

# Nasscom Associative Data Analyst

Supermarket Sales  
Analysis in R



# Introduction

## Overview of the project :

- Supermarket sales analysis is the science of forecasting future trends based on existing data, was once dismissed as simply fancy fortune-telling. Today, it's making inroads into the retail sector.
- Grocery retailers and supermarkets, who generally work on razor-thin profit margins, have started turning to this data-driven science to help them get ahead of the competition



- Supermarket sales analysis is now a key driver of profit.
- Retailers are leveraging predictive technology tools to unleash empower their customer-facing and operational functions.
- This form of analytics not only tells the retailer of what you did last summer but can predict, with surprising accuracy, what you will be doing this summer.



# Majorly it has five areas of application:

- Promotions
- Shopper targeting
- Marketing Campaign Management
- Pricing
- Inventory management





# Dataset

## Attribute information:

**Invoice id:** Computer generated sales slip invoice identification number

**Branch:** Branch of supercenter (3 branches are available identified by A, B and C).

**City:** Location of supercenters

**Customer type:** Type of customers, recorded by Members for customers using member card and Normal for without member card.

**Gender:** Gender type of customer





# Dataset

**Total:** Total price including tax

**Date:** Date of purchase (Record available from January 2019 to March 2019)

**Time:** Purchase time (10am to 9pm)

**Product line:** General item categorization groups - Electronic accessories, Fashion accessories, Food and beverages, Health and beauty, Home and lifestyle, Sports and travel

**Unit price:** Price of each product in \$

**Quantity:** Number of products purchased by customer



# Dataset

**Payment:** Payment used by customer for purchase (3 methods are available – Cash, Credit card and Ewallet)

**COGS:** Cost of goods sold

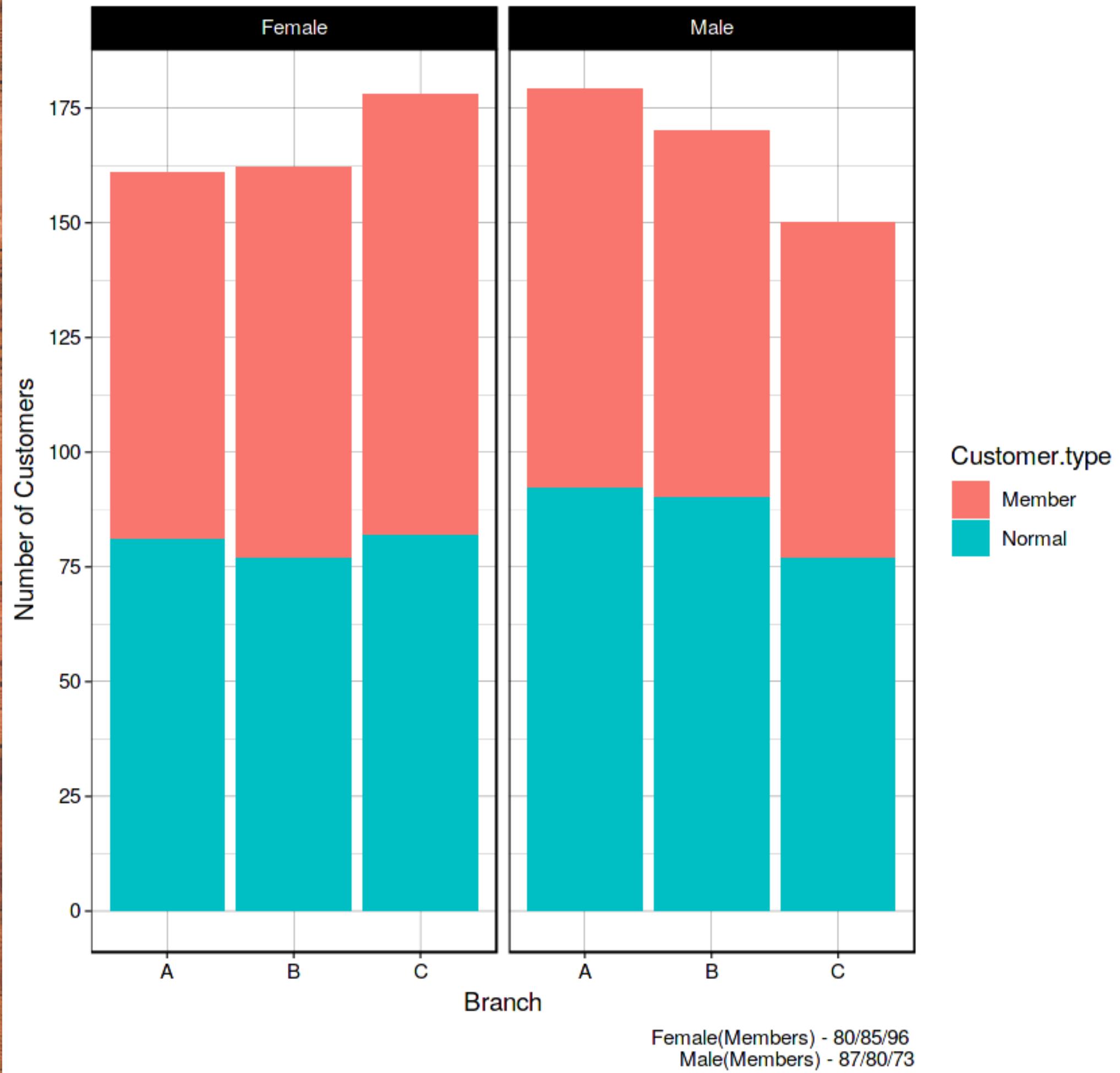
**Gross margin percentage:** Gross margin percentage

