|  |  |  |
| --- | --- | --- |
| NicePng_national-emblem-of-india_9441929.png | **PRODUCT SALES ANALYSIS** |  |

**NAAN MUDHALVAN PROJECT REPORT**

Submitted by

**KAVIPRIYA A (730321243011)**

**MANOVATHI J (730321243014)**

**PRIYADHARSHINI G (730321243019)**

**SWEATHA B (730321243028)**

**FIFTH SEMESTER**

**ARTIFICIAL INTELLIGENCE AND DATA SCIENCE**

**BUILDERS ENGINEERING COLLEGE , KANGEYAM**

**ANNA UNIVERSITY :: CHENNAI 600 025**

**NOVEMBER 2023**



**BONAFIDE CERTIFICATE**

Certified that this is a Bonafide record of work done by **KAVIPRIYA A(730321243011) , MANOVATHI J(730321243014)** and **PRIYADHARSHINI G(730321243019) , SWEATHA B(730321243028)** in NAAN MUTHALVAN during the Academic year 2023- 24 for fifth Semester.

|  |  |
| --- | --- |
|  |  |
| **STAFF -INCHARGE** | **HEAD OF THE DEPARTMENT** |
|  |  |
| Submitted for the Naan mudhalvan | viva voce held on\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
|  |  |
| **INTERNAL EXAMINER** | **EXTERNAL EXAMINER** |

**ABSTRACT**

The analysis of product sales is a pivotal endeavor for businesses seeking to thrive in an ever-evolving marketplace. This process entails a comprehensive examination and comprehension of product performance over a specified period, with the ultimate goal of gleaning invaluable insights to bolster decision-making, streamline inventory management, refine marketing strategies, and augment overall profitability. In an era marked by data-driven decision-making, this analysis stands as a linchpin in the competitive landscape, serving as a compass that guides companies toward success. This abstract delves into the multifaceted realm of product sales analysis, shedding light on the core objectives and the profound impact it can have on the strategic direction of a business. With data as its cornerstone, product sales analysis equips organizations with the knowledge necessary to adapt, innovate, and excel in a dynamic market environment.

|  |
| --- |
| **TABLE OF CONTENT** |
| **CHAPTER TITLE PAGE NO.** |
| **NO.** |
| 1. **PROBLEM STATEMENT 5** |
| **2 DESIGN THINKING 6** |
| **3 DATASET DEFINITION 8** |
| **4 DATA PREPROCESSING AND 10**  **FEATURE EXTRACTION** |
| **5 PROPOSED ALGORITHM 14**    **6 PROPOSED INNOVATION 18**  **TECHNIQUE**  **7 CONCLUSION AND 22**  **FUTURE SCOPE** |
|  |

**CHAPTER 1**

**PROBLEM STATEMENT**

The problem of product sales analysis involves examining and understanding the performance of products sold by a company or business over a specific period. This analysis aims to provide valuable insights to support decision-making, optimize inventory, improve marketing strategies, and enhance overall profitability.

**Key Challenges:**

1. **Data Collection and Quality:**

Ensuring that the data collected is accurate and comprehensive can be a challenge. Inaccurate or incomplete data can lead to misleading conclusions and hinder effective analysis.

**2. Data Integration:**

Companies often have data stored in various systems, and integrating this data can be complex. Combining data from different sources, such as point-of-sale systems, online platforms, and physical stores, can be challenging.

**3. Data Volume :**

For businesses with a high volume of sales, handling and processing large datasets can be a significant challenge. It may require powerful hardware and software infrastructure to perform the analysis efficiently.

**4.Seasonality and Trends :**

Understanding and accounting for seasonal variations and trends in product sales is important for accurate forecasting and decision-making.

**5. Customer Behavior Analysis:**

Connecting product sales to customer behavior and preferences can be challenging. Understanding why customers buy certain products and how to tailor marketing and inventory management accordingly can be complex.

**CHAPTER 2**

**DESIGN THINKING**

**ANALYSIS OBJECTIVES**

When The objective of product sales analysis is to gain valuable insights and make data-driven decisions to improve the performance and profitability of a product or product line.It including, Sales Performance Assessment, Market Segmentation,Demand Forecasting,Customer Behavior Analysis,Price Optimization, Geographic Analysis, Customer Feedback Analysis,Seasonal and Trend Analysis

1. **Market Segmentation :**

**Objective**: To divide the market into distinct segments based on demographics, psycho graphics, behavior, and other relevant factors.

**Metrics**: Customer segmentation rate (percentage of customers segmented) Conversion rate by segment,Market share by segment,Customer satisfaction scores by segment

1. **Demand Forecasting:**

**Objective:** To predict future demand for products or services accurately.

**Metrics**: Forecast accuracy (measured as the variance between predicted and actual demand), Sales growth rate,Inventory turnover ratio, Customer churn rate.

1. **Customer Behavior Analysis deate:**

**Objective:** To understand how customers interact with your products or services.

**Metrics**: Customer acquisition cost (CAC),Customer lifetime value (CLV),Churn rate,Repeat purchase rate,Customer satisfaction scores, Net Promoter Score (NPS) .

**4. Price Optimization :**

**Objective:** To determine the optimal pricing strategy that maximizes profit and customer value.

**Metrics:** Gross margin, Price elasticity, Customer willingness to pay (WTP), Pricing efficiency (measured as the ratio of price change to revenue change)

**5**. **Geographic Analysis:**

**Objective:** To assess how product sales vary geographically and make informed decisions based on regional performance.

**Metrics:** Sales by region,Market share by region,Regional growth rate,Regional customer demographic,Distribution and logistics costs by region

**6. Customer Feedback Analysis:**

**Objective:** To gather and analyze customer feedback to improve products, services, and customer experience.

**Metrics:** Customer feedback response rate, Net Promoter Score (NPS), Customer satisfaction scores,Number of product or service improvement suggestions received

**7. Seasonal and Trend Analysis:**

**Objective:** To identify patterns, trends, and seasonal variations in sales data.

**Metrics**: Sales seasonality index,Year-over-year sales growth,Trend identification (e.g., upward or downward), Predictive accuracy of seasonality and trend models

**8. Learn and Iterate:**

**Objective:** Document lessons learned and iterate for continuous improvement**.**

**Metrics:** Analyze feedback from all stages of the design process. Document insights, successes, and challenges. Track changes in awareness and participation over time to inform future iterations.

