

# ***Retail Apparel Sales Analysis***



- The business objectives are to clean and analyze the data, calculate key performance indicators (KPIs), and provide actionable insights to improve sales across customer segments, products, and regions.

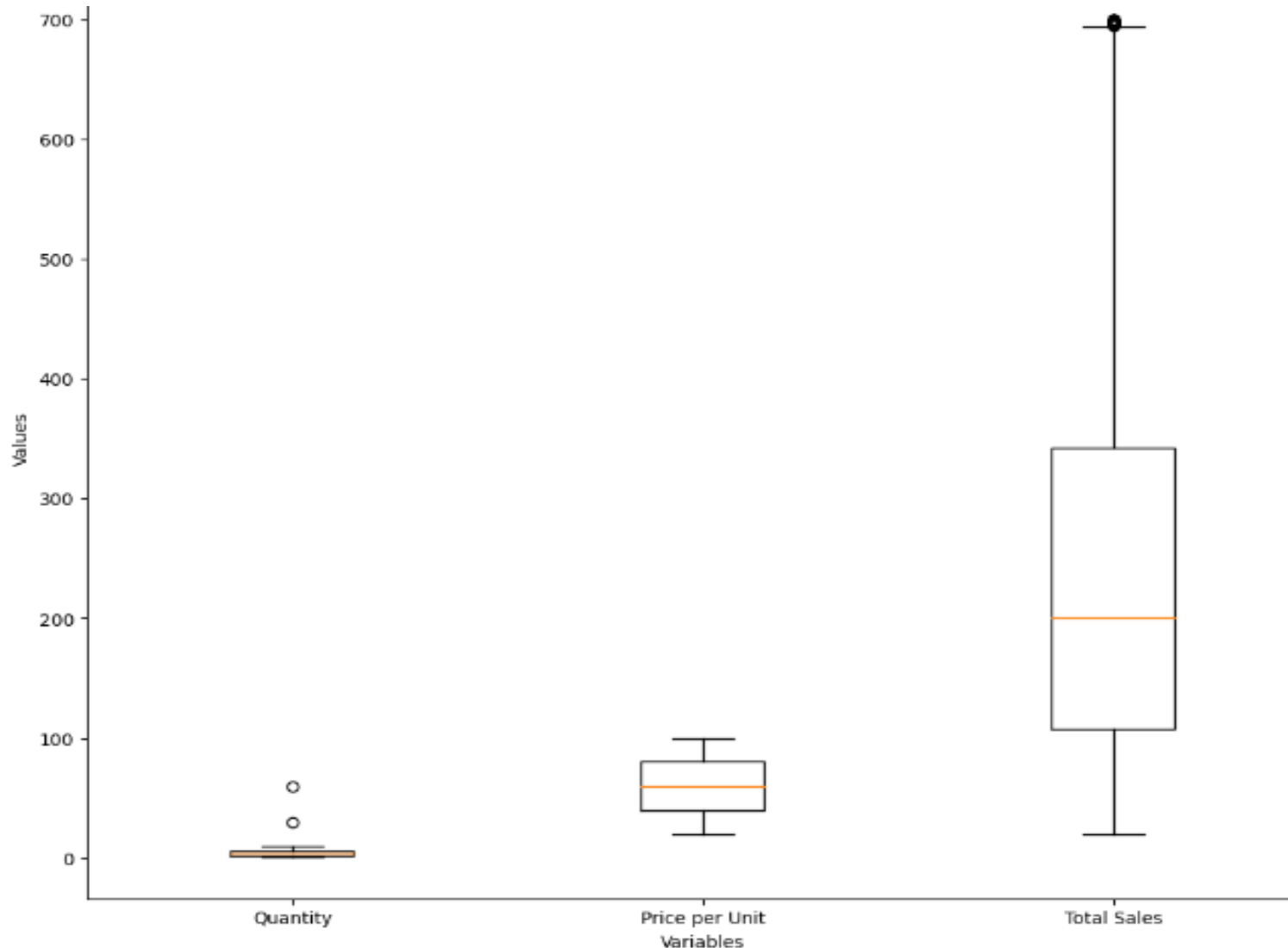
# Introduction

- This is about analyzing the sales performance in a retail apparel company.
- This dataset contains variables like Invoice Number, Customer ID, Product ID, Total Sales etc.
- 16898 rows and 12 columns are included in this dataset.
- Dataset contains data from January 1<sup>st</sup> 2021 to January 11<sup>th</sup> 2021.
- KPIs have been calculated such as Average Invoice Value, Rate of Returned Products, etc.
- Data Preprocessing, Data Analysis, Data Visualization parts have been done in this analysis.

