UNIVERSITY INSTITUTE of COMPUTING
Asia's Fastest Growing University

Sales Analysis

Submitted by

Boby Chauhan
25MCD10010
25MCD2/A

Submitted to

Mr. Amarjeet Arora
(E4999)

In partial fulfilment for the award of **the degree of Master of Computer Application**

CHANDIGARH UNIVERSITY
Discover. Learn. Empower.

Chandigarh University

Aug 2025 – Nov 2025

UNIVERSITY INSTITUTE of COMPUTING
Asia's Fastest Growing University

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to Mr. Amarjeet Arora for he continuous guidance, encouragement, and support throughout this project. Her insightful suggestions, constructive feedback, and expert guidance helped me understand the concepts of data handling, statistical analysis, and visualization more deeply.

We wish to reciprocate in full measure the kindness shown by Dr.Krishan Tuli (H.O.D, University Institute of Computing) who inspired us with his valuable suggestions in successfully completing the project work.

I have gained practical knowledge in managing and analyzing sales data. Working on this project allowed me to explore how real-world business datasets can be processed and visualized to generate actionable insights.

This project has provided me with hands-on experience in RStudio, enabling me to perform data aggregation, visualization, and interpretation of sales trends, product performance, regional sales, and customer contributions.

I am confident that the skills and knowledge I have gained through this project will be highly valuable for my career in data science and analytics.

This project has strengthened my skills, knowledge, and confidence, motivating me to aim higher in future.

Date: 27.10.2025

Place: Chandigarh University, Mohali, Punjab

By: Boby Chauhan , UID – 25MCD10010

UNIVERSITY INSTITUTE of COMPUTING
Asia's Fastest Growing University

ABSTRACT

The Sales Analysis project provides an in-depth study of sales data to understand sales revenue, product performance, customer behavior, and regional sales trends. The dataset contains 500 simulated sales records, including OrderID, Date, Customer, Region, Product, Quantity, UnitPrice, and TotalSales.

Using R and RStudio, the project performs data cleaning, aggregation, visualization, and interpretation. Visualizations such as line charts, area charts, bar charts, and pie charts help identify high-performing products, peak sales periods, top customers, and regional sales patterns.

The analysis enables data-driven decision-making, supports business strategies, inventory planning, and marketing campaigns, and demonstrates the practical application of R programming for real-world analytics.

INTRODUCTION

In today's competitive business environment, analyzing sales data is crucial for operational efficiency and strategic decision-making.

Organizations generate massive amounts of data from transactions, customer purchases, and product sales. Without proper analysis, this data remains underutilized and does not contribute to business growth.

The Sales Analysis project converts raw sales data into actionable insights. By evaluating historical sales, organizations can:

- Identify top-performing products and categories
- Understand customer purchasing behavior and preferences
- Analyze regional sales performance
- Forecast demand and optimize inventory
- Plan marketing campaigns and promotions effectively

The dataset contains records, including OrderID, Date, Customer, Region, Product, Quantity, UnitPrice, and TotalSales.

Using RStudio, this project applies tidyverse and ggplot2 for data cleaning, aggregation, and visualization.

Visual representation plays a key role. Line and area charts show daily sales trends, bar charts compare product and regional performance, and pie charts display customer and product contributions. These insights allow management to make informed decisions, improve sales strategies, and allocate resources efficiently.

UNIVERSITY INSTITUTE of COMPUTING
Asia's Fastest Growing University

OBJECTIVE

1. **Analyze Sales Revenue:**

    - Calculate total sales revenue across products, regions, and customers.
    - Identify high-performing products and regions to determine key revenue drivers**.**
    - Understand revenue trends over time to forecast future sales potential.

2. **Understand Customer Behavior:**

    - Segment customers based on purchase patterns, frequency, and total spending.
    - Identify top customers contributing to overall revenue.
    - Understand customer preferences to improve marketing strategies and promotions.

3. **Evaluate Product Performance:**

    - Compare performance of different products using sales volume, revenue, and profitability.
    - Determine best-selling and low-performing products**.**
    - Provide insights for inventory management, procurement, and production planning**.**

4. **Regional Sales Analysis:**

    - Compare sales across different regions to identify strong and weak markets.
    - Support regional marketing campaigns and sales strategies**.**
    - Help in allocation of resources and distribution planning**.**

5. **Visualize Sales Data Effectively:**

    - Use line charts, area charts, bar charts, and pie charts to represent sales trends, product performance, and customer contribution.
    - Make complex datasets easier to interpret for managers and stakeholders.
    - Facilitate data-driven decisions through clear graphical representation.