**CHAPTER 3**

**DATASET DEFINITION**

**Product Sales Analysis Data set:**

A product sales analysis data set is a structured repository of data that encompasses various attributes related to the sales of products by a company or business. This data set typically includes the following components:

**Product Information:**

* Product ID or SKU
* Product name
* Product category or type
* Product description
* Manufacturer or supplier information

**Sales Transaction Data:**

* Date of sale
* Sales order or invoice number
* Customer ID or name
* Quantity sold
* Unit price
* Total sales revenue

**Customer Information:**

* Customer demographics (e.g., age, gender, location)
* Customer contact information (optional)
* Customer segmentation (if available)

**Geographic Information:**

* Location of sale (if sales occur in multiple locations)
* Regional or geographic data (for geographic analysis)

**Marketing and Promotions Data:**

* Marketing campaign information
* Discounts or promotions applied
* Source of sale (e.g., online, in-store)

**Time-Related Data:**

* Time of sale (hour, day, month)
* Seasonality data (for seasonal and trend analysis)
* Historical sales data (for trend analysis)

**Customer Feedback and Reviews:**

* Customer reviews and ratings (if applicable)
* Feedback and comments related to products

**Inventory Data (optional):**

* Stock levels
* Restocking or replenishment information

**Returns and Refunds Data (if applicable):**

* Date and reason for returns
* Refund amounts

**Additional Metrics (for analysis):**

* Gross margin
* Net profit
* Customer acquisition cost (CAC)
* Customer lifetime value (CLV)
* Inventory turnover ratio

**CHAPTER 4**

**DATA PREPROCESSING AND FEATURE EXTRACTION**

**DATA PREPROCESSING**

Data preprocessing and data cleaning are essential steps in preparing a product sales analysis dataset for meaningful and accurate analysis. These processes involve identifying and addressing issues in the data to ensure that it is accurate, consistent, and suitable for analysis. Here are some common data preprocessing and data cleaning tasks for a product sales analysis dataset:**Data Cleaning:Handling Missing Values:** Identify and handle missing values in the dataset. Depending on the situation, you can either remove rows with missing values, replace them with appropriate values (e.g., mean, median, mode), or use advanced imputation methods.

**Data Type Conversion**: Ensure that data types are consistent and appropriate for analysis. For example, dates should be in date format, and numerical values should be in the correct numeric format.

**Deduplication:** Check for and remove duplicate records in the dataset to avoid double-counting or inaccuracies in analysis.

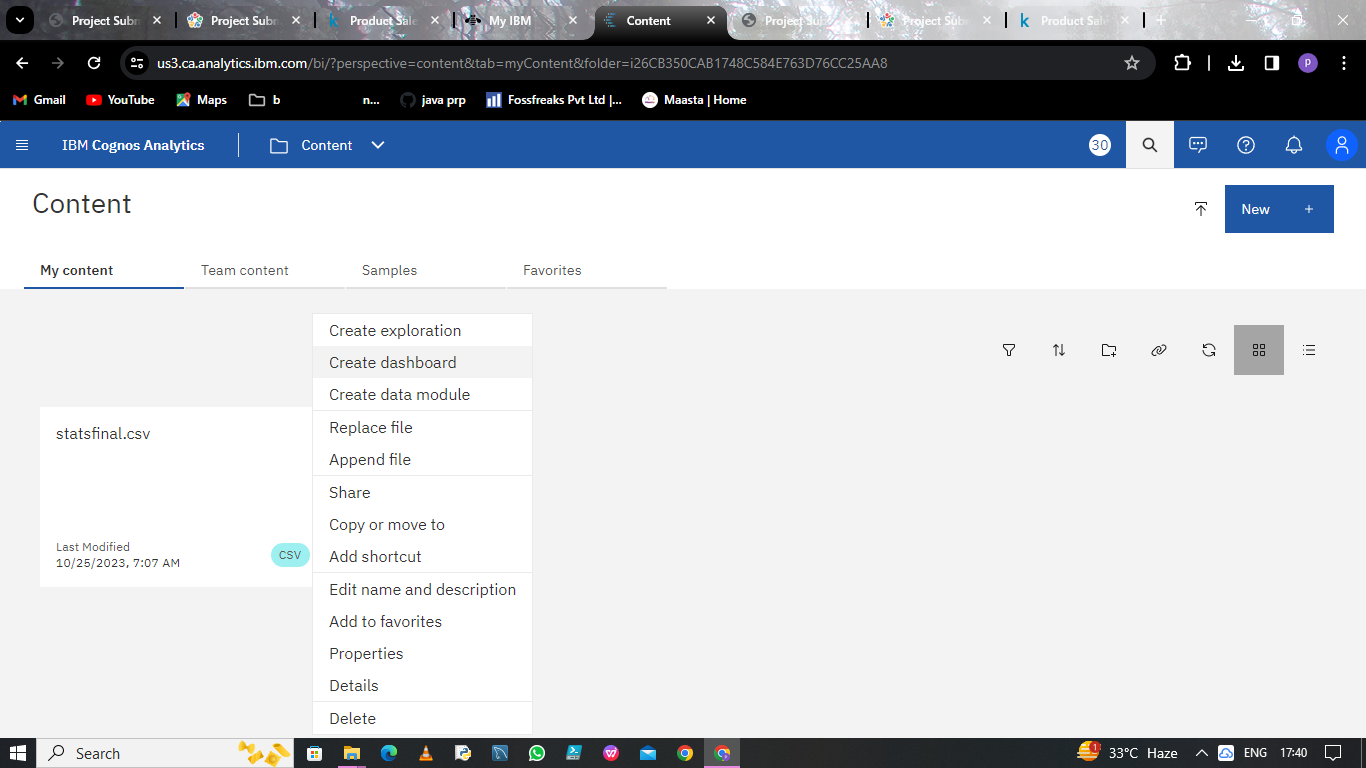
**Outlier Detection and Handling:**Identify outliers in numerical data that may skew analysis results. You can choose to remove, transform, or handle outliers based on the context.

