DATA MANIPULATION AND TRANSFORMING ASSIGNMENT

By Ankit Mittal
MABSPG24055

LOADING AND VIEWING DATASET

```r
library(dplyr)
library(ggplot2)
library(readr)

# Loading the Superstore dataset
superstore <- read_csv("C:/Users/mitta/OneDrive/Desktop/Superstore.csv")

# Viewing the dataset
View(superstore)
```

DATA CLEANING

```r
# 1. Data Cleaning

# Convert date columns to Date format
superstore <- superstore %>%
  mutate(
    `Order Date` = as.Date(`Order Date`, format = "%Y/%m/%d"),
    `Ship Date` = as.Date(`Ship Date`, format = "%Y/%m/%d")
  )
View(superstore)
# Check for missing values
colSums(is.na(superstore))
```

Here we have cleaned the data, converting date column into date format and also checking missing values

DATA TRANSFORMING

```r
# 2. Data Transformation

# Add a new column for Order Processing Time (days between Order and Ship Date)
superstore <- superstore %>%
  mutate(Order_Processing_Time = as.numeric(`Ship Date` - `Order Date`))
View(superstore)
```

```r
# Filter out high-discount sales (> 0.5)
high_discount_sales <- superstore %>% filter(Discount > 0.5)
View(superstore)
```

DATA TRANSFORMING

```r
#Category wise sales
sales_by_category <- superstore %>%
  group_by(Category) %>%
  summarise(
    Total_Sales = sum(Sales, na.rm = TRUE),
    Total_Profit = sum(Profit, na.rm = TRUE),
    Average_Discount = mean(Discount)
  ) %>%
  arrange(desc(Total_Sales))

print(sales_by_category)          #printing city wise sales
```

DATA TRANSFORMING

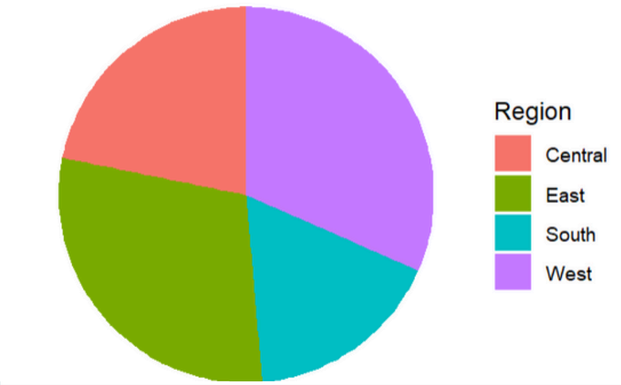
```r
# Summarize sales by region
sales_by_region <- superstore %>%
  group_by(Region) %>%
  summarise(
    Total_Sales = sum(Sales, na.rm = TRUE),
    Total_Profit = sum(Profit, na.rm = TRUE),
    Average_Discount = mean(Discount)
  ) %>%
  arrange(desc(Total_Sales))

print(sales_by_region)          #printing region wise sales
```

DATA ANALYSIS

1.Region wise sales analysis

```
#Region wise sales
ggplot(sales_by_region,aes(x="",y=Total_Sales,fill= Region))+
  geom_bar(width=1,stat="Identity")+
  coord_polar(theta="y")+
  labs(title="Region wise Sales")+
  theme_void()                    #Remove axis labels and background for pie chart
```



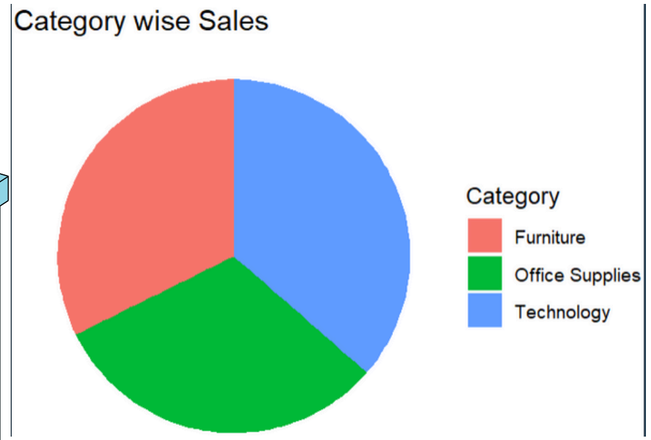
1.Maximum sales is from West Region (i.e. 725457.8) and its recorded profit is also the highest (i.e. 108418.45), it might due to lowest discount offered in this region (i.e. 10.93%).