6. **Support Strategic Decision-Making:**

    - Provide insights that can influence marketing campaigns, promotions, and product launches**.**
    - Enable better inventory control, demand forecasting, and business planning.
    - Transform raw sales data into actionable information for informed decision-making**.**

7. **Demonstrate Practical Skills in R:**

    - Apply R programming, tidyverse, lubridate, and ggplot2 for data manipulation and visualization.
    - Gain hands-on experience in cleaning, aggregating, and visualizing real-world datasets**.**

UNIVERSITY INSTITUTE of COMPUTING
Asia's Fastest Growing University

LITERATURE REVIEW

Sales analysis is widely studied in retail, e-commerce, and business analytics, highlighting that effective analysis improves profitability, inventory management, and customer engagement. Key insights:

I. **Importance of Data Visualization:**

Data visualization converts complex datasets into intuitive charts and dashboards, helping managers interpret trends, detect anomalies, and analyze seasonal variations.

Tools like ggplot2 and tidyverse allow managers to quickly identify underperforming products or regions, supporting strategic and timely decision-making. Visualization also improves communication of insights to stakeholders who may not be technically skilled.

II. **Customer Segmentation and Marketing:**

Segmenting customers based on purchase frequency, revenue contribution, product preferences, and geographic location enhances targeted marketing. Research shows effective segmentation increases conversion rates, repeat purchases, and loyalty program success.

By analyzing past sales data, businesses can create personalized promotions and allocate marketing resources efficiently, maximizing ROI.

III. **Regional Performance Analysis:**

Analyzing sales performance across regions helps businesses identify high-performing markets and plan expansion strategies. It also reveals low-performing areas that may require promotions or operational changes.

Regional analysis supports inventory allocation, logistics planning, and marketing campaigns, helping maximize revenue potential.

IV. **Feature Outcome:**

The project implements several features with measurable outcomes:

- *Daily Sales Trend Analysis***:** Detects high-demand and slow sales periods.
- *Product Performance Analysis:* Identifies top-selling and low-performing products.
- *Regional Sales Comparison:* Highlights strong and weak performing areas.
- *Customer Contribution Insights:* Identifies customers generating the highest revenue.
- *Visual Dashboards:* Provides easy-to-read graphical reports for management decision-making.

V. Seasonal Sales Forecasting:

Forecasting sales based on historical data helps predict future demand and seasonal trends. Accurate forecasting assists businesses in preparing inventory, marketing strategies, and workforce allocation to meet anticipated demand.

This reduces the risk of overstocking or stockouts during peak or off-peak seasons, ensuring smoother business operations.

VI. Profit Margin and Pricing Evaluation:

Analyzing profit margins enables organizations to determine which products or services generate the highest returns. Such insights support pricing optimization, discount strategies, and cost control.

By understanding revenue versus cost behavior, businesses can prioritize profitable products and eliminate underperforming ones

VII. Customer Retention and Loyalty Analysis:

Sales data enables tracking of customer purchase behavior and loyalty patterns. Identifying repeat customers and understanding their preferences helps design loyalty programs and personalized marketing campaigns.

This approach improves customer satisfaction, retention rate, and long-term revenue growth.

VIII. Sales Channel Performance:

Comparing performance across different sales channels—such as online stores, retail outlets, and distributors—reveals where sales are most effective.

This analysis guides management in reallocating resources, optimizing marketing efforts, and improving underperforming channels.

IX. Inventory Turnover and Demand Optimization:

Integrating sales data with inventory management helps monitor stock movement and turnover rates.

Efficient demand analysis ensures that fast-moving products remain well-stocked while minimizing excess inventory of slow sellers. This balance reduces operational costs and enhances overall supply chain efficiency.

UNIVERSITY INSTITUTE of COMPUTING
Asia's Fastest Growing University

REQUIREMENTS

1. Software Requirements:

a) **R** – Open-source programming language for statistical computing and data analysis.
b) **RStudio IDE** – Integrated development environment for writing, running, and debugging R code with visualization support.

