AJIO

Fashion Collections using R

Created by -VRINDA GOEL

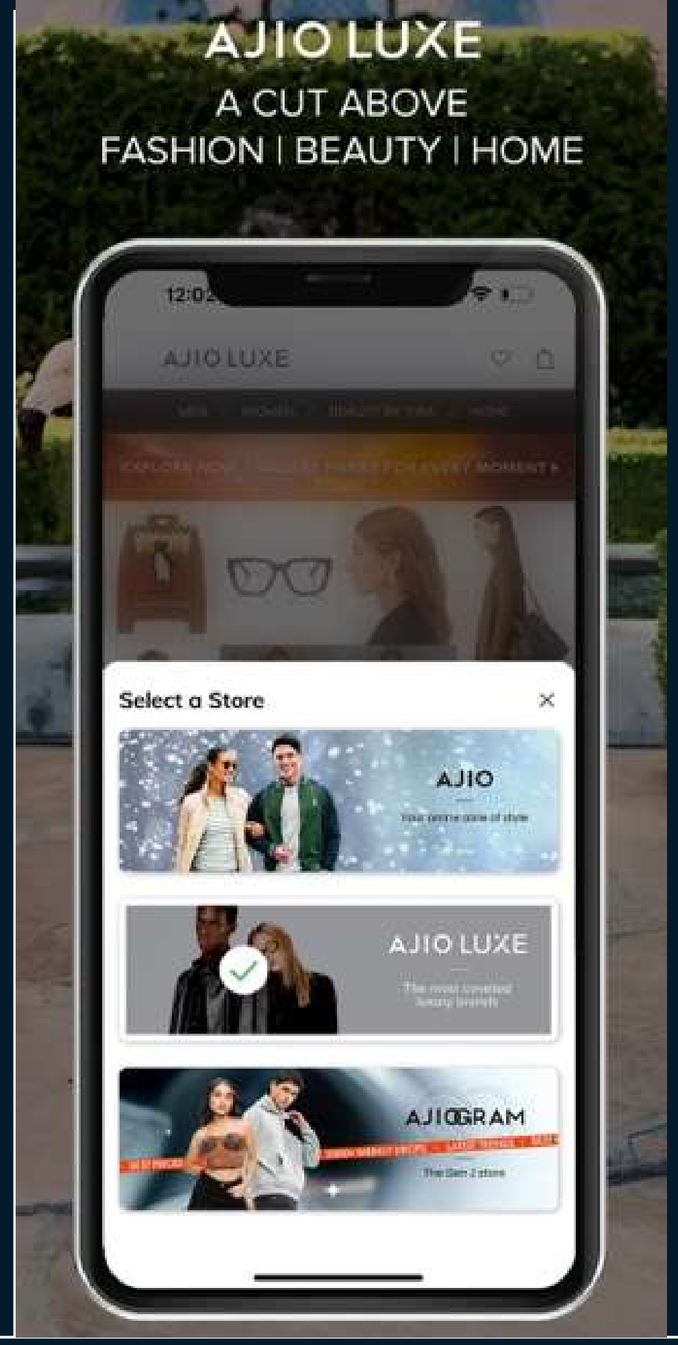
INTRODUCTION

This project aims to analyze AJIO's dataset to identify key trends in pricing, discounts, product categories, and regional offerings. By using descriptive statistics, visualizations, and statistical methods like correlation, the study explores relationships among variables and provides actionable insights to enhance business strategies and customer engagement.

OBJECTIVE

The primary objective of this project is to analyze AJIO's product data to understand various aspects such as pricing, discounts, product characteristics, and seller attributes. Specific goals include:

- Exploring the distribution of prices and discounts.
- Identifying trends in product attributes, such as material, category, and shipping time.
- Analyzing relationships between numerical variables like price and discount.
- Evaluating brand popularity and country-level insights.



RESEARCH METHODOGLY

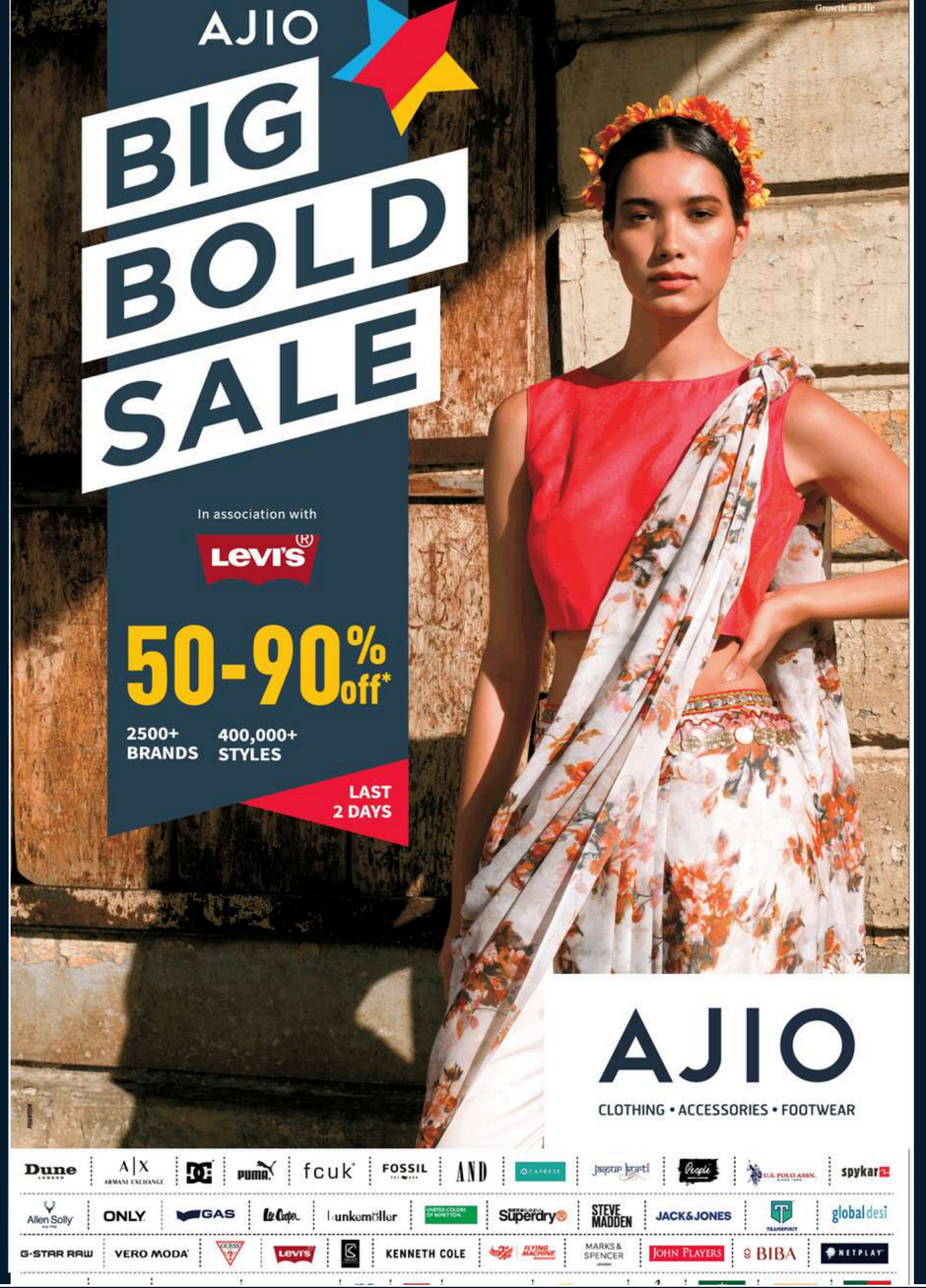
- Data Collection: Data was obtained from AJIO, focusing on product listings with attributes like price, discount, category, and brand.
- Data Cleaning:Missing values were handled using na.omit().Duplicates were removed using the distinct() function.Column names were standardized for consistency.
- Exploratory Data Analysis (EDA):Univariate, bivariate, and multivariate analyses were conducted using descriptive statistics and visualizations.Variables were analyzed for their relationships and distributions.
- Statistical Analysis: Techniques like correlation and regression analysis were applied to uncover deeper insights.

Data and Variables

Data Description

The dataset contains product-related information with the following key variables:

- price_usd: Price of the product.
- discount_usd: Discount offered on the product.
- product_type: Category or type of the product.
- product_material: Material of the product.
- brand_name: Brand offering the product.
- shipping_time: Average time for delivery.
- country: Origin of the seller.



