Exploratory Data Analysis of

Retail Store Inventory

BY: Khushi Goyal (24034)

INTRODUCTION

Purpose and Scope

- To analyze retail store inventory and sales data for insights into product performance, regional trends, and sales forecasting.
- Focus on categories, regions, sales, inventory, pricing, and seasonal impacts.

OBJECTIVE

1. Understand the sales trends and inventory distribution.
2. Identify correlations between variables like pricing, discounts, and sales.
3. Analyze regional and category-wise performance.
4. Provide actionable insights for inventory management and sales strategies.

Research Methodology

1.Data Collection:

Used retail store inventory data, focusing on variables like product categories, inventory levels, and demand forecasts.

2. Data Preprocessing:

Cleaned and formatted data, removed outliers, and created aggregated metrics.

3. Tools and Techniques:

- Software: R (with dplyr and ggplot2).
- Analysis: Descriptive statistics, correlation analysis, and visualization.

4. Data Analysis:

- Univariate: Summary of individual variables.
- Bivariate: Relationships explored using scatter plots, heatmaps, and bar charts.

Data and Variables

Dataset Overview

- Rows: Number of records
- Columns: Key attributes like:
- Categorical Variables: Region, Category, Seasonality
- Numerical Variables: Units Sold, Price, Discount, Inventory Level
- Time-Related Data: Date

Key Focus Variables

- Units Sold: Sales performance metric.
- Price: Impact on sales.
- Region: Geographic trends.
- Category: Product-level analysis

Key Findings

Filter

	Avg_Units_Sold	Median_Units_Sold	Avg_Price	Avg_Discount
1	136.4649	107	55.13511	10.00951

Summary Statistics

```
# Summary statistics for numerical columns
summary_stats <-retail_store_inventory1 %>%
  summarise(
    Avg_Units_Sold = mean(`Units.Sold`, na.rm = TRUE),
    Median_Units_Sold = median(`Units.Sold`, na.rm = TRUE),
    Avg_Price = mean(Price, na.rm = TRUE),
    Avg_Discount = mean(Discount, na.rm = TRUE)
  )
print(summary_stats)
```

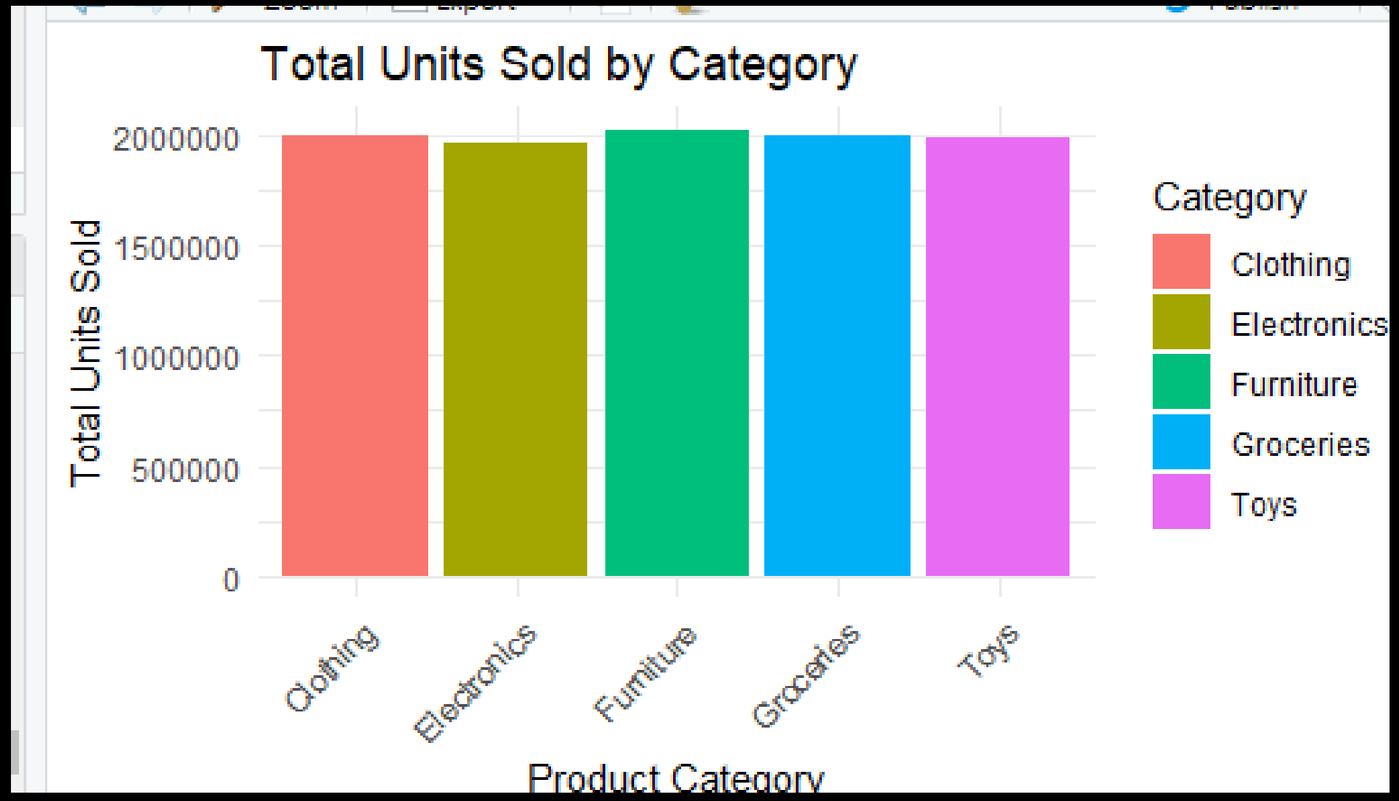
Sales By Category

```
# Sales by Category
category_sales <-
retail_store_inventory1%>%
  group_by(Category) %>%
  summarise(Total_Units_Sold =
sum(`Units.Sold`, na.rm = TRUE))
View(category_sales)
```

Filter

row names	Category	Total_Units_Sold
2	Electronics	1960432
3	Furniture	2025017
4	Groceries	2000482
5	Toys	1990485