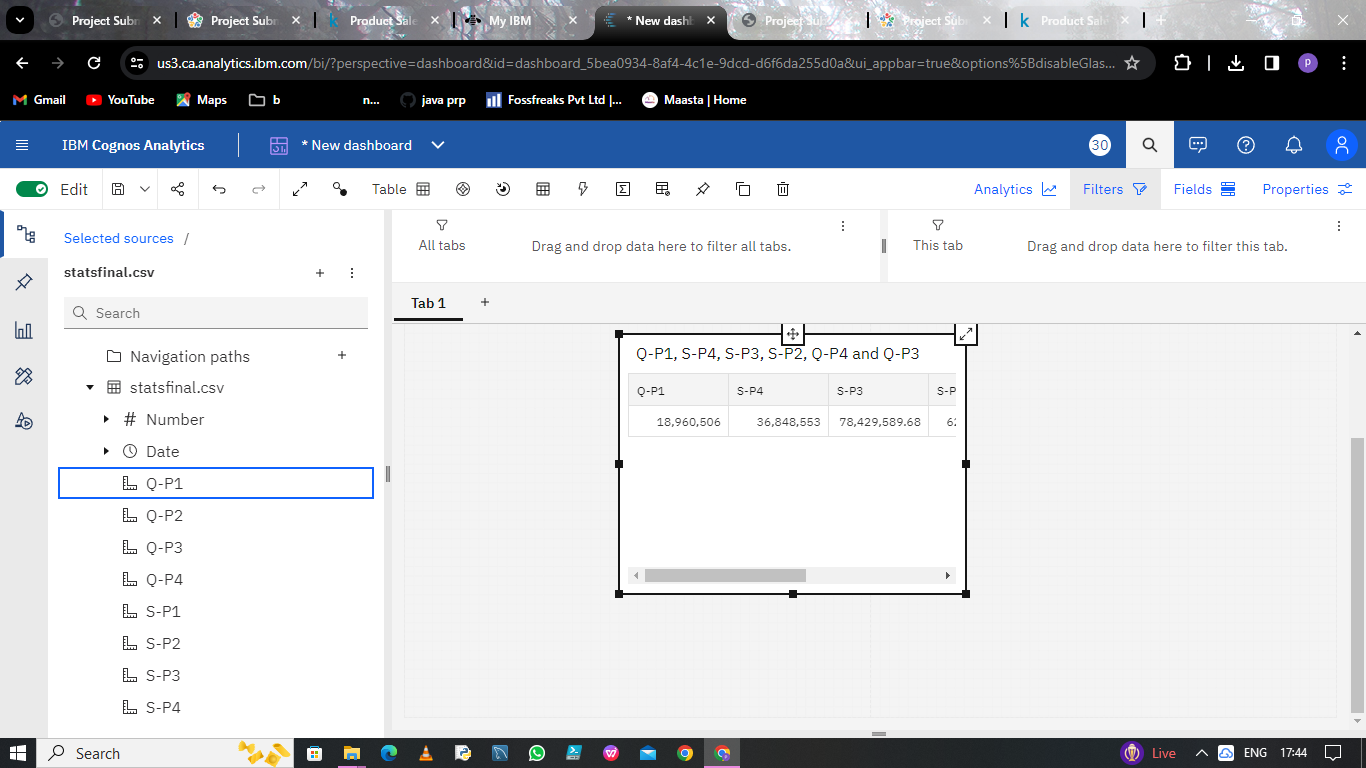
**Normalization and Scaling**: Normalize or scale numerical features if needed, especially when using algorithms sensitive to feature scales, such as distance-based methods or neural networks.

**Feature Extraction:**

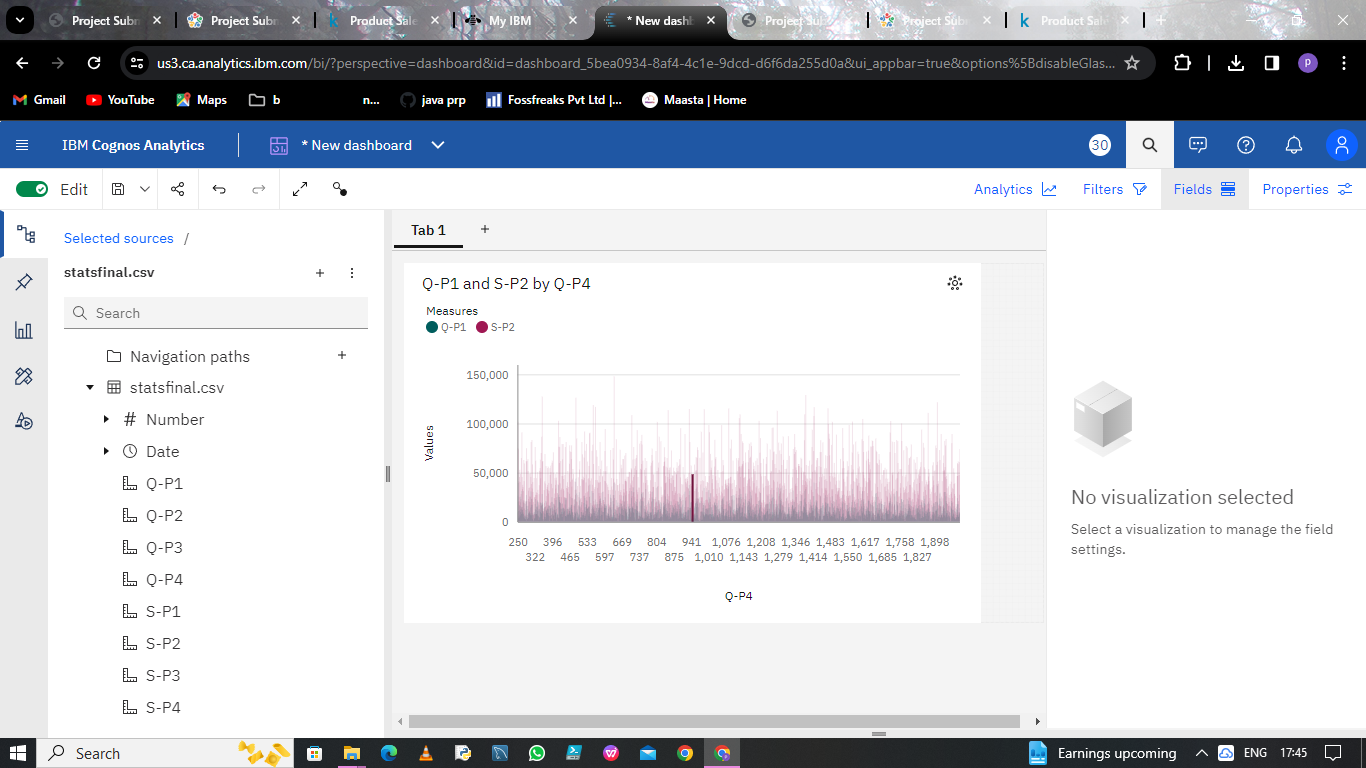
Feature extraction is a critical step in the preparation of a product sales analysis dataset. It involves selecting and transforming relevant information from the dataset to create new features (variables) that capture essential aspects of the data. Feature extraction helps to reduce dimensionality, improve model performance, and uncover hidden patterns in the sales data.**Dimensionality Reduction:** Use techniques like principal component analysis (PCA) to reduce the dimensionality of your dataset while preserving important information.**Handling Categorical Variables:Dummy Variables:** Create dummy variables for categorical variables, especially if you plan to use them in regression or other modeling techniques.**Ordinal Encoding:** Encode ordinal categorical variables in a way that preserves their order and meaningfulness.**Data Imputation:**If missing data is prevalent and you decide not to remove records with missing values, you may use data imputation techniques to estimate missing values, such as mean imputation, median imputation, or machine learning-based imputation methods.**Data Standardization:**Standardize data if necessary, particularly for variables with different units or measurement scales.**Text Data Processing:**If the survey dataset contains text responses, text data preprocessing may involve tasks such as text cleaning, tokenization, stemming, and sentiment analysis.

