I
E & C
T
IIT Guwahati

Advanced Certification Programme in Data Science Business Analytics

Week 2
Data Analysis Using R

Topics Covered

- Retail Sales Data Analysis
- Data Manipulation and Visualisation
- Time Series Analysis of Sales
- Sales Distribution by Country
- Product Performance Analysis
- Sales Analysis by Quantity Sold
- Customer Segmentation
- Sales Performance by Product Category
- Month-Wise Sales Distribution
- Customer Purchasing Patterns
- Q & A

Retail Sales Data Analysis

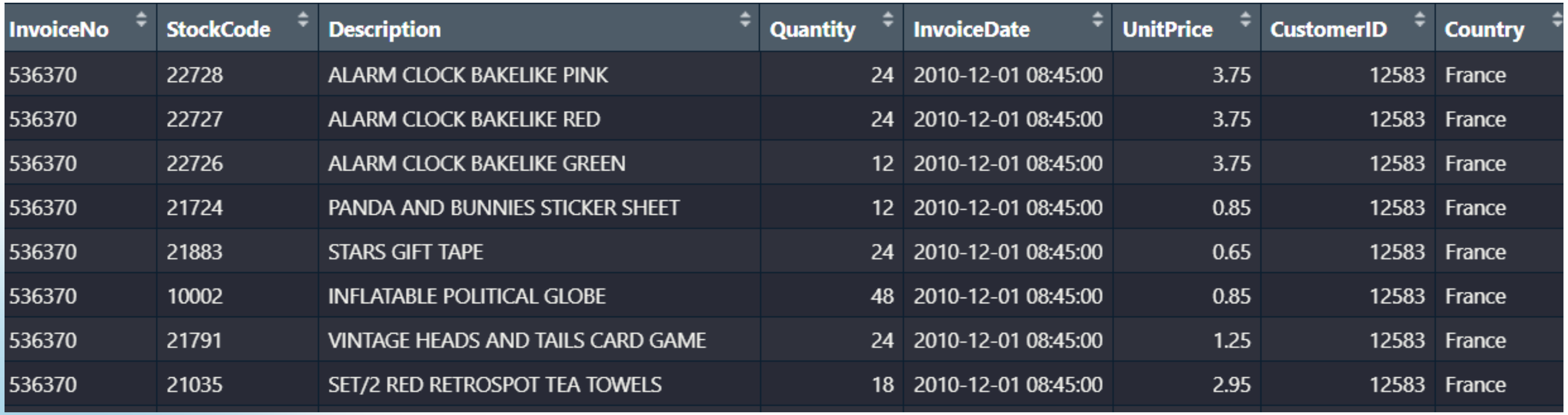
Key Transaction Metrics for Business Insights

Column	Description
Invoice number	Unique ID for each transaction
Invoice date	Date and time of transaction
Customer ID	Unique ID for each customer
Description	Product details
Quantity	Items sold per transaction
Unit price	Price per unit
Total price	Quantity × Unit price
Country	Transaction location

Data Manipulation and Visualisation

Practical Data Applications

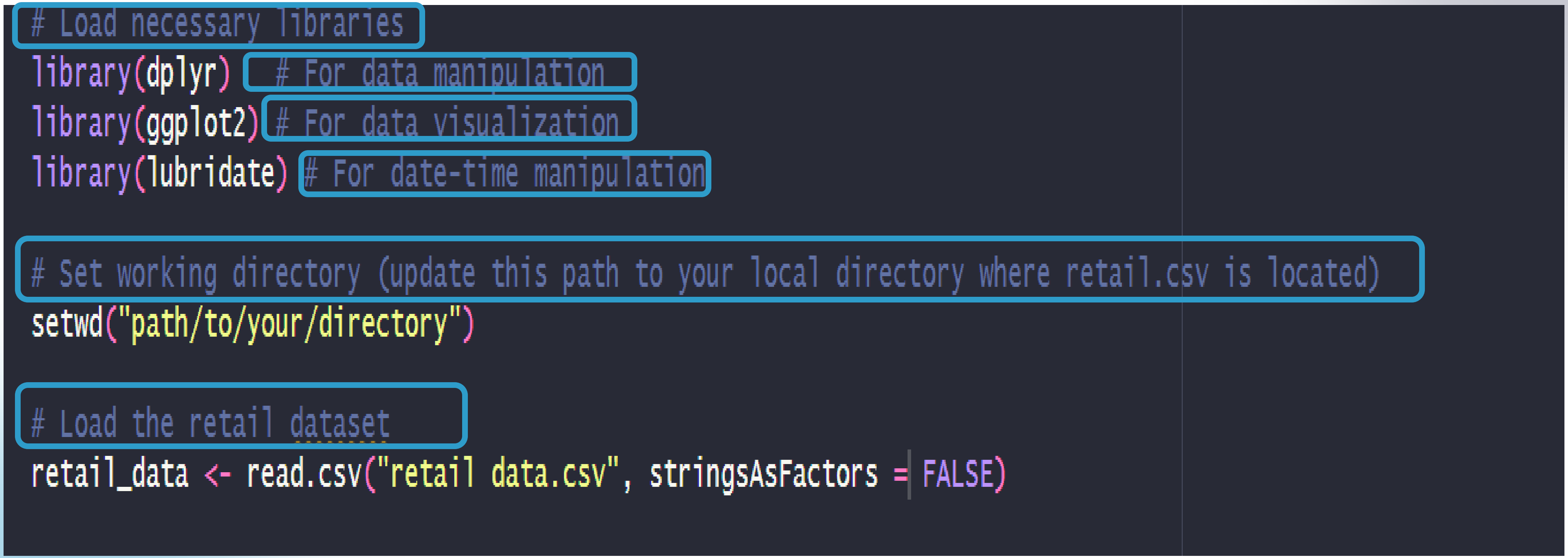
- Learn to manipulate and visualise retail sales data using R
- Create effective visualisations for better data interpretation
- Focus on data cleaning, aggregation and generating insights



InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
536370	22728	ALARM CLOCK BAKELIKE PINK	24	2010-12-01 08:45:00	3.75	12583	France
536370	22727	ALARM CLOCK BAKELIKE RED	24	2010-12-01 08:45:00	3.75	12583	France
536370	22726	ALARM CLOCK BAKELIKE GREEN	12	2010-12-01 08:45:00	3.75	12583	France
536370	21724	PANDA AND BUNNIES STICKER SHEET	12	2010-12-01 08:45:00	0.85	12583	France
536370	21883	STARS GIFT TAPE	24	2010-12-01 08:45:00	0.65	12583	France
536370	10002	INFLATABLE POLITICAL GLOBE	48	2010-12-01 08:45:00	0.85	12583	France
536370	21791	VINTAGE HEADS AND TAILS CARD GAME	24	2010-12-01 08:45:00	1.25	12583	France
536370	21035	SET/2 RED RETROSPOT TEA TOWELS	18	2010-12-01 08:45:00	2.95	12583	France

Data Manipulation and Visualisation

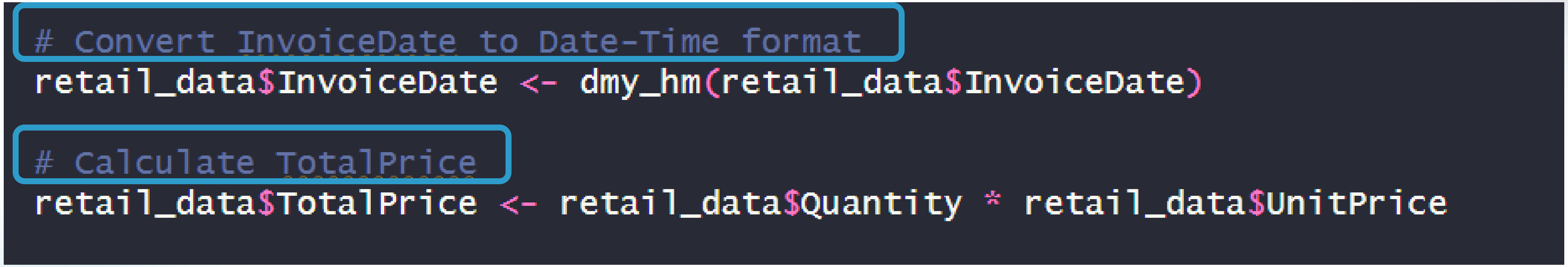
Import Libraries and Load Data



```r
# Load necessary libraries
library(dplyr)  # For data manipulation
library(ggplot2)# For data visualization
library(lubridate) # For date-time manipulation

# Set working directory (update this path to your local directory where retail.csv is located)
setwd("path/to/your/directory")

# Load the retail dataset
retail_data <- read.csv("retail data.csv", stringsAsFactors = FALSE)
```

