

# Product Sales Analysis with Machine Learning

## Phase 5

### **Team Members:**

Diya Arshiya S (2021115033) [diya.arshiya@gmail.com](mailto:diya.arshiya@gmail.com)

Dhivyadharshini S K (2021115030) [dhivyadharshini0907@gmail.com](mailto:dhivyadharshini0907@gmail.com)

Mukesh Raja K (2021115065) [mukeshrajatmr2021@gmail.com](mailto:mukeshrajatmr2021@gmail.com)

Mukilarasan V (2021115066) [mukilarasan.v@gmail.com](mailto:mukilarasan.v@gmail.com)

Karthik V (2021115321) [karthiksk9360@gmail.com](mailto:karthiksk9360@gmail.com)

### **Table of Contents**

1. Introduction
2. Data preprocessing
3. Visualization
4. Output
5. Insights
6. Conclusion

## **Introduction**

The purpose of this report is to document the data analysing performed on the dataset contained in the "statsfinal.csv" file. Data analysing is a crucial step in data analysis. It gives the visualized form of data.

## **Project Objective:**

The project aims to analyze and visualize sales data to derive actionable insights for inventory management and marketing strategies.

## **Design Thinking Process:**

**Empathize:** Understand the need for data analysis in inventory management and marketing.

**Define:** Define the specific objectives, such as identifying top-selling products, peak sales periods, and customer preferences.

**Ideate:** Develop data analysis and visualization strategies.

**Prototype:** Create code for data cleaning, formatting, and visualization.

**Test:** Evaluate the code's effectiveness in generating insights.

**Product Analysis:** Functions are defined to identify the top-selling products, products with the highest sales, and average sales on the 31st day of each month.

**Peak Sales Periods:** The code identifies the peak sales month and year for both total unit sales and total revenue.

**Actionable Insights:** The insights derived from the analysis can guide inventory management and marketing strategies. For example, knowing the top-selling products can help in stock management, while understanding peak sales periods can optimize marketing campaigns.

**Presentation:** The results and insights are typically presented in a report or presentation, including the visualizations and actionable recommendations.

### **Analysis Objectives:**

- Identify top-selling products.
- Determine products with the highest total unit sales and total revenue.

- Discover peak sales periods for different products.
- Understand customer preferences for specific products.
- Analyze sales trends by year.
- Insights and Recommendations:
- The insights derived from this analysis can help in optimizing inventory management and marketing strategies.

For instance:

- Knowing the top-selling products can guide inventory stocking decisions.
- Identifying peak sales periods can help in timing marketing campaigns and promotions.
- Understanding customer preferences can aid in tailoring marketing efforts to specific product categories.
- Sales trends by year can provide insights into long-term growth or decline of product categories.

## **Development Phases:**

### **Data preprocessing**

#### **Data Overview**

We begin by loading the dataset using the Python library **`pandas`**. The dataset is read from the *statsfinal.csv* file, and some initial information about the dataset is displayed using the **`info()`** method.

#### **Data Cleaning**

##### **i. Identifying the missing values:**

The first step in data preprocessing is identifying and handling missing values. Missing values can disrupt the analysis and modeling process. In this dataset, we identify missing values using the **`isnull().sum()`** method, which counts the number of missing values in each column.

##### **ii. Dropping Rows with Missing Values**

Even though there are no missing values, it is good practice to drop rows with missing data when necessary. This can be done using the **`dropna()`** method.

##### **iii. Removing Duplicates**

Duplicate rows can also affect the accuracy of analysis. To remove duplicate rows, the **`drop\_duplicates()`** method is used.

