Final Report:
Developing a Research Report
Data Analytics, Fall 2022

ANALYSIS OF CUSTOMER SPENDING HABITS TO IMPROVE SALES PERFORMANCE

Name: **TULKINOV SHAVKATJON 李克涵**
Student ID: **502022145004**

Abstract: This quantitative research study analyzes the customer spending habits of a business in terms of revenue, profit, and product sales across different gender and geographic segments. The study uses data obtained from Kaggle and employs various visualization techniques such as line graphs and bar charts to represent the data. The findings suggest that male customers generate higher total revenue and profits than female customers. Moreover, the highest total profits were generated by customers aged 25-45, mostly from male customers. The highest profit-generating product category was accessories, while bikes showed negative profit in 2015 but improved in 2016. Tires and tubes were the most sold items, followed by bottles and cages and helmets. The analysis also revealed that Germany generated the highest total profit, followed by the United States, while France had the lowest profit. The study concludes that the business can use the visual representation of revenue, profit, and sales data to identify areas of potential improvement in their sales and marketing strategies.

Keywords: Spending habits; Sales performance; Revenue analysis; Product categories.

1. INTRODUCTION

1.1 Background and Context

In the highly competitive business world, understanding customers' spending habits has become a critical factor in boosting sales performance. Analyzing customer spending habits can provide businesses with valuable insights that can help them improve their product offerings, customer service, and overall customer experience. As such, this topic has garnered significant interest among researchers and business professionals alike.

Numerous studies have explored the relationship between customer spending habits and sales performance, highlighting the importance of understanding customer behavior. For instance, in a study by KPMG (2021), it was found that businesses that analyze their customer data can increase their sales by up to 15%. Similarly, in a study by McKinsey & Company (2020), it was discovered that businesses that personalize their offerings based on customer behavior can increase their revenue by up to 15%. To analyze customer spending habits effectively, businesses can use a range of methods, including customer surveys, transaction data analysis, and social media monitoring. For instance, in a study by Hoang and Nguyen (2021), it was found that transaction data analysis can help businesses identify patterns in customer spending habits, enabling them to offer tailored promotions and discounts. The benefits of analyzing customer spending habits extend beyond sales performance. Businesses can also use the insights gained to improve their customer retention rates and reduce customer churn. For example, in a study by Bain & Company (2019), it was found that businesses that use customer data to personalize their offerings can reduce customer churn by up to 10%.

In conclusion, analyzing customer spending habits is crucial for improving sales performance and enhancing customer experience. Businesses that leverage the insights gained from customer data analysis can make more informed decisions, leading to increased revenue, improved customer retention, and enhanced brand loyalty.

1.2 Problem Statement

The research problem statement for my project: **'Analysis of Customer Spending Habits to Improve Sales Performance'.** While analyzing customer spending habits is critical for improving sales performance and enhancing customer experience, many businesses struggle to effectively collect, analyze, and utilize customer data. This leads to missed opportunities for personalization and tailored offerings, which can negatively impact revenue and customer retention rates. Therefore, the research problem has been chosen for this particular project. In this particular project, a detailed visual representation of total revenue by gender, total profit vs gender by country, total profit vs year by product categories, and Total number of items sold in each sub-category. Analyzing consumer behavior is key to success when it comes to commerce business models so this analysis offers powerful ways into understanding your customer base better.

1.3 Research Questions

The research questions developed to be answered after the analysis are as follows:

- What is the gender-based differences in total revenue for the business, and how do they vary by product category and country?
- How do the profits vary across product categories and countries, and what is the impact of gender on these profits?
- What is the trend of total profit over the years, and how does this vary across product categories?
- What is the relationship between the total number of items sold and the sub-categories, and how does this vary across different gender and geographic segments?
- How can the business use the visual representation of revenue, profit, and sales data to identify areas of potential improvement in their sales and marketing strategies?

2. RESEARCH METHODOLOGY

The research will employ a quantitative research design, which utilizes numerical data and statistical analysis to provide detailed answers to the research questions. The data for this study has been collected from Kaggle (www.kaggle.com), a platform for data scientists and machine learning enthusiasts to share and find datasets. The use of Kaggle's diverse and high-quality datasets will enable the study to provide comprehensive insights into the research topic.

Our study uses the collected data to analyze, visualize in R language, relying our knowledge from the course, Data Analytics. This study comprises of analysis generated by categories. Thus, descriptive statistics are given in tables showing profit and statistics values along with line graphs and bar graphs.

