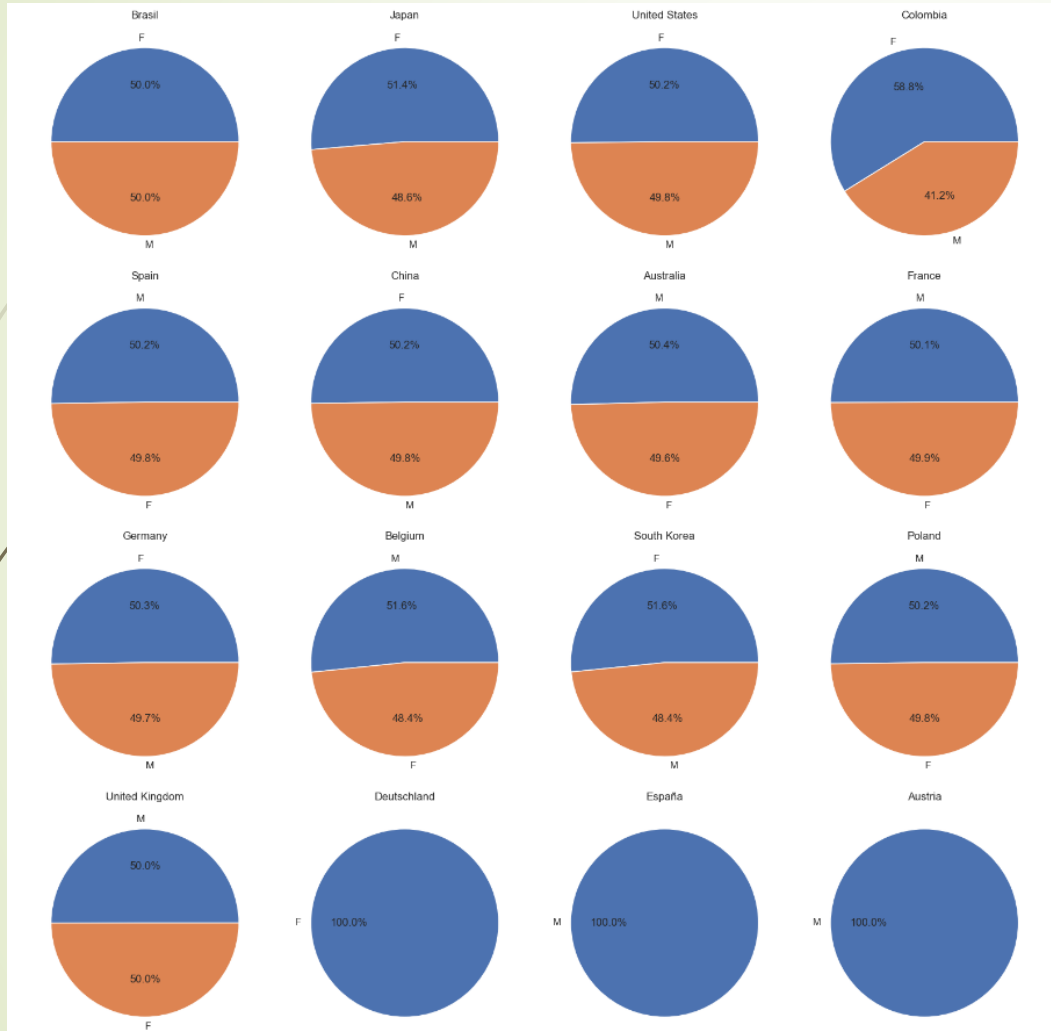

Inference
- In the majority of countries, the male-female ratio is balanced, except in Germany, Spain, and Austria where it is entirely dominated by one gender (but the number of users is very low).
- This could indicate varying cultural, social, or economic factors influencing gender participation or representation in different regions.

Order Fulfilment Analysis

Order vs number of items

Most of the order has only 1 item (70%)

Inference

- The majority of orders consisting of only one item may indicate a pattern of individual or single-item purchases rather than bulk orders or larger transactions.
- This insight suggests potential opportunities for strategies aimed at increasing order size or encouraging customers to purchase complementary products.

Shipping, Delivery and Return duration

Inference

- Average Shipping Duration: 13 Hours

  - Orders are swiftly processed and dispatched within approximately 12 hours, reflecting efficient logistics.