**Gross income:** Gross income

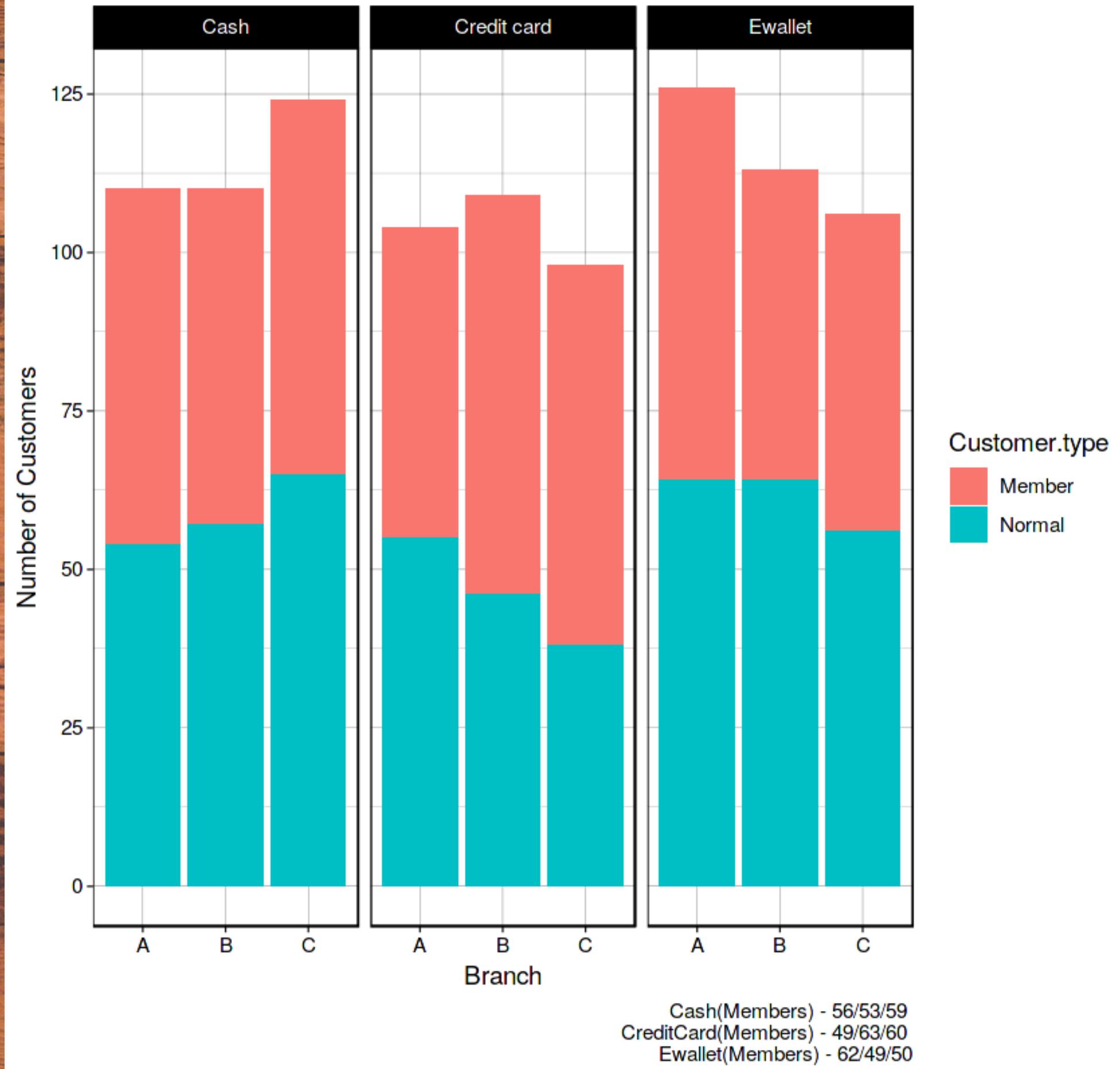
**Rating:** Customer stratification rating on their overall shopping experience.