Total Units Sold by Category




```
ggplot(category_sales, aes(x = Category, y = Total_Units_Sold, fill = Category)) +
 geom_bar(stat = "identity") +
 labs(title = "Total Units Sold by Category",
 x = "Product Category",
 y = "Total Units Sold") +
 theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

It identify which product categories contribute most to sales, guiding inventory planning and promotional efforts.

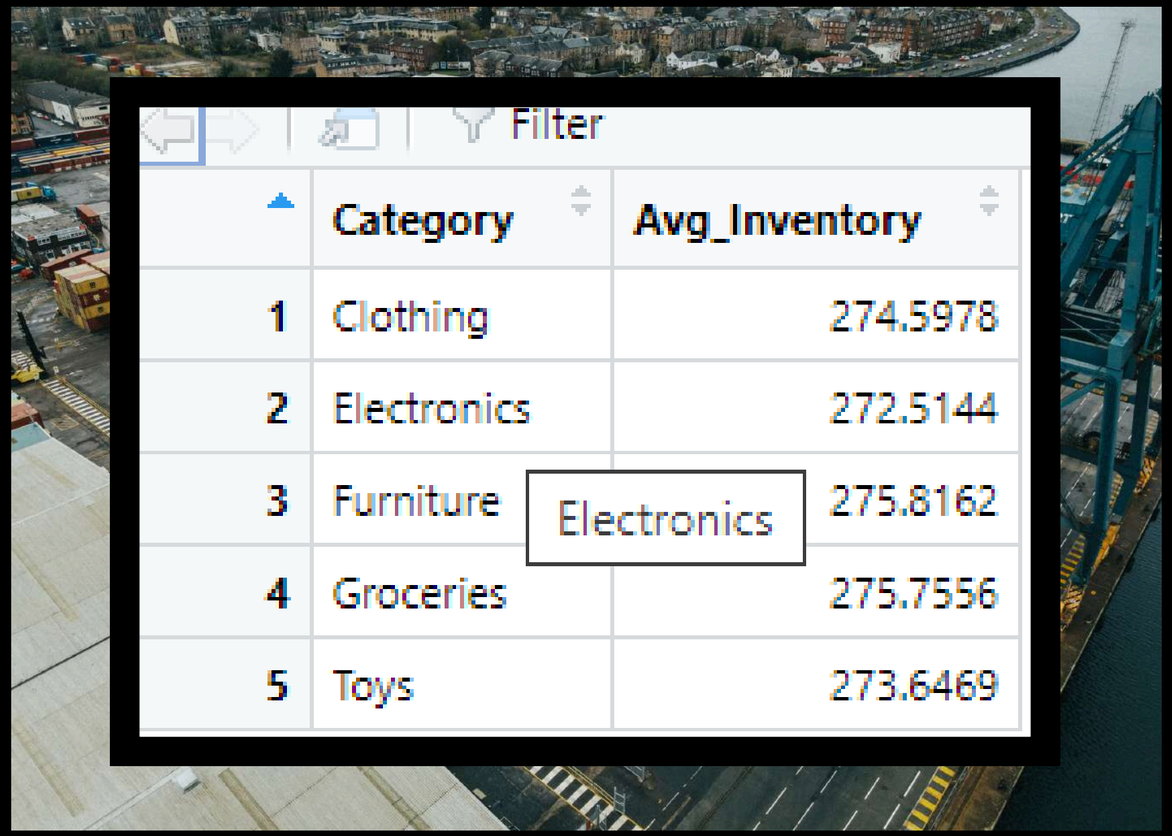
Regional Sales Performance