3. DATA, ANALYSIS AND RESULTS

This Sales Data dataset offers a unique insight into the spending habits of customers from various countries across the globe. With detailed information on customer age, gender, product category, quantity, unit cost and price, as well as revenue generated through sales of products listed in this dataset, we will explore and discover patterns in consumer behavior. Also, with this dataset sales managers can gain valuable insights about the changes in consumer demand for specific products over time – find out which Products had better margin or however see how different promotions impacted overall sales performance from different categories and sub-categories. ***Data input (codes for R) shown in the references has been used for analysis and visualization of the data used in this project (see the last page)***. The data frame consists of 34865 rows and 16 columns. The rows give the observation and the columns represent the various variables used in the research such as date, customer age, gender, product category/subcategory revenue and so on as shown in table 1:

Table 1: Summarized data frame:

index	Date	Year	Month	Customer.Age	Customer.Gender	Country	State	Product.Category	Sub.Category	Quantity	Unit.Cost	Unit.Price	Cost	Revenue	Column1
0	02/19/16	2016	February	29	F	United States	Washington	Accessories	Tires and Tubes	1	80.00	109.000000	80	109	NA
1	02/20/16	2016	February	29	F	United States	Washington	Clothing	Gloves	2	24.50	28.500000	49	57	NA
2	02/27/16	2016	February	29	F	United States	Washington	Accessories	Tires and Tubes	3	3.67	5.000000	11	15	NA
3	03/12/16	2016	March	29	F	United States	Washington	Accessories	Tires and Tubes	2	87.50	116.500000	175	233	NA
4	03/12/16	2016	March	29	F	United States	Washington	Accessories	Tires and Tubes	3	35.00	41.666667	105	125	NA
5	04/08/16	2016	April	29	F	United States	Washington	Accessories	Tires and Tubes	1	66.00	78.000000	66	78	NA
6	04/17/16	2016	April	29	F	United States	Washington	Accessories	Tires and Tubes	2	52.00	60.000000	104	120	NA
7	04/17/16	2016	April	29	F	United States	Washington	Accessories	Tires and Tubes	1	60.00	68.000000	60	68	NA
8	06/22/16	2016	June	29	F	United States	Washington	Accessories	Tires and Tubes	2	8.00	10.000000	16	20	NA
9	06/22/16	2016	June	29	F	United States	Washington	Accessories	Tires and Tubes	2	2.50	3.000000	5	6	NA
10	06/24/16	2016	June	29	F	United States	Washington	Accessories	Helmets	2	17.50	21.000000	35	42	NA
11	07/01/16	2016	July	29	F	United States	Washington	Accessories	Tires and Tubes	3	40.00	48.666667	120	146	NA
12	07/22/16	2016	July	29	F	United States	Washington	Accessories	Tires and Tubes	2	46.00	60.000000	92	120	NA
13	07/22/16	2016	July	29	F	United States	Washington	Accessories	Tires and Tubes	1	9.00	10.000000	9	10	NA
14	08/18/15	2015	August	29	F	United States	Washington	Accessories	Helmets	3	151.67	181.333333	455	544	NA
15	08/30/15	2015	August	29	F	United States	Washington	Accessories	Tires and Tubes	3	70.00	72.666667	210	218	NA

The summarized table shows only the part of the whole table where the rest of the file is given in the main Excel file. The main excel file has been used for data visualization after we have completed data manipulation in R programming language. The analysis and the results have been given in the following. Here in the Figure 1, The summarized graph of total revenue by customer gender has been shown:

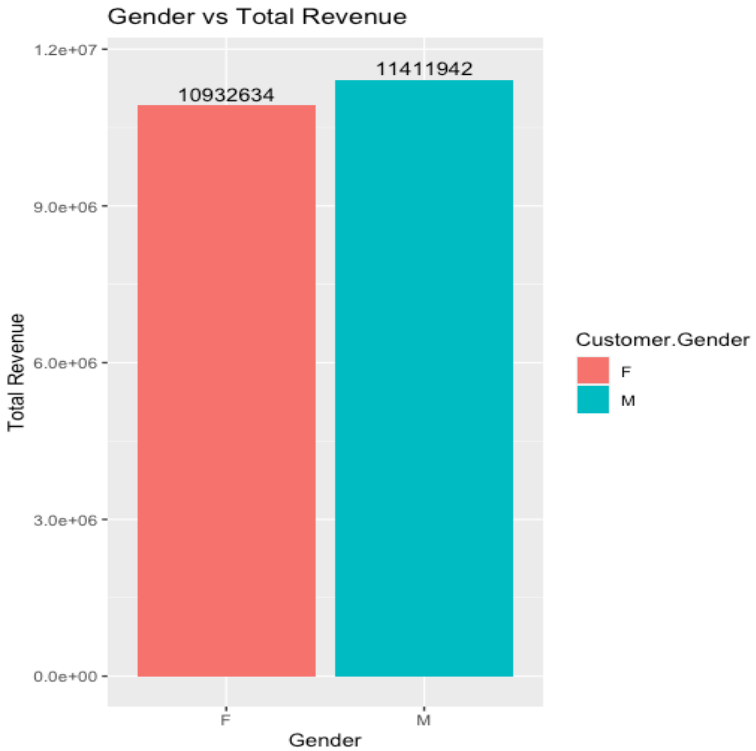

Figure 1: Total revenue by customer gender.

