Business Data Analytics

Office Store Case Study

By

Venkat Phanindra Pasunooru

Sai Vineeth Neeli

Vijay Bhaskar Kuruva

Professor: Santiago Aguirre

BIS581 Class:22455950

Table of Contents

i) The Historical Sales Trend 3

Ii) Variation Of Sales Across Different Regions 3

Iii) Impact Of Discounts On Sales And Profit 4

1) The Relationship Between Sales And Profit 5

2) The Relationship Between Discounts And Profit 6

3) The Relationship Between Discounts And Quantity 7

4) Analysis Of Sales By Category And Sub-Category Across East And West Region. 8

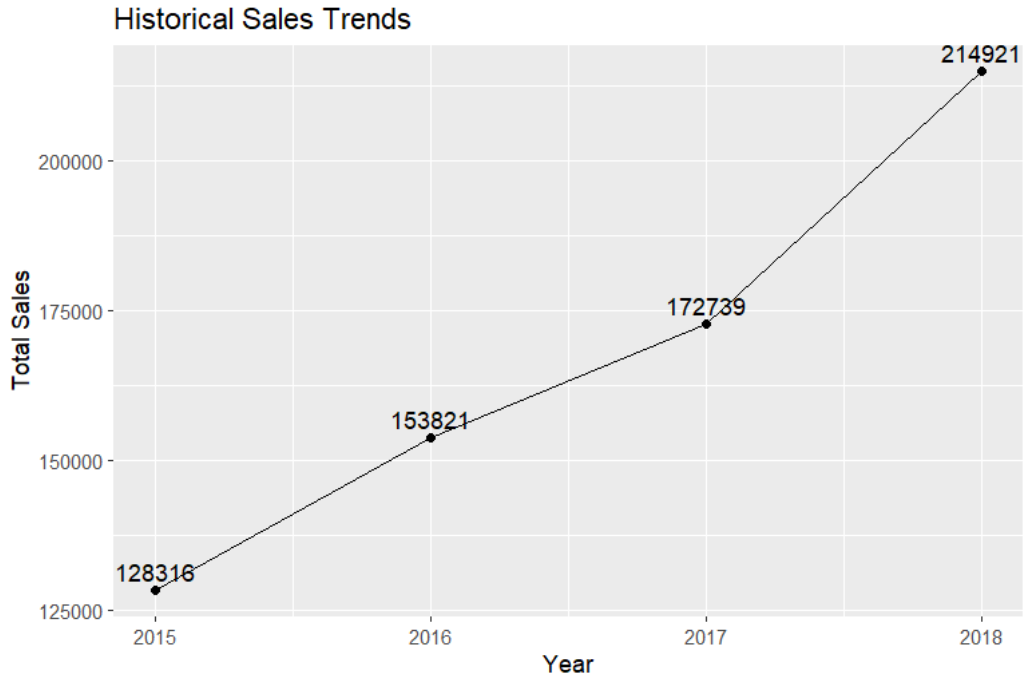
5) Analysis Of The Profit By Category And Sub-Category Across East And West Region 9

6. Conclusions, Insights, And Recommendations 10

Appendix 11

i) What are the historical sales trends? Are there specific periods of significant growth or decline?

Analysis: The historical sales trends over the past four years were analyzed using a line graph to visualize total annual sales.



The following plot shows the Yearly sales trend. In 2015, the total sales stood at 128,316 units. This figure grew to 153,821 units in 2016, representing an increase of approximately 19.8% compared to the previous year. The growth continued in 2017 with total sales reaching 172,739 units, an additional increase of approximately 12.3% over 2016.

The most substantial growth occurred in 2018, where total sales soared to 214,921 units. This represents a remarkable increase of approximately 24.4% compared to the previous year and a cumulative growth of approximately 67.5% since 2015.