2. Package Requirements:

a) **tidyverse** – Collection of packages for data manipulation and visualization.
b) **dplyr** – used for easy, fast, and readable data manipulation to filter, select, arrange, mutate, and summarize data efficiently.
c) **ggplot2** – Creates high-quality charts such as line charts, bar charts, area charts, and pie charts.

3. Dataset Requirements:

a) **Sales_Analysis.csv** – Contains 500 records with columns:

- OrderID, Date, Customer, Region, Product, Quantity, UnitPrice, TotalSales
- Enables detailed analysis of sales trends, product performance, and customer contribution.

4. System Requirements:

a. **Operating System:** Windows 10/11 or macOS
b. **RAM:** Minimum 4GB
c. **Storage:** Minimum 50GB
d. **Internet:** Required for downloading R, RStudio, and packages

CODE

 # Sales Analysis Dashboard in R

```
# 1. Load required libraries
install.packages(c("tidyverse", "dplyr", "ggplot2))
library(tidyverse)
library(dplyr)
library(ggplot2)

# 2. Read the CSV file
sales <- read.csv("Sales_Analysis.csv")
sales$Date <- as.Date(sales$Date)

# 3. Data Overview
head(sales)
summary(sales)
total_sales <- sum(sales$TotalSales)
cat("Total Sales Revenue: ", total_sales, "\n")
total_quantity <- sum(sales$Quantity)
cat("Total Quantity Sold: ", total_quantity, "\n")
customer_summary <- sales %>%
  group_by(Customer) %>%
  summarise(TotalOrders = n(),
            TotalSales = sum(TotalSales),
            TotalQuantity = sum(Quantity))
print(customer_summary)

# 4. Sales Trend over Time
sales_trend <- sales %>%
  group_by(Date) %>%
  summarise(DailySales = sum(TotalSales))
ggplot(sales_trend, aes(x=Date, y=DailySales)) +
  geom_line(color="blue", size=1) +
geom_area(fill="lightblue", alpha=0.4) +
  ggtitle("Daily Sales Trend") +
xlab("Date") + ylab("Sales Revenue") +
  theme_minimal()

# 5. Sales by Product (Bar Plot)
sales_by_product <- sales %>%
  group_by(Product) %>%
  summarise(TotalSales = sum(TotalSales))
```

```
ggplot(sales_by_product, aes(x=Product, y=TotalSales, fill=Product)) +
  geom_bar(stat="identity") +
  ggtitle("Total Sales by Product") +
  xlab("Product") + ylab("Sales Revenue") +
  theme_minimal()
```

 # 6. Sales by Region (Bar Plot)

```
sales_by_region <- sales %>% group_by(Region) %>%
  summarise(TotalSales = sum(TotalSales))
ggplot(sales_by_region, aes(x=Region, y=TotalSales, fill=Region)) +
  geom_bar(stat="identity") +
  ggtitle("Total Sales by Region") +
  xlab("Region") + ylab("Sales Revenue") +
  theme_minimal()
```

#7.Product Distribution (Pie Chart)

```
sales_product_dist <- sales %>% group_by(Product) %>%
  summarise(TotalSales = sum(TotalSales))
ggplot(sales_product_dist,aes(x="", y=TotalSales,
 fill=Product)) + geom_bar(stat="identity", width=1) +
coord_polar("y", start=0) +
ggtitle("Sales Distribution by Product") +
theme_void()
```

#8 . Customer Segment Distribution (Pie Chart)

```
customer_dist <- sales %>% group_by(Customer) %>%
summarise(TotalSales = sum(TotalSales))
ggplot(customer_dist, aes(x="", y=TotalSales, fill=Customer)) +
geom_bar(stat="identity", width=1) +
coord_polar("y", start=0) +
ggtitle("Sales Distribution by Customer") +
theme_void()
```

#9. Total Quantity by Product (Bar Plot)

```
quantity_by_product <- sales %>% group_by(Product) %>%
summarise(TotalQuantity = sum(Quantity))
ggplot(quantity_by_product, aes(x=Product, y=TotalQuantity, fill=Product)) +
geom_bar(stat="identity") +
ggtitle("Total Quantity of Product sold ") +
xlab("Product") + ylab("Quantity") +
theme_minimal()
```