```
#Regional Sales Performance
region_sales <-retail_store_inventory1
%>%
group_by(Region) %>%
summarise(Total_Units_Sold =
sum(`Units.Sold`, na.rm = TRUE))
```

Filter

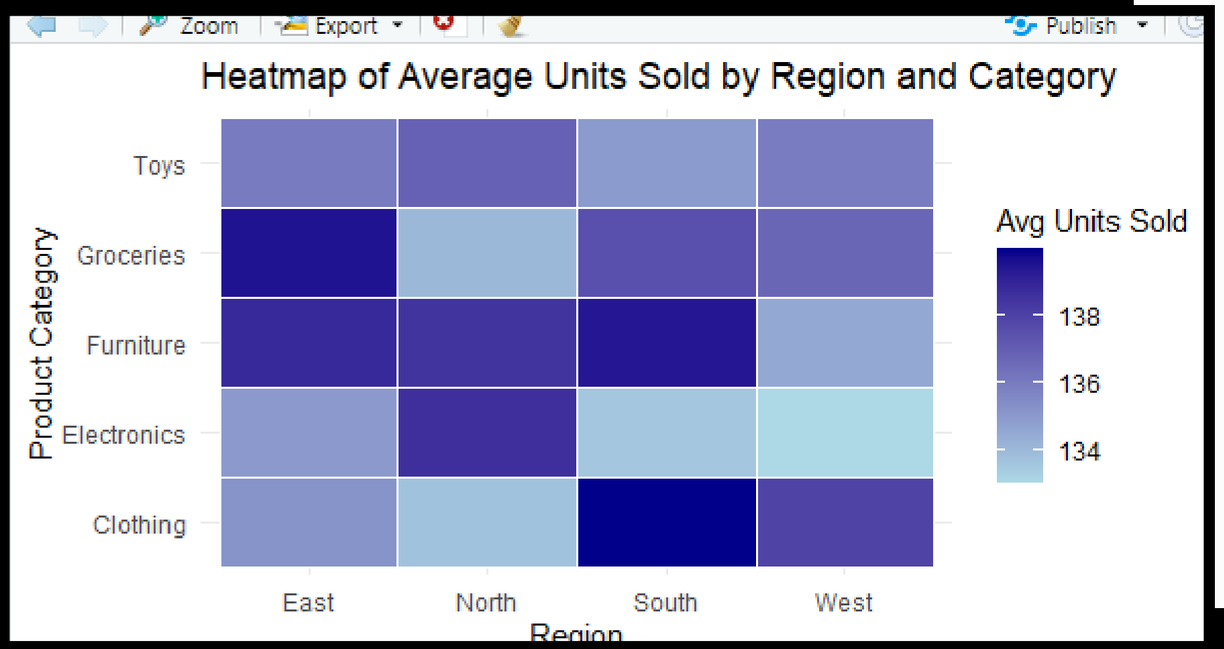
	Region	Total_Units_Sold
1	East	2511265
2	North	2484966
3	South	2507799
4	West	2471552

Inventory Level Analysis

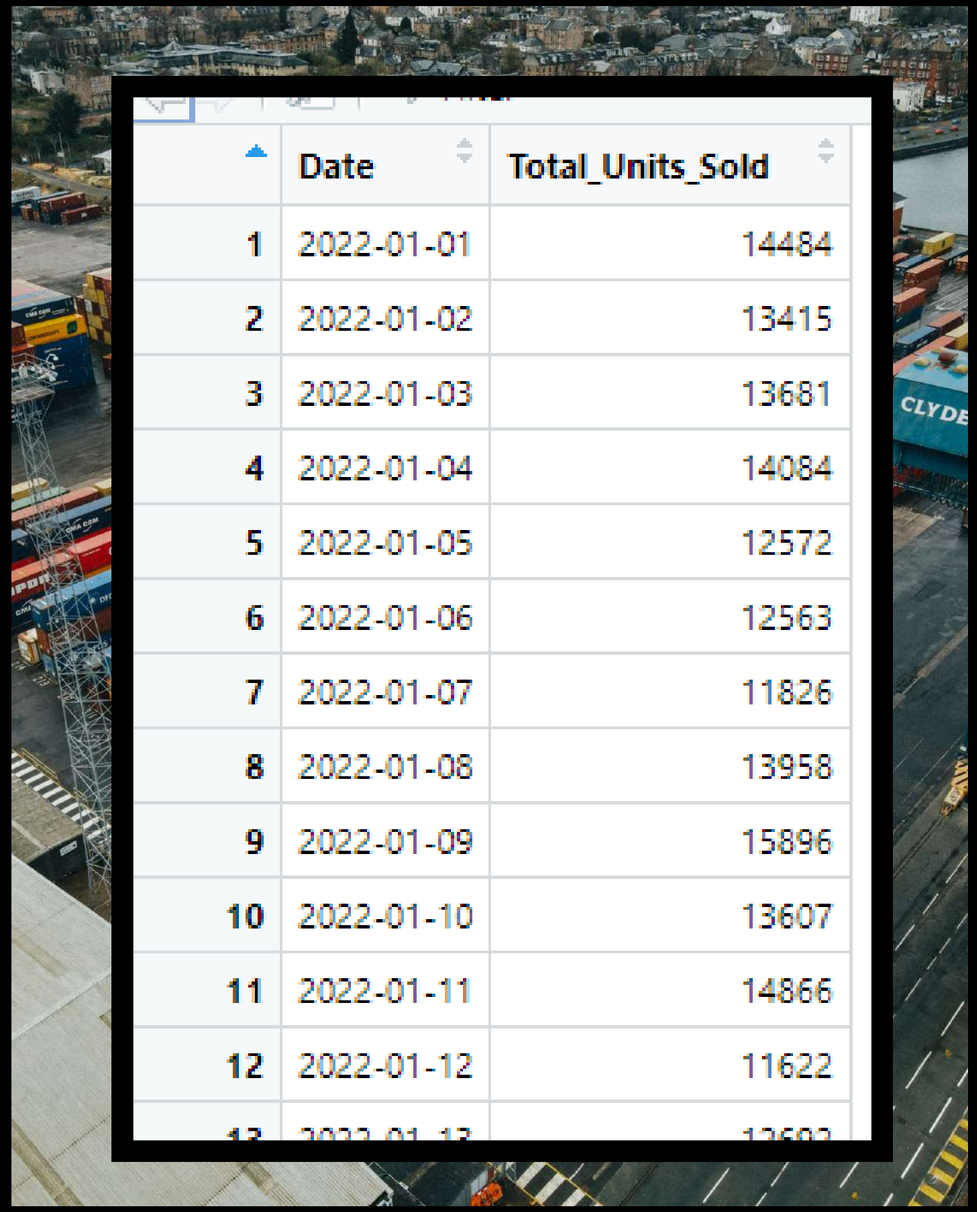



```
#Inventory Level Analysis
inventory_analysis <-
retail_store_inventory1 %>%
  group_by(Category) %>%
  summarise(Avg_Inventory =
mean(`Inventory.Level`, na.rm = TRUE))
View(inventory_analysis)
```

Average Units Sold by Region and Category



- Target High-Performing Combinations: Focus marketing and inventory on Groceries in the East and South regions to maximize revenue.
- Improve Low-Performing Areas: Investigate why Clothing and Electronics perform poorly in certain regions. Adjust strategies like pricing, promotions, or visibility.



	Date	Total_Units_Sold
1	2022-01-01	14484
2	2022-01-02	13415
3	2022-01-03	13681
4	2022-01-04	14084
5	2022-01-05	12572
6	2022-01-06	12563
7	2022-01-07	11826
8	2022-01-08	13958
9	2022-01-09	15896
10	2022-01-10	13607
11	2022-01-11	14866
12	2022-01-12	11622
13	2022-01-13	12602

Time Trend of Units Sold