# Data Preprocessing

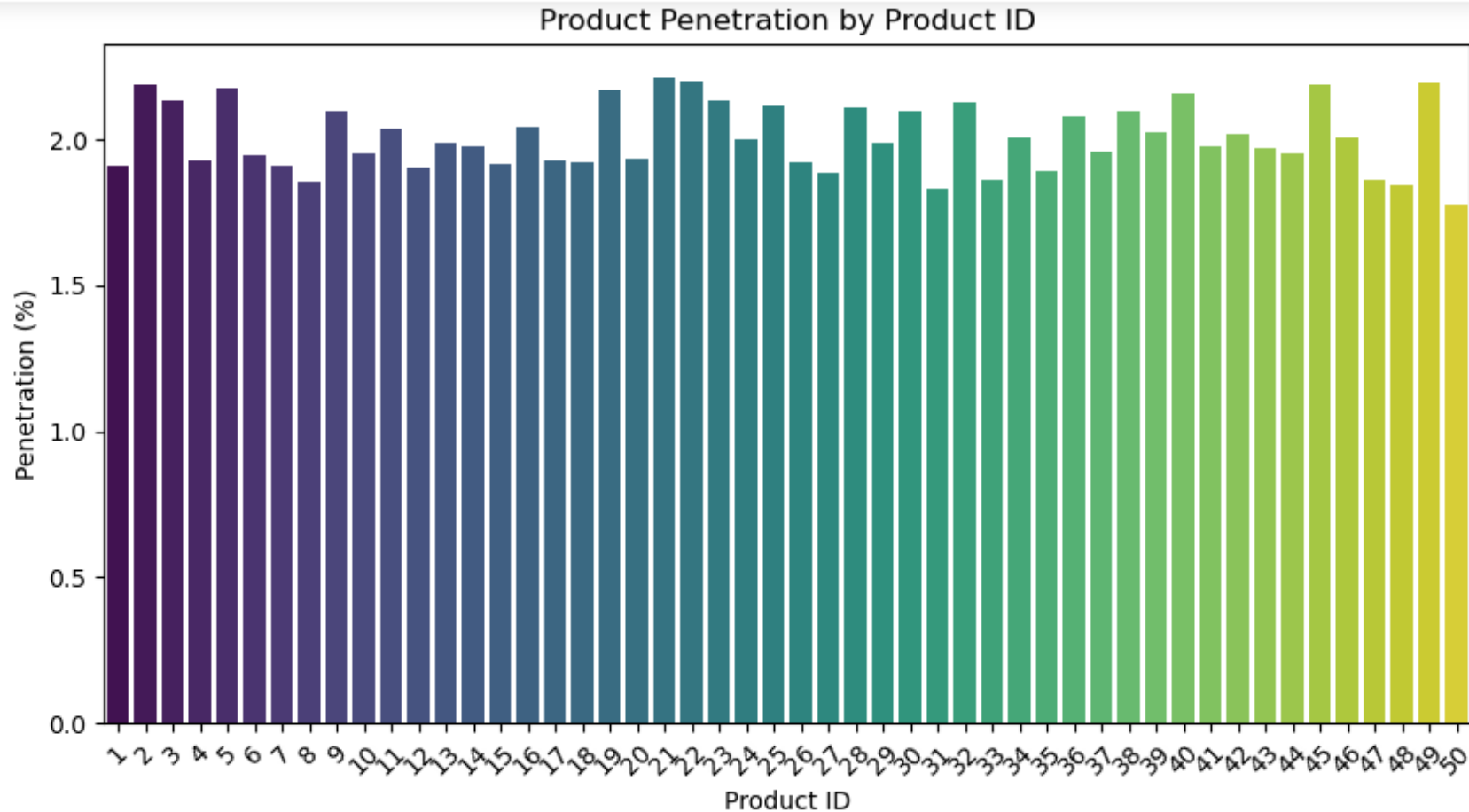
- The Date Time column has been converted into proper datetime format(Year, Month, Day) for the easiness of date-based operations.
- A row with an invalid Date Time was removed from the dataset.
- 346 null values from the Customer Age Group column have been dropped.
- No duplicate values were found in this dataset.
- There were misspelled values in the 'Product Category' and 'Sales Channel' variables and found using the `value_counts()` method in pandas and renamed to the correct name.