UNIVERSITY INSTITUTE of COMPUTING
Asia's Fastest Growing University

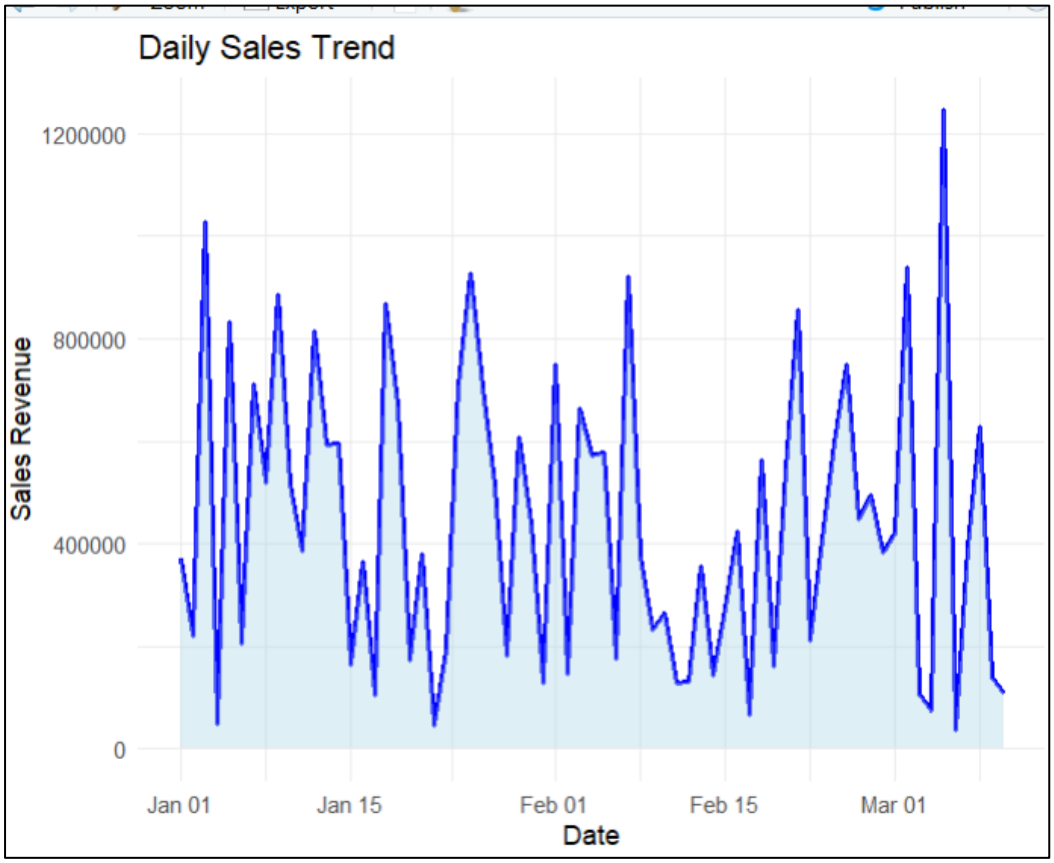
RESULT/OUTPUT

#3 Data Overview

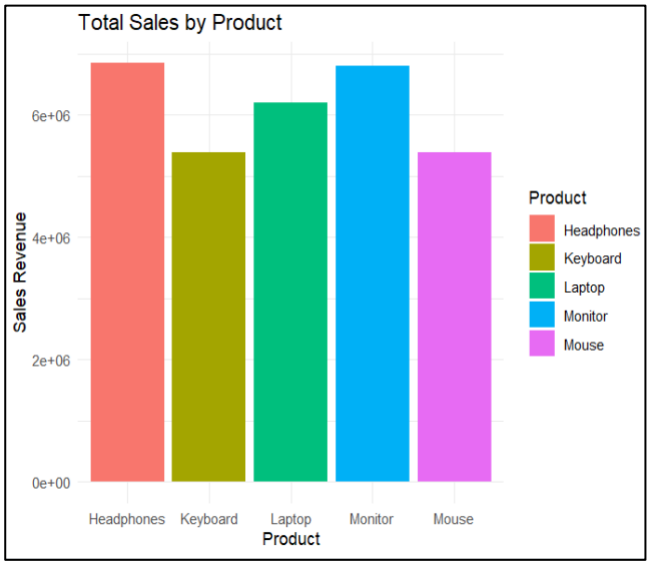
```
R 4.5.1 · C:/Users/boby/OneDrive/Desktop/B/mini project/r/mini project for R/
> # 3. Data Overview
> head(sales)
  OrderID       Date Customer Region    Product Quantity UnitPrice TotalSales
1       1 2025-02-20      Bob  North Headphones        5      8000      40000
2       2 2025-02-17  Charlie   East     Laptop        3      8000      24000
3       3 2025-01-28      Bob  North      Mouse        9      8000      72000
4       4 2025-02-05    Alice   West Headphones        5      8000      40000
5       5 2025-03-10      Eve   West     Laptop       10       500       5000
6       6 2025-01-16      Bob  North     Laptop        6       500       3000
> summary(sales)
    OrderID           Date              Customer            Region            Product
 Min.   :  1.0   Min.   :2025-01-01   Length:500         Length:500         Length:500
 1st Qu.:125.8   1st Qu.:2025-01-16   Class :character   Class :character   Class :character
 Median :250.5   Median :2025-02-04   Mode  :character   Mode  :character   Mode  :character
 Mean   :250.5   Mean   :2025-02-03