```
#Time Trend of Units Sold (if Date is available)
retail_store_inventory1$Date <-
as.Date(retail_store_inventory1$Date)
time_trend <-retail_store_inventory1 %>%
  group_by(Date) %>%
  summarise(Total_Units_Sold = sum(`Units.Sold`,
na.rm = TRUE))
View(time_trend)
```

Top Performing Categories

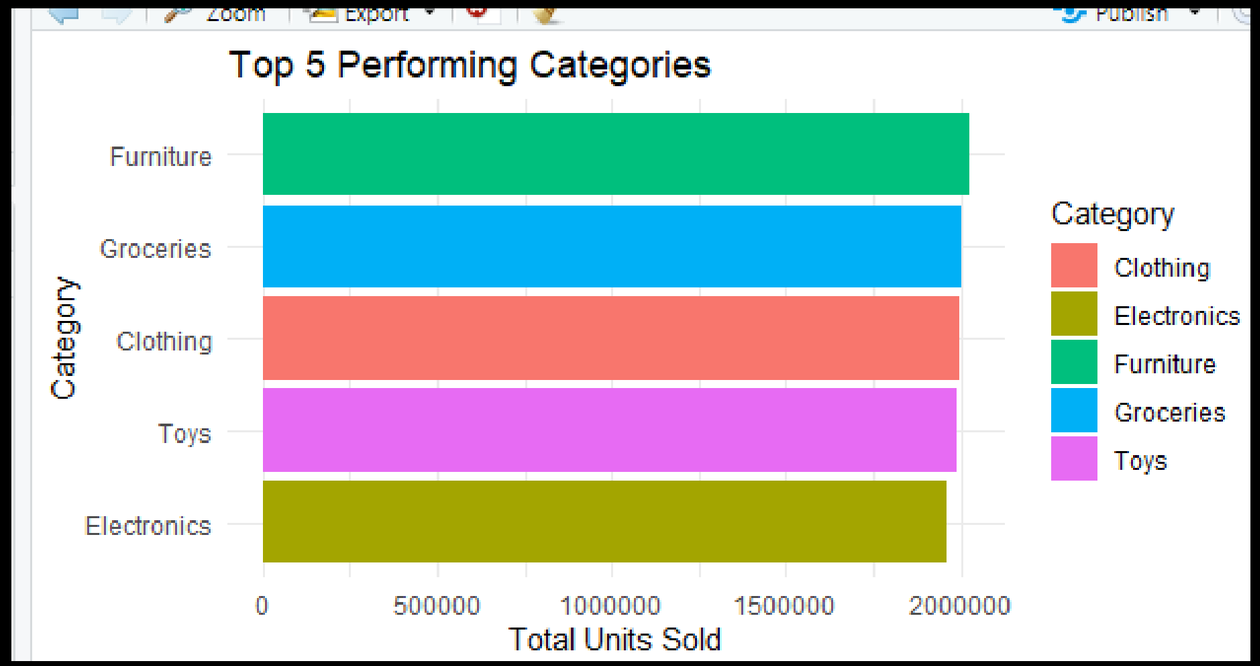
Filter

	Category	Total_Units_Sold
1	Furniture	2025017
2	Groceries	2000482
3	Clothing	1999166
4	Toys	1990485
5	Electronics	1960432

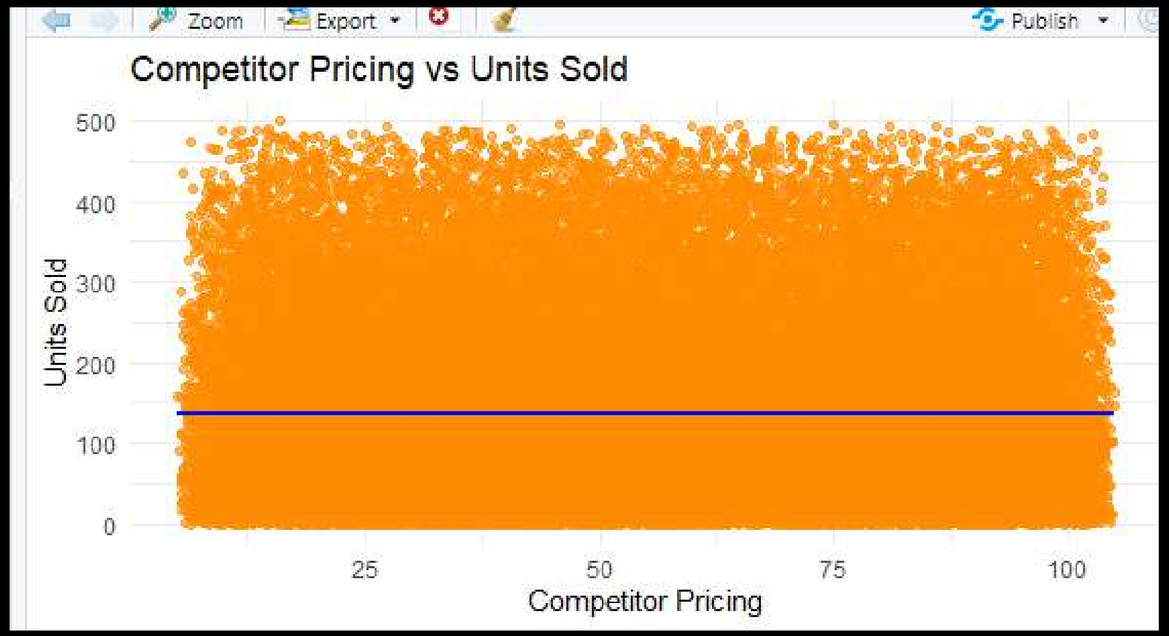
```
#Top Performing Categories
top_categories <-
retail_store_inventory1%>%
  group_by(Category) %>%
  summarise(Total_Units_Sold =
sum(`Units.Sold`, na.rm = TRUE)) %>%
  arrange(desc(Total_Units_Sold))
View(head(top_categories))
```

Top Performing Categories

- Focus on High Performers: Allocate more inventory and marketing resources to Furniture and Groceries to sustain their strong performance.