Data Manipulation and Visualisation

Data Processing in R

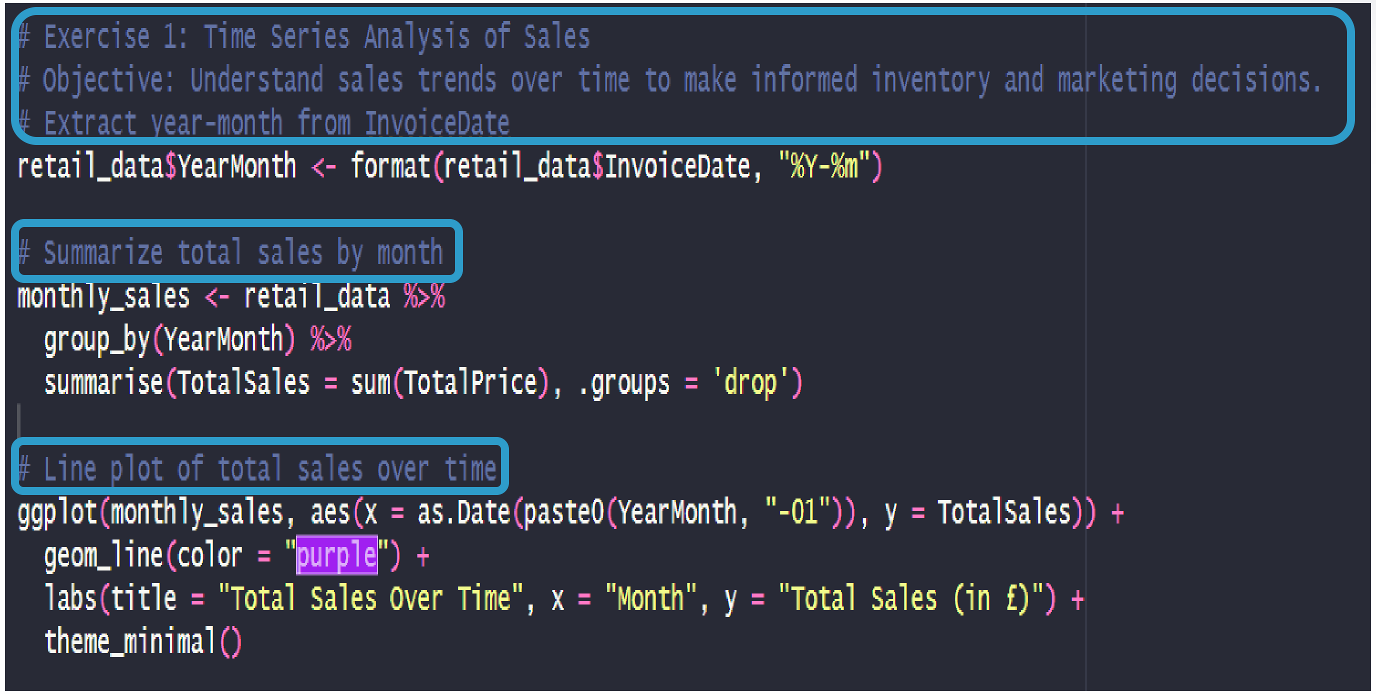


```
# Convert InvoiceDate to Date-Time format
retail_data$InvoiceDate <- dmy_hm(retail_data$InvoiceDate)

# Calculate TotalPrice
retail_data$TotalPrice <- retail_data$Quantity * retail_data$UnitPrice
```

Time Series Analysis of Sales

Identifying Trends for Smarter Business Decisions

- Use insights to optimise inventory and marketing strategies
- Analyse sales trends over time for better decision-making

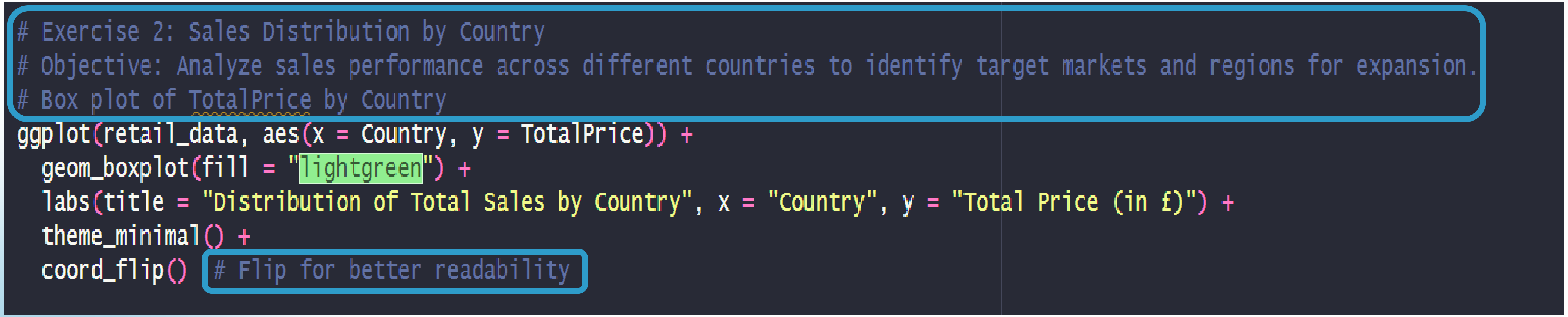


```
# Exercise 1: Time Series Analysis of Sales
# Objective: Understand sales trends over time to make informed inventory and marketing decisions.
# Extract year-month from InvoiceDate
retail_data$YearMonth <- format(retail_data$InvoiceDate, "%Y-%m")

# Summarize total sales by month
monthly_sales <- retail_data %>%
  group_by(YearMonth) %>%
  summarise(TotalSales = sum(TotalPrice), .groups = 'drop')

# Line plot of total sales over time
ggplot(monthly_sales, aes(x = as.Date(paste0(YearMonth, "-01")), y = TotalSales)) +
  geom_line(color = "purple") +
  labs(title = "Total Sales Over Time", x = "Month", y = "Total Sales (in £)") +
  theme_minimal()
```

