An Analysis on Customer Return Behavior and Predictive Modeling

Project Report submitted to the **SRM INSTITUTE OF SCIENCE AND TECHNOLOGY** in partial fulfilment of the requirements for the award of the Degree of

MASTER OF BUSINESS ADMINISTRATION

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

Submitted by

Mano Barathi. M

(Reg. No.RA2352008010065)

Under the guidance of

Dr. V. M. Shenbagaraman

Chair-Information System and Technology Management

FACULTY OF MANAGEMENT

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

KATTANKULATHUR- 603 203

APRIL 2025

An Analysis on Customer Return Behavior and Predictive Modeling

A Project Report submitted to the **SRM INSTITUTE OF SCIENCE AND TECHNOLOGY** in partial fulfilment of the requirements for the award of the Degree of

MASTER OF BUSINESS ADMINISTRATION

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

Submitted by

Mano Barathi.M

(Reg. No.RA2352008010065)

Under the guidance of

Dr. V. M. Shenbagaraman

Chair-Information System and Technology Management

FACULTY OF MANAGEMENT

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
KATTANKULATHUR- 603 203

APRIL 2025

OFFER LETTER

Date: April 19, 2025

Dear

Mano barathi M

We are pleased to offer you an internship opportunity at **Elevate Labs** as a **Data Analyst Intern**. This internship is designed to help you enhance your skills and gain practical experience in the field of data analysis. Below are the details of your internship:

Position: Data Analyst Intern
Duration: 1 Month
Start Date: April 21, 2025
Location: Remote

Benefits:
Completion Certificate and Letter of Recommendation (LOR) upon successful completion of the internship with full time offer based on performance.

During the internship, you will work closely with our team on projects that will enhance your understanding of data analytics, visualization, and reporting. We expect you to be proactive, committed, and eager to learn throughout this period. To confirm your acceptance of this offer, please email your confirmation to elevatelabshire@gmail.com by April 20, 2025. If you have any questions, feel free to reach out to us

We look forward to having you on board!

Best Regards
maheswgwd
CEO & FOUNDER
(ELEVATE LABS)

MSME
MICRO, SMALL & MEDIUM ENTERPRISES
OUR STRENGTH
Ministry of MSME, Govt. of India

elevatelabshire@gmail.com

BONAFIDE CERTIFICATE

This is to certify that the Project Report entitled **"**An Analysis on Customer Return Behavior and Predictive Modeling " in partial fulfillment of the requirements for the award of the Degree of **Master of Business Administration** in **Artificial Intelligence and Data Science** is a record of original training undergone by **Mano barathi. M (Reg. No.RA2352008010065)** during the year **2024-2025** of his study in the Faculty of Management, **SRM IST,** Kattankulathur under my supervision and the report has not formed the basis for the award of any Degree/Fellowship or other similar title to any candidate of any University.

Place:

Date:

Signature of Guide

Dr. V. M. Shenbagaraman

Chair – Information Systems And Technology Management

Faculty of Management SRM IST

Kattankulathur

INTERNAL EXAMINER

EXTERNAL EXAMINER

DEAN-FOM

DECLARATION

I, **Mano Barathi.M,** hereby declare that the Project Report, entitled **"**An Analysis on Customer Return Behavior and Predictive Modeling at Elevate Labs**"**, submitted to the **SRM IST** in partial fulfillment of the requirements for the award of the Degree of Master of Business Administration is a record of original training undergone by me during the period **19-04-2025 to 19-05-2025** under the supervision and guidance **Dr. V. M. Shenbagaraman**, SRM IST, Kattankulathur and it has not formed the basis for the award of any Degree/Fellowship or other similar title to any candidate of any University.

Place: **Signature of the Student**

Date:

ACKNOWLEDGEMENT

First and foremost, I offer my sincerest gratitude to our Chancellor, SRM University, for his academic support and the facilities provided to carry out the project work at the Institute. His wide vision and concern for students have been inspirational.

I express my heartfelt thanks to our Dean, Faculty of Management, SRM IST, Kattankulathur who provided all the facilities for carrying out this project.