# Detecting and removing outliers



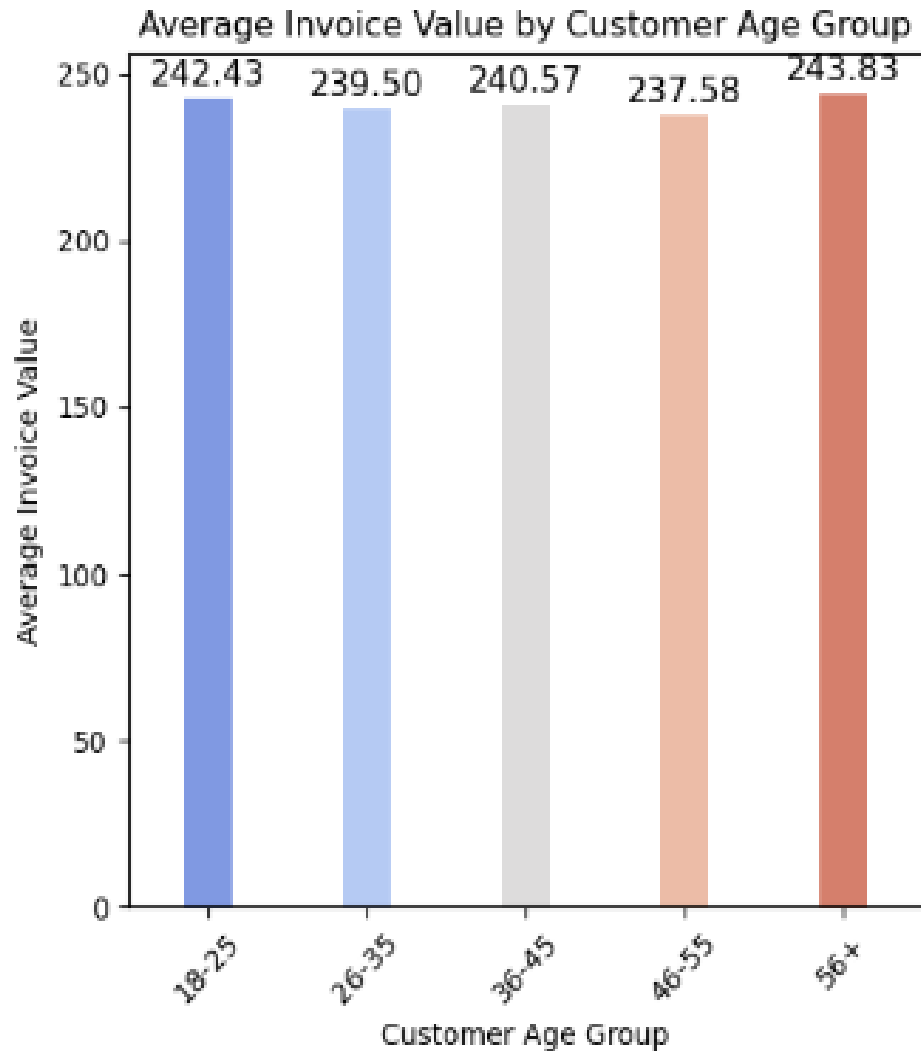
- Boxplots were used to detect the outliers of numerical variables.
- Outliers in the Quantity and Total Sales.
- IQR method has been used to detect the outliers.
- The numerical dataset without the outliers have been taken.
- Renamed the numerical variables with a new name and concatenated to the original dataset.
- The old numerical name columns and the null values from the dataset have been removed.
- Again, renamed the numerical variables to previous names and reorder it as previous dataset.