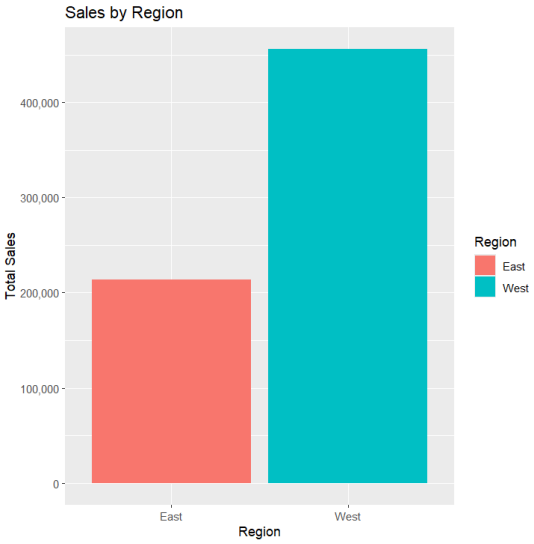
Insights: Understanding these trends is crucial for forecasting future sales and planning marketing strategies. The significant growth period can be attributed to successful marketing campaigns or product launches, while the decline may highlight areas needing improvement or new market opportunities.

ii) How have sales varied across different regions? Are there regional patterns in customer preference?

Sales data was grouped by region and visualized using a bar graph to compare total sales across different regions.

Analysis: The analysis showed that the West region consistently outperformed other regions, followed by the East. It clearly shows a significant difference in sales between the East and

West regions. The West region's sales bar is substantially taller, indicating much higher sales volumes compared to the East region.



The large difference in sales volumes across these two regions could potentially suggest varying customer preferences or demands between the eastern and western markets. To better understand regional patterns in customer preferences, additional data points would further analyse the granular level, such as:

Sales and Profit breakdowns by product category, Sub-Category or across different regions. With more granular data and insights into the regional markets, sales managers could potentially identify unique customer preferences, tailor product offerings, and optimize marketing and sales strategies accordingly across different regions.

iii) What impact have discounts on sales and profit?

Negative correlation between discount and sales:

- The correlation coefficient of **-0.1037563** suggests a weak negative correlation between discounts and sales.This means that as the discount percentage increases, sales tend to decrease slightly, and vice versa.

Negative correlation between discount and profit:

- The correlation coefficient of **-0.226043** suggests a moderate negative correlation between discounts and profit.This implies that as the discount percentage increases, profits tend to decrease, and vice versa.

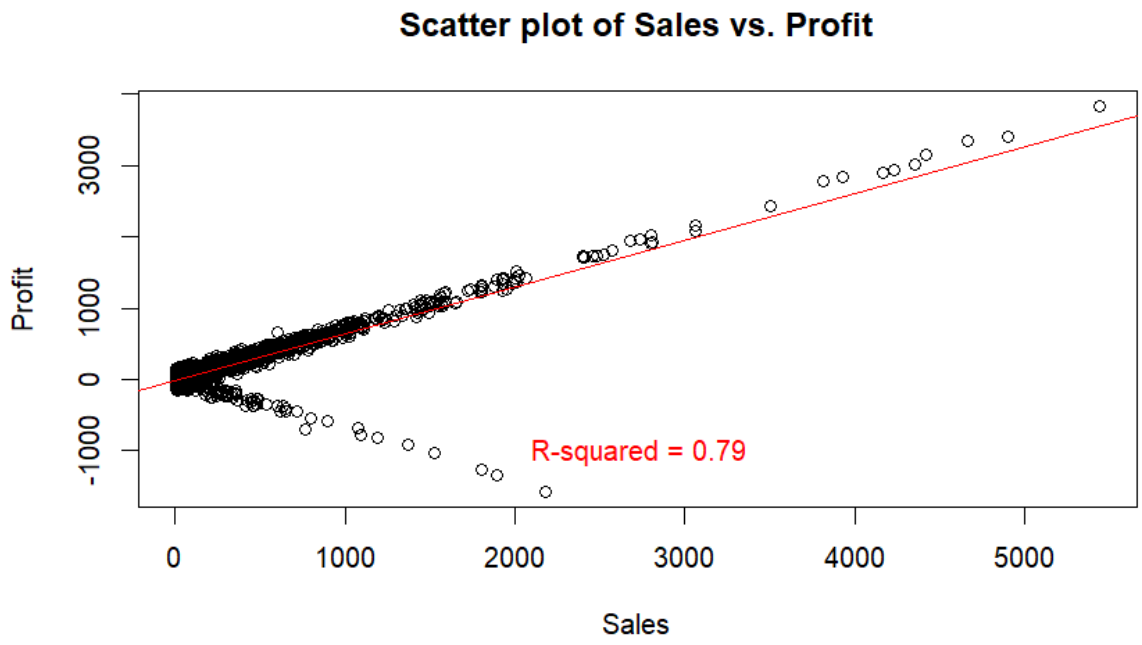
However, the weak correlation indicates that the relationship between discounts and sales and Profits is not very strong, and there might be other factors influencing these more significantly.