Variables for Analysis

- **Dependent Variable: price_usd.**
- **Independent Variables: discount_usd, product_material, product_category, shipping_time.**

STATISTICS TOOLS AND TECHNIQUES FOR DATA ANALYSIS

Descriptive Analysis

1. **Summarized data using functions like summary() for an overview of the dataset.**
2. **Visualizations (e.g., histograms and bar plots) to understand distributions.**

Correlation Analysis

- **Pearson correlation coefficient was calculated to measure the strength of the linear relationship between price and discount (cor(price_usd, discount_usd)).**

EXPLORATORY DATA ANALYSIS

1. **Univariate Analysis: Histograms and bar charts revealed trends in discounts, product types, and top-selling countries.**
2. **Bivariate Analysis: Scatter plots and correlation analysis highlighted the relationship between price and discount.**
3. **Multivariate Analysis: Faceted plots and grouped summaries showed how product attributes like material, shipping time, and condition impacted pricing.**

Ques:- Find the Average discount of products by country?

(filtered_data) | AJIO.R | head(avg_discount_usd_by_country)

Filter

	country	avg_discount_usd
1	Macau	1446.4848
2	Andorra	618.2333
3	Russia	333.4899
4	Guernsey	320.6850
5	Japan	188.9787
6	Qatar	185.8321

20.1.R | 20.R | Data Manipulation and Transformation ... | 20.2.R | AJIO.R | head(avg_discount_usd_by_country) | 21.R | 21.1.R | Untitled

Source on Save | Run

```
39 #Find the Average discount of products by country
40 avg_discount_usd_by_country<-Ajio_data%>%
41   group_by(country) %>%
42   summarize(avg_discount_usd=mean(discount_usd,na.rm=TRUE)) %>%
43   arrange(desc(avg_discount_usd))
44 # View the result
45 View(head(avg_discount_usd_by_country))
46
47 #Identify the Most Common product type offered by brands
48 most_common_products<-Ajio_data%>%
49   count(product_type) %>%
50   arrange(desc(n)) %>%
51   slice(5)
52 print(most_common_products)
53
54 #What are the top 10 most expensive products?
55 top_10_expensive_products<-Ajio_data %>%
```

39:50 (Top Level)

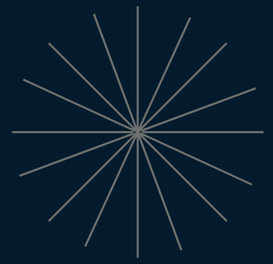
Console | Background Jobs

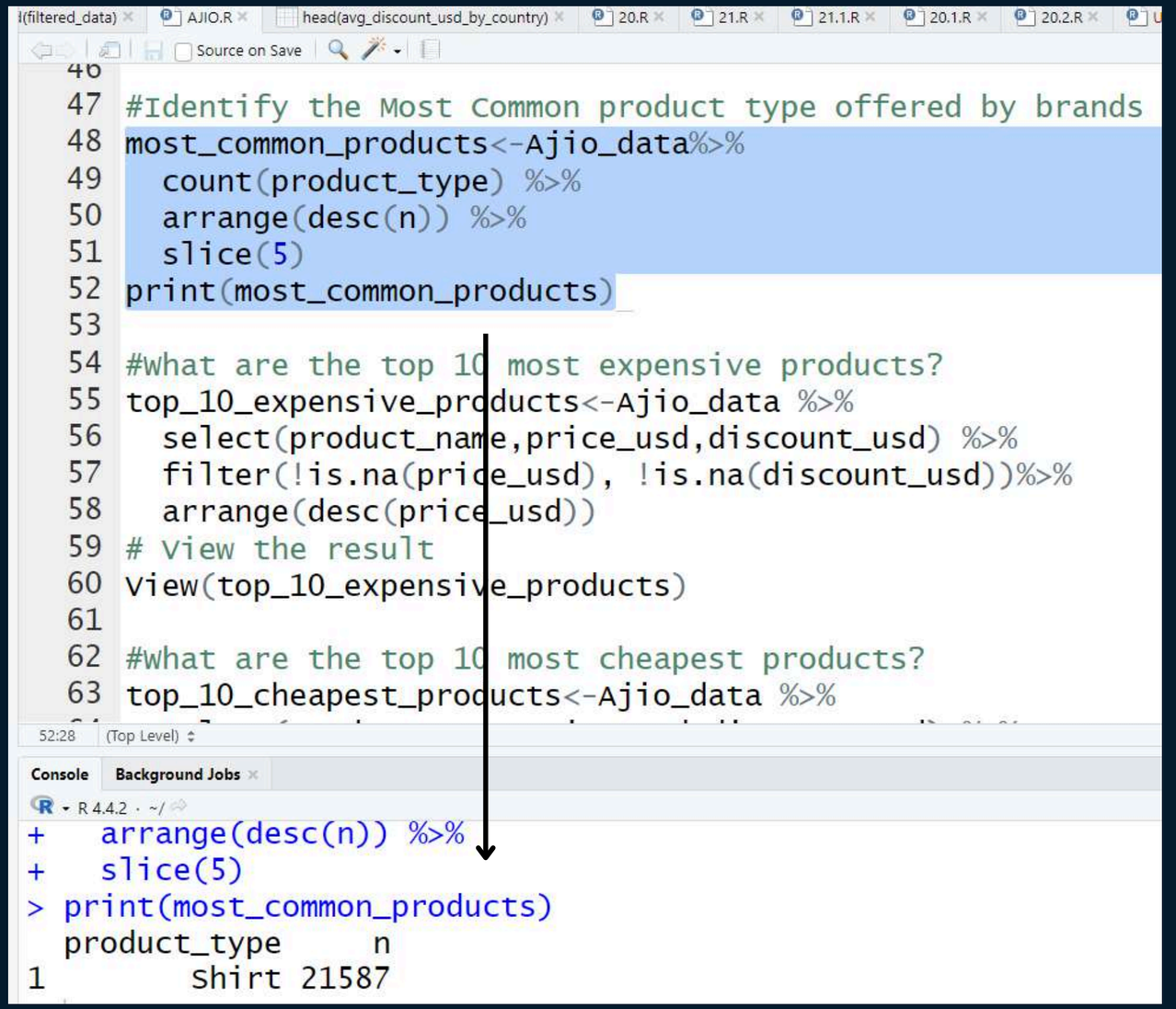
R 4.4.2 · ~/

```
> avg_discount_usd_by_country<-Ajio_data%>%
+   group_by(country) %>%
+   summarize(avg_discount_usd=mean(discount_usd,na.rm=TRUE)) %>%
+   arrange(desc(avg_discount_usd))
```

Ques:-

Identify the most common product type offered by brands?





Ques:-

What are the top 10 most expensive products?

```
filtered_data) | AJIO.R | head(avg_discount_usd_by_country) | 20.R | 21.R | 21.1.R | 20.1.R | 20.2.R
Source on Save
53
54 #What are the top 10 most expensive products?
55 top_10_expensive_products<-Ajio_data %>%
56   select(product_name,price_usd,discount_usd) %>%
57   filter(!is.na(price_usd), !is.na(discount_usd))%>%
58   arrange(desc(price_usd))
59 # View the result
60 View(top_10_expensive_products)
61
62 #What are the top 10 most cheapest products?
63 top_10_cheapest_products<-Ajio_data %>%
64   select(product_name,price_usd,discount_usd) %>%
65   filter(!is.na(price_usd), !is.na(discount_usd))%>%
66   arrange(price_usd)
67 # View the result
68 View(top_10_cheapest_products)
69
70 #Find the top 10 Highest-Brand Name by product type
58:27 (Top Level)
Console | Background Jobs
R 4.4.2 · ~/
> top_10_expensive_products<-Ajio_data %>%
+   select(product_name,price_usd,discount_usd) %>%
+   filter(!is.na(price_usd), !is.na(discount_usd))%>%
+   arrange(desc(price_usd))
```

head(filtered_data) | AJIO.R | top_10_expensive_products | 20.R | 21.R

Filter

	product_name	price_usd	discount_usd
1	Nautilus watch Patek Philippe Metallic in Steel	632610.00	105435.00
2	Nautilus white gold watch Patek Philippe Silver in White gold	457960.22	76326.70
3	Nautilus white gold watch Patek Philippe Silver in White gold	428859.06	71476.51
4	Nautilus gold watch Patek Philippe Black in Gold	303411.56	50568.59
5	Nautilus watch Patek Philippe Silver in Steel	291011.50	48501.92
6	Watch Richard Mille Silver in Titanium	249163.61	41527.27
7	Platinum watch Ulysse Nardin Blue in Platinum	230040.00	38340.00
8	Watch Patek Philippe Silver in Steel	183796.73	30632.79
9	Yellow gold watch Girard Perregaux Gold in Yellow gold	179718.75	29953.13
10	Daytona platinum watch Rolex Silver in Platinum	178081.97	29680.33
11	Nautilus pink gold watch Patek Philippe Gold in Pink gold	168480.34	28080.06
12	Yellow gold watch Blancpain Yellow in Yellow gold	163083.56	27180.59
13	Panthère white gold watch Cartier Metallic in White gold	140580.00	23430.00
14	World time white gold watch Patek Philippe White in White ...	139631.80	12693.80
15	Day	135340.70	12303.70
16	Daytona pink gold watch Rolex Gold in Pink gold	134190.00	22365.00
17	Platinum watch Breguet Black in Platinum	127949.65	21324.94
18	Crocodile jacket Hermès Black size 50 FR in Crocodile	121410.00	20235.00
19	Watch Graff Black in Titanium	121410.00	20235.00

owing 1 to 19 of 886,778 entries, 3 total columns

```
onsole | Background Jobs
R 4.4.2 · ~/
  select(product_name,price_usd,discount
  filter(!is.na(price_usd), !is.na(disco
  arrange(desc(price_usd))
# View the result
View(top_10_expensive_products)
```