The results of Figure 1 show that the total revenue generated by male customers is higher than the total revenue generated by female customers. Specifically, male customers generated a total revenue of 11411942 USD, while female customers generated a total revenue of 10932634 USD. This suggests that male customers are more significant contributors to the business's revenue stream than female customers. To better understand the gender-based

differences in revenue, it may be helpful to analyse the sales data by specific product categories and regions. Further analysis can be conducted to explore the reasons behind this difference and identify potential strategies to attract more female customers and increase their contribution to the overall revenue. Additionally, the business could consider conducting surveys or focus groups to gain insights into the purchasing behaviour and preferences of their male and female customers. By gaining a deeper understanding of their customer base, the business can tailor their sales and marketing strategies to better meet the needs of both male and female customers, ultimately leading to increased revenue and customer satisfaction.

The line graph below in Figure 2 provides a visualization of the total profit generated from male (M) and female (F) customers, grouped by age:

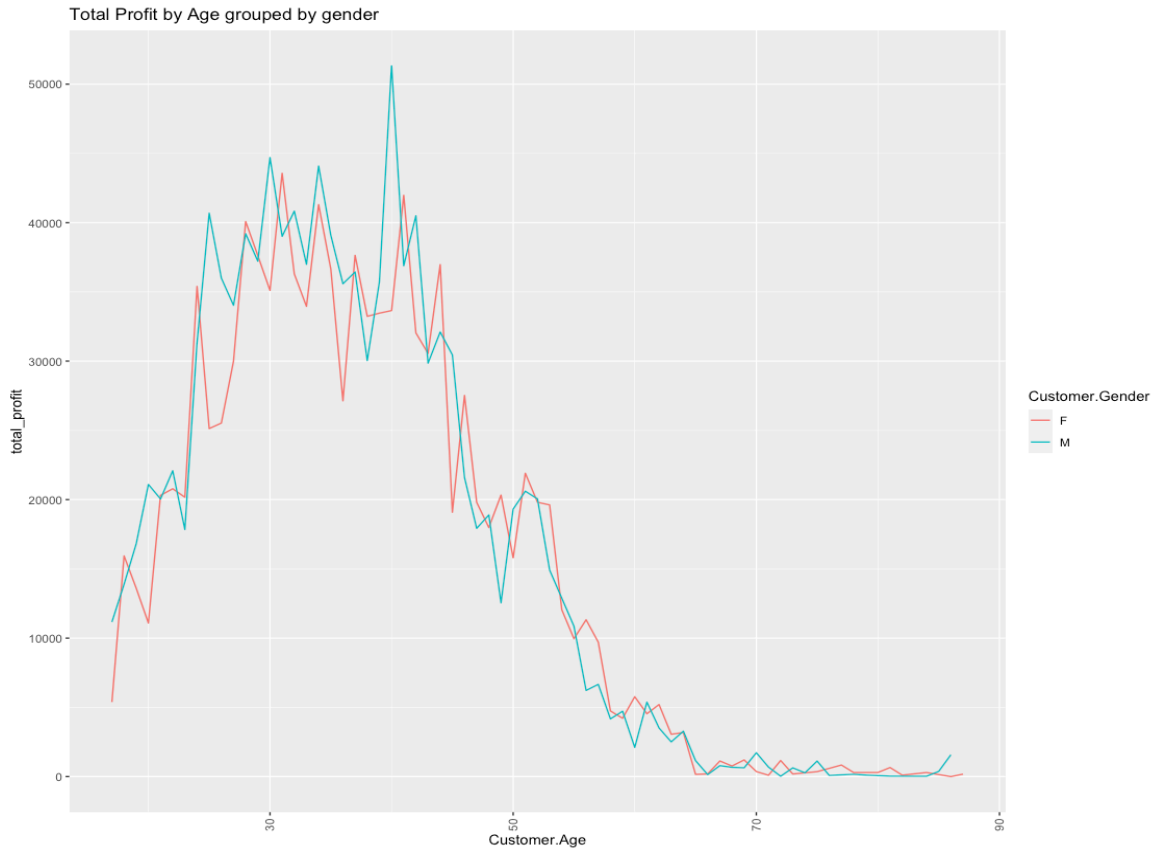

Figure 2: Total profit generated from customers grouped by age.