I take this opportunity to express my profound gratitude and deep regards to my guide Dr.V.M.Shenbagaraman, for the exemplary guidance, monitoring, and constant encouragement throughout this project.

I also take this opportunity to express my sincere gratitude to Elevate Labs for providing me with a valuable internship opportunity. The guidance, resources, and support extended by the organization were instrumental in enhancing my practical skills and successfully completing this project. I owe my wholehearted thanks and appreciation to the entire staff of the company for their cooperation and assistance during my project.

I thank God Almighty for showering his perennial blessing on me and for giving me the courage to pursue this project work successfully. I owe a lot to my parents, who encouraged and helped me at every stage of my personal and academic life, and longed to see this achievement come true.

M.Mano Barathi

(Reg. No.RA2352008010065)

CONTENTS

Chapter – I

INTRODUCTION

1.1 INTRODUCTION:

In the modern digital age, the role of data has transformed dramatically. From being a passive byproduct of business operations, data has evolved into a powerful asset that drives critical decisions across all industries. With companies generating massive volumes of data daily, the ability to effectively analyze and interpret this information has become central to success. The process of data analysis involves extracting meaningful patterns, identifying relationships, and translating raw data into actionable insights. As a data analyst intern, my experience lies at the intersection of these dynamic capabilities, aiming to contribute toward smarter and evidence-based decision-making.

The Role of Data Analysis in Modern Business

Data analysis plays a crucial role in helping businesses understand their operations, customers, and market environment. It involves the use of statistical tools, algorithms, and visualization techniques to uncover hidden patterns and trends. Analysts transform raw, unstructured data into understandable formats, supporting decision-makers with reliable insights. Whether it's forecasting demand, improving customer satisfaction, reducing operational costs, or enhancing product performance, data analysis empowers organizations to move beyond intuition and base their strategies on solid evidence.

During my internship, I came to understand that a data analyst serves not only as a technical contributor but also as a bridge between business objectives and technological solutions. A strong analyst must be able to communicate findings clearly and influence strategic planning, contributing to both short-term goals and long-term visions of a company.

Importance of Data-Driven Decisions

One of the most significant advantages of data analysis is its ability to support data-driven decision-making. Rather than relying on assumptions, managers and leaders can refer to actual performance metrics and predictive models. This reduces uncertainty and promotes a culture of accountability and precision within the organization.

For example, in a sales-focused business, data analysis helps identify which products are performing well, what times of the year experience peak sales, or which regions have high conversion rates. This allows management to allocate resources more effectively, plan marketing campaigns better, and optimize inventory management. Similarly, in customer service, analyzing feedback and behavior patterns can highlight areas where customer experience can be improved.

From my experience during the internship, I noticed that companies that leverage data tend to respond more quickly to market changes, assess risks more accurately, and seize opportunities more efficiently.

Applications Across Industries

The applications of data analysis span virtually every industry. In healthcare, patient records and clinical trials are analyzed to improve diagnostics and treatment plans. In finance, transaction data is used to detect fraud and guide investment decisions. E-commerce platforms rely on user data to personalize product recommendations and improve the user experience. Manufacturing sectors analyze machinery data to predict maintenance needs and minimize downtime.

My internship was based in a business environment where sales, customer, and financial data were the core focus. I worked on datasets involving customer payments, revenue tracking, and monthly performance dashboards. This exposed me to the real-world impact of data analysis in helping companies evaluate their financial health and take corrective action where necessary.

Moreover, with the rise of AI and automation, data analysis is no longer limited to descriptive statistics. It now includes predictive and prescriptive analytics, allowing companies not just to understand past and present performance but also to forecast future outcomes and recommend optimal strategies.

The Power of Business Intelligence Tools

To manage, analyze, and visualize large volumes of data efficiently, organizations increasingly rely on programming languages like Python, which offer powerful libraries and tools tailored for data science. During my internship, I extensively used Python to handle the end-to-end data analysis process, from data cleaning and transformation to modeling and visualization.

Python's rich ecosystem—including libraries such as **Pandas**, **NumPy**, **Matplotlib**, **Seaborn**, and **Scikit-learn**—enabled me to perform robust data manipulation, generate meaningful statistical insights, and build predictive models like logistic regression. These tools proved essential for analyzing return patterns, creating risk profiles, and identifying high-return product categories.