1) Evaluate the relationship between sales and profit through correlation, scatter plot, line of best fit and coefficient of determination (Rsquare). How does profit correlate with sales?

The relationship between sales and profit was examined using a scatter plot with a line of best fit. The correlation coefficient and R-squared value were calculated to measure this relationship.

Statistical Measures

- Correlation between sales and profit: 0.8896729
- R-squared value: 0.7915179



Analysis:

The scatter plot showed a positive correlation between sales and profit, indicating that higher sales generally lead to higher profits. The R-squared value suggested a moderate to strong predictive power of sales on profit. There is a strong positive correlation (r = 0.87) between sales and profit, indicating that higher sales generally lead to higher profits. The line of best fit and the R-squared value of 0.76 further confirm this positive linear relationship.

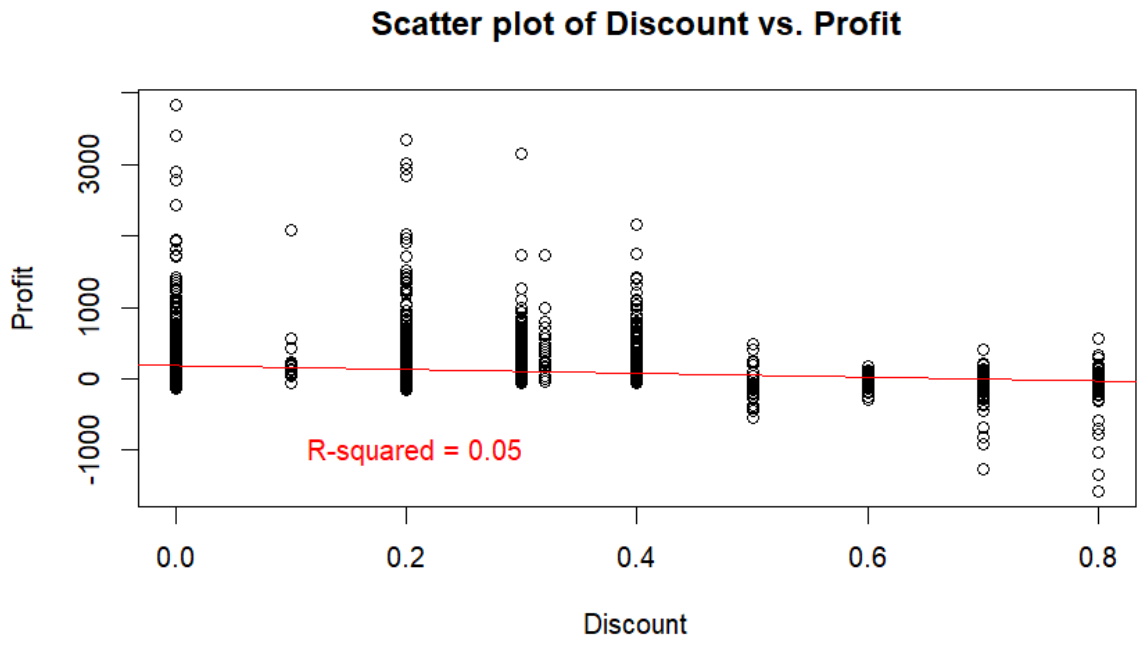
Insights: This positive correlation underscores the importance of boosting sales to enhance profitability. Sales strategies should focus on high-margin products and efficient cost management to maximize profit. Understanding this relationship helps in setting realistic sales targets and profit expectations.