**Tax:** 5% tax fee for customer buying

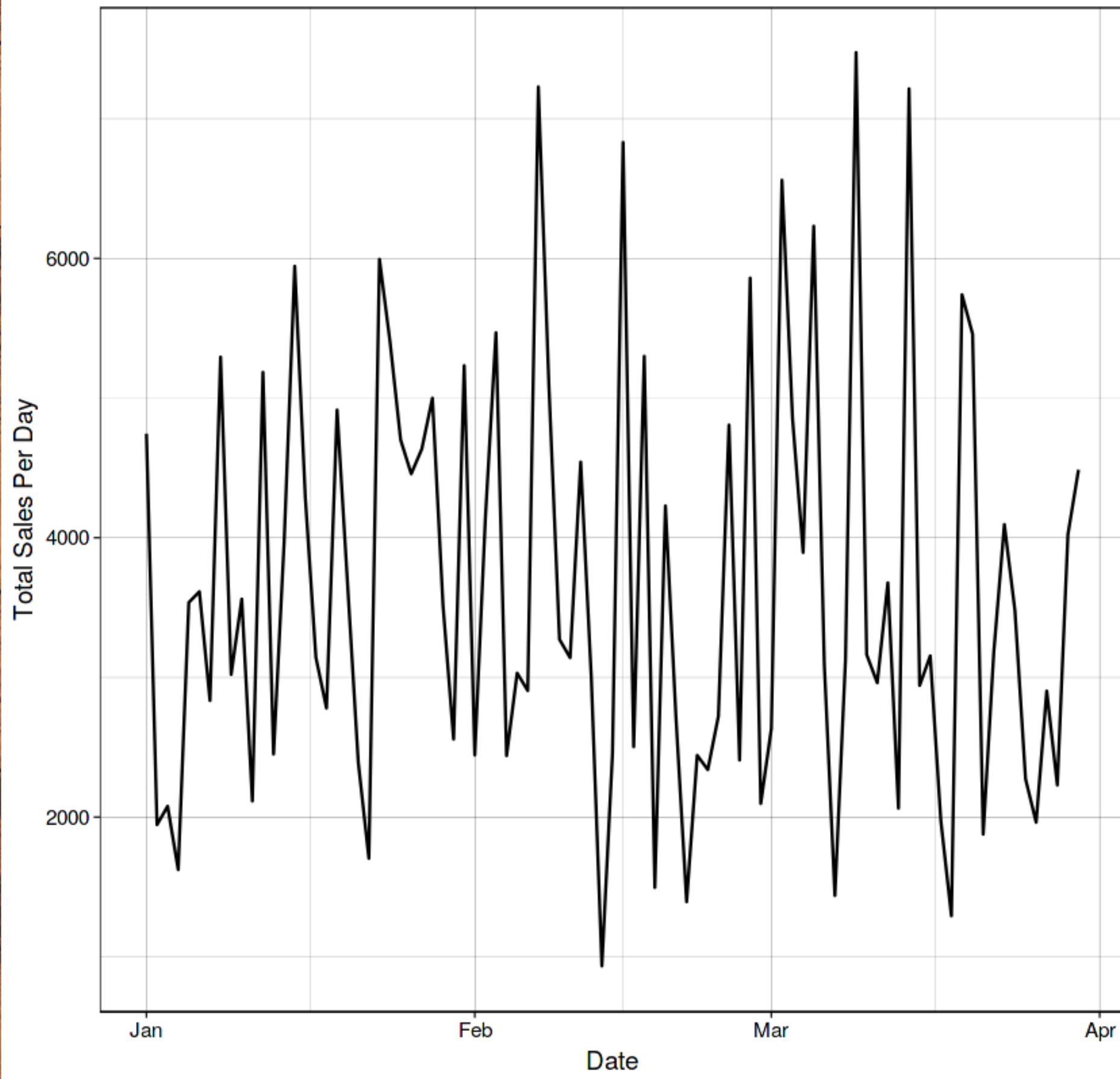
## Distribution of Customers in Each Branch Based on Gender