The analysis shows that the highest total profit is generated from customers aged 25-45, with the majority of the profit coming from male customers. After the age of 45, the profit declines sharply, with a continued decrease for customers in the 50-65 age group. Customers aged 60 and above generate the minimum profit or no profit at all. It is worth noting that there is a significant difference in profit generated between male and female customers across all age groups. Further investigation into the reasons behind this gender-based difference may be necessary to develop effective strategies to address this issue. Additionally, the business could consider offering personalized products or services tailored to the specific needs and preferences of customers in the 25-50 age group to further increase their profits. These findings suggest that the business should focus on attracting and retaining customers within the age group of 25-50 to maximize profits. It may be beneficial for the business to consider

implementing marketing strategies that specifically target this age group to maintain their loyalty and encourage repeat purchases. Furthermore, the business could investigate the reasons behind the decline in profit among older customers and identify ways to improve their shopping experience to potentially increase profits from this demographic. Overall, the data suggests that a targeted approach to sales and marketing efforts could lead to increased profits for the business.

Figure 3 provides insights into the total profit generated by product categories in 2015 and 2016:

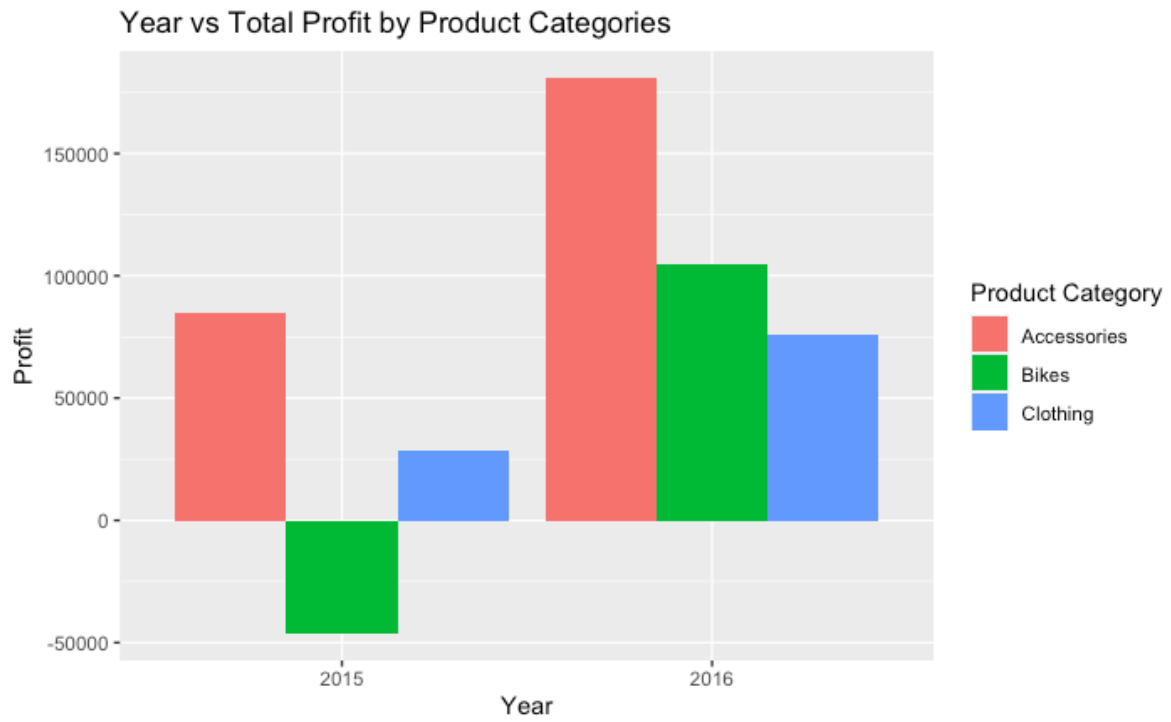

***Figure 3**:* Total Profit generated by product categories in 2015 and 2016.

The analysis shows that in both years, the highest total profit was generated from accessories. This suggests that accessories are a significant contributor to the company's overall profitability. However, in 2015, the company's total profit generated from bikes was negative, indicating not only a lack of profit but also a loss of nearly 50,000 USD. This could be attributed to a variety of factors such as low demand, high production costs, or pricing strategies. Nonetheless, the company managed to turn things around in 2016, with the profit generated from bikes being the second-highest after accessories, making up more than 100,000 USD. This indicates that the company was able to successfully address the issues that led to the negative profit in 2015 and turn it into a profitable category in 2016. Another interesting observation is the significant increase in profit generated from clothing in 2016 compared to 2015. The profit generated from clothing almost doubled in 2016, indicating a successful shift in the company's focus or strategy to capture a larger market share in the clothing category. Overall, the analysis of Figure 3 suggests that the company's product mix and profitability can vary significantly across different product categories and time periods. The company should monitor these trends closely and adjust their strategies accordingly to maximize profitability across all product categories.