Ques:-

what are top most cheapest products?

```r
#What are the top 10 most cheapest products?
top_10_cheapest_products<-Ajio_data %>%
  select(product_name,price_usd,discount_usd) %>%
  filter(!is.na(price_usd), !is.na(discount_usd))%>%
  arrange(price_usd)
# View the result
View(top_10_cheapest_products)

#Find the top 10 Highest-Brand Name by product type
top_10_highest_brand_name<-Ajio_data%>%
  select(product_type,brand_name,price_usd)%>%
  filter(!is.na(brand_name), !is.na(price_usd))%>%
  arrange(desc(brand_name))%>%
  slice(1:10)
#view the result
View(top_10_highest_brand_name)
```

```
> View(top_10_expensive_products)
> top_10_cheapest_products<-Ajio_data %>%
+   select(product_name,price_usd,discount_usd) %>%
+   filter(!is.na(price_usd), !is.na(discount_usd))%>%
+   arrange(price_usd)
```

	product_name	price_usd	discount_usd
1	Cloth low trainers Puma Blue size 9 UK in Cloth	6.13	1.02
2	Leather purse Michael Kors White in Leather	11.00	1.00
3	Leather purse Coach Burgundy in Leather	12.10	1.10
4	Purse Kate Spade Black in Plastic	12.78	2.13
5	Superstar ostrich low trainers Adidas Gold size 10 US in Ostr...	12.78	2.13
6	Purse Ksubi Orange in Plastic	13.20	1.20
7	Leather wallet Coach Pink in Leather	13.20	1.20
8	Wool cap Alcott Red size M International in Wool	14.06	2.34
9	T	14.09	2.35
10	Shirt JORDAN White size M International in Cotton	14.30	1.30
11	T	14.30	1.30
12	Shirt Kenneth Cole Burgundy size XL International in Cotton	14.30	1.30
13	Silk tie Dior Homme Multicolour in Silk	14.30	1.30
14	Silk tie Michael Kors Beige in Silk	14.30	1.30
15	Silk tie Michael Kors Multicolour in Silk	14.30	1.30
16	Silk tie Oscar De La Renta Multicolour in Silk	14.30	1.30
17	Jewellery Jewelry Unlimited Silver in Metal	14.30	1.30
18	Watch Jewelry For Less Black in Rubber	14.30	1.30
19	Leather belt B	14.30	1.30

Showing 1 to 19 of 886,778 entries, 3 total columns

```
> top_10_cheapest_products<-Ajio_data %>%
+   select(product_name,price_usd,discount_
+   filter(!is.na(price_usd), !is.na(discou
+   arrange(price_usd)
> View(top_10_cheapest_products)
```

Ques:-

Find the top 10 Highest Brand Name by Product Type?

ques2.R | mtcars | AJIO.R | top_10_highest_brand_name

Filter

	product_type	brand_name	price_usd
1	Sweatshirt	Ølaf	77.56
2	Straight pants	Études Studio	98.41
3	Trousers	Études Studio	89.46
4	Linen blouse	Études Studio	52.50
5	Sweatshirt	Études Studio	83.07
6	T-shirt	Études Studio	76.68
7	T-shirt	Études Studio	49.84
8	Sweatshirt	Études Studio	83.07
9	Pull	Études Studio	76.68
10	Wool trousers	Études Studio	115.02

ques2.R | mtcars | AJIO.R | 20.R | 21.R | 21.1.R | 20.1.R | 20.2.R | Untitled1*

Source on Save

```
#What are the top 10 most cheapest products?
top_10_cheapest_products<-Ajio_data %>%
  select(product_name,price_usd,discount_usd) %>%
  filter(!is.na(price_usd), !is.na(discount_usd))%
  arrange(price_usd)
# View the result
View(top_10_cheapest_products)

#Find the top 10 Highest-Brand Name by product typ
top_10_highest_brand_name<-Ajio_data%>%
  select(product_type,brand_name,price_usd)%>%
  filter(!is.na(brand_name), !is.na(price_usd))%>%
  arrange(desc(brand_name))%>%
  slice(1:10)
#view the result
View(top_10_highest_brand_name)
```

(Top Level)

Background Jobs

R 4.4.2 · ~/

```
top_10_highest_brand_name<-Ajio_data%>%
  select(product_type,brand_name,price_usd)%>%
  filter(!is.na(brand_name), !is.na(price_usd))%>%
  arrange(desc(brand_name))%>%
  slice(1:10)
```

Ques:-

Find the Top 10 Lowest Brand Name by Product Type?

ques2.R | mtcars | AJIO.R | top_10_lowest_brand_name

Filter

	product_type	brand_name	price_usd
1	Rabbit short vest	#FR2	70.29
2	Jacket	#FR2	31.95
3	T-shirt	#FR2	93.50
4	T-shirt	#FR2	109.91
5	Jacket	#FR2	103.76
6	Sweatshirt	#FR2	68.62
7	Long gloves	#FR2	48.56
8	Purse	#FR2	166.14
9	Purse	#FR2	20.45
10	Fox stole	#FR2	85.63

ques2.R | mtcars | AJIO.R | 20.R | 21.R | 21.1.R | 20.1.R | 20.2.R | Untitled1*

Source on Save

```
78
79 #Find the Top 10 lowest-Brand Name by product type
80 top_10_lowest_brand_name<-Ajio_data%>%
81   select(product_type,brand_name,price_usd)%>%
82   filter(!is.na(brand_name), !is.na(price_usd))%>%
83   arrange(brand_name)%>%
84   slice(1:10)
85 #view the result
86 View(top_10_lowest_brand_name)
87
88 #What is the correlation between price and discount
89 correlation <- cor(Ajio_data$price_usd, Ajio_data$
90                    = "pearson")
91 print(correlation)
92
93 #what is the range of price and discount
94 range_price_usd <- range(Ajio_data$price_usd)
95 print(paste("Range of Price:", diff(range_price_us
```

84:14 (Top Level)