What stood out most to me was Python’s flexibility and efficiency in handling complex datasets and producing visualizations that clearly communicated insights. The ability to automate data workflows and customize analytical functions made Python an invaluable tool throughout the project. It empowered me to convert raw data into actionable business intelligence, which informed strategic recommendations for reducing product return rates.

1.2 OBJECTIVE OF THE STUDY:

1. Purpose of the Project

The main purpose of this data analysis project is to apply analytical techniques to a real-world dataset in order to derive actionable insights that support business decision-making. With the growing importance of data in every industry, this project serves as a platform to understand how structured and unstructured data can be transformed into meaningful conclusions.

2. Importance of Data Analysis

Data analysis is essential for identifying patterns, spotting anomalies, understanding relationships, and making evidence-based decisions. In this project, the objective is to go beyond traditional reporting and uncover trends that can help organizations improve performance, reduce costs, understand customer behavior, or predict future outcomes. It bridges the gap between data and strategic planning.

3. Learning and Skill Development

A key objective of this project is to enhance technical and analytical skills using tools such as Python, SQL, Excel, Tableau, or Power BI. The project also aims to develop soft skills like critical thinking, business understanding, problem-solving, and effective communication of results. It offers hands-on experience in the data lifecycle — from collection and cleaning to modeling and visualization.

4. Real-World Problem Solving

Through this project, the goal is to approach a real-world problem — whether it's in sales, marketing, human resources, finance, or operations — and analyze relevant datasets to answer business questions. The focus will be on identifying key performance indicators (KPIs), measuring efficiency, understanding customer behavior, or forecasting trends based on the chosen project theme.

5. Effective Communication of Insights

One of the final objectives is to present the findings in a way that is clear, visual, and easily understandable by both technical and non-technical audiences. Dashboards, visualizations, reports, and summaries will be used to communicate complex data insights effectively. The ability to explain results and suggest actionable steps based on the data is a key learning outcome of this project.

6. Outcome and Value Addition

The end goal is to deliver a complete, structured analysis that provides valuable insights to support smarter decisions. The expected outcome includes a cleaned dataset, a set of visual dashboards or models, and a summary of insights with recommendations. This project aims to replicate the role of a data analyst in a professional setting, preparing me for real-world analytics tasks and interviews.

1.3 IMPORTANCE OF THE STUDY

In today's competitive business environment, ensuring high product quality and operational efficiency is crucial for sustaining growth and customer satisfaction. The significance of this study lies in its ability to integrate data analytics and visualization to enhance quality control processes and streamline operations. The key points of importance are discussed below:

1.Improves Product Quality:

The core objective of this study is to identify the primary contributors to product defects and determine their root causes. By using structured data collection and analysis techniques, this study helps detect specific areas where quality lapses occur. Understanding these recurring defects enables the organization to take preventive and corrective actions. As a result, product quality improves significantly, leading to higher customer satisfaction, fewer returns, and improved brand credibility.

2. Data-Driven Decisions:

Rather than relying on assumptions or traditional trial-and-error methods, this study promotes a data-driven approach. By utilizing advanced data visualization tools such as Tableau or Python's visualization libraries, complex data is transformed into clear and actionable visual insights. These visuals facilitate faster and more accurate decision-making by highlighting patterns, outliers, and correlations that are not easily visible through tabular data alone. Decision-makers can now rely on evidence-based insights for quality improvement measures, resulting in more reliable and sustainable outcomes.

3.Boosts Operational Efficiency:

By analyzing the Management Information System (MIS) data, this study provides a clearer picture of daily operational performance. Anomalies such as spikes in defect rates or delays

in processing can be quickly identified and addressed. This proactive monitoring helps reduce downtime, improve resource utilization, and streamline workflow processes. Consequently, the company achieves greater productivity and cost efficiency.