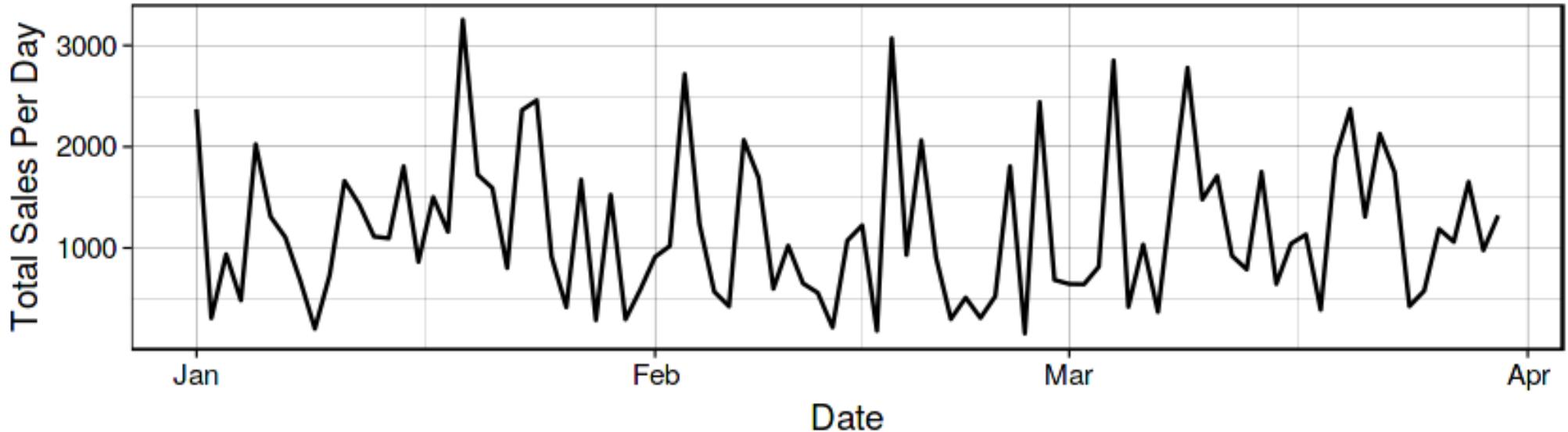
## Distribution of Customers in Each Branch Based on Payment Mode