Sales Distribution by Country

Understand Regional Sales Patterns

- Visualise total price distribution using a box plot by country
- Identify key target markets and expansion opportunities
- Evaluate sales performance across various countries

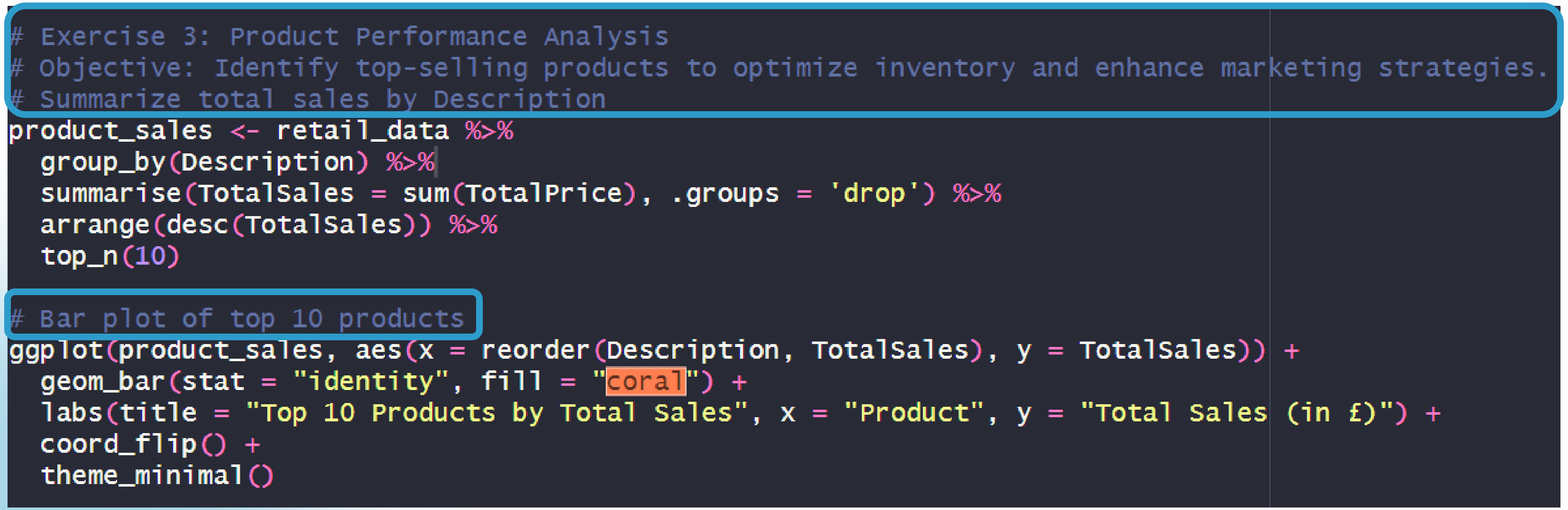


```
# Exercise 2: Sales Distribution by Country
# Objective: Analyze sales performance across different countries to identify target markets and regions for expansion.
# Box plot of TotalPrice by Country
ggplot(retail_data, aes(x = Country, y = TotalPrice)) +
  geom_boxplot(fill = "lightgreen") +
  labs(title = "Distribution of Total Sales by Country", x = "Country", y = "Total Price (in £)") +
  theme_minimal() +
  coord_flip() # Flip for better readability
```

Product Performance Analysis

Maximise Sales Through Product Insights

- Determine best-selling products for better inventory management
- Analyse product performance to refine marketing strategies