2.Lowest sales is recorded from South Region but as it offered discount(i.e. 14.72%) lower than that offered by Central region (i.e. 24.03%) so its Profit is not the lowest one (i.e.46749.43>39706.36).

DATA ANALYSIS

2. Category wise sales analysis

```
#Category wise sales
ggplot(sales_by_category,aes(x="",y=Total_Sales,fill= Category))+
  geom_bar(width=1,stat="Identity")+
  coord_polar(theta="y")+
  labs(title="Category wise Sales")+
  theme_void()                    #Remove axis labels and background for pie chart
```



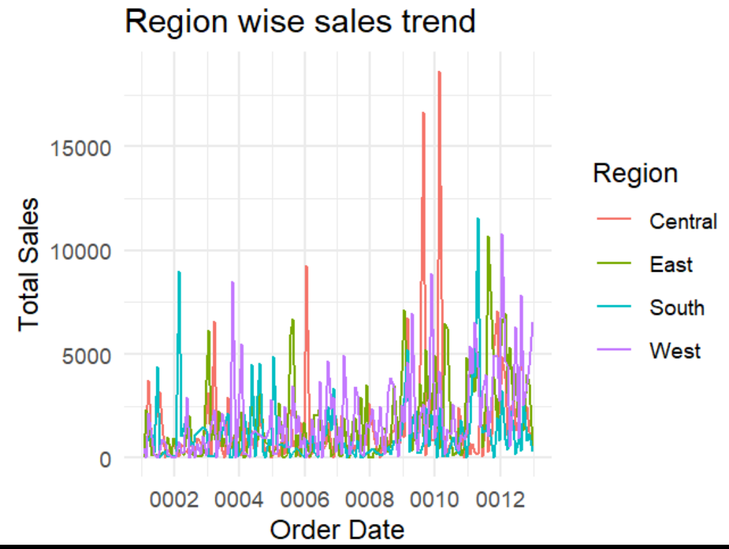
Highest profit is from Technology Category of products due to highest sales.

Profit from technology category=145455

DATA ANALYSIS

3. Sales trend over time

```
# Sales trend over time
ggplot(superstore, aes(x = `Order Date`, y = Sales, color = Region)) +
  geom_line(stat = "summary", fun = "sum") +
  labs(title = "Region wise sales trend", x = "Order Date", y = "Total Sales") +
  theme_minimal()
```



Central Region has recorded highest sales over time as compared to other regions.