 3rd Qu.:375.2   3rd Qu.:2025-02-22
 Max.   :500.0   Max.   :2025-03-10
    Quantity        UnitPrice       TotalSales
 Min.   : 1.000   Min.   :  500   Min.   :   500
 1st Qu.: 3.000   1st Qu.: 1200   1st Qu.:  4800
 Median : 5.000   Median : 5000   Median : 20000
 Mean   : 5.404   Mean   :11972   Mean   : 61255
 3rd Qu.: 8.000   3rd Qu.: 8000   3rd Qu.: 50000
 Max.   :10.000   Max.   :50000   Max.   :500000
> total_sales <- sum(sales$TotalSales)
> cat("Total Sales Revenue: ", total_sales, "\n")
Total Sales Revenue:  30627500
> total_quantity <- sum(sales$Quantity)
> cat("Total Quantity Sold: ", total_quantity, "\n")
Total Quantity Sold:  2702
```

```
> customer_summary <- sales %>%
+   group_by(Customer) %>%
+   summarise(TotalOrders = n(),
+             TotalSales = sum(TotalSales),
+             TotalQuantity = sum(Quantity))
> print(customer_summary)
# A tibble: 5 × 4
  Customer TotalOrders TotalSales TotalQuantity
  <chr>          <int>      <int>         <int>
1 Alice            108    6678000           592
2 Bob               93    6725300           519
3 Charlie           96    5978900           552
4 Eve              104    5712900           519
5 John              99    5532400           520
```