Below, Figure 4 provides a visual representation of the total number of items sold in each sub-category:

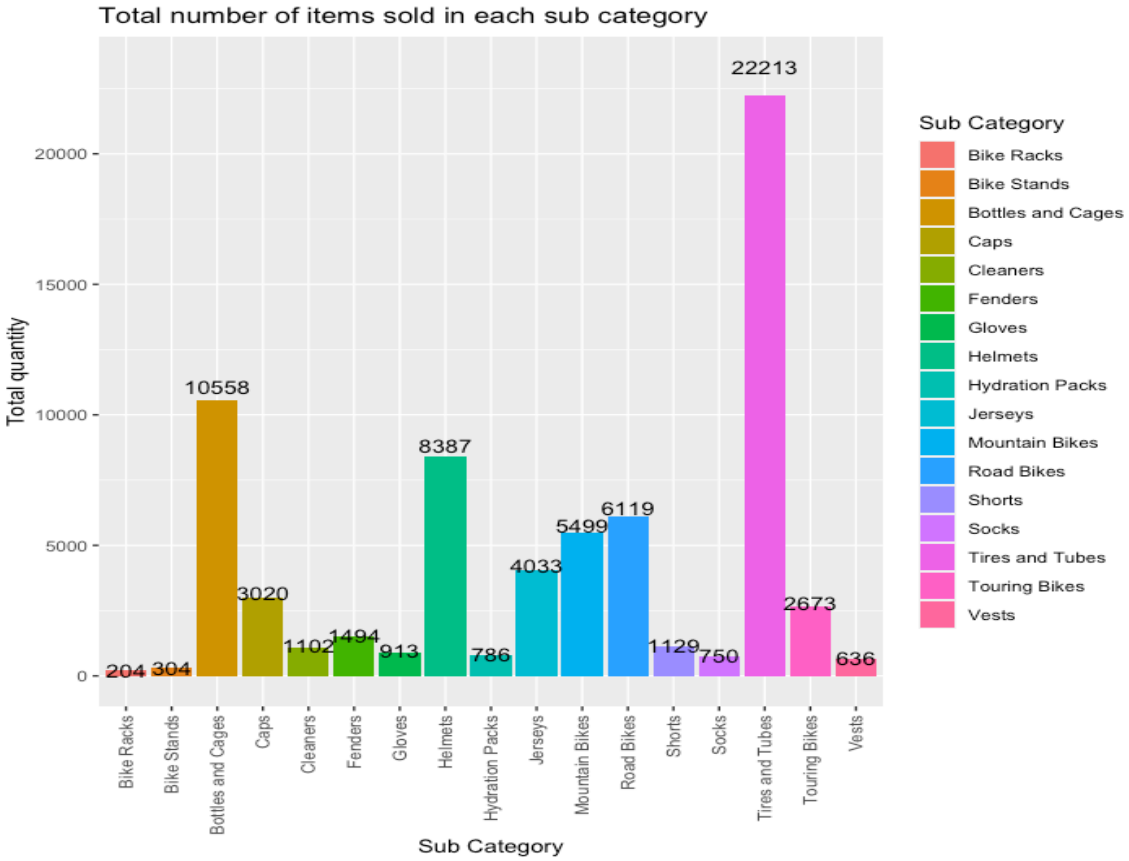

Figure 4: Total number of items sold in each sub-category.