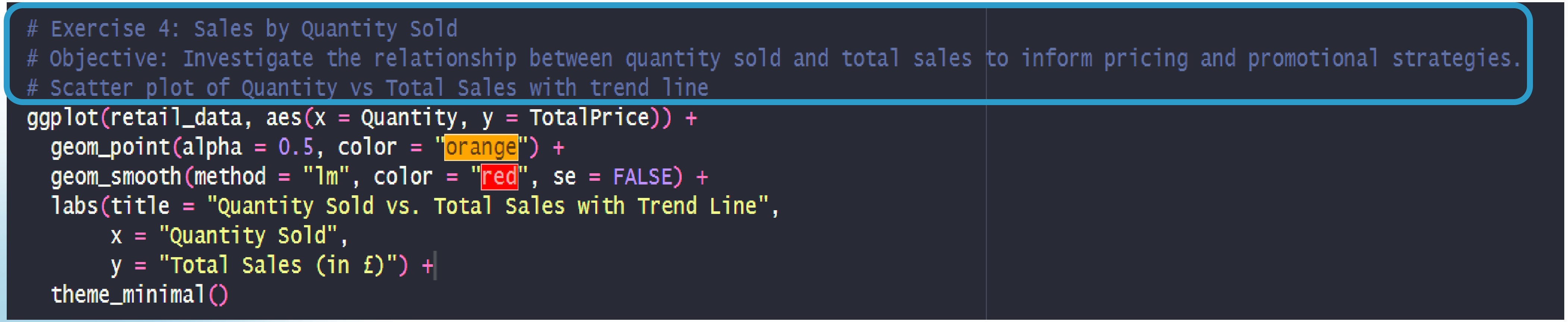


```r
# Exercise 3: Product Performance Analysis
# Objective: Identify top-selling products to optimize inventory and enhance marketing strategies.
# Summarize total sales by Description
product_sales <- retail_data %>%
  group_by(Description) %>%
  summarise(TotalSales = sum(TotalPrice), .groups = 'drop') %>%
  arrange(desc(TotalSales)) %>%
  top_n(10)

# Bar plot of top 10 products
ggplot(product_sales, aes(x = reorder(Description, TotalSales), y = TotalSales)) +
  geom_bar(stat = "identity", fill = "coral") +
  labs(title = "Top 10 Products by Total Sales", x = "Product", y = "Total Sales (in £)") +
  coord_flip() +
  theme_minimal()
```

Sales Analysis by Quantity Sold

Measuring Performance Through Sales Volume

- Examine how quantity sold affects total sales performance
- Identify trends to improve pricing and promotions
- Use visual analysis to uncover sales patterns

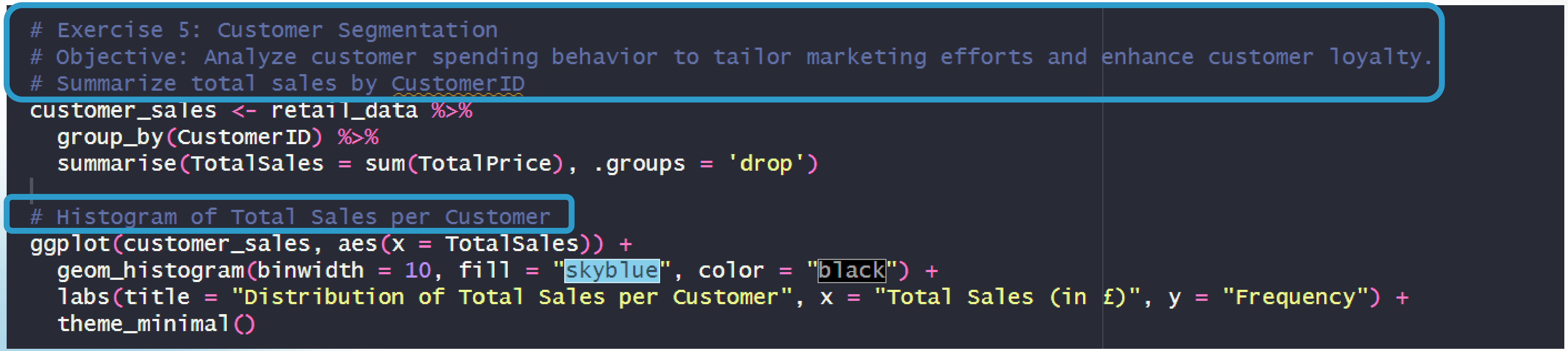


```
# Exercise 4: Sales by Quantity Sold
# Objective: Investigate the relationship between quantity sold and total sales to inform pricing and promotional strategies.
# Scatter plot of Quantity vs Total Sales with trend line
ggplot(retail_data, aes(x = Quantity, y = TotalPrice)) +
  geom_point(alpha = 0.5, color = "orange") +
  geom_smooth(method = "lm", color = "red", se = FALSE) +
  labs(title = "Quantity Sold vs. Total Sales with Trend Line",
       x = "Quantity Sold",
       y = "Total Sales (in £)") +
  theme_minimal()
```