# Product Penetration



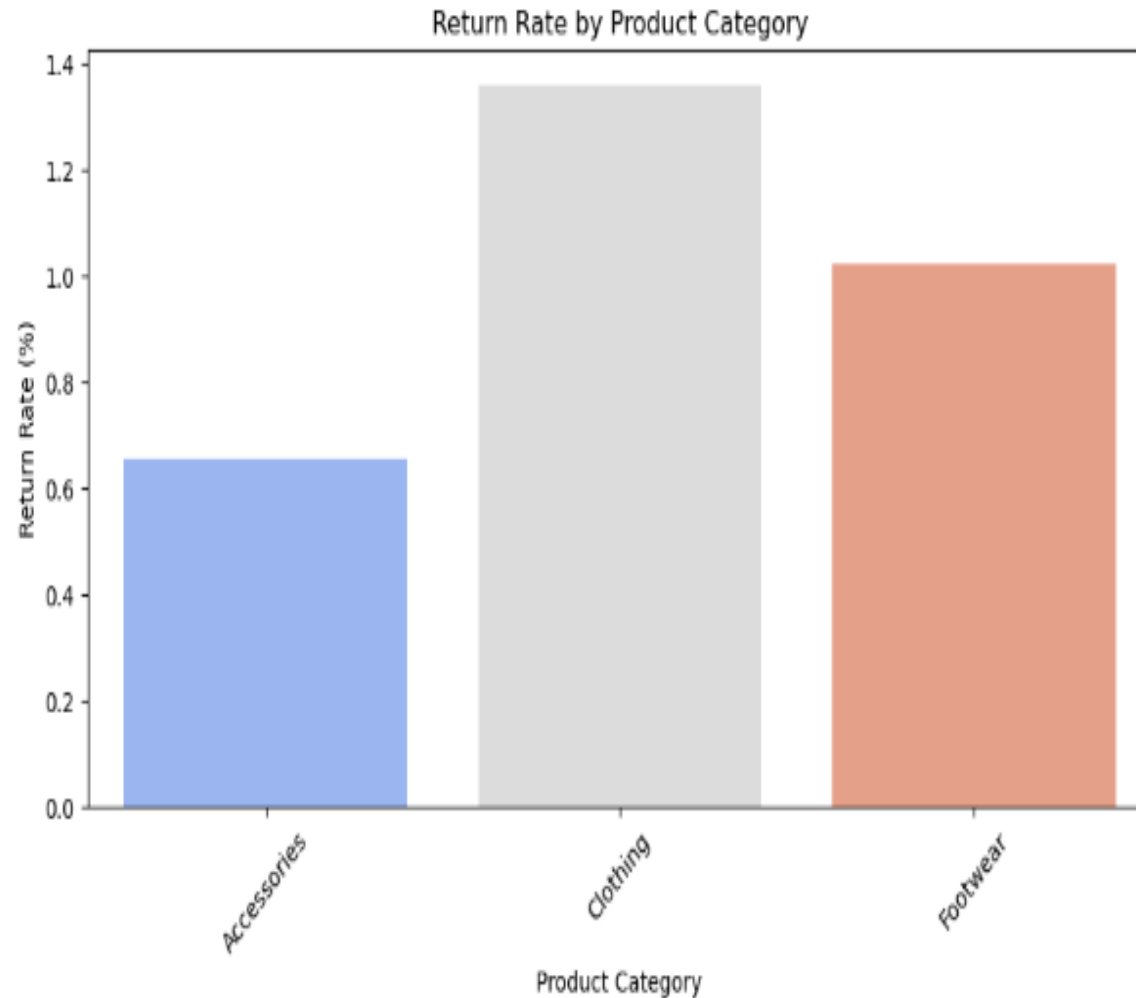
- The products over 2% penetration are most preferred products of customers.
- The Product ID 50 has the lowest penetration just above 1.5%
- Lower penetration can happen due to lack of awareness.
- There is no significant upward or downward trends and products may have similar segments.