- Growth Opportunities: Explore strategies to enhance sales of Clothing, Toys, and Electronics, such as regional promotions, better visibility, or new product offerings.



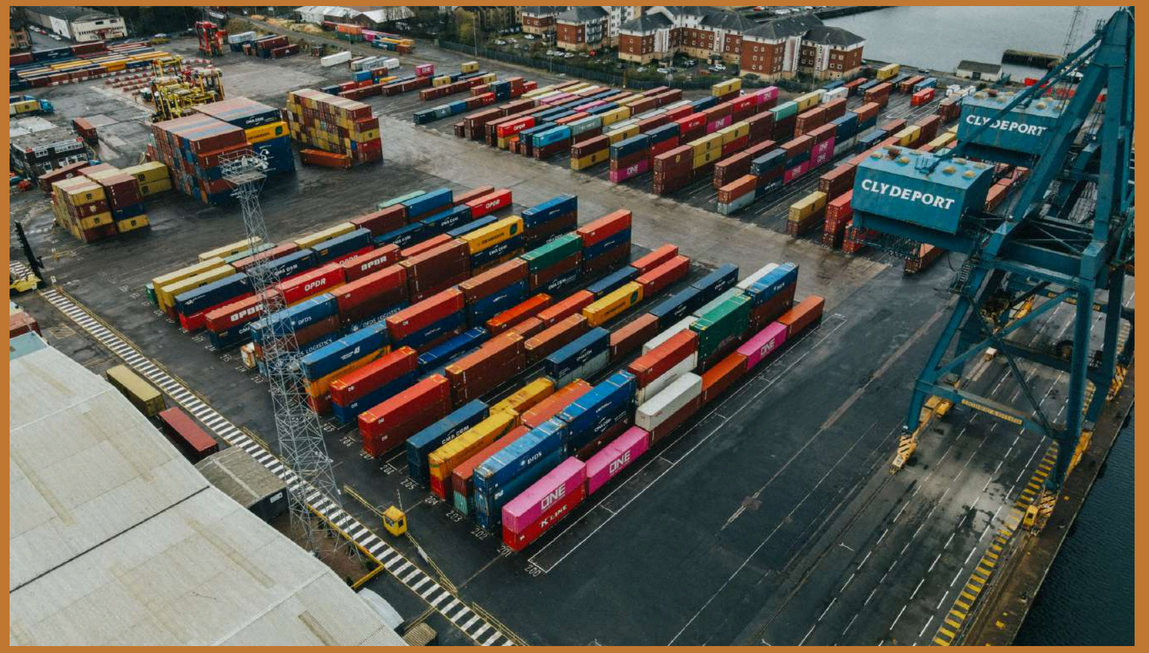
Competitor Pricing vs Sales



- A downward-sloping trend suggests customers are price-conscious and prefer lower-priced alternatives.
- Focus on offering better value through discounts or promotions during price-sensitive periods.

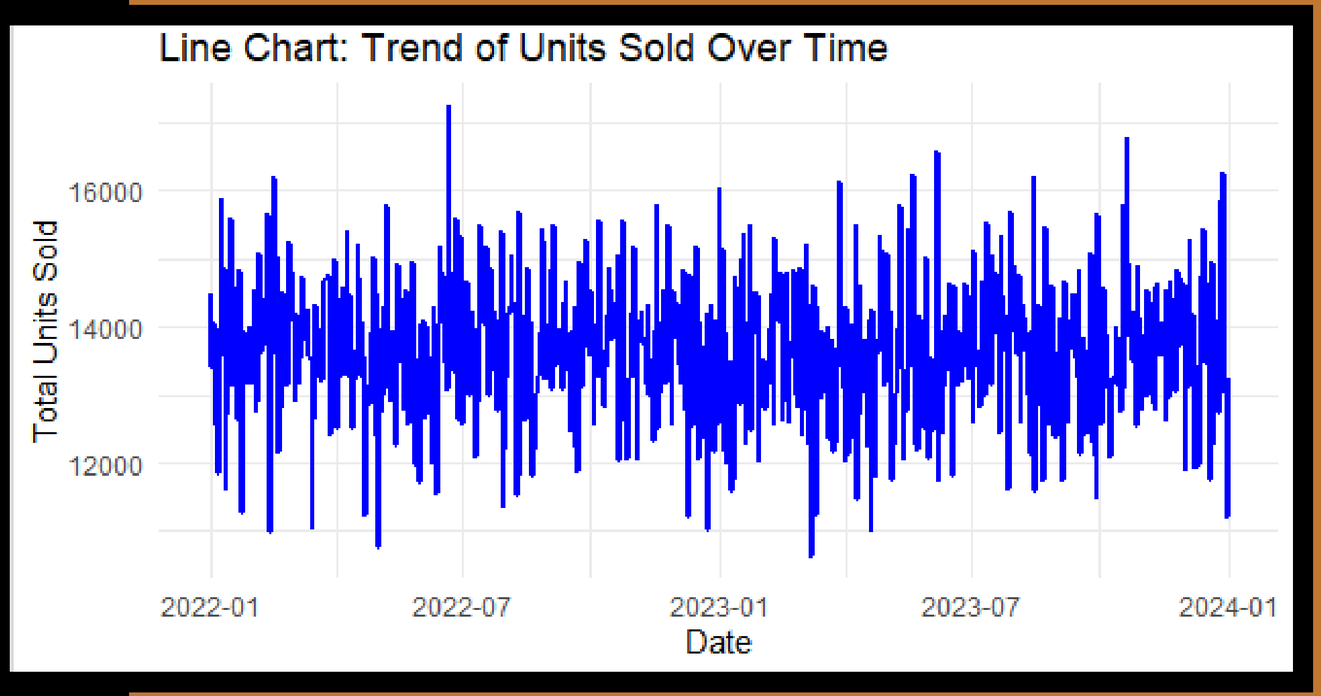
Price by Seasonality

- Higher or lower median prices during certain seasons might indicate increased demand or specific seasonal promotions.
- For instance, holiday seasons may show higher median prices due to increased demand.

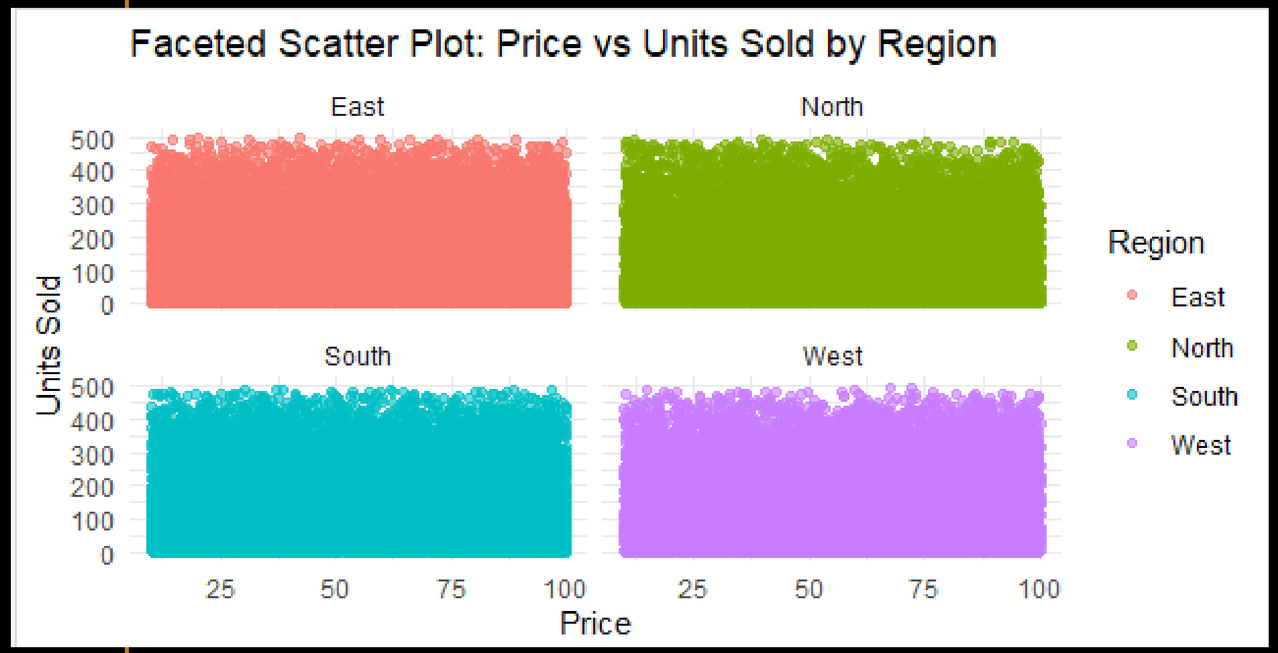


Trend of Units Sold Over Time

- The chart provides a clear visual representation of how the business has performed over time, helping identify growth opportunities or areas of concern.
- Plan promotional campaigns for periods with historically lower sales to boost revenue.