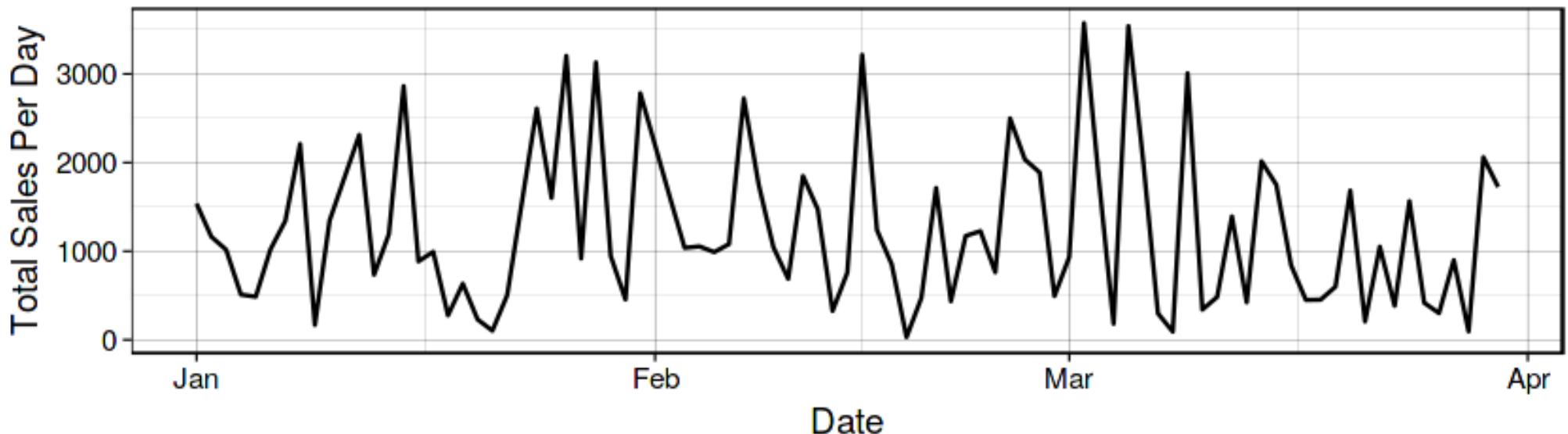
Time Series Relationship for the Total Sales per day



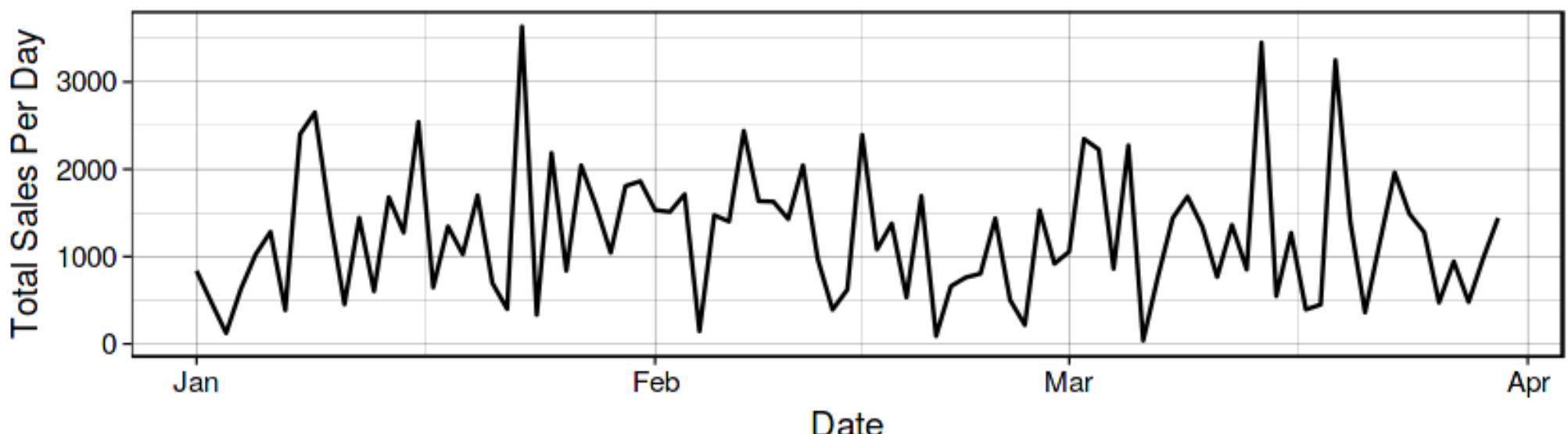
Time Series Relationship for the Total Sales per day in Branch A