**Documentation:**Maintain clear documentation of all preprocessing steps and the reasons for each transformation. This documentation is essential for transparency and reproducibility.**Privacy and Security:**Ensure that personally identifiable information (PII) is handled appropriately and that data privacy and security guidelines are followed**.**

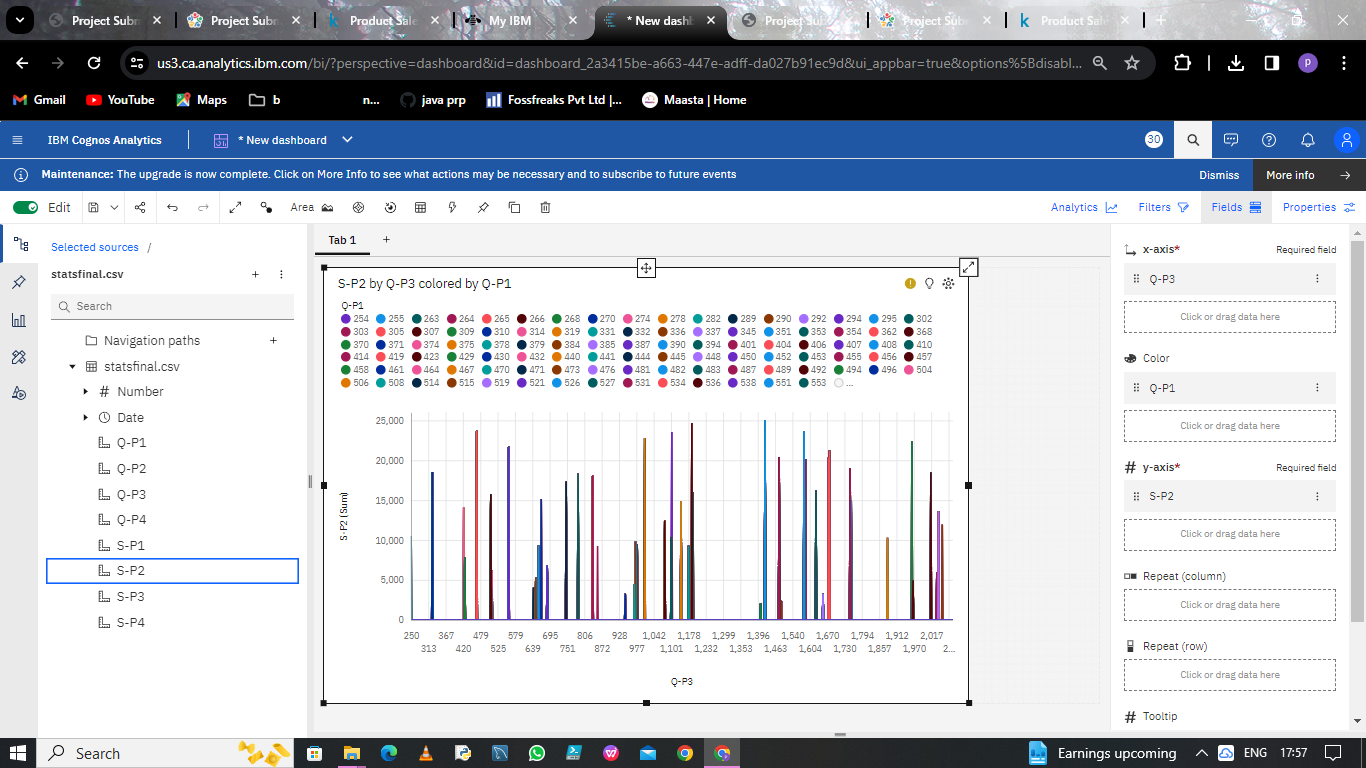
****

**Fig4.1**

**Fig 4.2**

****

**Fig 4.3**

****

**Fig 4.4**

**CHAPTER 5**

**PROPOSED ALGORITHM**

**REGULARIZATION**

Regularization is a technique used in machine learning, particularly in regression and other predictive modeling tasks, to prevent overfitting and improve the generalization of models. When applied to a product sales analysis dataset, regularization can be useful to ensure that the models developed for analysis are robust and not overly influenced by noise or irrelevant features. Below, I'll describe how regularization can be applied to a product sales analysis dataset:

**Linear Separation:Data Preprocessing:**Before applying regularization, you should prepare your dataset through data cleaning, feature engineering, and data splitting (into training and testing sets).

**Select a Model:** Choose a suitable model for your product sales analysis. Common choices include linear regression, logistic regression, decision trees, random forests, or neural networks, depending on the nature of your analysis and the specific questions you want to answer.

**Regularization Techniques:**

**a. L1 Regularization (Lasso):** Lasso regularization adds a penalty term to the loss function, which encourages the model to set some feature coefficients to zero. This effectively performs feature selection, eliminating less important features.In a product sales analysis, Lasso can help identify which product attributes or factors have the most significant impact on sales.

**b. L2 Regularization (Ridge):** Ridge regularization adds a penalty term to the loss function that discourages feature coefficients from becoming excessively large. It helps prevent model overfitting and promotes more balanced weights across features. In product sales analysis, Ridge can help in scenarios where multiple features are correlated, or when it's important to avoid large coefficients that might lead to unstable predictions.

**c. Elastic Net Regularization:** Elastic Net combines both L1 and L2 regularization. It allows you to balance feature selection (as in Lasso) and coefficient stability (as in Ridge).This is particularly useful when dealing with a dataset where there might be many features with varying importance in the sales analysis.