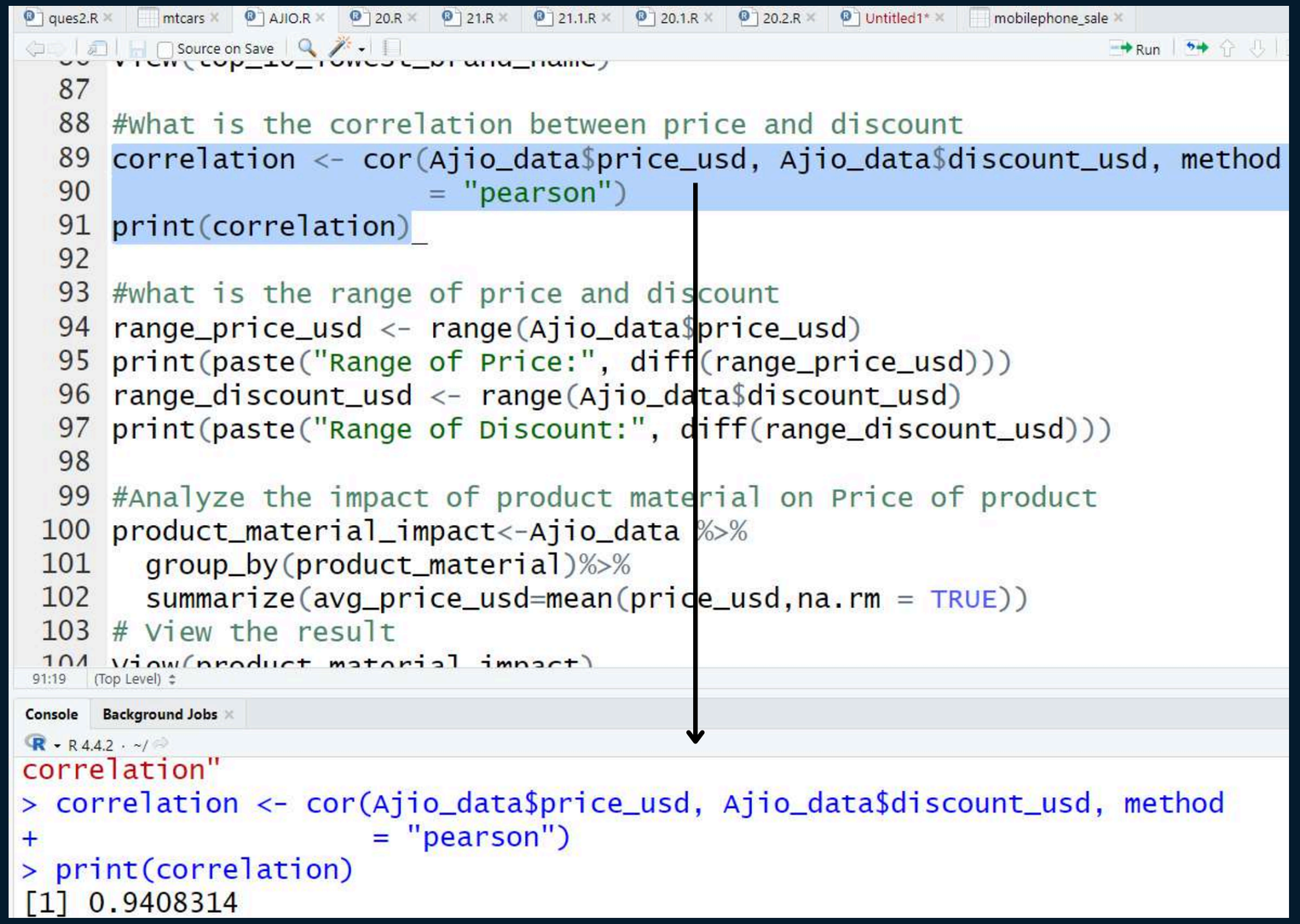
Console | Background Jobs

R 4.4.2 · ~/

```
> top_10_lowest_brand_name<-Ajio_data%>%
+   select(product_type,brand_name,price_usd)%>%
+   filter(!is.na(brand_name), !is.na(price_usd))%>%
+   arrange(brand_name)%>%
+   slice(1:10)
```

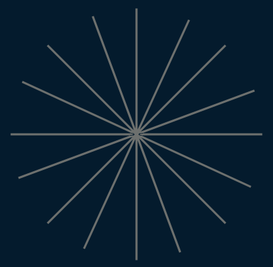
Ques:-

What is the correlation between price and discount?



Ques:-

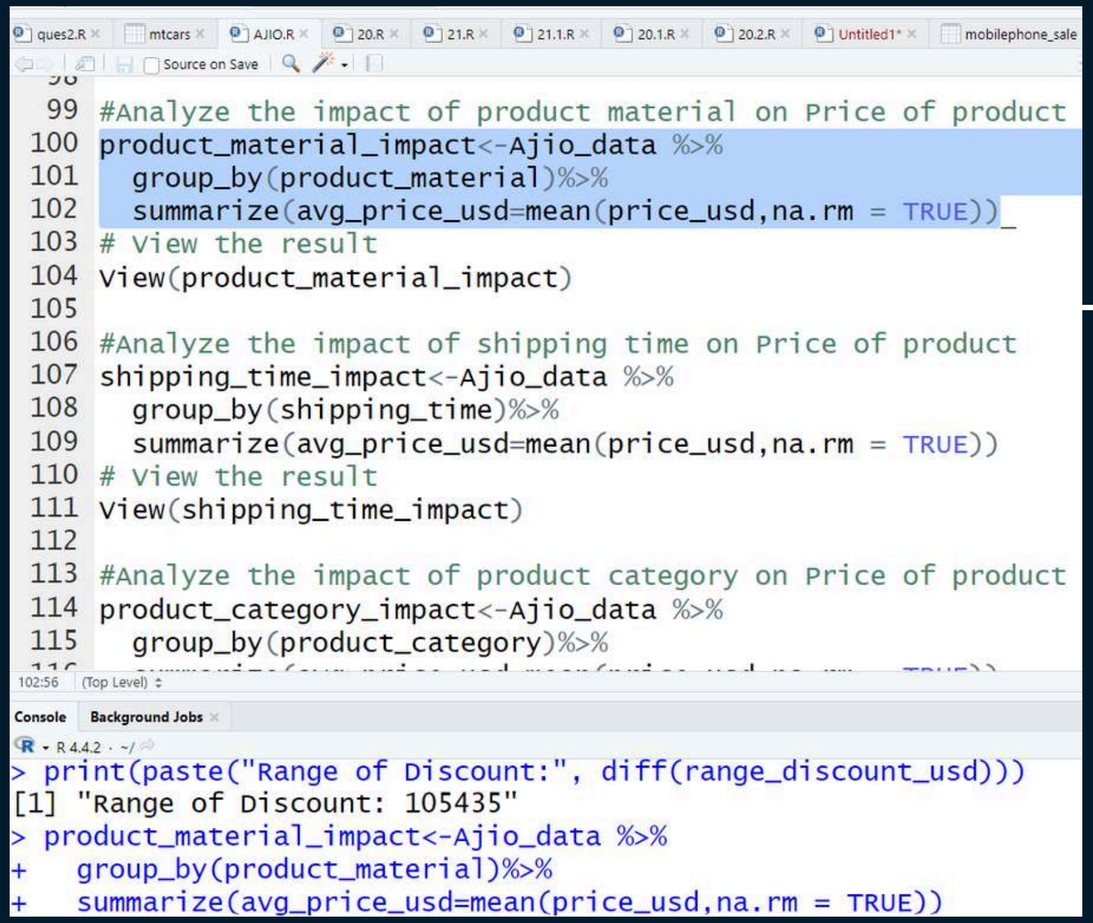
What is the range of price and product?

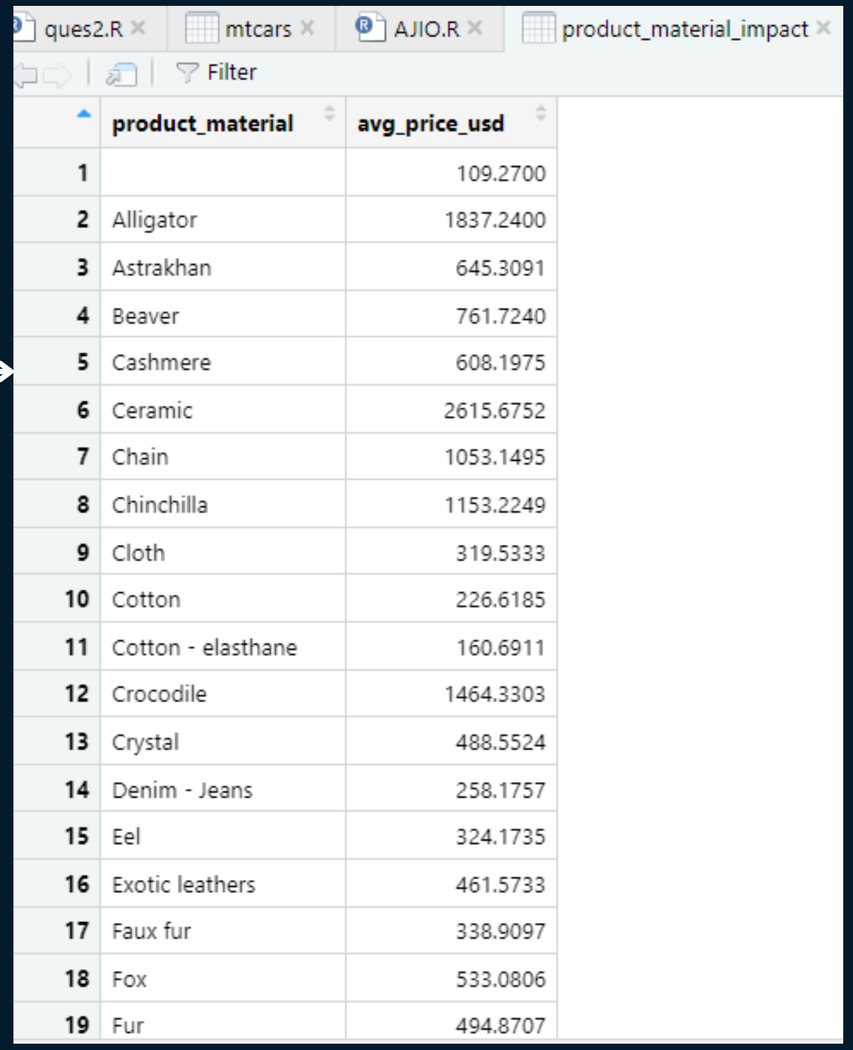


```
92
93  #what is the range of price and discount
94  range_price_usd <- range(Ajio_data$price_usd)
95  print(paste("Range of Price:", diff(range_price_usd)))
96  range_discount_usd <- range(Ajio_data$discount_usd)
97  print(paste("Range of Discount:", diff(range_discount_usd)))
98
99  #Analyze the impact of product material on Price of product
100 product_material_impact<-Ajio_data %>%
101   group_by(product_material)%>%
102   summarize(avg_price_usd=mean(price_usd,na.rm = TRUE))
103 # View the result
104 View(product_material_impact)
105
106 #Analyze the impact of shipping time on Price of product
107 shipping_time_impact<-Ajio_data %>%
108   group_by(shipping_time)%>%
109   summarize(avg_price_usd=mean(price_usd,na.rm = TRUE))
```

```
> print(paste("Range of Price:", diff(range_price_usd)))
[1] "Range of Price: 632603.87"
> range_discount_usd <- range(Ajio_data$discount_usd)
> print(paste("Range of Discount:", diff(range_discount_usd)))
[1] "Range of Discount: 105435"
```