The analysis reveals that tires and tubes were the highest sold sub-category, with a total of 22,213 items sold. This suggests that there is a high demand for tires and tubes among the company's customer base. Bottles and cages were the second-highest sold sub-category, with a total of 10,558 items sold. This indicates that customers also value water bottles and cages as important accessories for their bikes. Helmets were the third-highest sold sub-category, with a total of 8,387 items sold. This suggests that the company's efforts to promote the importance of wearing a helmet to ensure safety and avoid accidents have been successful. Road bikes were the next highest sold sub-category, with a total of 6,119 items sold, followed by mountain bikes with 5,499 items sold. This suggests that the demand for bikes is relatively high among the company's customer base. Jerseys, caps, and touring bikes also had significant sales figures, with 4,033, 3,020, and 2,673 items sold, respectively. This indicates that customers value both apparel and different types of bikes for their cycling needs. The sub-categories with the least number of items sold were bike racks and bike stand, with 204 and 304 items sold, respectively. This suggests that these accessories may not be as popular among the company's customer base or that the company may need to work on improving its marketing efforts for these products. Overall, the analysis of Figure 4 highlights the importance of understanding customer preferences and needs when developing marketing strategies for different sub-categories. By identifying which sub-categories are the most popular among its customer base, the company

can better tailor its product offerings and promotional efforts to maximize sales and profitability.

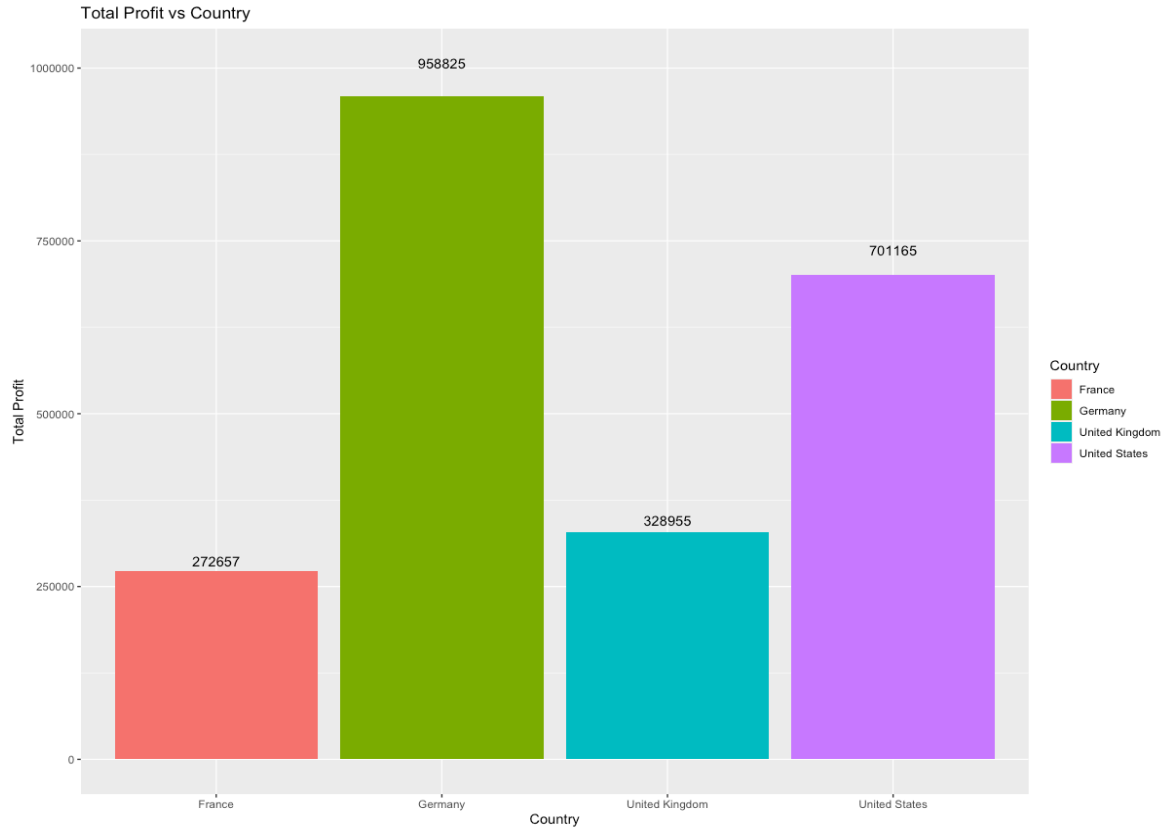

***Figure 5**:* Graph of Total Profit generated by country.

From the graph shown in Figure 5, it can be seen that the highest profit generated was in Germany, with a total of 958825 USD, which was more than triple the profit generated in France, which was the lowest among the countries listed, with a total of 272657 USD. The United States came in second with a profit of 701165 USD, which was nearly double the profit generated in the United Kingdom at 328955 USD. Overall, it is clear that there is a significant difference in the profit generated by country, with Germany and the United States being the most profitable, and France being the least profitable. It may be worthwhile for the company to investigate the reasons behind the discrepancies in profit generated by country and consider strategies to improve sales performance in less profitable regions.