#### iv. Data Formatting

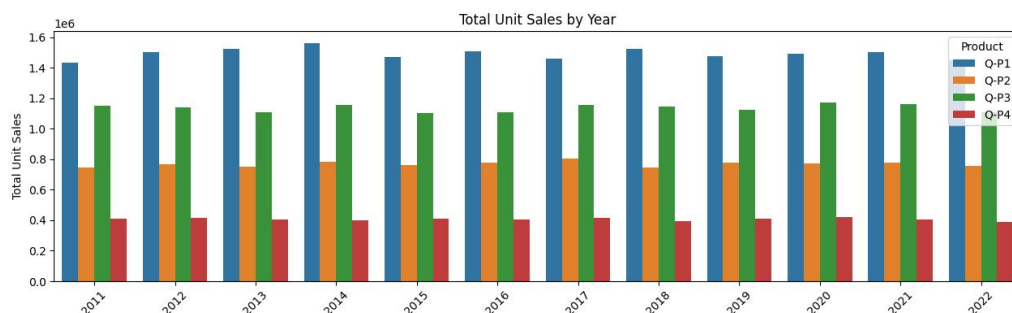
The next preprocessing step involves formatting the data, specifically by separating the date into separate columns for "Day," "Month," and "Year." This is achieved by applying a lambda function to split the "Date" column.

#### v. Data Reduction

In some cases, certain data points may need to be removed due to inconsistencies or insufficient data. In this dataset, data for the years 2010 and 2023 are removed as they have insufficient data. Additionally, incorrect date entries for September 31st and November 31st are also removed.

### Visualization

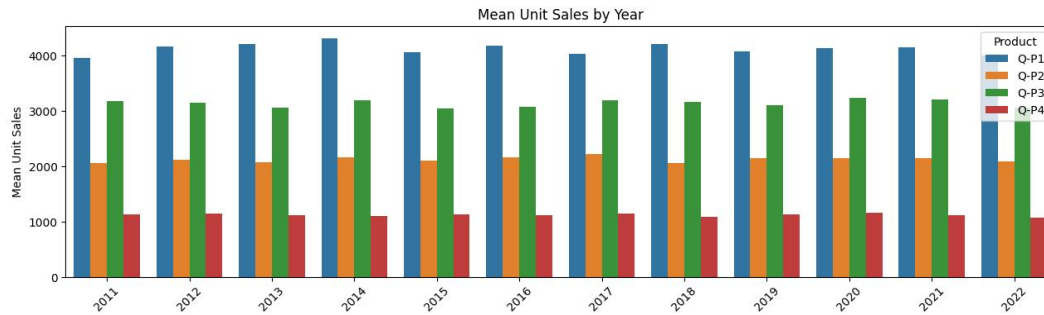
**Total Unit sales of products P1, P2, P3, P4 from 2011 to 2022**



#### Insights

Total unit sales have been relatively consistent from 2011 to 2022. Product Q-P2 consistently leads in total unit sales.

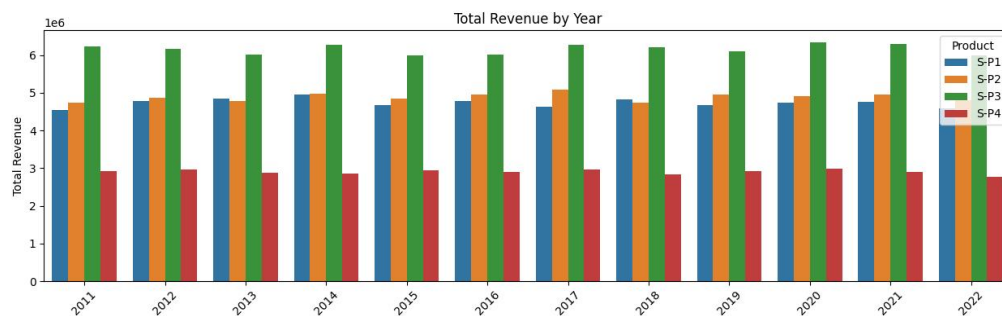
**Mean Unit sales of products P1, P2, P3, P4 from 2011 to 2022**



## Insights

The mean unit sales for all products show a gradual increase over the years. Product Q-P4 has the highest mean unit sales in recent years.

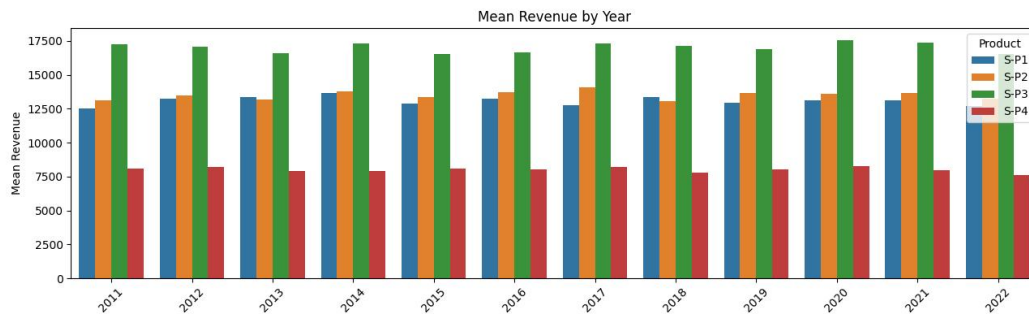
## Total revenue earned from products P1, P2, P3, P4 from 2011 to 2022