Ques:- Analyze the impact of product material on price of product





	product_material	avg_price_usd
1		109.2700
2	Alligator	1837.2400
3	Astrakhan	645.3091
4	Beaver	761.7240
5	Cashmere	608.1975
6	Ceramic	2615.6752
7	Chain	1053.1495
8	Chinchilla	1153.2249
9	Cloth	319.5333
10	Cotton	226.6185
11	Cotton - elasthane	160.6911
12	Crocodile	1464.3303
13	Crystal	488.5524
14	Denim - Jeans	258.1757
15	Eel	324.1735
16	Exotic leathers	461.5733
17	Faux fur	338.9097
18	Fox	533.0806
19	Fur	494.8707

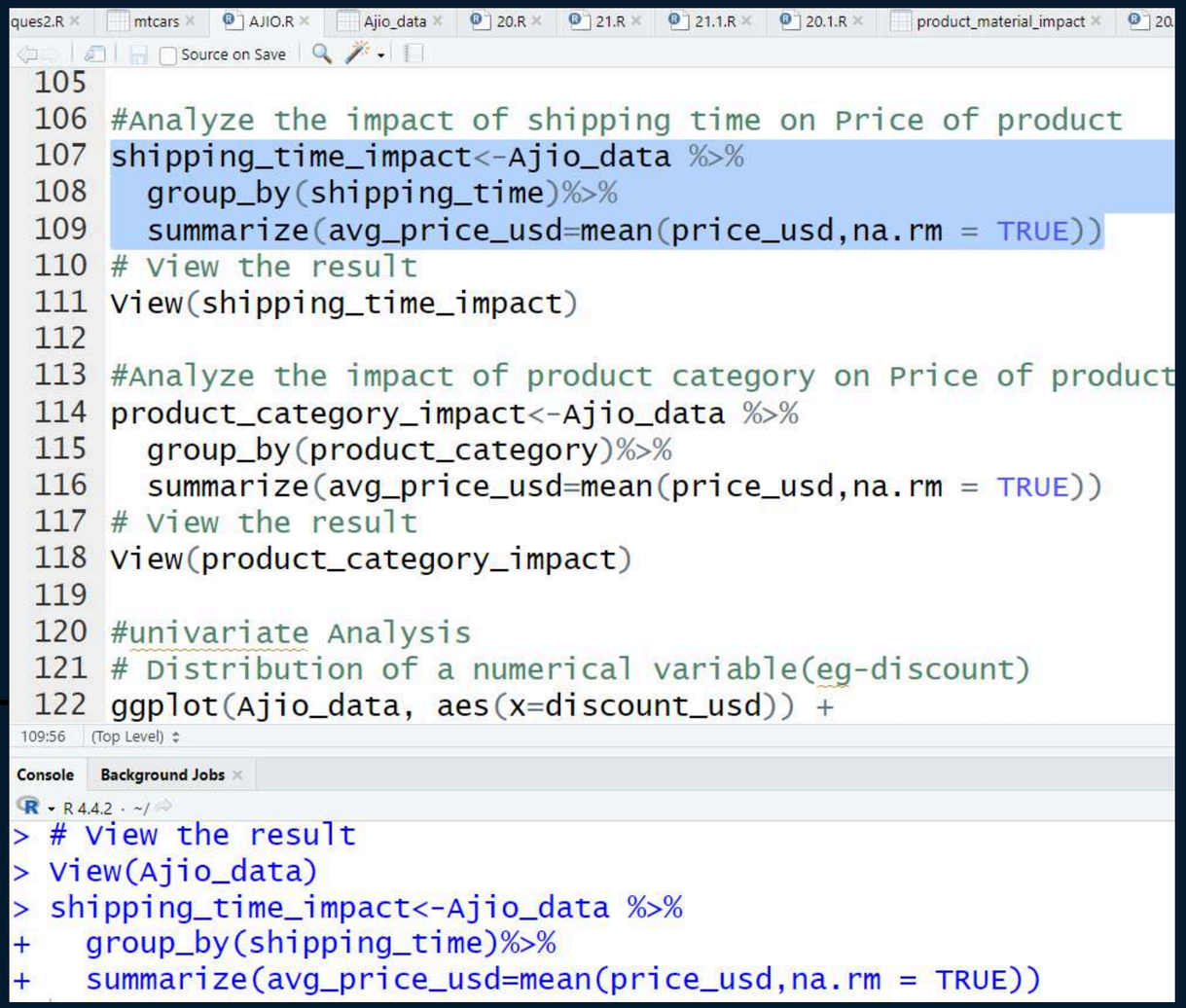
Ques:-

Analyze the impact of Shipping Time on price of product

ques2.R | mtcars | AJIO.R | shipping_time_impact

Filter

	shipping_time	avg_price_usd
1		464.2124
2	1-2 days	357.2077
3	3-5 days	427.3262
4	6-7 days	328.2024
5	More than 7 days	263.8120