**Hyperparameter Tuning:** Regularization techniques have hyperparameters (e.g., alpha in Lasso and Ridge) that control the strength of regularization. These hyperparameters need to be tuned to find the optimal balance between bias and variance.

**Model Training:** Train your chosen model with the regularization technique applied. The regularization penalty will influence the model's coefficient values during the training process.

**Evaluation:** Evaluate the model's performance using appropriate metrics (e.g., Mean Squared Error for regression tasks, accuracy for classification tasks). Regularization should help prevent overfitting, leading to better generalization to new data.

**Interpretation:**Regularized models may provide more interpretable results, as they emphasize the most important features and reduce the impact of noise. In a product sales analysis, this can help identify key factors influencing sales performance.

**Feature Selection:** Post-regularization, you can inspect the coefficients of the features to understand which features have been retained as important predictors and which have been penalized or eliminated.

**Fine-tuning and Iteration:** Regularization strength and model choice may require fine-tuning through experimentation and iteration. You can adjust hyperparameters and explore different models as needed to achieve the best results.

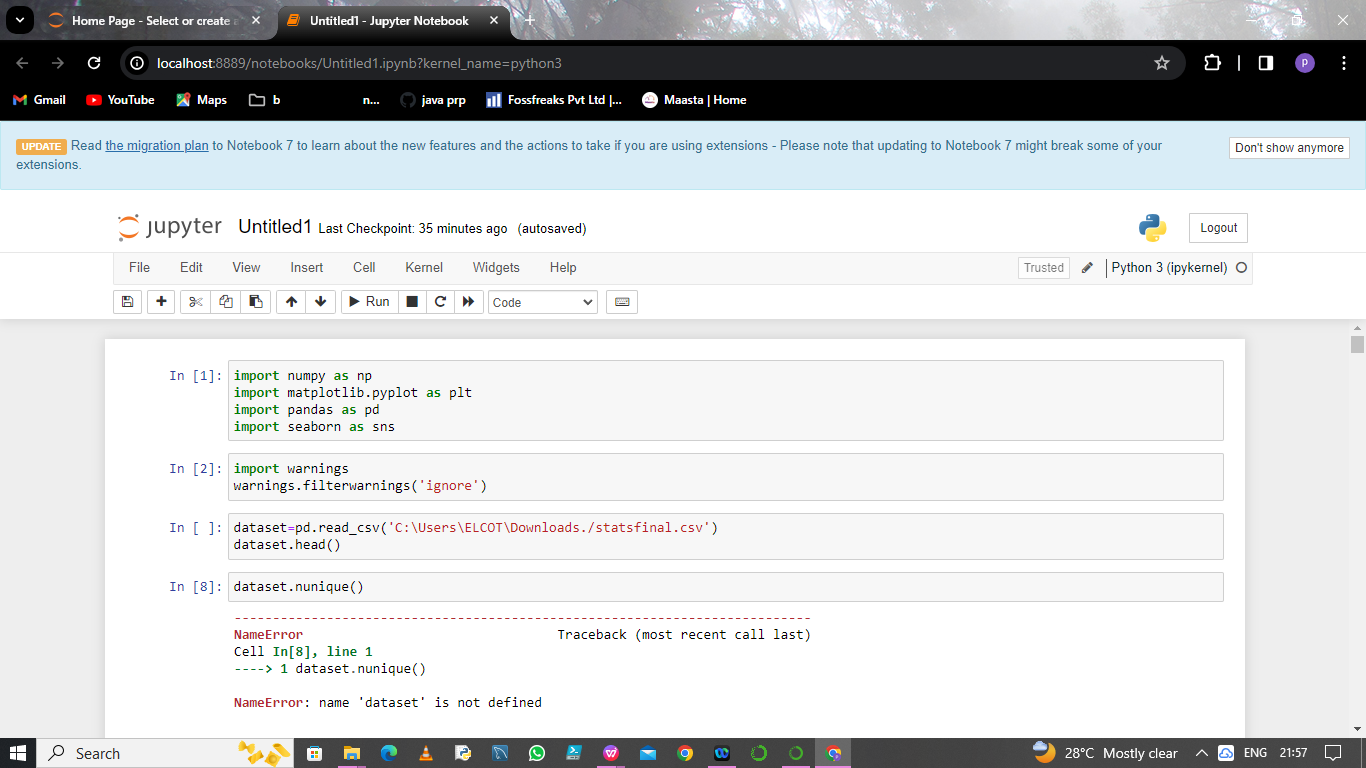


Fig 6.1

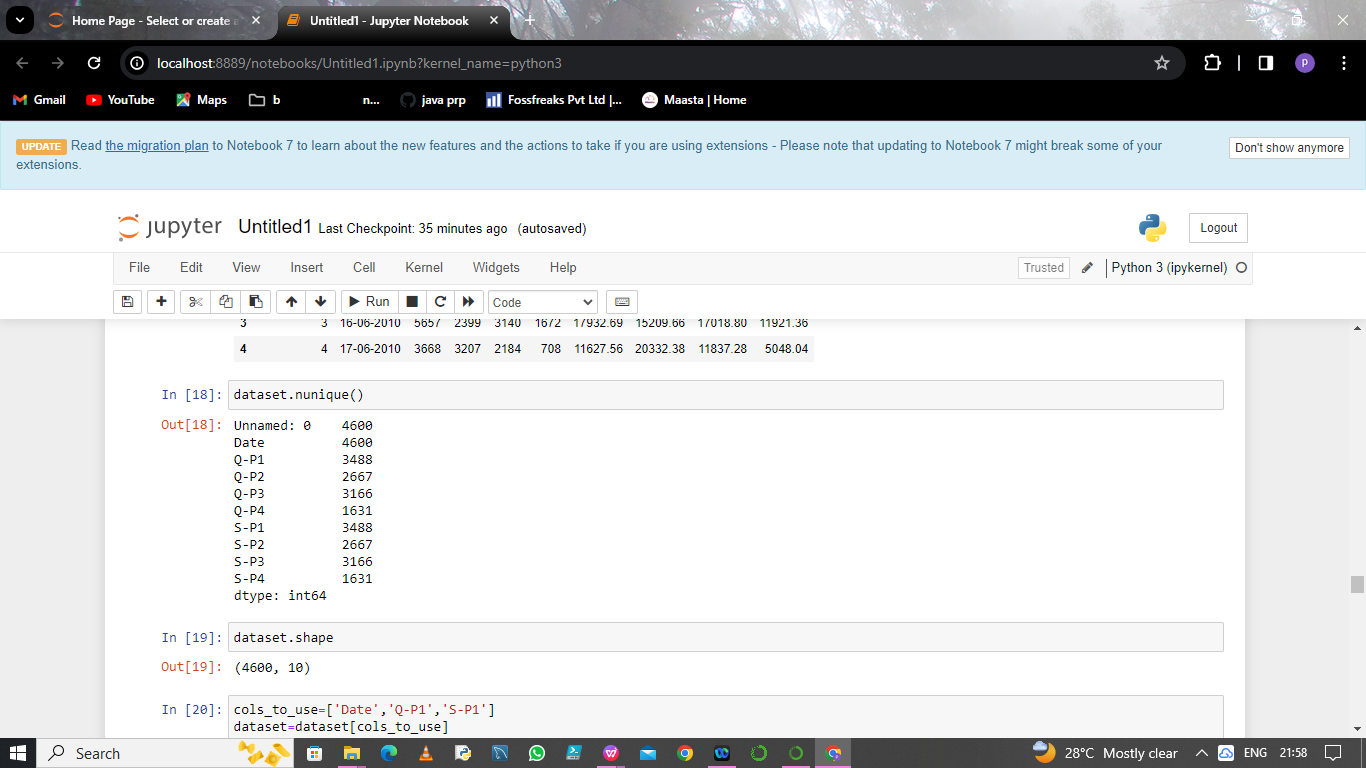


Fig 6.2

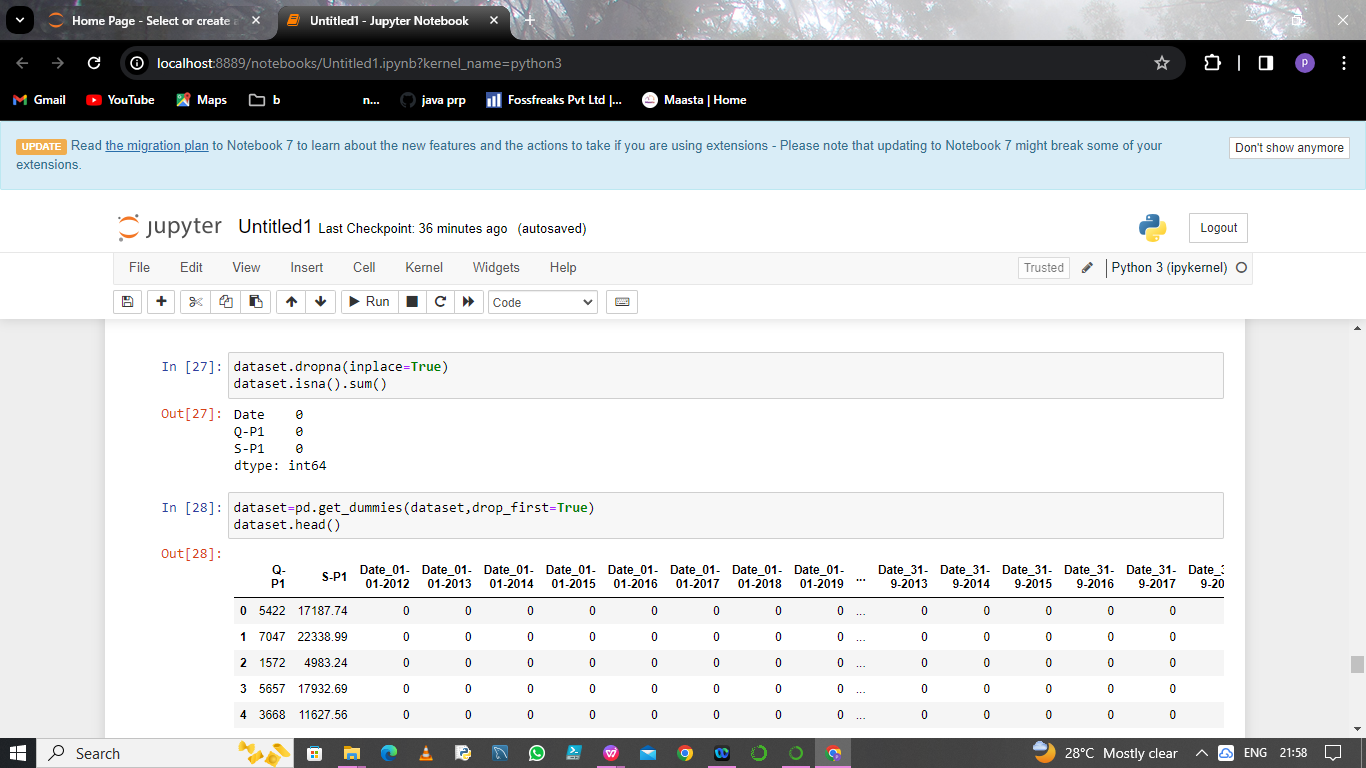


Fig 6.3

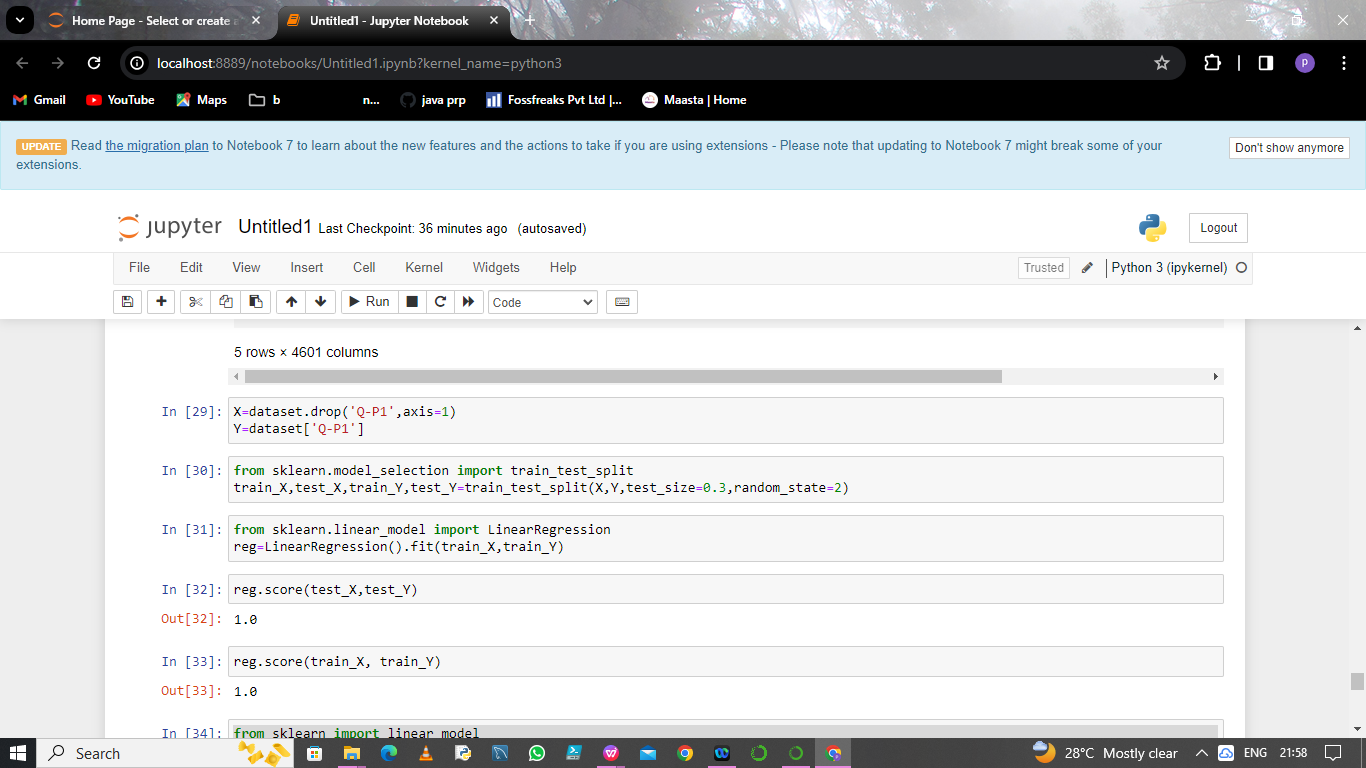


Fig 6.4

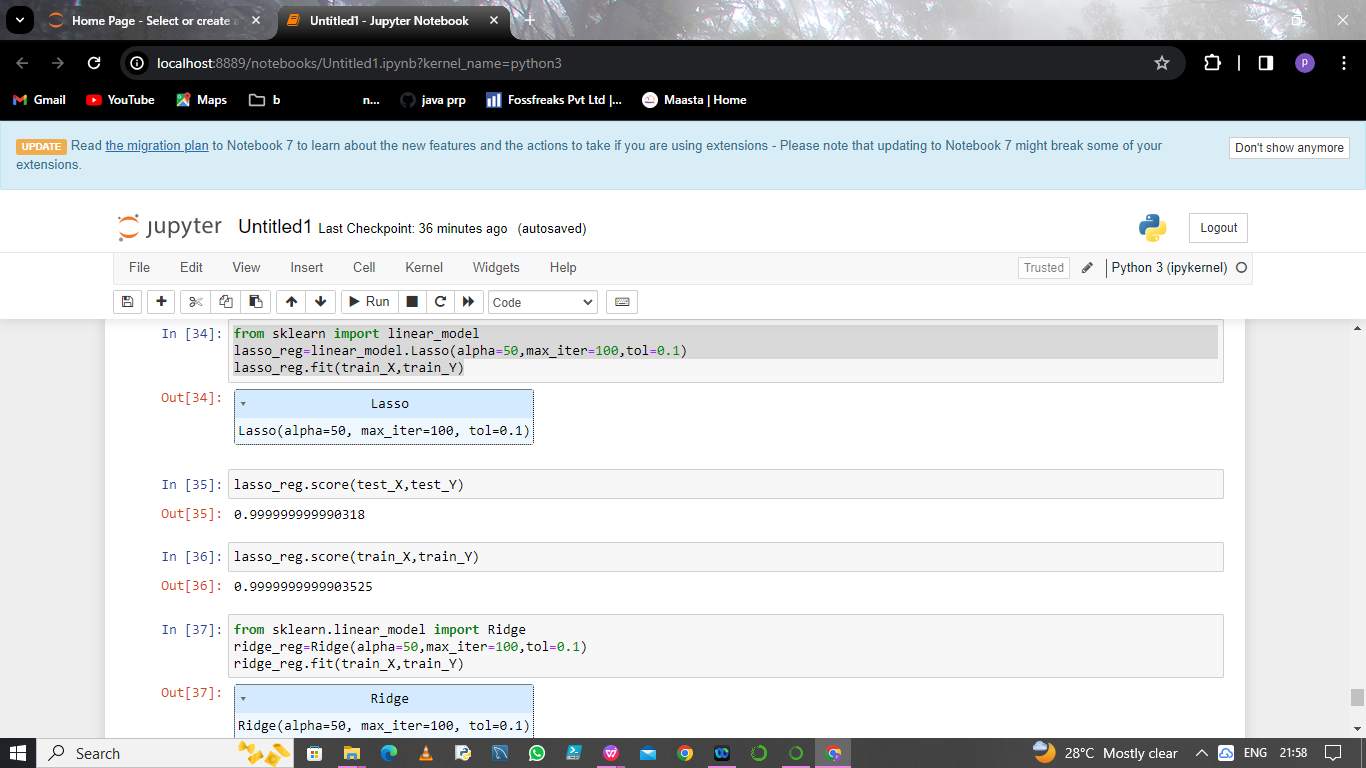


Fig 6.5

**CHAPTER 7**

**PROPOSED INNOVATION TECHNIQUE**

While developing machine learning models you must have encountered a situation in which the training accuracy of the model is low but the validation accuracy or the testing accuracy is too low. This is the case which is popularly known as overfitting in the domain of machine learning also this is the last thing a machine learning practitioner would like to have in his model. In this article, we will learn about a method known as regularization which helps us to solve the problem of overfitting. But before that let’s understand what is underfitting and overfitting.While developing machine learning models you must have encountered a situation in which the training accuracy of the model is low but the validation accuracy or the testing accuracy is too low. This is the case which is popularly known as overfitting in the domain of machine learning also this is the last thing a machine learning practitioner would like to have in his model. In this article, we will learn about a method known as regularization which helps us to solve the problem of overfitting. But before that let’s understand what is underfitting and overfitting.



Fig 7.1

## What are Overfitting and Underfitting?

* **Overfitting** is a phenomenon that occurs when a machine learning model is constrained to the training set and not able to perform well on unseen data. That is when our model learns the noise in the training data as well. This is the case when our model memorizes the training data instead of learning the patterns in it.