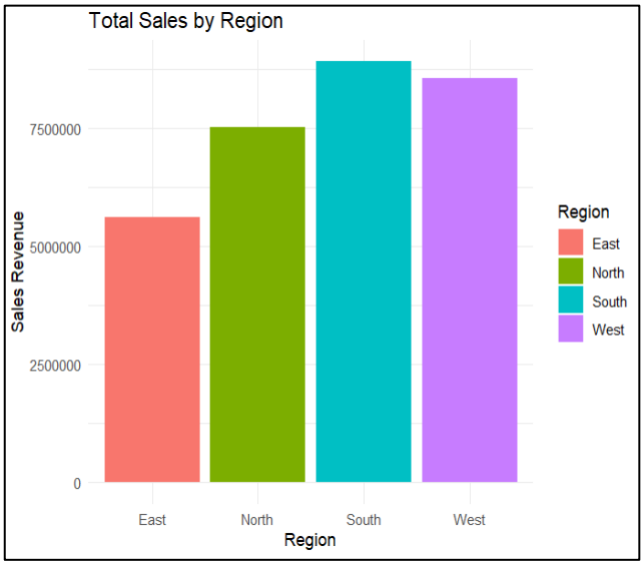
#4 Sales Trend over Time



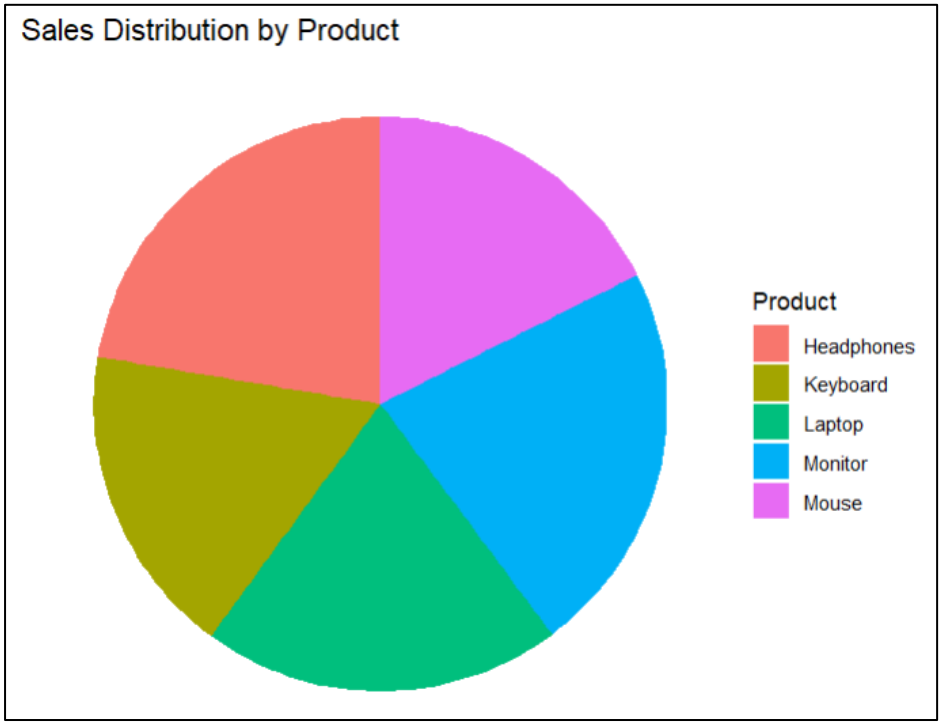
 # 5. Sales by Product (Bar Plot)



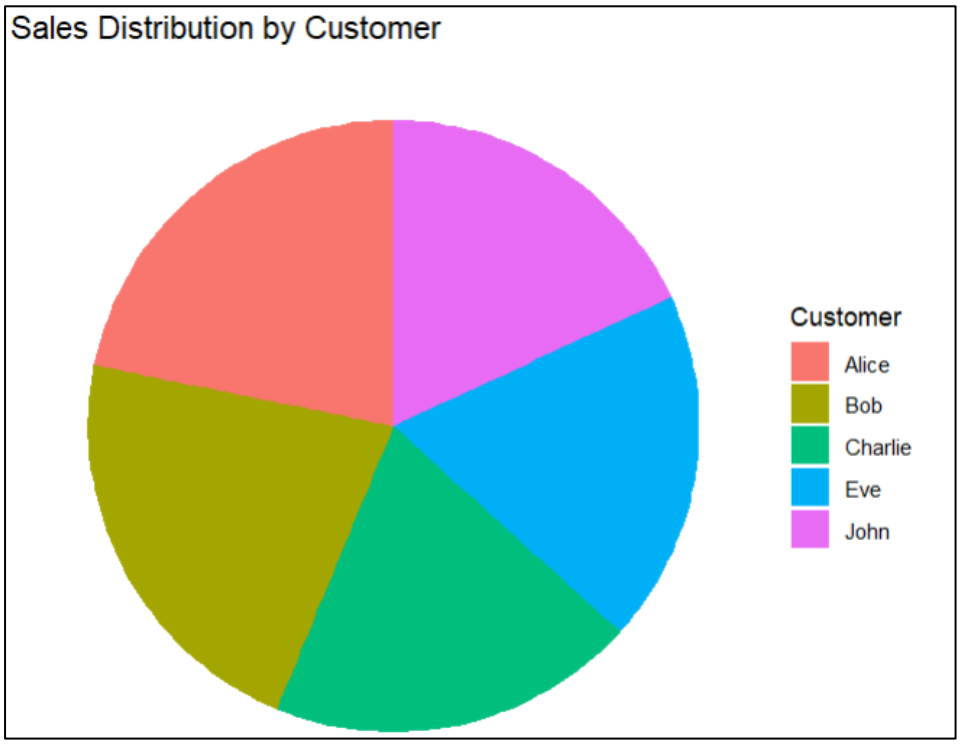
 # 6. Sales by Region (Bar Plot)