DATA ANALYSIS

4. Top 10 profitable products

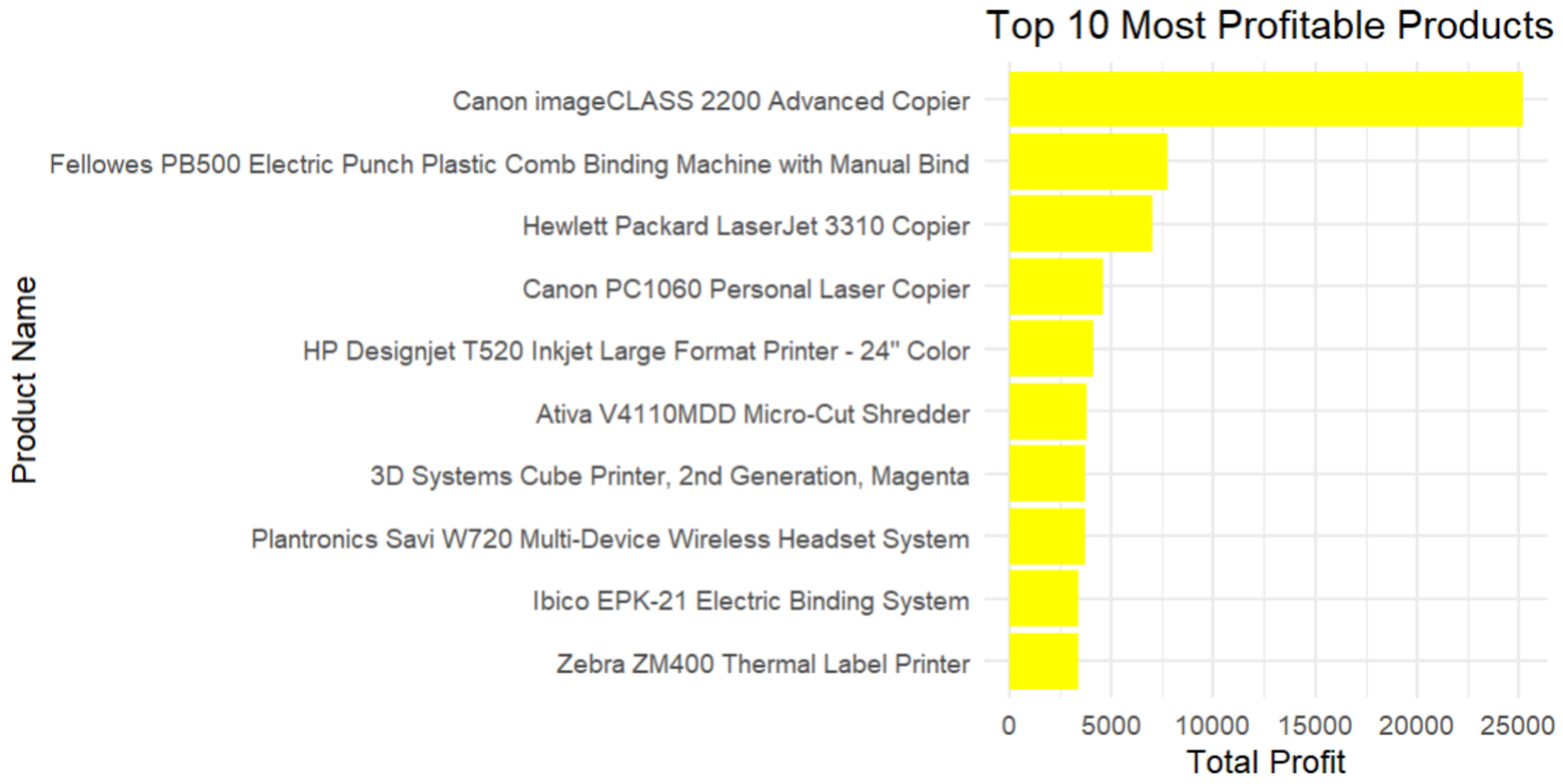
```r
# Top 10 profitable products
top_products <- superstore %>%
  group_by(`Product Name`) %>%
  summarise(Total_Profit = sum(Profit, na.rm = TRUE)) %>%
  arrange(desc(Total_Profit)) %>%
  slice_head(n = 10)
print(top_products)
#Visualization
ggplot(top_products, aes(x = reorder(`Product Name`, Total_Profit), y = Total_Profit)) +
  geom_bar(stat = "identity", fill = "yellow") +
  coord_flip() +
  labs(title = "Top 10 Most Profitable Products", x = "Product Name", y = "Total Profit") +
  theme_minimal()
```

Result➡

	`Product Name`	Total_Profit
	<chr>	<dbl>
1	"Canon imageCLASS 2200 Advanced Copier"	25200.
2	"Fellowes PB500 Electric Punch Plastic Comb Binding Machine with Manual Bind"	7753.
3	"Hewlett Packard LaserJet 3310 Copier"	6984.
4	"Canon PC1060 Personal Laser Copier"	4571.
5	"HP Designjet T520 Inkjet Large Format Printer - 24\" Color"	4095.
6	"Ativa V4110MDD Micro-Cut Shredder"	3773.
7	"3D Systems Cube Printer, 2nd Generation, Magenta"	3718.
8	"Plantronics Savi W720 Multi-Device Wireless Headset System"	3696.
9	"Ibico EPK-21 Electric Binding System"	3345.
10	"Zebra ZM400 Thermal Label Printer"	3344.

DATA ANALYSIS

4. Top 10 profitable products visualization



Canon imageCLASS 2200 Advanced Copier has recorded the highest profit of 25200

CONCLUSION

By addressing these findings, the business can strengthen its profitability, streamline operations, and improve customer satisfaction while exploring new growth opportunities.