4. DISCUSSION

The results of this project have provided valuable insights into customer spending habits and their impact on sales performance. From the visual representation of revenue by gender, it is clear that male customers generated higher revenue than female customers. This raises the question of why this difference exists and what can be done to increase revenue from female customers. Perhaps there is a need to re-evaluate the marketing strategies targeted towards women to increase their interest in the products. Also suggests that the company may want to consider targeting their marketing efforts towards male customers to further increase revenue. Moreover, Figure 2 shows that the age group between 25 and 45 generates the highest total profit, with male customers being the primary contributors. This information can help the

company to target their marketing and sales efforts towards this demographic in order to maximize profit. Additionally, the sharp decline in profit after the age of 45 indicates a need for further research into the buying behaviors of this age group to identify any potential reasons for the decline. The analysis of profit generated by product category and year revealed some interesting trends as well. Accessories were the highest-selling product in both 2015 and 2016, and Figure 3 shows that the company's profit generated from bikes had a significant negative impact in 2015. However, the profit from bikes increased substantially in 2016, which could suggest that changes were made to improve the product or marketing strategy. This raises the question of what changes were made and how they affected the profit generated from bikes. This information can help the company to focus their inventory and marketing efforts towards these profitable product categories. Lastly, the analysis of profit generated by country showed that Germany and the United States were the most profitable regions, while France was the least profitable. This suggests that the company may want to focus their sales and marketing efforts towards these profitable regions and investigate the reasons behind the discrepancies in profit generated by country.

Here we can list the summarized the answers for the research questions of this project:

- The analysis revealed that male customers generated higher revenue compared to female customers in both overall revenue and profit by gender in product categories and countries. Additionally, Germany had the highest total profit generated, while France had the lowest among the countries analyzed.
- Profit varied significantly across product categories and countries, with accessories being the highest profit-generating category in both 2015 and 2016. Gender had a significant impact on profit, with male customers generating more profit than female customers, particularly in the 25-45 age range.
- The trend of total profit over the years was not significant, but there were fluctuations across product categories. For example, profits from bikes were negative in 2015 but increased significantly in 2016. Clothing sales also saw a considerable increase in profit from 2015 to 2016.
- The total number of items sold varied significantly across sub-categories, with tires and tubes being the highest-selling category, followed by bottles and cages and helmets. There were also differences across gender and geographic segments.
- The visual representation of revenue, profit, and sales data can be used to identify areas of potential improvement in sales and marketing strategies. For example, targeting the 25-45 age range, which generates the highest profits, or focusing on selling more accessories, which are the highest profit-generating category, could potentially improve overall sales performance. Additionally, analyzing sales data by country and gender could help identify areas for targeted marketing efforts.

Overall, the results provide valuable insights into the company's customer spending habits and their impact on sales performance which can be used to inform future business decisions. It highlights the importance of analyzing customer data to identify areas of improvement, develop targeted marketing and sales strategies, and capitalize on opportunities for growth.

5. CONCLUSION

In conclusion, this project aimed to analyze customer spending habits to improve sales performance. Through the use of quantitative research design, we were able to obtain and analyze data from Kaggle to answer our research questions. The results of this study have provided valuable insights into customer spending habits and profit generated by different factors. We found that male customers generate higher total revenue than female customers, with male customers generating $11,400 compared to $10,900 for female customers. Furthermore, the highest total profit was generated by customers aged 25-45, mostly from male customers. Total profit generated by product categories showed that accessories were the highest-selling item in both 2015 and 2016, followed by bikes in 2016, which had negative profit in 2015. Clothing also showed a significant increase in profit from 2015 to 2016. The graph of total profit generated by country showed that Germany had the highest profit generated, with triple the amount generated in France, followed by the United States and the United Kingdom. The graph of total number of items sold in each sub-category showed that tires and tubes were the highest-selling item, followed by bottles and cages, and helmets.

Overall, the results of this study provide important insights for businesses looking to improve their sales performance. By understanding customer spending habits and identifying top-selling items and demographics, businesses can make informed decisions regarding product development, marketing strategies, and customer targeting. The findings of this study can be used to optimize sales performance and drive growth in the business.

However, it is important to note that this study has its limitations. The data used in this study was obtained from Kaggle, and as such, may not be representative of the broader population. Additionally, the study only analyzed profit generated by certain factors, and other factors such as marketing expenses and production costs were not considered. Future research could address these limitations by using a more representative sample and analyzing additional factors that impact sales performance. Nonetheless, the findings of this study have significant implications for businesses looking to improve their sales performance and remain competitive in today's marketplace.