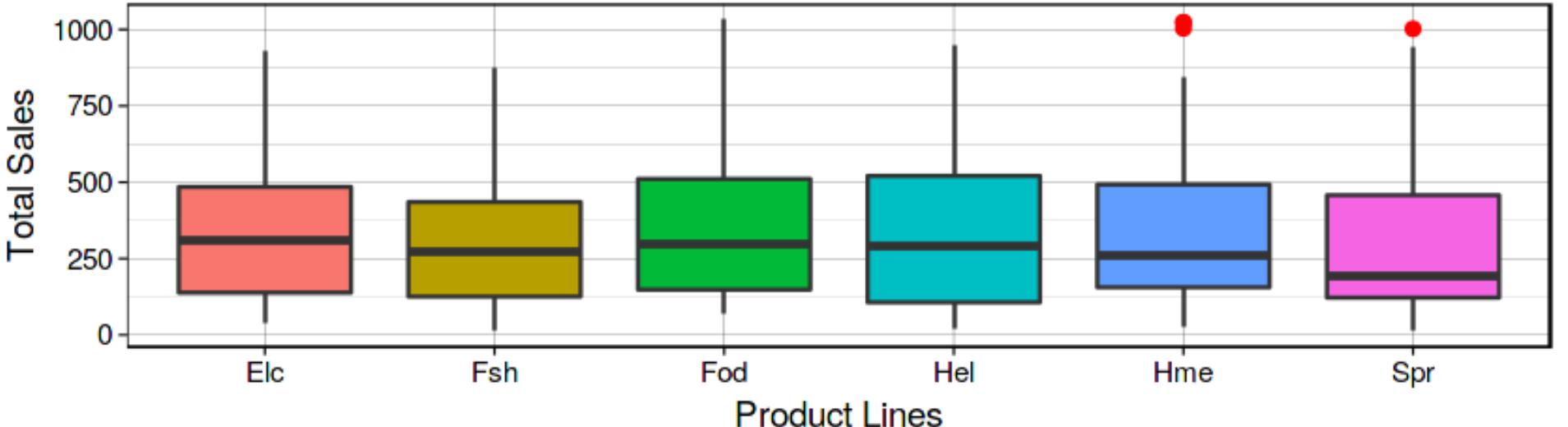
Time Series Relationship for the Total Sales per day in Branch B



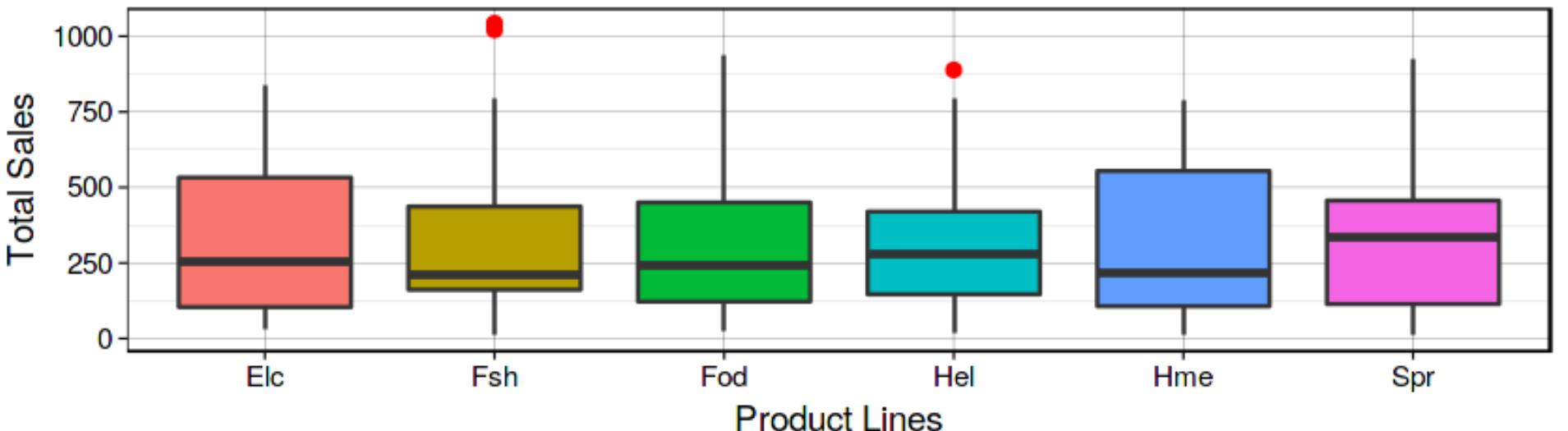
Time Series Relationship for the Total Sales per day in Branch C