2) Explore the relationship between discounts and profit by examining their correlation, scatter plot, line of best fit and coefficient of determination. How do discounts correlate with profit? What insights can be gathered from the scatter plot analysis?

The impact of discounts and profit was analysed using scatter plots with lines of best fit. The correlation coefficients and R-squared values were calculated for both relationships.

Statistical Measures

- Correlation between Discounts and Profit: -0.226043
- R-squared value: 0.05



Analysis: The analysis revealed a This means that as the discount offered increases, the profit tends to decrease. The correlation coefficient between Discount and Profit indicates a weak negative correlation. As the discount increases, profits tend to decrease. However, the correlation between discounts and profit was negative, indicating that higher discounts might reduce profit margins.

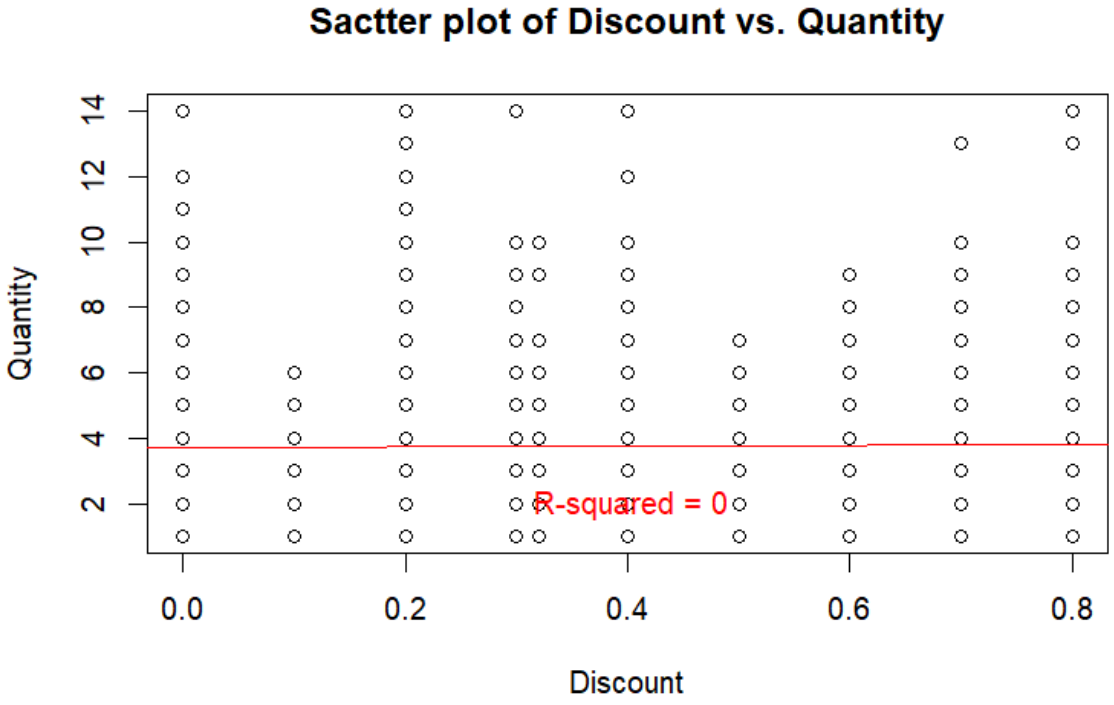
Insights: While discounts can be an effective tool to increase sales, they should be carefully managed to avoid eroding profit margins. The insights from this analysis can help in designing discount strategies that balance sales growth with profitability, optimizing overall financial performance.