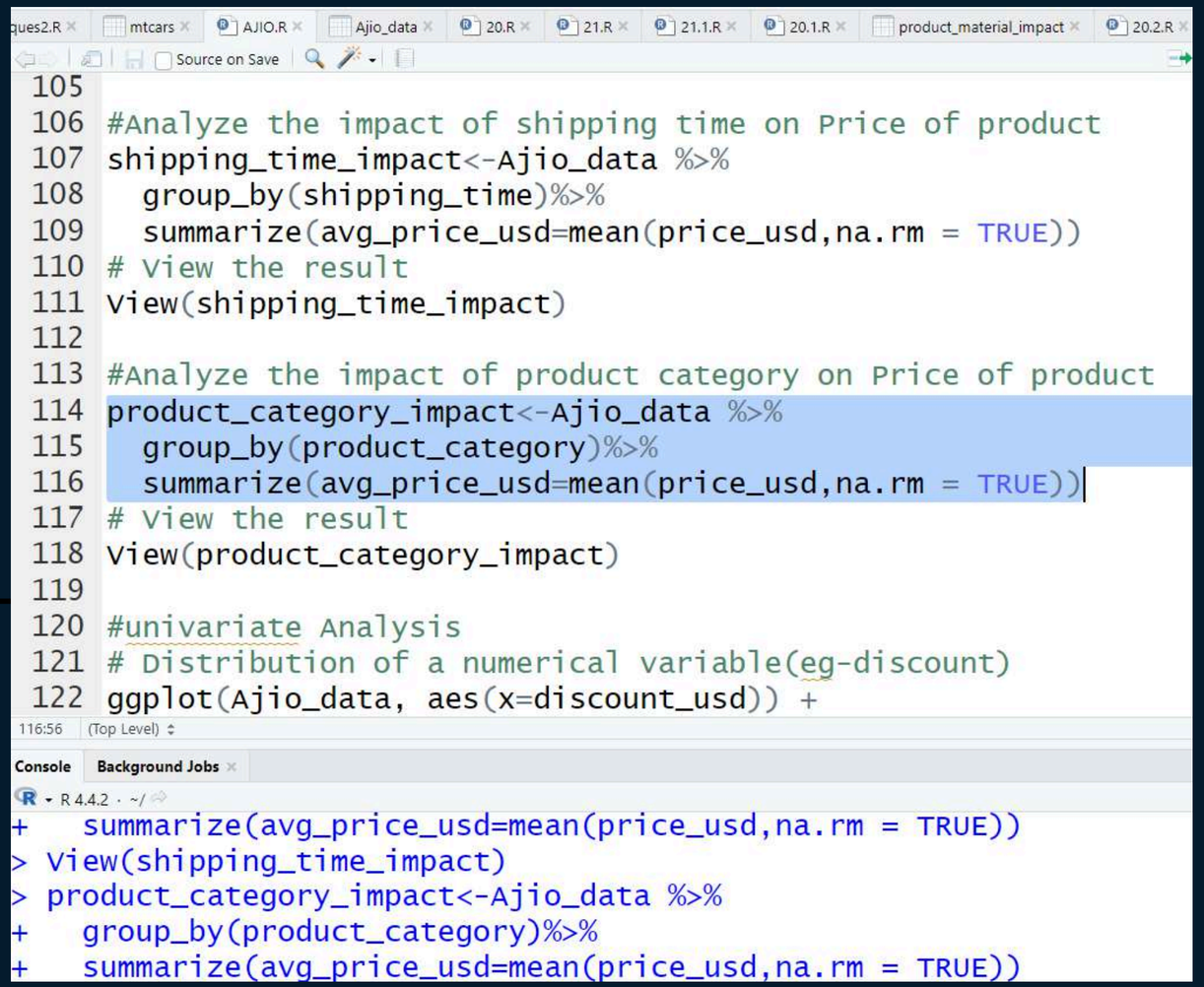
Ques:-

Analyze the impact of Product Category on price of product

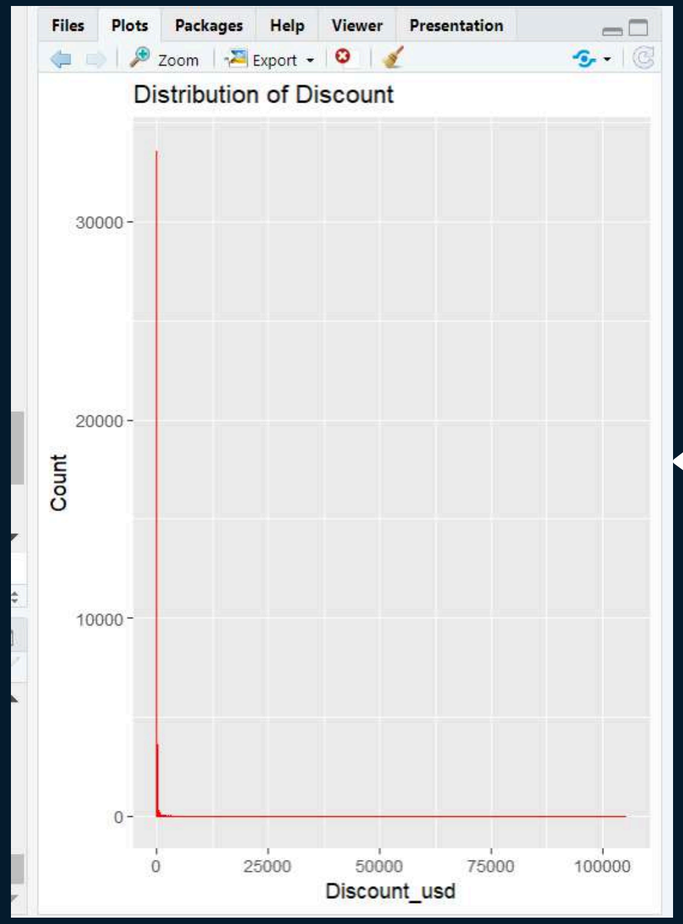
2.R | mtcars | AJIO.R | product_category_impact

Filter

	product_category	avg_price_usd
1		166.7388
2	Men Accessories	820.5507
3	Men Clothing	302.3383
4	Men Shoes	330.5511
5	Women Accessories	403.7646
6	Women Clothing	319.2131
7	Women Shoes	338.2185



Ques:- Histogram of single variable discount




```
119
120 #univariate Analysis
121 # Distribution of a numerical variable(eg-discount)
122 ggplot(Ajio_data, aes(x=discount_usd)) +
123   geom_histogram(binwidth=0.5, fill="blue",
124                  color="red") +
125   labs(title="Distribution of Discount",
126        x="Discount_usd", y="Count")
127
128 # Calculate the count of each product material
129 country_counts <- Ajio_data %>%
130   count(country) %>%
131   arrange(desc(n)) %>%
132   slice(1:5)  # Keep only the top 5 product material
133 # View the result
134 View(country_counts)
135
136 # Create the bar plot for the top 5 country
```

```
> ggplot(Ajio_data, aes(x=discount_usd)) +
+   geom_histogram(binwidth=0.5, fill="blue",
+                  color="red") +
+   labs(title="Distribution of Discount",
+        x="Discount_usd", y="Count")
```

Ques:-

Calculate the count of each product material?

ques2.R | mtcars | AJIO.R | country_counts

Filter

	country	n
1	Italy	233336
2	France	124924
3	United States	123073
4	United Kingdom	75129
5	Spain	45382

ues2.R | mtcars | AJIO.R | Ajio_data | 20.R | 21.R | 21.1.R | 20.1.R | product_materi

Source on Save

```
127
128 # Calculate the count of each product material
129 country_counts <- Ajio_data %>%
130   count(country) %>%
131   arrange(desc(n)) %>%
132   slice(1:5)  # Keep only the top 5 product materia
133 # View the result
134 View(country_counts)
135
136 # Create the bar plot for the top 5 country
137 ggplot(country_counts, aes(x = reorder(country, n)
138   geom_bar(stat = "identity", fill = "pink") +
139   labs(title = "Top 5 Country", x = "Country", y =
140   theme(axis.text.x = element_text(angle = 90, hjus
141   coord_flip()
142
143 #Bivariate Analysis
144 # Scatter plot between two numerical variables (pr
```

32:15 (Top Level)

Console | Background Jobs

R 4.4.2 · ~/

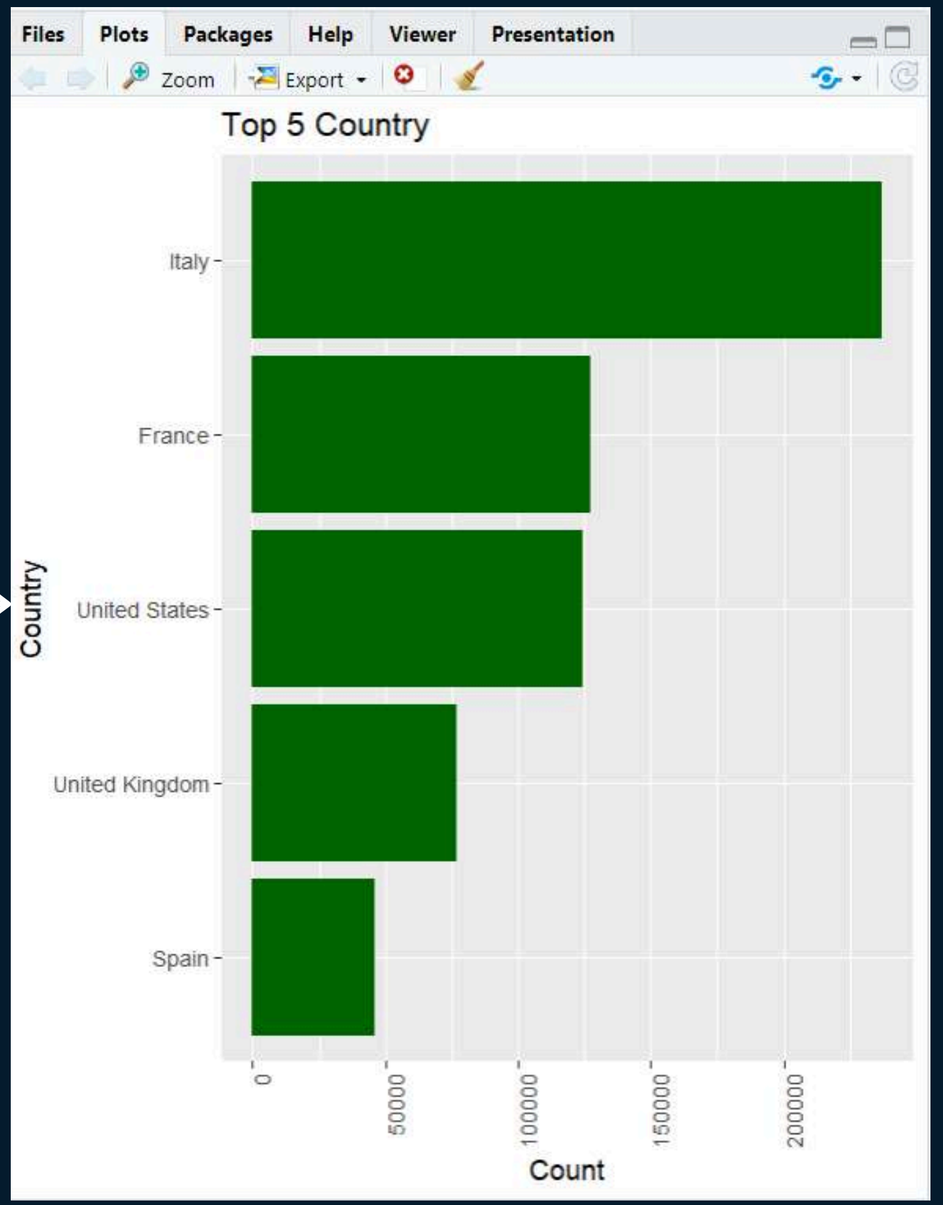
```
        x="Discount_usd", y="Count")
> country_counts <- Ajio_data %>%
+   count(country) %>%
+   arrange(desc(n)) %>%
+   slice(1:5)
```

Ques:-

Bar Plot for Top 5 Country

```r

# Calculate the count of each product material
country_counts <- Ajio_data %>%
  count(country) %>%
  arrange(desc(n)) %>%
  slice(1:5)  # Keep only the top 5 product material
# View the result
View(country_counts)

# Create the bar plot for the top 5 country
ggplot(country_counts, aes(x = reorder(country, n), y = n)) +
  geom_bar(stat = "identity", fill = "darkgreen") +
  labs(title = "Top 5 Country", x = "Country", y = "Count") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  coord_flip()

#Bivariate Analysis
# Scatter plot between two numerical variables (price and disc
```