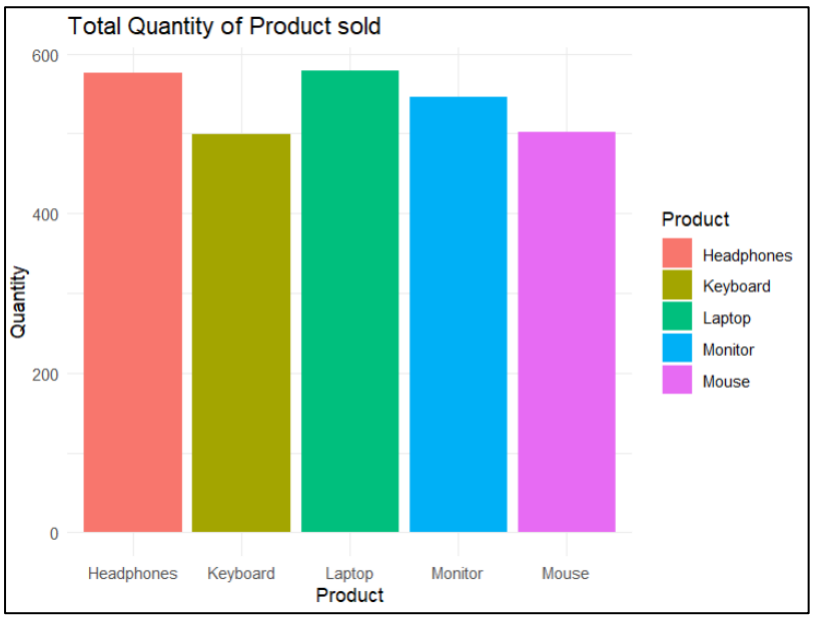
#7.Product Distribution (Pie Chart)



#8. Customer Segment Distribution (Pie Chart)



#9. Total Quantity of Product sold (Bar Plot)



CONCLUSION

The Sales Analysis project successfully demonstrates how data can be effectively collected, processed, and visualized to uncover meaningful insights about sales performance. Using RStudio and libraries such as dplyr, ggplot2, and tidyverse, this project provided a clear understanding of total sales revenue, quantity sold, and customer-wise and product-wise performance.

Through visual tools like bar charts and summary tables, it became easier to identify top-performing products, understand customer purchasing trends, and analyze overall business performance. These findings can help in making strategic business decisions, such as inventory planning, sales forecasting, and customer segmentation.

In conclusion, this project highlights the importance of data analytics in the modern business environment. It not only strengthened my technical understanding of R programming and data visualization but also improved my analytical thinking and problem-solving approach. Future enhancements could include integrating predictive analytics, real-time dashboards, and machine learning models for more advanced sales forecasting and decision-making.

UNIVERSITY INSTITUTE of COMPUTING
Asia's Fastest Growing University

REFERENCE

Git Hub mini project (https://github.com/bcchauhanbc9923-oss/R-lab-mini-project)

Wickham H. et al. (2019)

This book provides a comprehensive guide to data science workflows using R, covering data import, cleaning, transformation, visualization, and modeling. It is an essential resource for understanding tidy data principles and practical approaches to data analysis in R.

Hadley Wickham. *ggplot2:*

Focuses on creating high-quality visualizations in R using the ggplot2 package. It explains grammar of graphics concepts, advanced plotting techniques, and methods to effectively communicate insights through data visualization.

R Documentation: https://www.r-project.org/

The official R website provides detailed documentation for base R functions, packages, and installation guidance. It is a primary reference for understanding R programming language, syntax, and functions.

Lubridate Package Documentation: https://lubridate.tidyverse.org/

Lubridate is an R package designed to simplify date and time manipulation. The documentation provides instructions for parsing, extracting, and performing calculations on date-time objects, which is essential for time-based sales analysis.

Tidyverse Documentation: https://www.tidyverse.org/

Tidyverse is a collection of R packages for data science, including dplyr, ggplot2, tidyr, and readr. The documentation explains data manipulation, cleaning, and visualization techniques, enabling effective handling of large datasets.

Matloff N. (2011).

The Art of R Programming. No Starch Press.
Provides a deeper understanding of R programming, including data structures, functions, and debugging, which is essential for writing efficient and maintainable code for data analysis projects.