Customer Segmentation

Understand Spending Patterns for Targeted Marketing

- Examine customer spending habits to refine marketing strategies
- Segment customers based on total sales data for better engagement
- Identify key consumer groups to enhance customer loyalty

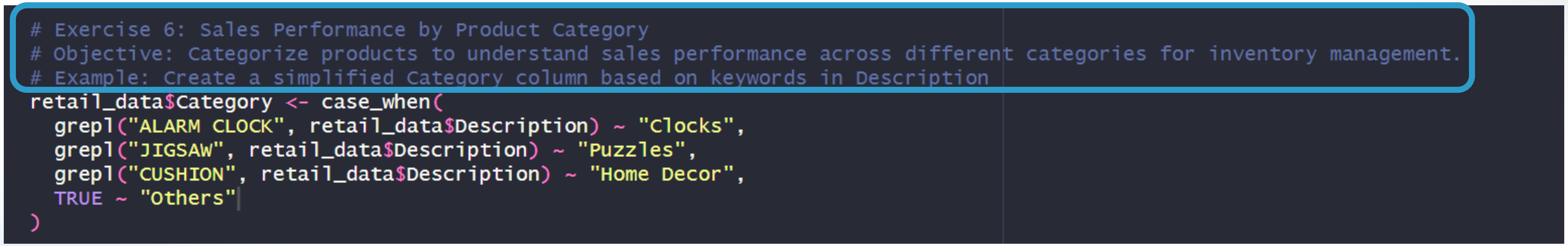


```
# Exercise 5: Customer Segmentation
# Objective: Analyze customer spending behavior to tailor marketing efforts and enhance customer loyalty.
# Summarize total sales by CustomerID
customer_sales <- retail_data %>%
  group_by(CustomerID) %>%
  summarise(TotalSales = sum(TotalPrice), .groups = 'drop')

# Histogram of Total Sales per Customer
ggplot(customer_sales, aes(x = TotalSales)) +
  geom_histogram(binwidth = 10, fill = "skyblue", color = "black") +
  labs(title = "Distribution of Total Sales per Customer", x = "Total Sales (in £)", y = "Frequency") +
  theme_minimal()
```

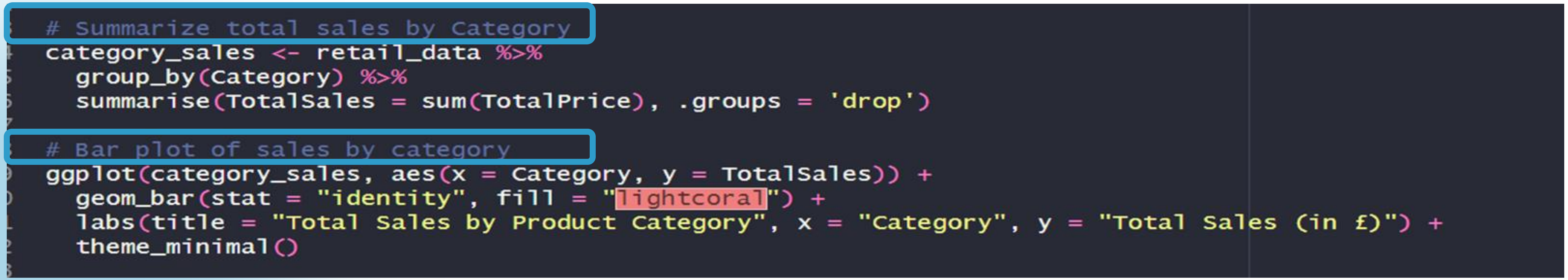
Sales Performance by Product Category

Analyse Product Categories for Better Inventory Management

- Categorise products to evaluate sales trends across different categories
- Assess category-wise sales data to improve stock management



```r
# Exercise 6: Sales Performance by Product Category
# Objective: Categorize products to understand sales performance across different categories for inventory management.
# Example: Create a simplified Category column based on keywords in Description
retail_data$Category <- case_when(
  grepl("ALARM CLOCK", retail_data$Description) ~ "Clocks",
  grepl("JIGSAW", retail_data$Description) ~ "Puzzles",
  grepl("CUSHION", retail_data$Description) ~ "Home Decor",
  TRUE ~ "Others"
)
```