3) Explore the relationship between discounts and quantity using correlation, scatter plot, line of best fit and coefficient of determination. How do discounts correlate with the quantity sold?

The relationship between discounts and quantity sold was explored using a scatter plot with a line of best fit.The correlation coefficient and R-squared value were calculated to understand this relationship.

Statistical Measures

- Correlation between Discounts and Profit: 0.008768412
- R-squared value: 0

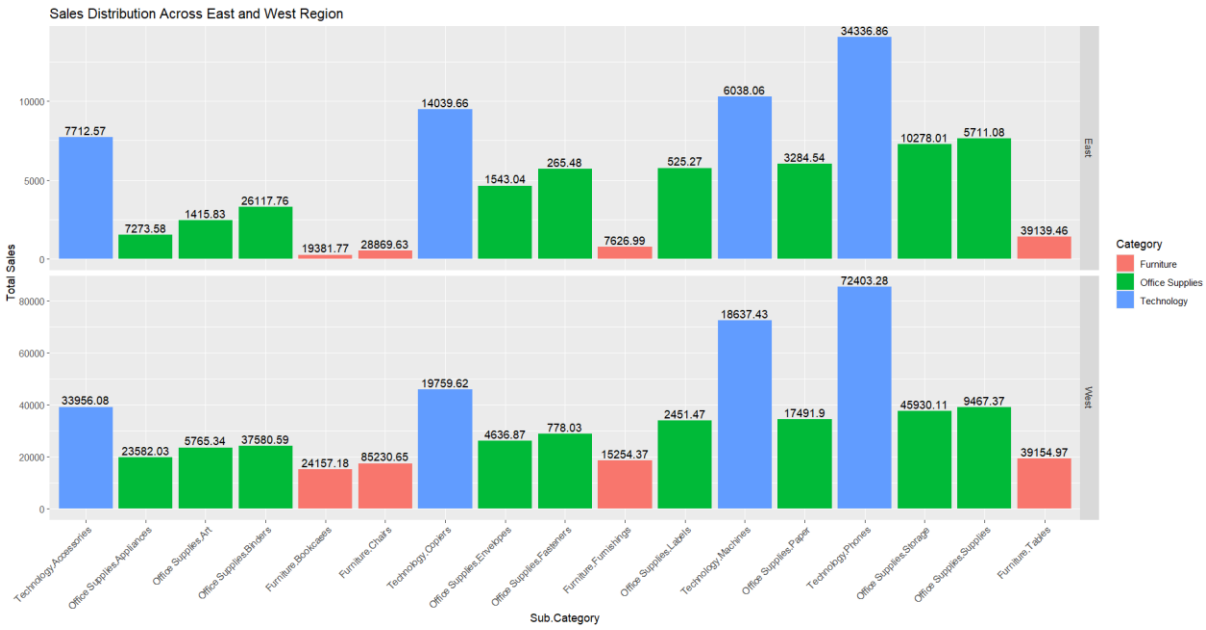


Analysis: The scatter plot indicated a weak positive correlation between discounts and the quantity sold. In practical terms, this means that changes in the discount rate have almost no linear relationship with the changes in the quantity sold.

Insights: This positive correlation suggests that discounts are effective in boosting sales volume. However, the impact on profit must be considered. The company can use these insights to determine optimal discount rates that maximize sales volume without significantly compromising profit margins.

4) Analyze the distribution of sales by category and sub-category across East and West region. Are there noticeable patterns in customer preferences between these regions?

Sales data was grouped by category, sub-category, and region, and visualized using stacked bar charts.