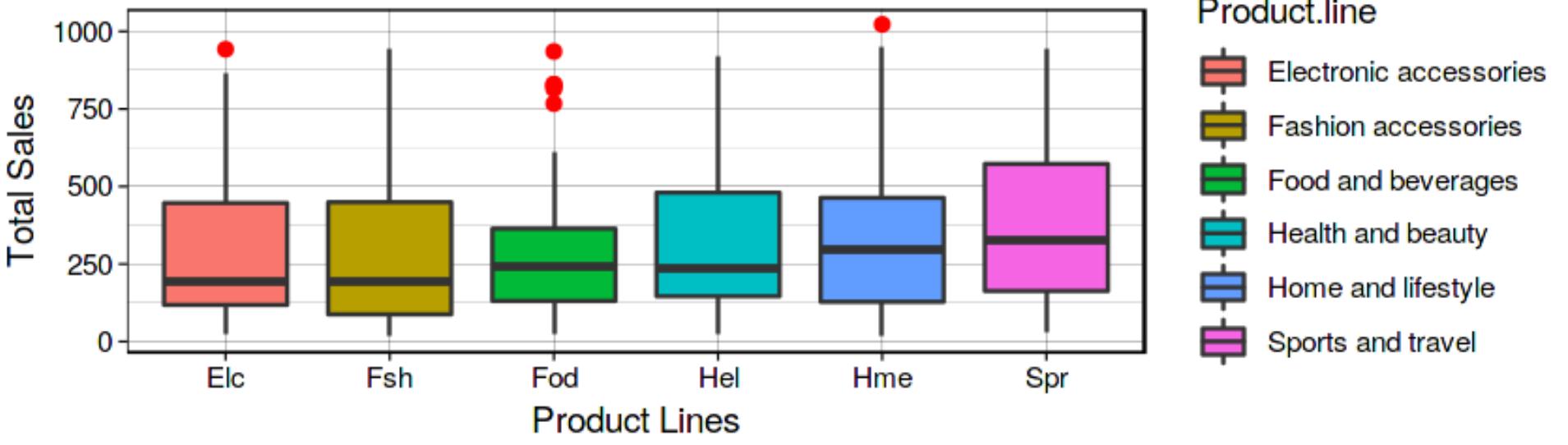
### January Sales Based on Product Lines



### February Sales Based on Product Lines



### March Sales Based on Product Lines





# Analysis

## In terms of scope

The feature invoice id is of no use for us. So let us drop the feature. We also need to check if there are any missing values in the data set.

## In terms of resources

try and arrive at some notable insights. Lets take a look into the branches customers rated high/low.

## In terms of time

Additionally, we converted the Date to a standardized format





# Summary

We have processed our data and cleaned it as per our requirements. Below are some plots based on the dataset. They include:

- GENDER COUNT
- RATING BASED ON BRANCH
- SALE OF THE PRODUCT PER HOUR
- MONTHLY INSIGHT OF THE DATA BASED ON BRANCH AND QUANTITY
- LASTLY A MONTHLY SUMMARY OF THE SALES PER HOUR OF OUR PRODUCT

The whole graph plotting and predictions were done on the basis of following dataset columns :

- |                 |                           |
|-----------------|---------------------------|
| • BRANCH        | • DATE                    |
| • CITY          | • TIME                    |
| • CUSTOMER.TYPE | • PAYMENT                 |
| • GENDER        | • COGS                    |
| • PRODUCT.LINE  | • GROSS.MARGIN.PERCENTAGE |
| • UNIT.PRICE    | • GROSS.INCOME            |
| • QUANTITY      | • RATING                  |
| • TAX.5.        |                           |
| • TOTAL         |                           |





**Thank  
You**