## Insights

The revenue earned from product is consistent over the years. The product S-P1 earned more revenue

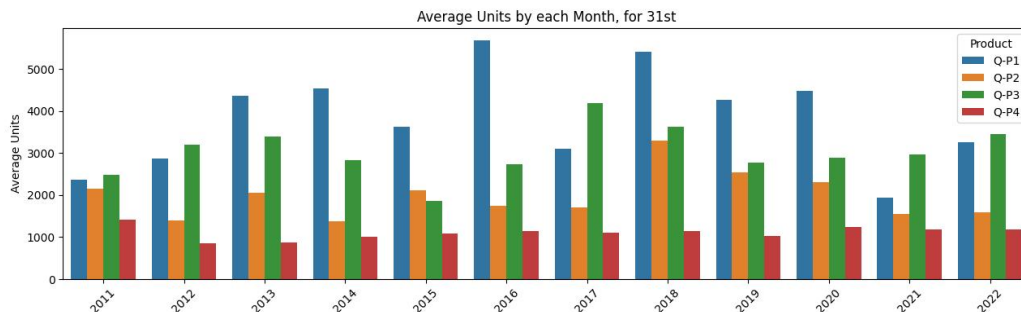
## Mean revenue earned from products P1, P2, P3, P4 from 2011 to 2022



### Insights

The revenue earned from product is consistent over the years. The product S-P1 earned more revenue

## Average units of product sold by each month, for 31<sup>st</sup> from 2011 to 2022



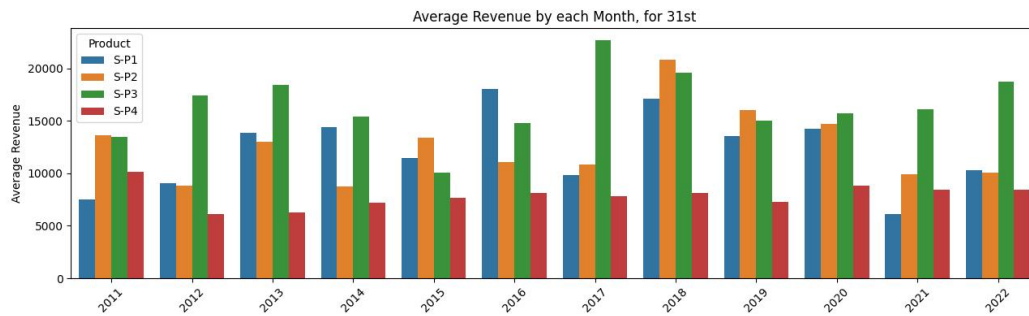
### Insights

The units of product sold in each month is gradually increasing and decreasing over the years. The product Q-P1 is sold more