- Investigate low-sales periods to uncover underlying issues, such as inventory shortages or pricing misalignment.



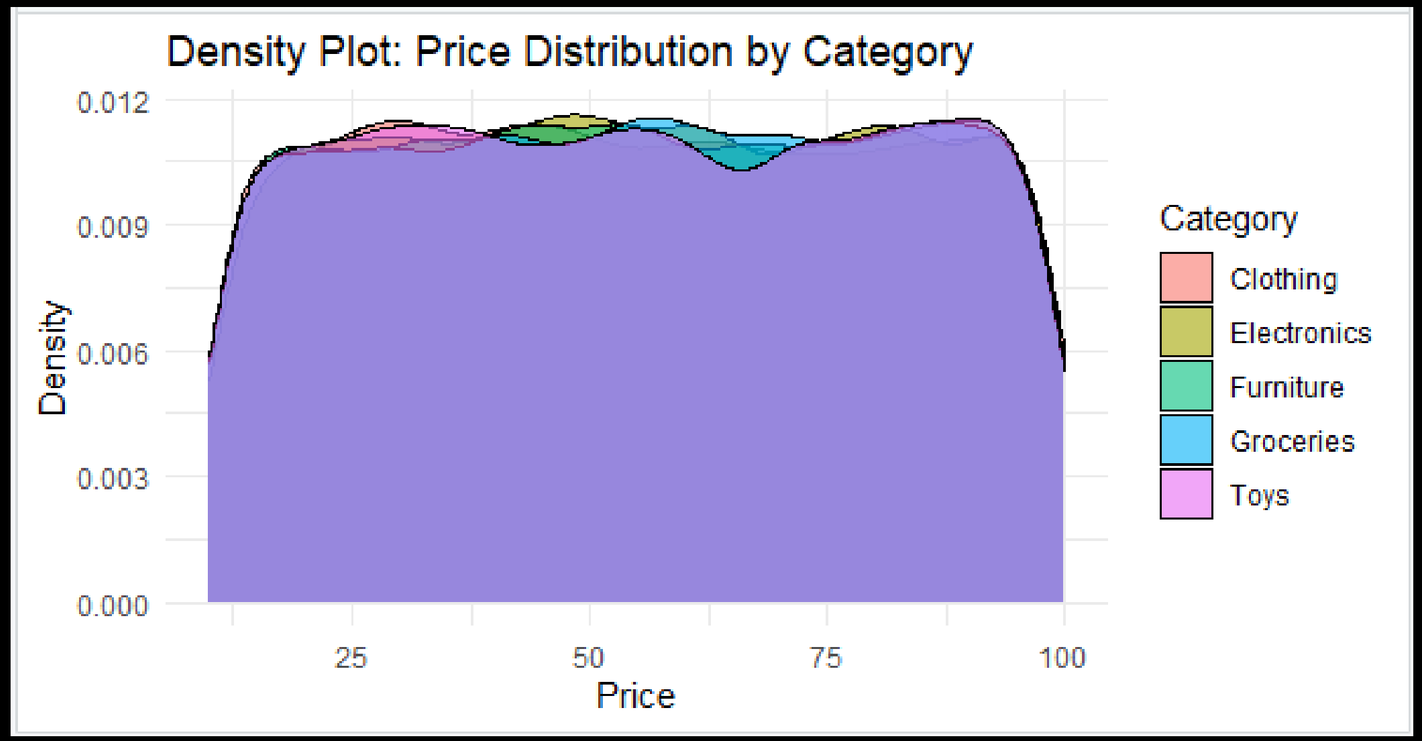
Price vs Units Sold by Region



- East (red), North (green), South (blue), and West (purple):

1. No significant differences are visible between the regions in terms of price or units sold.
2. All regions exhibit similar variability in the relationship between price and units sold, with no clear pattern indicating that higher prices correspond to higher or lower units sold.

Price Distribution by Category



- The height of the curves reflects the density of prices within specific ranges. For example, higher density means a larger number of products in that price range for a given category.
- • Some categories, such as Clothing (pink) and Groceries (blue), appear to contribute more consistently across the price range.

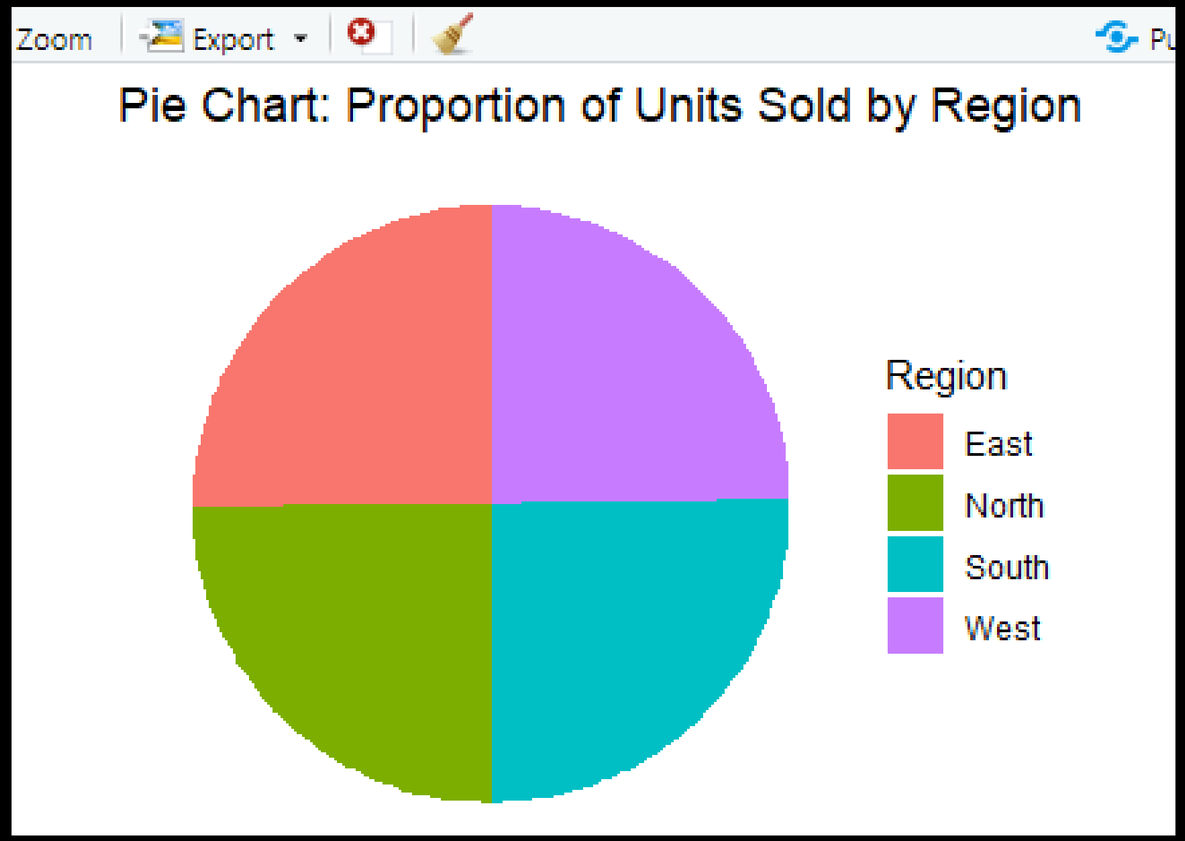
Total Units Sold by Region