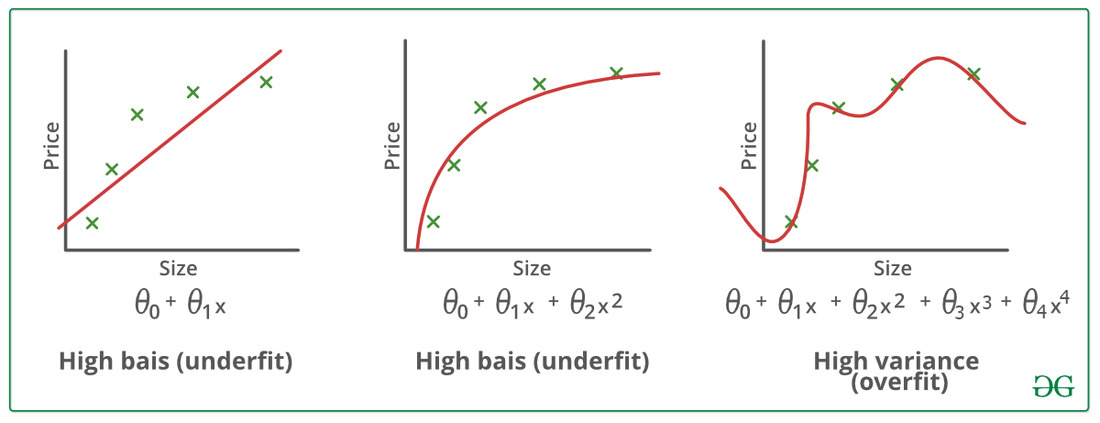
## **Underfitting** on the other hand is the case when our model is not able to learn even the basic patterns available in the dataset. In the case of the underfitting model is unable to perform well even on the training data hence we cannot expect it to perform well on the validation data. This is the case when we are supposed to increase the complexity of the model or add more features to the feature set.

## What are Bias and Variance?

**Bias** refers to the errors which occur when we try to fit a statistical model on real-world data which does not fit perfectly well on some mathematical model. If we use a way too simplistic a model to fit the data then we are more probably face the situation of **High Bias** which refers to the case when the model is unable to learn the patterns in the data at hand and hence performs poorly.

**Variance** implies the error value that occurs when we try to make predictions by using data that is not previously seen by the model. There is a situation known as **high variance** that occurs when the model learns noise that is present in the data.

Fig 7.2

we cho

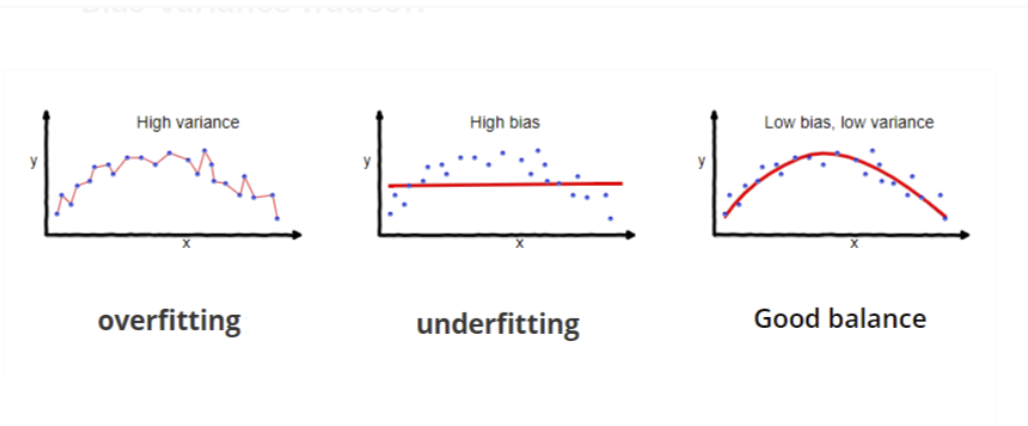
Finding a proper balance between the two that is also known as the

Bias-Variance Tradeoff can help us prune the model from getting overfitted to the training data.

## Regularization in Machine Learning

Regularization is a technique used to reduce errors by fitting the function appropriately on the given training set and avoiding overfitting. The commonly used regularization techniques are :

1. Lasso Regularization – L1 Regularization
2. Ridge Regularization – L2 Regularization
3. Elastic Net Regularization – L1 and L2 Regularization.



**CHAPTER 8**

**CONCLUSION AND FUTURE SCOPE**

**CONCLUSION**

A product sales analysis dataset and suggesting potential future scopes. Keep in mind that the conclusions and future scope will vary based on the specific data, industry, and business goals. Here's a step-by-step guide:**Objectives:Data Overview:** Start with a brief overview of the dateset, including the time frame, product categories, and any other relevant information.

**Descriptive Statistics:** Calculate basic statistics such as mean, median, standard deviation, and percentiles for sales, prices, and quantities sold.

**Trend Analysis:** Identify sales trends over time to determine if there are any seasonality patterns, long-term growth, or fluctuations.

**Product Performance:** Analyze the performance of individual products, including best-sellers, low-performing products, and product life cycles.

**Customer Segmentation:** Segment customers based on their purchasing behavior, such as frequency, recency, and monetary value.

**Geographic Analysis:** Examine sales trends by geographic region to identify potential opportunities or challenges in specific markets.

**Price Elasticity:**  Analyze how changes in pricing affect sales volume and revenue. This can help optimize pricing strategies.

**Inventory Management:** Evaluate inventory turnover rates, identifying slow-moving and fast-moving items. This can assist in better inventory management.

**Market Basket Analysis:** Discover associations between products often purchased together to optimize product placement and cross-selling strategies.

**Customer Churn Analysis:** Identify customers who have stopped making purchases and develop strategies to retain them.

**Marketing Channel Analysis:** Assess the effectiveness of different marketing channels (online, offline, social media, etc.) in driving sales.

"**FUTURE SCOPE”:**

Based on the conclusions, you can suggest potential areas for further exploration and action:

**Optimizing Product Portfolio:** If certain products are underperforming, consider discontinuing them or improving their marketing.

**Customer Retention:**Develop strategies to retain existing customers and increase their lifetime value.

**Market Expansion:** Identify untapped markets or segments for potential growth opportunities.

**Pricing Strategy:** Refine pricing strategies to maximize revenue and profitability.

**Inventory Optimization:** Implement better inventory management techniques to minimize carrying costs.

**Marketing Strategy:** Adjust marketing channels and campaigns based on their effectiveness.

**Data Enrichment:** Collect additional data, such as customer demographics or competitor information, to enhance analysis.