- Average Delivery Duration: 2.5 Days

  - Customers typically receive their orders in about 2 days and 12 hours, showcasing relatively prompt delivery times.

- Average Return Duration: 1.5 Days

  - Returns are efficiently processed within approximately 1 day and 12 hours, ensuring a smooth customer experience.
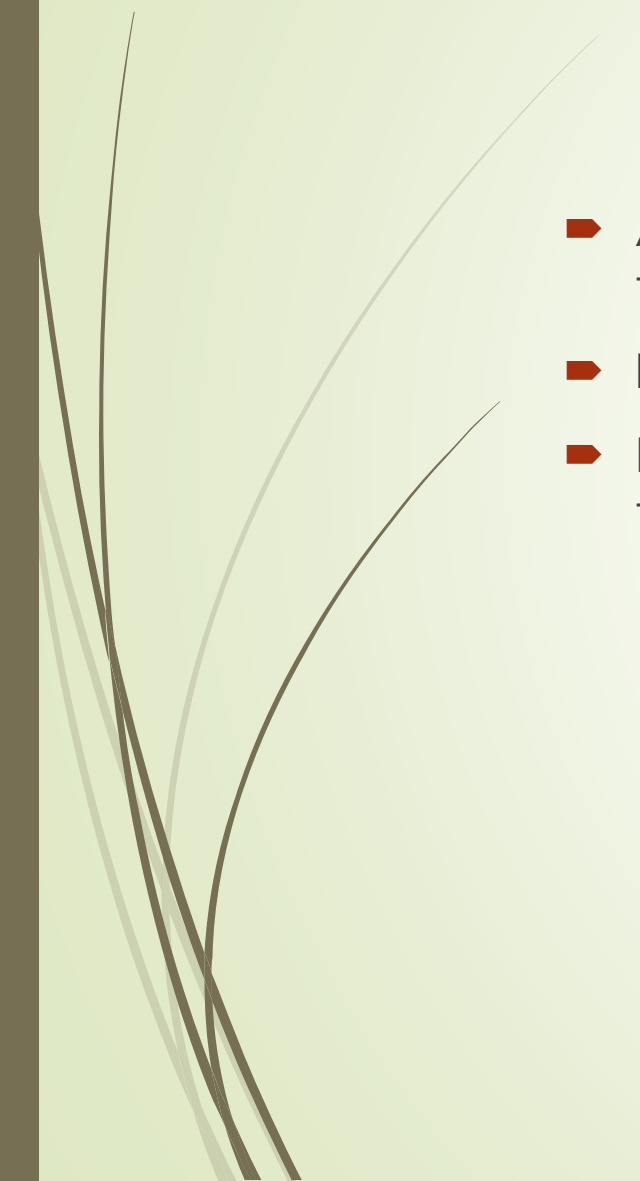
Top Product from ordered (number of orders).

name	category
Michael Kors Men's 3 Pack Brief	Underwear
Mens Print Hot Body Boxer Swimsuit Gary Majdel...	Swim
Haggar Men's Tonal Stria Pleat Front Cuff Dres...	Pants
Thorlo Men's Lt Walking Mini Crew Sock	Socks
Joe's Jeans Men's Rebel Relaxed Fit Jean	Jeans
Wrangler Rugged Wear Men's Unlined Denim Jacke...	Outerwear & Coats
Knothe - Mens Waffle Weave Robe Tan 29649	Sleep & Lounge
Bayside Apparel Adult USA-Made Long-Sleeve Poc...	Tops & Tees
Elegant Men's Two button Three piece Strip Suit	Suits & Sport Coats
Life is Good Men's King Of The Grill Short Sle...	Active

These are the top product ordered

Insights and suggestions

- Address the significant proportion of canceled and returned items through targeted sales strategies and consumer engagement.

- Improve average delivery times to enhance overall customer satisfaction.

- Encourage bulk ordering and promote complementary items to increase the number of items per order and maximize sales potential.

- END

- **Ajay Krishnan**
- bsajaykrishnanat@gmail.com
- LinkedIn : https://www.linkedin.com/in/ajzkrish/
- Github : https://github.com/ajay-ajzkrish