```
pie_data <- retail_store_inventory1 %>%
  group_by(Region) %>%
  summarise(Total_Units_Sold =
sum(`Units.Sold`, na.rm = TRUE)) %>%
  mutate(Percentage = Total_Units_Sold /
sum(Total_Units_Sold) * 100)
```

Filter

	Region	Total_Units_Sold	Percentage
1	East	2511265	25.17412
2	North	2484966	24.91049
3	South	2507799	25.13938
4	West	2471552	24.77602

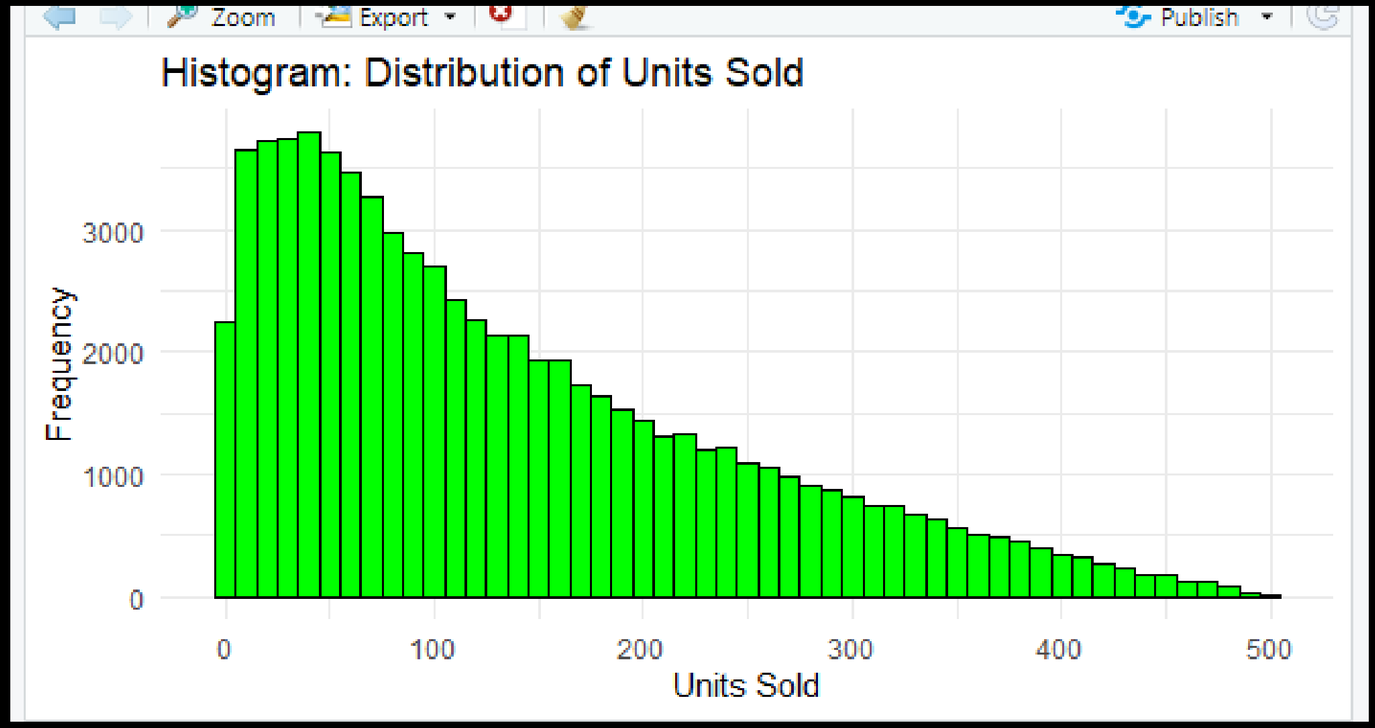
Proportion of Units Sold by Region



- The pie chart visually displays the relative contribution of each region to the total units sold.
- Focus inventory and marketing efforts on high-performing regions to sustain or increase sales.
- Investigate and address factors limiting sales in underperforming regions, such as promotions, product availability, or regional preferences.

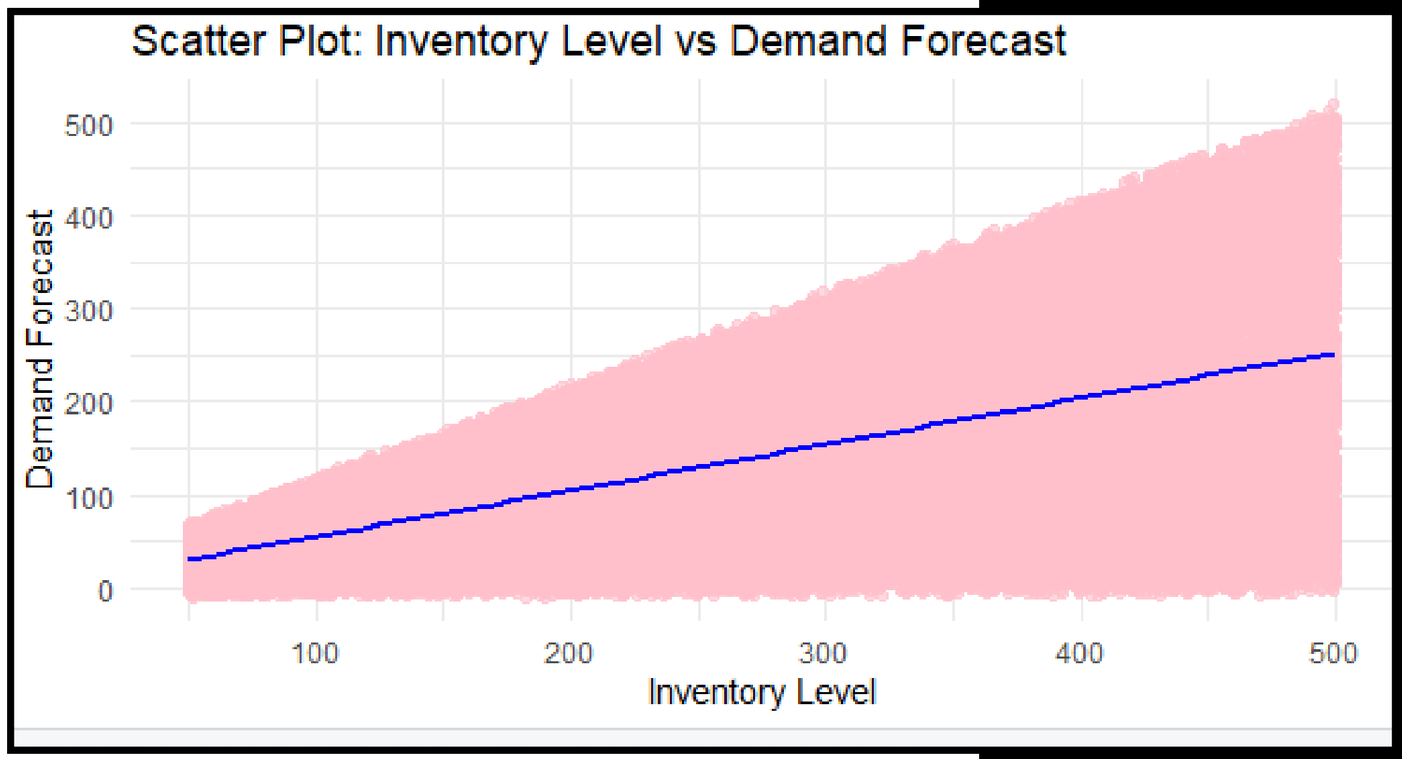
Distribution of Units Sold

- The x-axis represents the number of units sold, and the y-axis shows the frequency of occurrences for each range (binwidth = 10).
- Understanding the most frequent sales ranges can help optimize inventory levels to meet demand.
- High frequency in lower bins suggests potential underperformance in certain products, requiring further analysis.