4. Refines Quality Control:

Quality control processes benefit significantly from the insights generated through this study. The ability to pinpoint specific defects, their frequency, and their sources allows for refining inspection criteria, adjusting manufacturing parameters, and retraining staff. Additionally, the continuous feedback loop created by this analysis fosters a culture of ongoing quality improvement. By staying ahead of potential issues and maintaining high standards, the organization strengthens its competitive position in the market.

In summary, this study plays a vital role in enhancing the overall quality and performance of operations through systematic defect analysis, data visualization, and actionable recommendations. Its contribution extends beyond immediate problem-solving, fostering long-term improvements in product reliability and customer trust.

1.4 SCOPE OF THE STUDY

The scope of this study encompasses a focused analysis of defects and operational inefficiencies within a particular production or process module. It integrates various analytical techniques and tools to deliver comprehensive insights and recommendations. The major areas included in the scope are as follows:

1. **Defect Analysis:**

A primary area of focus is the systematic identification and examination of defects occurring in a selected module or product line. This involves collecting data on defect types, frequency, severity, and location. By categorizing defects and analyzing trends, the study identifies the most critical issues affecting product quality. Root cause analysis techniques such as the 5 Whys or Fishbone Diagram may be used to trace issues back to their origins—whether they stem from machine errors, human oversight, or material issues.

2. **Advanced Data Visualization:**

   The study utilizes advanced data visualization tools to convert complex datasets into intuitive and meaningful visual formats such as dashboards, bar charts, heatmaps, and trend lines. These tools help stakeholders quickly interpret data, spot emerging patterns, and make faster decisions. Visualization is not only used for reporting defects but also for exploring the relationship between variables such as time of operation, shift schedules, or supplier batches and the occurrence of defects.

3. **MIS Data Review:**

   The Management Information System (MIS) provides rich, structured data on daily operations, including production volumes, machine performance, defect counts, and resource usage. The study includes an in-depth review of this data to uncover inefficiencies and irregularities. By correlating operational metrics with quality issues, the study aims to find hidden connections that might otherwise be overlooked. This ensures a holistic view of the operational landscape and supports long-term improvement strategies.

4. **Optimization Recommendations:**

   Based on the findings from defect analysis and MIS review, the study delivers practical, data-backed recommendations to improve quality and operational performance. These may include changes in process workflows, machine calibration schedules, training programs for staff, or improvements in raw material handling. Each recommendation is

supported by data evidence and is aimed at reducing defect rates, improving process stability, and increasing customer satisfaction.

In conclusion, the scope of this study is both analytical and solution-oriented. It not only examines current quality control challenges but also provides a framework for continuous improvement. By integrating defect analysis with MIS data insights and advanced visualization, the study offers a powerful approach to driving excellence in quality assurance and operational efficiency.

1.5 PERIOD OF THE STUDY:

- For the first 15 days of my internship, I worked on tasks assigned by Elevate labs. These tasks are designed to help brush up on basics and reinforce the foundational skills in data analysis

- Tasks will be submitted daily at the end of each day (EOD). Please use the provided Excel sheet to submit your work. This will be an essential phase to get familiar with the tools, techniques, and processes that you'll be using throughout your internship.

- In the following weeks, you will be working on 1-2 major projects that will help you apply the skills you've learned so far.

These projects are designed to be industry-relevant and will add significant value to your resume.

• Focus on quality work and apply your knowledge to real-world problems.

• These projects will give you hands-on experience and help build a strong portfolio. This phase is crucial for your professional growth and will showcase your abilities to potential employers.