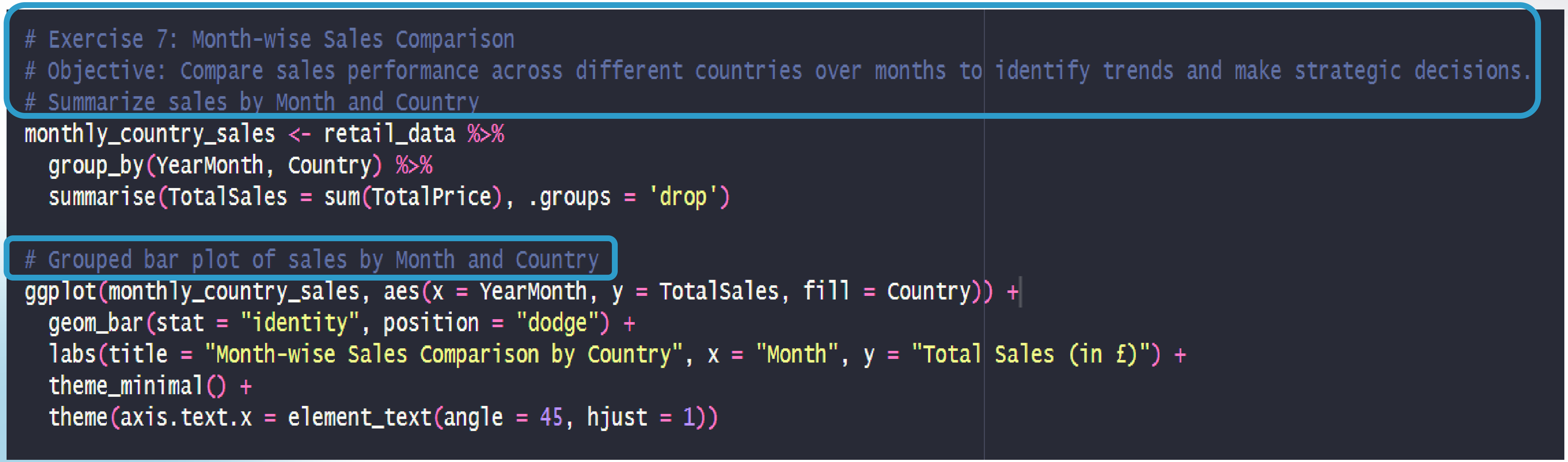


```r
 # Summarize total sales by Category
 category_sales <- retail_data %>%
   group_by(Category) %>%
   summarise(TotalSales = sum(TotalPrice), .groups = 'drop')

 # Bar plot of sales by category
 ggplot(category_sales, aes(x = Category, y = TotalSales)) +
   geom_bar(stat = "identity", fill = "lightcoral") +
   labs(title = "Total Sales by Product Category", x = "Category", y = "Total Sales (in £)") +
   theme_minimal()
```

Month-wise Sales Comparison

Evaluate Sales Trends Across Countries Over Time

- Compare sales performance across different countries over months
- Identify trends to support data-driven strategic decisions