Acknowledgement: I would like to express my heartfelt gratitude to Professor Pei Lei for his invaluable guidance and support in teaching me how to use the R programming language. Without his patient and expert instruction, the current project would not have been possible. His passion for the subject and dedication to teaching have been truly inspiring, and I feel fortunate to have had the opportunity to learn from him. I am deeply grateful for his encouragement, feedback, and willingness to go above and beyond in assisting me throughout the project.
Once again, thank you, Professor Pei Lei, for your exceptional teaching and unwavering support.

References

Bain & Company. (2019). Personalization Matters: How Brands Can Increase Customer Loyalty. Retrieved from https://www.bain.com/insights/personalization-matters-how-brands-can-increase-customer-loyalty/

Hoang, D. T., & Nguyen, H. T. (2021). Analyzing customer purchasing behavior in the e-commerce industry using transaction data analysis. Journal of Open Innovation: Technology, Market, and Complexity, 7(2), 46.

KPMG. (2021). Unlocking Value from Data: The Business Impact of Data-Driven Insights. Retrieved from https://assets.kpmg/content/dam/kpmg/pdf/2021/03/unlocking-value-from-data.pdf

McKinsey & Company. (2020). How Personalization Drives Revenue Growth for Retailers. Retrieved from https://www.mckinsey.com/industries/retail/our-insights/how-personalization-drives-revenue-growth-for-retailers

DATA INPUT

I have analyzed and created visualization by writing these codes in R programming language below:

demo<-read.csv("SalesForCourse_quizz_table.csv")
library(ggplot2)
library(dplyr)
#Data Manipulation

```r
#Data Manipulation
demo1<-demo %>%
  transmute(Date,Year,Customer.Age,Customer.Gender,Country,Product.Category,Sub.Category,Quantity,Cost,Revenue,Profit=Revenue-Cost)%>%
  filter(Year %in% c(2015,2016) )
#Data Visualization
library(lubridate)
demo1$Date<- format(as.Date(demo1$Date, "%m/%d/%y"), "%Y-%m-%d")
demo1$month <- month(as.Date(demo1$Date))
demo2<-demo1 %>%
  group_by(Year,month,Product.Category)  %>%
  transmute(Year,month,Product.Category,total_profit=sum(Profit))%>%
  distinct(.keep_all = TRUE)
#BAR
ggplot(demo2,aes(x = as.factor(Year), y = total_profit,fill=Product.Category)) +
  geom_bar(stat = "identity",position="dodge") +
  labs(x = "Year", y = "Profit", fill = "Product Category")+
  ggtitle("Year vs Total Profit by Product Categories")
#Revenue generated from female and male customers
demo1%>%
  group_by(Customer.Gender) %>%
  summarize(total_revenue=sum(Revenue))%>%
  ggplot(aes(x = Customer.Gender, y = total_revenue,fill=Customer.Gender)) +
  geom_bar(stat = "identity",position="dodge") +
  labs(x = "Gender", y = "Total Revenue", fill = "Customer.Gender")+
  ggtitle("Gender vs Total Revenue")+
  geom_text(aes(label=total_revenue),position = position_stack(vjust = 1.02))
#Total profit generated from all the categories in each country
demo1%>%
  group_by(Country) %>%
  summarize(total_profit=sum(Profit))%>%
  ggplot(aes(x = Country, y = total_profit, fill=Country)) +
  geom_bar(stat = "identity",position="dodge") +
  labs(x = "Country", y = "Total Profit", fill = "Country")+
  ggtitle("Total Profit vs Country")+
  geom_text(aes(label=total_profit),position = position_stack(vjust = 1.05))
#Total number of items sold in each sub category
demo1 %>%
  group_by(Sub.Category) %>%
  summarize(total_quantity=sum(Quantity))%>%
  ggplot(aes(x = Sub.Category, y = total_quantity, fill=Sub.Category)) +
  geom_bar(stat = "identity",position="dodge") +
  labs(x = "Sub Category", y = "Total quantity", fill = "Sub Category")+
  ggtitle("Total number of items sold in each sub category")+
  geom_text(aes(label=total_quantity),position = position_stack(vjust = 1.05))+
  theme(axis.text.x=element_text(angle=90,vjust=0.5,hjust=1))
#Total Profit by Age grouped by gender
demo1 %>%
  group_by(Customer.Gender,Customer.Age) %>%
  summarize(total_profit=sum(Profit))%>%
  ggplot(aes(x = Customer.Age, y = total_profit, color=Customer.Gender)) +
  geom_line() +
  labs(x = "Customer.Age", y = "total_profit", fill = "Customer.Gender")+
  ggtitle("Total Profit by Age grouped by gender")+
  theme(axis.text.x=element_text(angle=90,vjust=0.5,hjust=1))
```