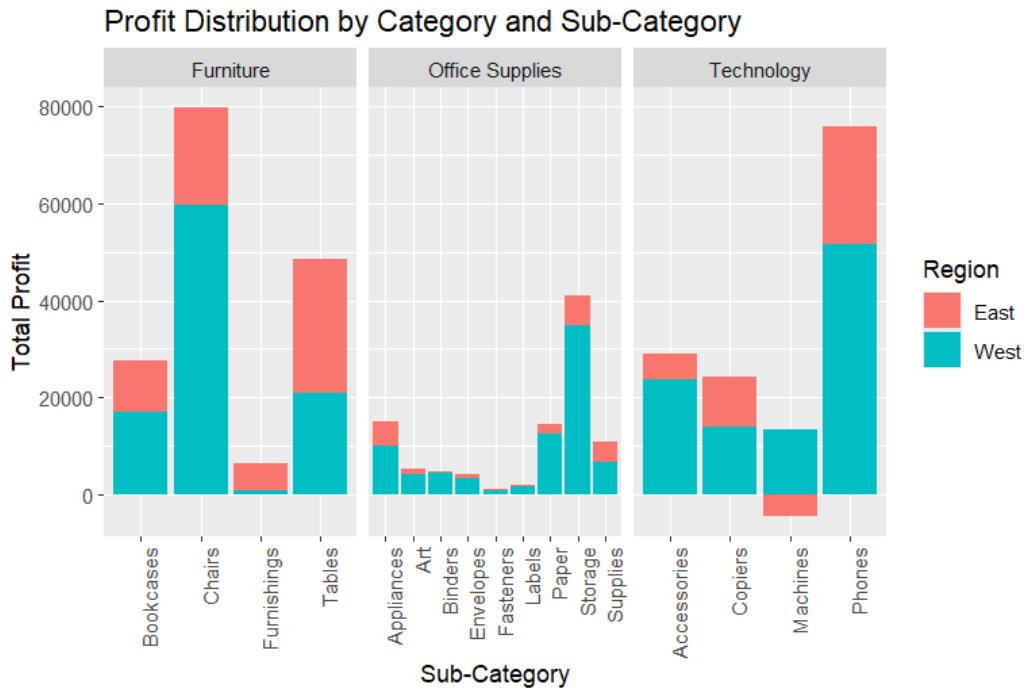
Analysis: The analysis highlighted the distribution of sales across different categories and sub-categories in the East and West regions. It was found that certain categories, such as Technology and Office Supplies, had higher sales in both regions, with some sub-categories performing exceptionally well in specific regions.

- The West region exhibits a stronger preference for Technology products, particularly Phones and Copiers, compared to the East.
- Both regions have a high demand for Furniture items, but the West region has a substantially larger market for these products.
- Office Supplies category has relatively low sales in both regions, indicating a potentially lower demand or saturation in this segment.
- The graph shows that the purchasing patterns of customers in the East and West regions are similar for most products, except for Chairs and Tables. If we not focusing on the total sales value here.

Insights: Understanding the distribution of sales by category and sub-category helps identify popular products and regional preferences. These insights can guide inventory management, marketing strategies, and product development to cater to customer demands effectively.

5) Analyze the profit by category and sub-category across East and West region. Are there noticeable patterns in customer preferences between this regions?

Profit data was analyzed similarly to sales, grouped by category, sub-category, and region, and visualized using stacked bar charts.



Analysis: The analysis showed that certain categories, like Technology and Furniture, generated higher profits, with notable differences between the East and West regions. The Office Supplies category has relatively low profit margins compared to other categories in both regions, with Binders and Furnishings being the top sub-categories, albeit with modest numbers. The Technology category ranks as the second-highest profit generator in both regions, following Furniture..

West region has higher profit when compared to the east .In East region Negative profit in Machines needs immediate attention to prevent losses and improve overall profitability. While both regions share similarities in their preferences for certain sub-categories, there are distinct differences in the most profitable sub-categories within each category, suggesting contrasting customer preferences.