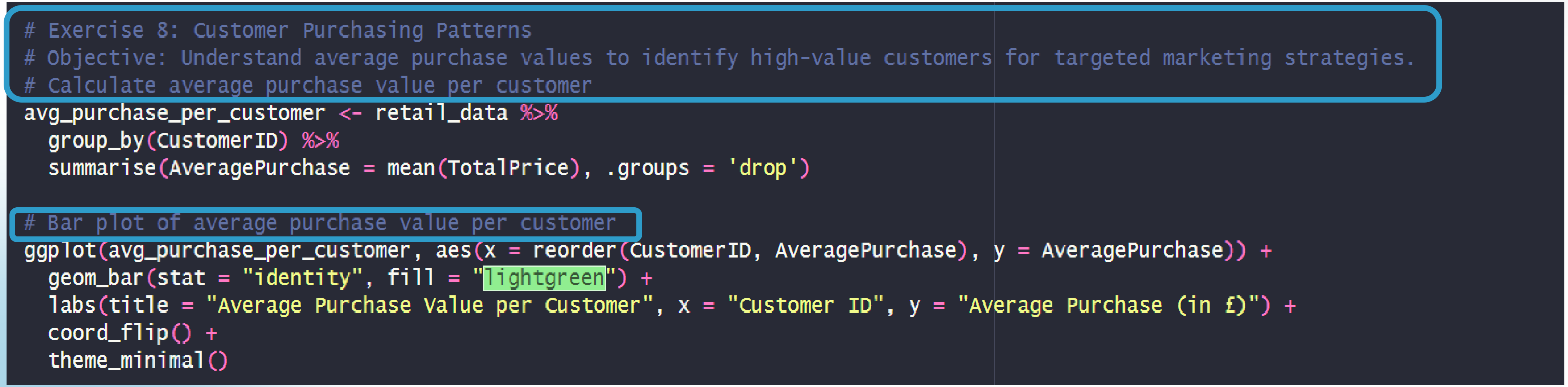


```
# Exercise 7: Month-wise Sales Comparison
# Objective: Compare sales performance across different countries over months to identify trends and make strategic decisions.
# Summarize sales by Month and Country
monthly_country_sales <- retail_data %>%
  group_by(YearMonth, Country) %>%
  summarise(TotalSales = sum(TotalPrice), .groups = 'drop')

# Grouped bar plot of sales by Month and Country
ggplot(monthly_country_sales, aes(x = YearMonth, y = TotalSales, fill = Country)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Month-wise Sales Comparison by Country", x = "Month", y = "Total Sales (in £)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Customer Purchasing Patterns

Identify High-Value Customers Through Purchase Analysis

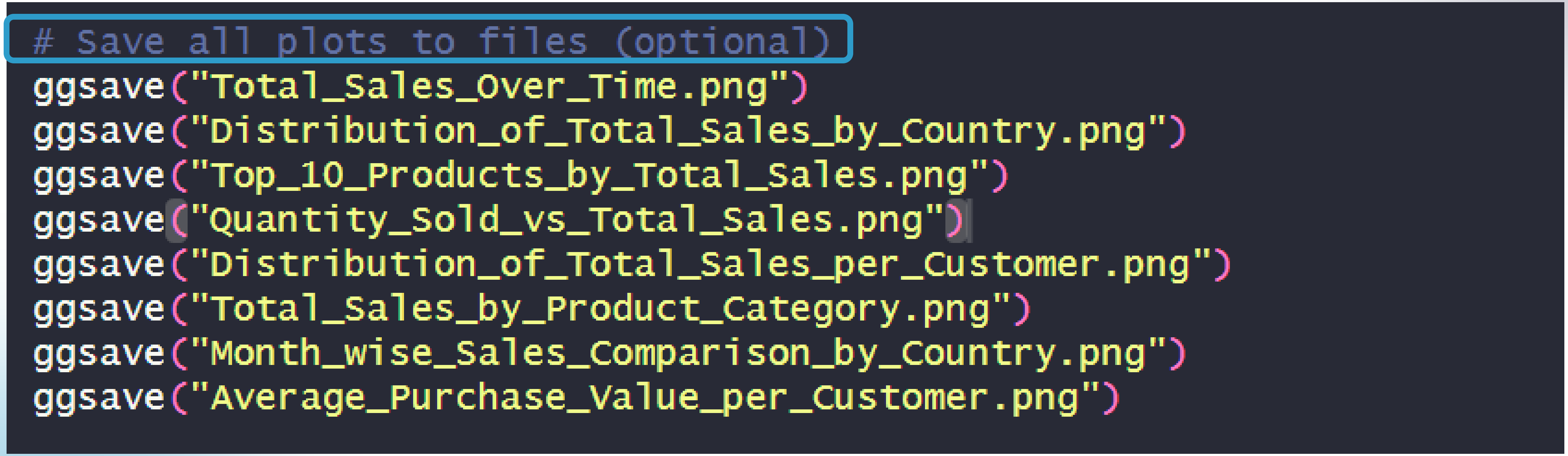
- Analyse average purchase values per customer
- Identify high-value customers for targeted marketing strategies



```
# Exercise 8: Customer Purchasing Patterns
# Objective: Understand average purchase values to identify high-value customers for targeted marketing strategies.
# Calculate average purchase value per customer
avg_purchase_per_customer <- retail_data %>%
  group_by(CustomerID) %>%
  summarise(AveragePurchase = mean(TotalPrice), .groups = 'drop')

# Bar plot of average purchase value per customer
ggplot(avg_purchase_per_customer, aes(x = reorder(CustomerID, AveragePurchase), y = AveragePurchase)) +
  geom_bar(stat = "identity", fill = "lightgreen") +
  labs(title = "Average Purchase Value per Customer", x = "Customer ID", y = "Average Purchase (in £)") +
  coord_flip() +
  theme_minimal()
```

Save all plots to files (optional)

Export Visualisations for Future Reference



```
# Save all plots to files (optional)
ggsave("Total_Sales_Over_Time.png")
ggsave("Distribution_of_Total_Sales_by_Country.png")
ggsave("Top_10_Products_by_Total_Sales.png")
ggsave("Quantity_Sold_vs_Total_Sales.png")
ggsave("Distribution_of_Total_Sales_per_Customer.png")
ggsave("Total_Sales_by_Product_Category.png")
ggsave("Month_wise_Sales_Comparison_by_Country.png")
ggsave("Average_Purchase_Value_per_Customer.png")
```

Q & A

Thank you