- Higher frequency in upper bins reflects strong-performing items, which may benefit from increased stock or promotion.



Total Units Sold by Region

- It indicates that higher demand forecasts are associated with higher inventory levels, suggesting alignment between forecasting and inventory planning.
- It evaluates how well inventory levels are managed based on demand predictions, highlighting potential areas for efficiency or corrective measures.



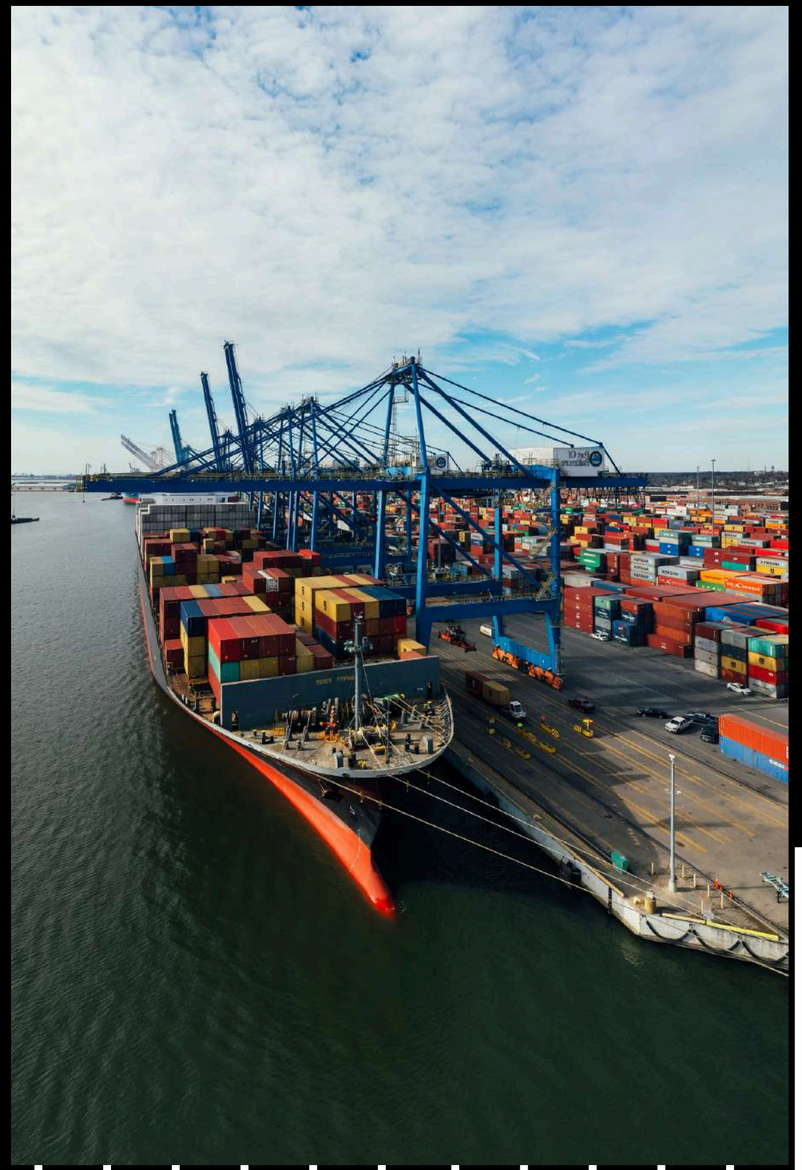
Findings

Key Insights:

- Category Trends: Certain product categories, like electronics, have higher average sales.
- Regional Performance: Sales vary significantly by region, with North performing best.
- Pricing and Discounts: Discounts are positively correlated with higher sales volumes.
- Inventory Management: Excess inventory in specific categories indicates potential overstocking issues.

Conclusion

- Pricing and discounts significantly influence sales.
- Regional analysis highlights performance differences.
- Inventory levels need optimization for demand forecasting.



Thank You