1.6 CHAPTERIZATION:

CHAPTER I – Introduction

1.1 Introduction

1.2 Objective of the Study

1.3 Importance of the Study

1.4 Scope of the Study

1.5 Period of the Study

1.6 Chapterization

CHAPTER II – Industry and Company Profile

2.1 Industry Profile

2.2 Company Profile

CHAPTER III – Research Methodology

3.1 Introduction

3.2 Research Design

3.3 Data Collection

3.4 Data Cleaning and Preprocessing

3.5 Tools and Techniques Used

3.6 Model Building Approach

3.7 Evaluation Metrics

3.8 Limitations of the Study

CHAPTER IV – Data Analysis and Model Evaluation

4.1 Project Overview and Problem Statement

4.2 Dataset Description

4.3 Data Preparation and Feature Engineering

4.4 Exploratory Data Analysis

4.5 Return Reason and Geography Analysis

4.6 Logistic Regression and Model Evaluation

4.7 Model Output Interpretation

Chapter – II

INDUSTRY AND COMPANY PROFILE

2.1 COMPANY PROFILE:

We are an MSME registered company dedicated to empowering students by providing valuable internship opportunities. Our mission is to bridge the gap between college life and industry, ensuring that students are equipped with practical skills and realworld experience to succeed in their future careers. At our company, we believe that the transition from academia to the professional world should be seamless. We offer internships across various industries and disciplines, giving students the chance to apply their academic knowledge in a dynamic, real-world environment. Our internships are designed not only to enhance students' resumes but also to foster personal and professional growth, preparing them for the challenges and opportunities that lie ahead in their careers.

1. Industry Overview: Educational Technology (EdTech)

The Educational Technology (EdTech) industry has witnessed significant growth over the past decade, driven by the increasing demand for personalized learning solutions and the integration of technology into educational practices. EdTech encompasses a broad range of applications, including online learning platforms, educational software, and cognitive training tools.

Key trends in the EdTech industry include:

- **Personalized Learning**: Leveraging data analytics and AI to tailor educational content to individual learners' needs.
- **Gamification**: Incorporating game elements into learning to enhance engagement and motivation.
- **Mobile Learning**: Providing access to educational resources through mobile devices, facilitating learning anytime and anywhere.
- **Data-Driven Insights**: Utilizing analytics to assess learner performance and inform instructional strategies.

The global EdTech market is projected to continue its upward trajectory, with increasing investments and innovations aimed at improving learning outcomes and accessibility.

2. Company Overview: Elevate Labs

Founded: 2010
Headquarters: San Francisco, California, USA
Website: www.elevateapp.com
Industry: Educational Software, Cognitive Training
Founders: Jesse Pickard, Karl Stenerud, Jeff Evans, Andy Mroczkowski
CEO: Jesse Pickard

Employees: Approximately 69 as of 2025

Funding: Raised over $13.6 million from investors including Sequoia Capital and Felicis Ventures.

Elevate Labs is a privately held company specializing in the development of cognitive training tools designed to enhance communication and analytical skills. Their flagship product, the **Elevate app**, offers personalized, game-based training programs that adapt over time based on user performance. The app focuses on improving various skills, including memory, comprehension, and processing speed.

In addition to Elevate, the company launched **Balance**, a personalized meditation app that has received accolades such as Google's Best App award. Balance aims to improve users' stress management, sleep quality, and overall mental well-being.

3. Mission and Vision

Elevate Labs is driven by a mission to help individuals lead healthier, more productive lives through the development of tools that enhance cognitive abilities. The company envisions a world where everyone has access to personalized training programs that foster continuous learning and self-improvement.

4. Company Culture and Work Environment

Elevate Labs operates as a fully distributed company, embracing remote work and fostering a culture of continuous learning and collaboration. The organization invests in cultivating a robust and healthy company culture, emphasizing employee well-being and professional growth. This commitment has earned Elevate Labs recognition as one of Built In's Best Places to Work in 2022.

5. Products and Services

- **Elevate App**: A cognitive training application that offers personalized, game-based exercises aimed at improving skills such as reading comprehension, writing, and math. The app adjusts its difficulty based on user performance, ensuring a tailored learning experience.
- **Balance App**: A meditation and mindfulness app that provides personalized audio sessions to help users manage stress, improve sleep, and enhance overall mental health.

Both applications are available on iOS and Android platforms and have garnered positive reviews for their user-friendly interfaces and effectiveness.

6. Market Position and Competitors

Elevate Labs operates in a competitive landscape, with notable competitors including Lumosity, BrainHQ, and Peak. Despite the competition, Elevate has distinguished itself through its emphasis on personalized learning experiences and its commitment to scientific research in cognitive training.

The company's focus on data-driven personalization and user engagement has contributed to its strong market position and user base growth.