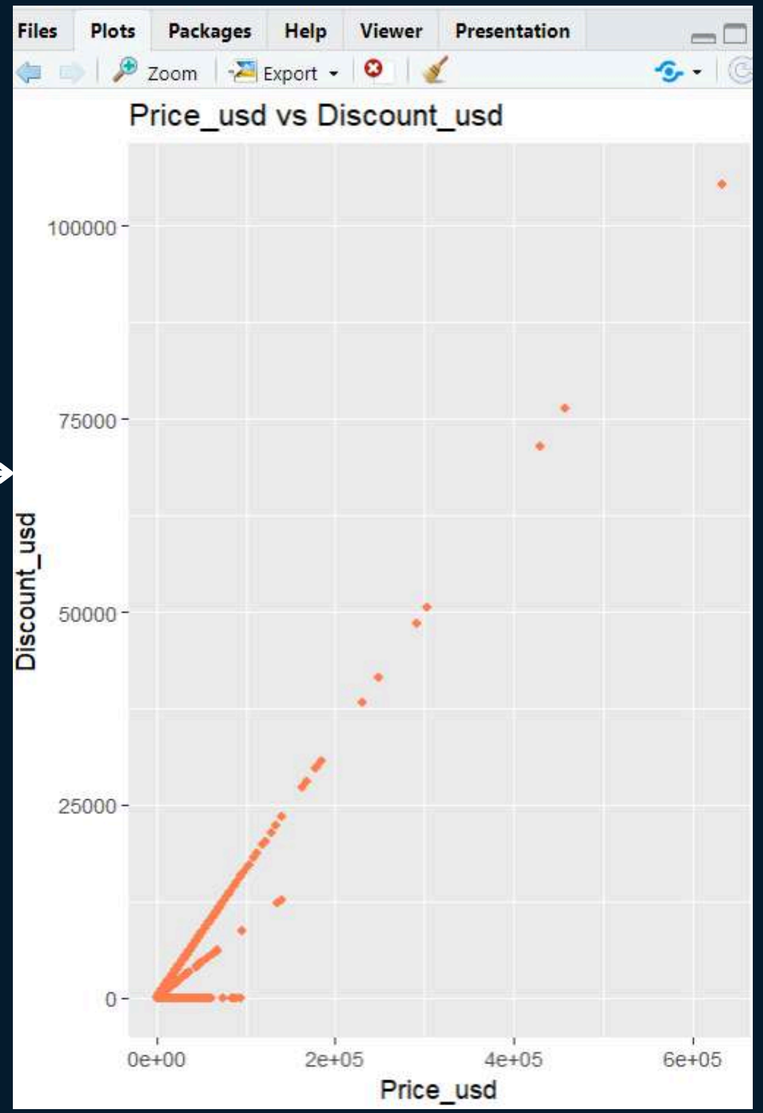
```r
ggplot(country_counts, aes(x = reorder(country, n), y = n)) +
  geom_bar(stat = "identity", fill = "darkgreen") +
  labs(title = "Top 5 Country", x = "Country", y = "Count") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  coord_flip()
```



Ques:- Scatter Plot between price and discount

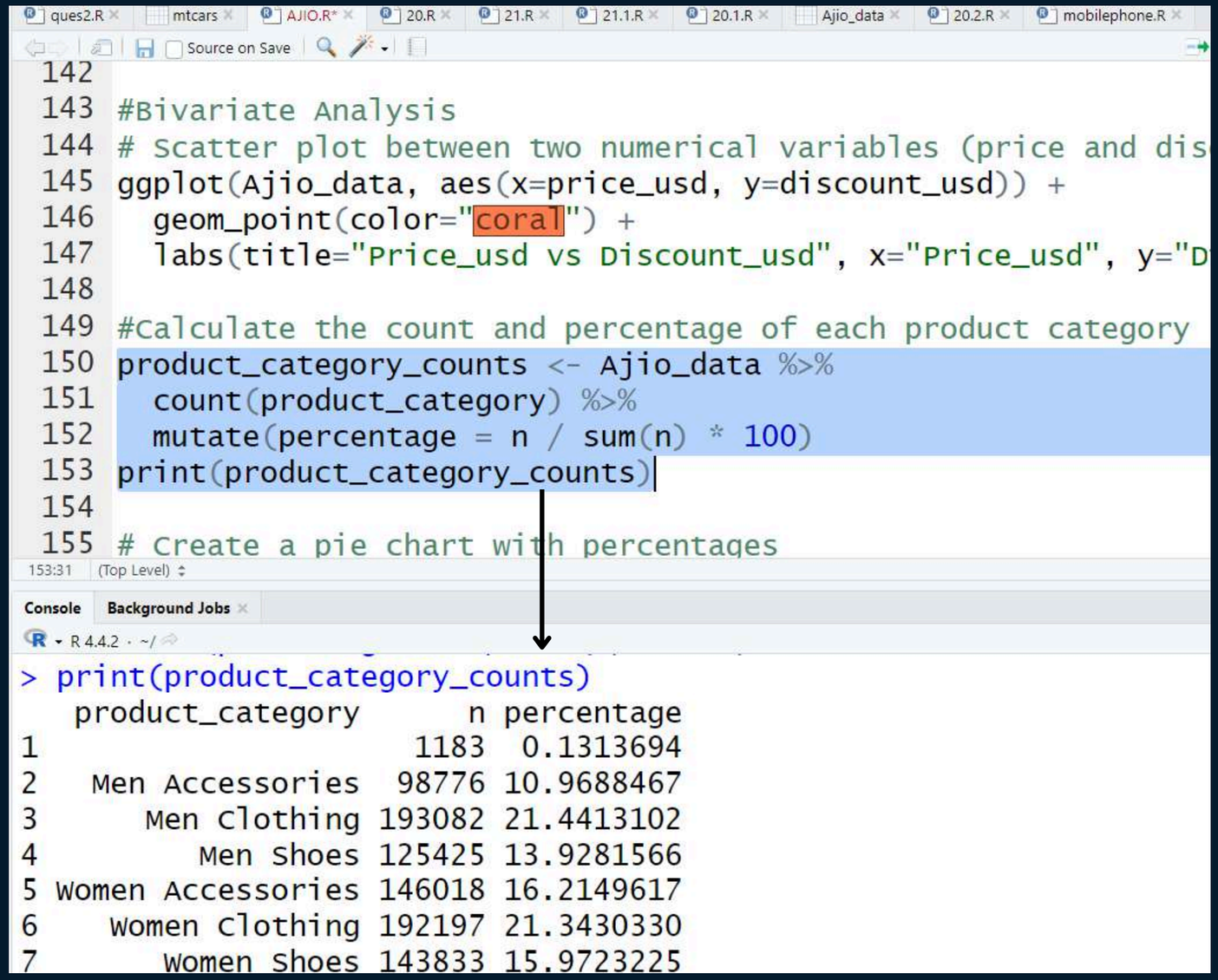
```r
#Bivariate Analysis
# Scatter plot between two numerical variables (price and discount)
ggplot(Ajio_data, aes(x=price_usd, y=discount_usd)) +
  geom_point(color="coral") +
  labs(title="Price_usd vs Discount_usd", x="Price_usd", y="Discount_usd")

#Calculate the count and percentage of each product category
product_category_counts <- Ajio_data %>%
  count(product_category) %>%
  mutate(percentage = n / sum(n) * 100)

# Create a pie chart with percentages
ggplot(product_category_counts, aes(x = "", y = percentage,
                                    fill = product_category)) +
  geom_bar(width = 1, stat = "identity") +km
  coord_polar(theta = "y") +
```

```
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  coord_flip()
ggplot(Ajio_data, aes(x=price_usd, y=discount_usd)) +
  geom_point(color="coral") +
  labs(title="Price_usd vs Discount_usd", x="Price_usd", y="Discount_usd")
```