## Average revenue earned from product sold by each month, for 31<sup>st</sup> from 2011 to



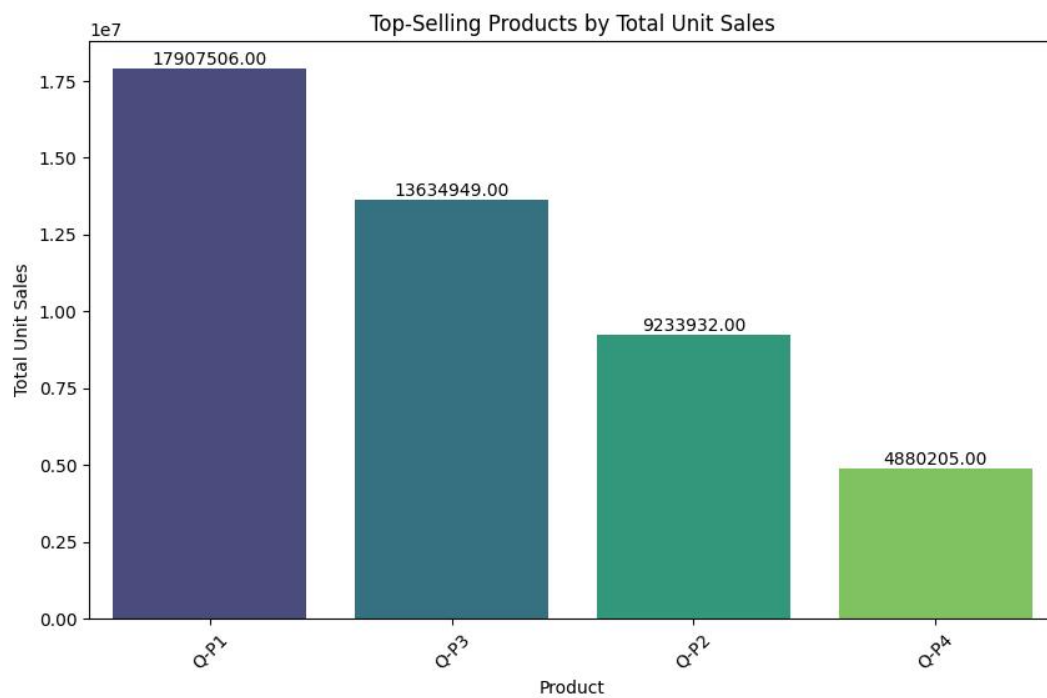
**2022**



## Insights

The average revenue earned in each month is gradually increasing and decreasing over the years. The product S-P3 earned much.

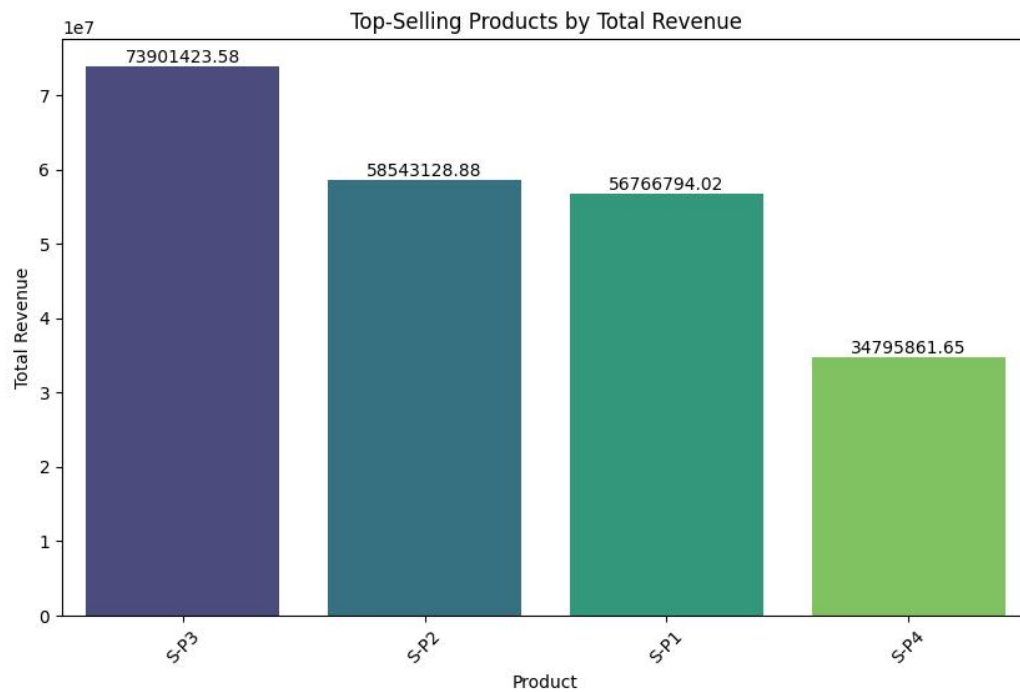
## Top selling products by total unit sales



## Insights

The product Q-P1 is the top selling product and Q-P4 is the least selling product based on unit sales

## Top selling products by total revenue



## Insights

The product S-P3 is the top selling product and S-P4 is the least selling product based on total revenue

## Conclusion

The given product dataset is analysed and visualized