7. Relevance to Data Analysis

Data analysis plays a pivotal role in Elevate Labs' operations. The company leverages data to:

- Monitor user engagement and performance.
- Personalize training programs based on individual progress.
- Assess the effectiveness of exercises and interventions.
- Inform product development and feature enhancements.

As a Data Analyst Intern, engaging with Elevate Labs provides an opportunity to work at the intersection of data science and educational technology, contributing to the development of tools that have a meaningful impact on users' cognitive development.

Chapter – III

RESEARCH METHODOLOGY

3.1 Introduction

Research methodology refers to the systematic process used to collect, analyze, and interpret data in order to answer specific research questions or test hypotheses. In this project, the focus is on understanding customer return behavior in an e-commerce context and predicting the likelihood of product returns using statistical modeling. This chapter outlines the research design, data sources, tools used, and techniques adopted for data analysis and model development.

3.2 Research Design

This study adopts a **quantitative research design**, primarily focusing on the statistical analysis of historical transaction data. The approach is **exploratory** and **predictive**, aiming to uncover patterns in return behavior and develop a logistic regression model to predict the probability of returns.

The research was conducted in several phases:

1. **Data Collection and Cleaning**: Importing raw Excel data into Python and preprocessing it.
2. **Exploratory Data Analysis (EDA)**: Identifying patterns, outliers, and correlations across key features such as product type, customer location, and overdue payments.
3. **Feature Engineering**: Converting categorical variables into numerical format using label encoding.
4. **Modeling**: Building and evaluating a logistic regression model to predict return likelihood.
5. **Insight Generation**: Interpreting the results and making business recommendations.

3.3 Data Collection

The dataset used in this study was obtained from **Elevate Labs**, as part of a hands-on internship experience. It includes **42 historical transaction records** from GREENS Circuits, a provider of

PCB manufacturing solutions. Despite its relatively small size, the dataset is rich in features relevant to return behavior.

Key columns in the dataset include:

- **Customer**: Identifier for the client
- **Balance**: Outstanding amount on the order
- **Days Overdue**: Days past the due date
- **Location**: City or region of the customer
- **Product List**: Type of product/service provided (e.g., Component Sourcing, Prototype PCB)
- **Return Status**: Indicates whether the product was returned
- **Return Reason**: Manually assigned reason for return (for post-analysis only)

3.4 Data Cleaning and Preprocessing

Raw data often contains missing values, inconsistent formats, and irrelevant entries. The following steps were taken to prepare the data for analysis:

- Removed currency symbols from the **Balance** column and converted values to float.
- Converted **Due Dates** into datetime objects and calculated **Days Overdue**.
- Standardized **Return Status** values to binary format (1 for 'Yes', 0 for 'No').
- Handled missing values by assigning default values or removing rows with excessive nulls.
- Added a new column for **Return Reason** using synthetic data, randomly assigned from predefined reasons for demonstration purposes.

3.5 Tools and Techniques Used

This project was conducted entirely in **Python**, a widely used programming language in data science and machine learning. The key libraries used include:

- **pandas**: Data manipulation and analysis
- **numpy**: Numerical computations
- **matplotlib & seaborn**: Data visualization
- **scikit-learn**: Machine learning modeling, preprocessing, and evaluation

Statistical Technique:

- **Logistic Regression**: A supervised machine learning algorithm used to estimate the probability of a binary outcome—in this case, whether a product will be returned or not.

The model was trained on preprocessed features, including encoded values for product type and location, as well as numerical values like balance and days overdue.

3.6 Model Building Approach

The modeling process involved the following steps:

1. **Label Encoding**: Categorical features such as product list and location were encoded into numerical values.
2. **Feature Scaling**: Input features were standardized using **StandardScaler** to improve model performance.
3. **Train/Test Split**: Data was divided into training (80%) and testing (20%) sets using train_test_split from scikit-learn.
4. **Logistic Regression**: The model was trained to classify return likelihood and evaluated using metrics such as **accuracy**, **precision**, **recall**, and **F1-score**.
5. **Probability Estimation**: The model also returned a probability score indicating how likely an order is to be returned.