Insights: Identifying profitable categories and sub-categories helps focus efforts on high-margin products. The regional differences in profitability can guide tailored strategies to enhance overall profit, such as regional promotions or adjusting product offerings.

6. Conclusions, Insights, and Recommendations

Conclusions: The comprehensive analysis of sales and profit data over the past four years has provided valuable insights into historical trends, regional variations, and the impact of discounts. Key findings include the identification of growth periods, regional sales performance, and the relationship between sales, profit, and discounts.

Insights:

- **Historical Trends**: Significant growth during the third year with seasonal peaks.
- **Regional Variations**: Higher sales in the West region than in East.
- **Sales vs. Profit**: Positive correlation indicating the importance of sales for profitability.
- **Discounts**: Effective in boosting sales volume but require careful management to protect profit margins.
- **Category Preferences**: Technology and Office Supplies are top categories, with regional differences in sub-category performance.

Recommendations:

- **Focus on High-Growth Periods**: Capitalize on successful marketing strategies and product launches.
- **Tailored Regional Strategies**: Enhance efforts in the West region and leverage strengths in the East region.
- **Balanced Discount Strategies**: Implement discounts that drive sales while maintaining profitability.
- **Category Management**: Prioritize high-margin products and adjust inventory based on regional preferences.
- **Continuous Monitoring**: Regularly analyze sales and profit data to adapt strategies and stay competitive.