# Average Invoice Value by Customer Age Group



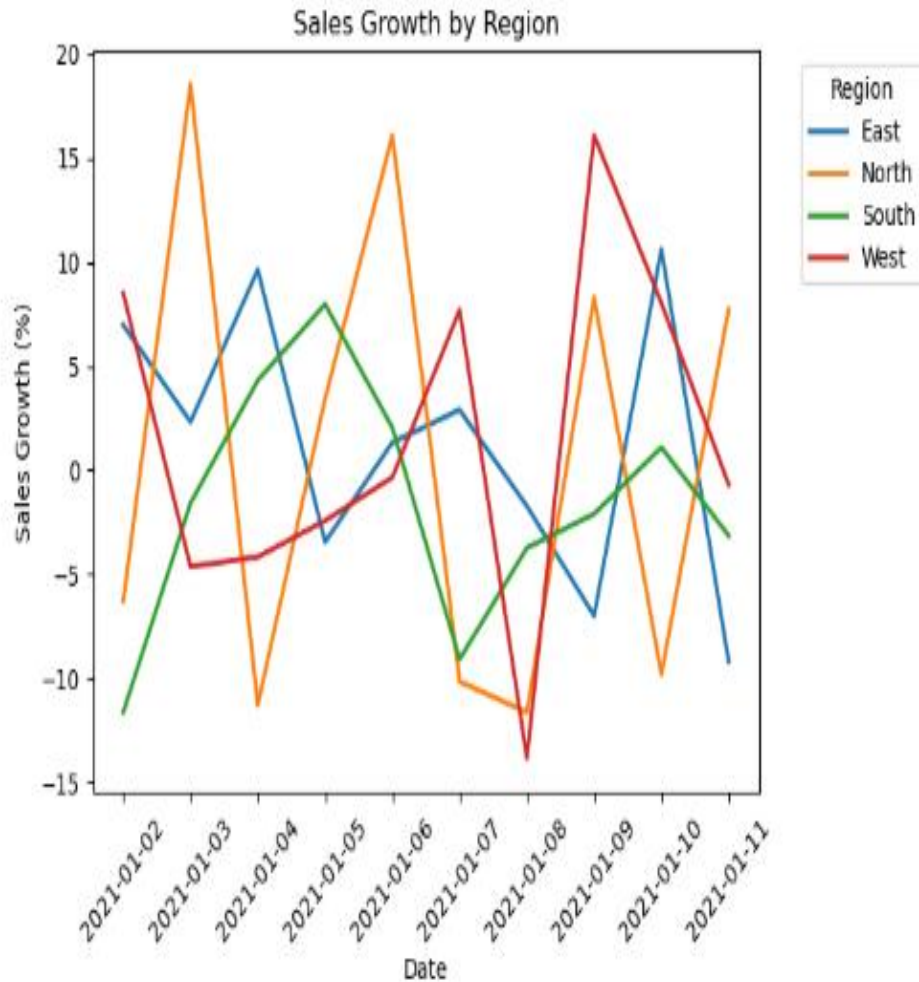
- Customer Age Group have been divided into 5 categories.
- The age group 56+ have the highest average per invoice this suggest older customers make larger purchase.
- 18-25 age group has the 2<sup>nd</sup> highest value.
- Young customers want to spend significant amount when they make purchases.
- 46-55 age group has the lowest values indicates less expensive items or smaller transactions.
- 26-35 and 36-45 age groups are middle aged adults and have a balance of spending power.

# Rate Of Returned Products



- Clothing has the highest return rate percentage.
- Clothing return rate maybe higher due to sizing issues, fabric damages, style preference.
- Accessories have the lowest return rate because they are less likely to have sizing issues.
- Customers may feel confident in purchasing accessories.
- Footwear returns can happen due to uncomfortable and sizing issues.

# Sales Growth by Region



- All four regions have significant fluctuations in sales growth over time.
- The North region have the most significant increases and decreases due to demand and supply disruptions.
- The South region have more smoother trend compared to other regions due to stable market conditions.
- East and West regions have fluctuations less extreme than the North region.
- The sharp peaks and drops can happen due to sales trend and promotions.



# Insights and Findings

- People prefer online shopping over in-store shopping, with 60.4% of sales coming from online purchases and 39.6% from in-store sales.
- There is a high positive relationship between the Quantity and Total Sales and moderate positive relationship between Price Per Unit and Total Sales.
- Most of the Invoices are for the Footwear and second for the Clothing and third for the Accessories.
- There were 1.1% of products have been returned from the sold products.
- Most of the sales are from the East region and least are from North region.
- Most of the sales are being happened before the 9<sup>th</sup> hour o the day and after the 9<sup>th</sup> hour the sales have gone down.
- Customers in the 46-55 age group have the highest total sales but the lowest average invoice value, possibly due to purchasing lower-priced items with high turnover.

***Thank You***