Ques:-

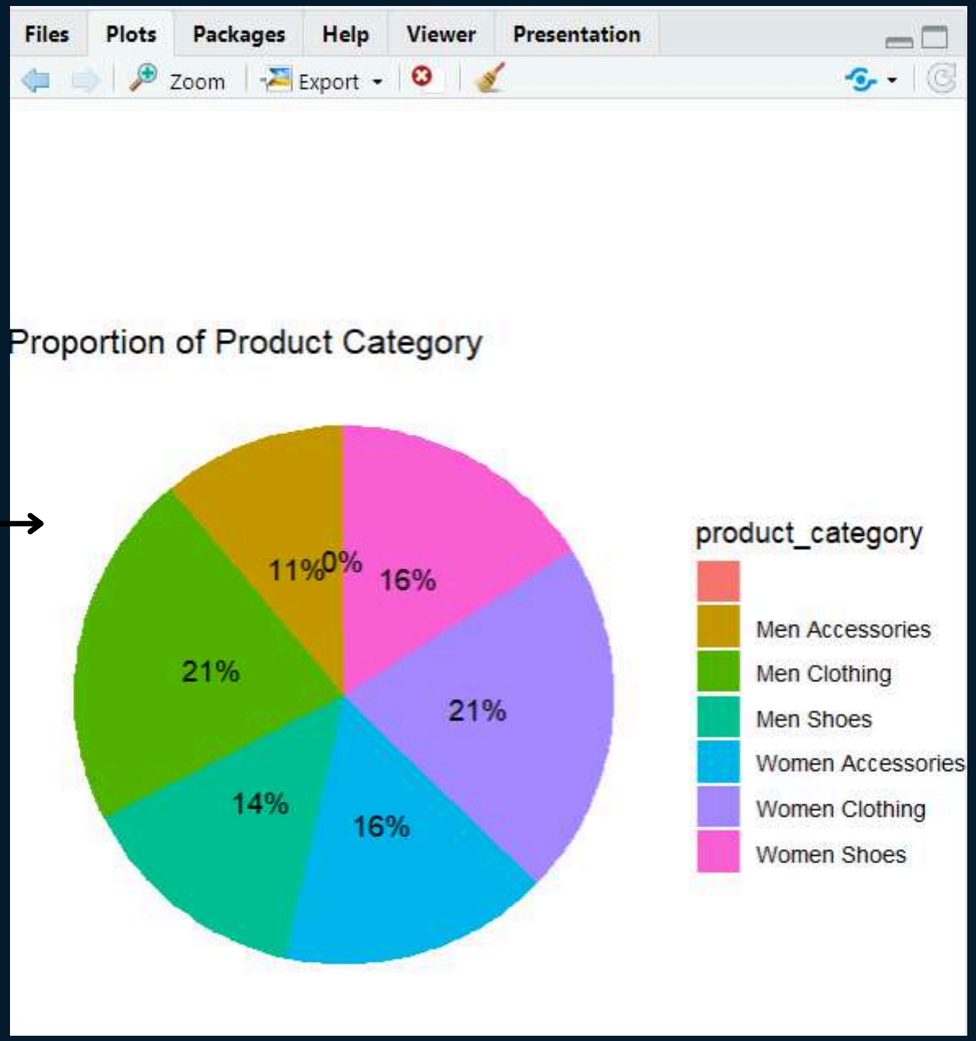
Calculate the count and percentage of each product category?



Ques:- Pie Chart of Product Category

```
  mutate(percentage = n / sum(n) * 100)
print(product_category_counts)

# Create a pie chart with percentages
ggplot(product_category_counts, aes(x = "", y = percentage,
                                    fill = product_category)) +
  geom_bar(width = 1, stat = "identity") +
  coord_polar(theta = "y") +
  geom_text(aes(label = paste0(round(percentage,0), "%")),
            position = position_stack(vjust = 0.5)) +
  labs(title = "Proportion of Product Category") +
  theme_void()

#Faceted plot to analyze multiple subsets
```

```
Error: object 'km' not found
> ggplot(product_category_counts, aes(x = "", y = percentage,
+                                     fill = product_category)) +
+   geom_bar(width = 1, stat = "identity") +
+   coord_polar(theta = "y") +
+   geom_text(aes(label = paste0(round(percentage,0), "%")),
+             position = position_stack(vjust = 0.5)) +
+   labs(title = "Proportion of Product Category") +
+   theme_void()
```

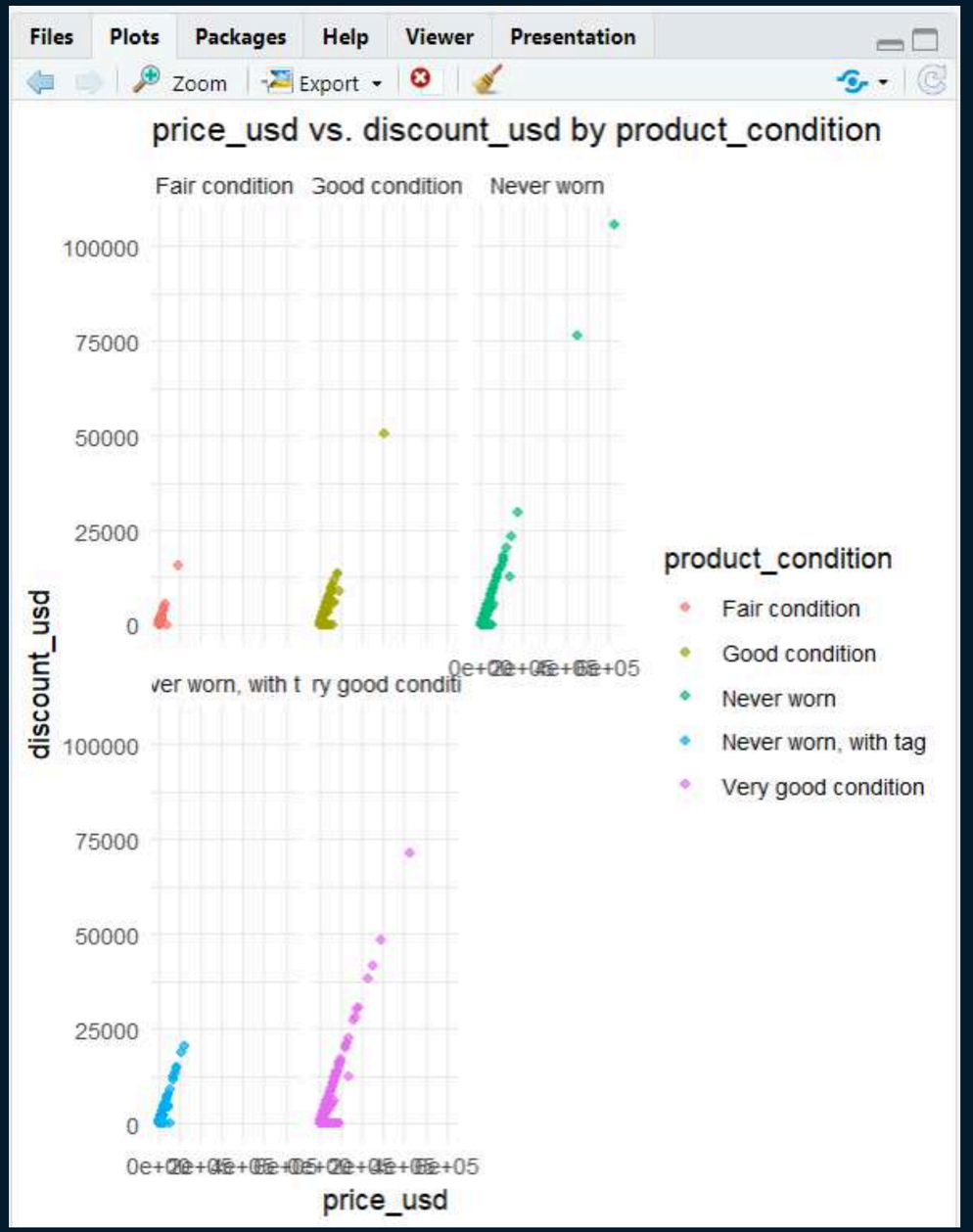


Ques:-

Faceted plot to analyze price vs discount across different product condition

```
  theme_void()

#Faceted plot to analyze multiple subsets
#(e.g.,price vs discount across different product condition)
ggplot(Ajio_data, aes(x = price_usd, y = discount_usd)) +
  geom_point(aes(color = product_condition), alpha =0.7) +
  facet_wrap(~ product_condition) +
  labs(title = "price_usd vs. discount_usd by product_conditi
  x = "price_usd", y = "discount_usd") +
  theme_minimal()
```

```
+              position = position_stack(vjust = 0.5)) +
+   labs(title = "Proportion of Product Category") +
+   theme_void()
> ggplot(Ajio_data, aes(x = price_usd, y = discount_usd)) +
+   geom_point(aes(color = product_condition), alpha =0.7) +
+   facet_wrap(~ product_condition) +
+   labs(title = "price_usd vs. discount_usd by product_condition",
+   x = "price_usd", y = "discount_usd") +
+   theme_minimal()
```



Findings

- **Pricing and Discounts:**

-The range of prices and discounts was significant, highlighting product diversity.

-A moderate positive correlation was found between price and discount.

- **Product Trends:**

-Certain materials and categories were associated with higher average prices.

-Shipping time influenced product pricing, with faster shipping times correlating with higher prices.

Brand and Country Insights:

-The most common product types and their associated brands were identified.

-Countries with the highest average discounts were determined, indicating regional pricing strategies.

Visual Insights:

-Scatter plots and faceted plots revealed trends in price vs. discount for different product conditions.

-Pie charts illustrated the distribution of product categories

Conclusion

This project provided valuable insights into AJIO's product data, aiding in better understanding market trends and customer preferences. The analysis highlights opportunities for optimizing pricing strategies and product offerings.

THANK YOU !