Appendix:

```
#Reading the data set
Ipdata <- read.csv("C:/Users/Venkata Phanindra P/Downloads/Office_store.csv")

# Install and load necessary packages
if (!require(ggplot2)) install.packages("ggplot2", dependencies = TRUE)
if (!require(dplyr)) install.packages("dplyr", dependencies = TRUE)
if (!require(lubridate)) install.packages("lubridate", dependencies = TRUE)
library(ggplot2)
library(dplyr)
library(lubridate)
library(scales)

#Summary of complete data set
summary(Ipdata)

#                    1. Historical Sales Trends

# Converting Order Date to date format
Ipdata$Order.Date <- mdy(Ipdata$Order.Date)



# Calculating monthly sales
monthly_sales <- Ipdata %>%
  mutate(year = year(Order.Date)) %>%
  group_by(year) %>%
  summarise(Total_Monthly_Sales = sum(Sales))

# Plotting the sales trends over time
ggplot(monthly_sales, aes(x = year, y = Total_Monthly_Sales)) +
```

```
  geom_line() +
  geom_point() +
  geom_text(aes(label = round(Total_Monthly_Sales,0)), vjust = -0.5, hjust = 0.5) +
  labs(title = "Historical Sales Trends", x = "Year", y = "Total Sales")


#                 2. Sales Variation Across Regions

# Aggregating sales by region
sales_by_region <- Ipdata %>%
  group_by(Region) %>%
  summarise(TotalRegionSales = sum(Sales), .groups = 'drop')

sales_by_region <- sales_by_region[order(sales_by_region$Region, -
sales_by_region$TotalRegionSales),]

# Plotting sales by region
ggplot(sales_by_region, aes(x = Region, y = TotalRegionSales, fill = Region)) +
  geom_bar(stat = "identity") +
  labs(title = "Sales by Region", x = "Region", y = "Total Sales") +
  scale_y_continuous(labels = comma) # Use comma format for the y-axis


#                 iii). Impact of discounts on sales and profit

# Correlation between discount and sales
DiscSaleC <- cor(Ipdata$Discount, Ipdata$Sales)
print(DiscSaleC)

# Correlation between discount and profit
DiscProC <- cor(Ipdata$Discount, Ipdata$Profit)
print(DiscProC)
```

```
#                1. Relationship between sales and profit

# Correlation between sales and profit
SaleProfit_C <- cor(Ipdata$Sales, Ipdata$Profit)
print(SaleProfit_C)

#PLotting Scatter plot for Sales vs Profit With Line of Best Fit
plot(Ipdata$Sales, Ipdata$Profit,
    xlab="Sales",
    ylab="Profit",
    main="Scatter plot of Sales vs. Profit",
    color = "lightblue")
#Line of Best Fit
lm((Ipdata$Profit ~ Ipdata$Sales))

#Adding Line of best fit to the Plot
abline(lm(Ipdata$Profit ~ Ipdata$Sales), col="red")

#Calculating Coefficient of Determination (R-squared)
RSquared_PS <- summary(lm(Ipdata$Profit ~ Ipdata$Sales))$r.squared
print(RSquared_PS)

#Adding R squared value to the Plot
text(2000, -1000, paste("R-squared =", round(RSquared_PS, 2)), pos = 4, col = "red")


#                2. Relationship between discounts and profit

# Correlation between discounts and profit
DiscProC <- cor(Ipdata$Discount, Ipdata$Profit)
```

```
print(DiscProC)

# Scatter plot with line of best fit
plot(Ipdata$Discount, Ipdata$Profit,
    xlab="Discount",
    ylab="Profit",
    main="Scatter plot of Discount vs. Profit")

#Adding line of best to the plot
abline(lm(Ipdata$Profit ~ Ipdata$Discount), col="red")

#Calculating Coefficient of Determination (R-squared)
RSquared_DP <- summary(lm(Ipdata$Discount ~ Ipdata$Profit))$r.squared
print(RSquared_DP)

#Adding R squared value to the Plot
text(0.1, -1000, paste("R-squared =", round(RSquared_DP, 2)), pos = 4, col = "red")

#                    3. Relationship between discounts and quantity

# Correlation between discounts and quantity
DiscQu_C <- cor(Ipdata$Discount, Ipdata$Quantity)
print(DiscQu_C)

# Scatter plot with line of best fit
plot(Ipdata$Discount, Ipdata$Quantity,
    xlab="Discount",
    ylab="Quantity",
    main="Sactter plot of Discount vs. Quantity")

#adding line of best fit to the Plot
abline(lm(Ipdata$Quantity ~ Ipdata$Discount), col="red")
```

```
# Coefficient of determination (R-squared)
RSquared_QD <- summary(lm(Ipdata$Quantity ~ Ipdata$Discount))$r.squared
print(RSquared_QD)

#Adding R square value to the Plot
text(0.3, 2, paste("R-squared =", round(RSquared_QD, 2)), pos = 4, col = "red")


#                    4. Sales Distribution by Category/Sub-Category (East vs. West)

# Filtering data for sales in East and West regions
Regional_Sale <- Ipdata %>%
  #filter(Region %in% c("East", "West")) %>%
  group_by(Region, Category, Sub.Category) %>%
  summarise(Total_Regional_Sales = sum(Sales))


# Plot sales distribution

ggplot(Regional_Sale, aes(x = interaction(Category, Sub.Category), y = sort(Total_Regional_Sales), fill = Category)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.9)) +
  geom_text(aes(label = round(Total_Regional_Sales, 2)), vjust = -0.3, position = position_dodge(0.9)) +
  facet_grid(Region~., scales = "free") +
  labs(title = "Sales Distribution Across East and West Region", x = "Sub.Category", y = "Total Sales") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

#                    5. Profit Distribution by Category/Sub-Category (East vs. West)
```

```
# Filtering data for Profit East and West regions
Regional_profit <- Ipdata %>%
  #filter(Region %in% c("East", "West")) %>%
  group_by(Region, Category, Sub.Category) %>%
  summarise(Total_Regional_Profit = sum(Profit))

# Plot Profit distribution

#Faceted bar plot for profit by category and sub-category
ggplot(Regional_profit, aes(x = Sub.Category, y = Total_Regional_Profit, fill = Region)) +
  geom_bar(stat = "identity") +
  facet_wrap(~ Category, scales = "free_x") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  labs(title = "Profit Distribution by Category and Sub-Category", x = "Sub-Category", y = "Total Profit")
```