3.7 Evaluation Metrics

The performance of the model was assessed using the following metrics:

- **Confusion Matrix**: Provides a summary of prediction results.
- **Accuracy**: Proportion of total correct predictions.
- **Precision**: Percentage of correctly predicted positive returns out of all positive predictions.
- **Recall**: Proportion of actual returns correctly identified.
- **F1 Score**: Harmonic mean of precision and recall, useful in imbalanced datasets.

3.8 Limitations of the Study

While the project achieved meaningful insights, several limitations must be acknowledged:

- **Small Dataset**: With only 42 records, the model may not generalize well to larger populations.
- **Synthetic Return Reasons**: Some return reasons were randomly assigned due to missing or incomplete records.
- **Simplified Features**: Advanced attributes like delivery time, order size, or customer reviews were not available.

Chapter – IV

DATA ANALYSIS AND MODEL EVALUATION

1. Project Overview

In the rapidly evolving landscape of e-commerce, product returns pose a significant challenge to both profitability and customer satisfaction. Frequent returns not only increase operational costs but also indicate potential flaws in product quality, delivery timelines, or customer expectations. This project aims to analyze customer return behavior by leveraging data analytics and predictive modeling techniques to identify the underlying causes of returns and assess return risks across various product categories and customer locations.

The analysis involved collecting and cleaning historical order data, conducting exploratory data analysis (EDA), and building a logistic regression model to predict the probability of product returns. The insights generated from this project serve as a foundation for developing strategic interventions to minimize returns, optimize logistics, and enhance customer experience.

Problem Statement

Product returns are a critical pain point in e-commerce, impacting inventory management, customer trust, and operational efficiency. Despite having transactional data, many companies lack a structured approach to identifying return patterns or predicting high-risk orders before fulfillment. The key challenges include:

- Identifying which **products**, **locations**, or **order attributes** are most associated with returns.
- Understanding **why customers return products**—whether due to quality, delivery delays, or incorrect shipments.
- Developing a model to **predict the likelihood of return** for new or existing orders to enable preventive actions.

This project addresses these challenges by applying data analytics and logistic regression modeling to uncover return behavior trends and predict return probabilities, thereby equipping the business with actionable insights to reduce return rates and improve customer satisfaction.

In the highly competitive and fast-paced electronics manufacturing industry, minimizing return rates is crucial for sustaining profitability, optimizing operational workflows, and maintaining high levels of customer satisfaction. At GREENS Circuits, a premier provider of PCB manufacturing solutions, frequent returns not only affect the bottom line but also signal underlying issues in product quality, customer communication, or logistics.

This project seeks to:

- Analyze return behavior based on customer segments, locations, and product categories.
- Develop a predictive model that estimates the probability of return for each order using transactional data.
- Provide actionable insights to assist the business in identifying high-risk transactions and implementing preventive strategies.

GREENS Circuits provides services including:

- Multi-layer, double-layer, and single-layer PCBs
- SMT and through-hole assembly
- Prototype testing
- Final box builds
- Component sourcing

2. Dataset Description

The dataset includes 42 historical customer order records. Though small in size, it captures key transactional and categorical details required for preliminary trend analysis and model building.

Key Columns:

- **Customer:** Name or identifier of the client placing the order.
- **Balance:** The amount due on the order invoice. Initially in currency format with symbols and commas.
- **Due Date:** The scheduled date for payment completion.
- **Days Overdue:** Number of days payment is past due. Missing values were interpreted as zero (i.e., not overdue).
- **Location:** Geographic location of the customer (e.g., Chennai, Hyderabad, Pune).
- **Product List:** The type of PCB product or service, including component sourcing, SMT, or final box build.
- **Return Status:** Indicates whether the order was returned (Yes/No).
- **Return Reason:** Explains the cause of the return (used only for post-analysis, not prediction).

3. Data Cleaning and Preparation

The dataset was cleaned and standardized using Python to prepare it for modeling and visualization.

Key Cleaning Steps:

- **Balance:** Removed special characters (e.g., '₹', ',') and converted to numeric float type.
- **Due Date:** Parsed into datetime format. Unreadable dates were set as NaT (Not a Time).
- **Days Overdue:** Missing entries were replaced with 0, assuming no delay in payment.
- **Return Status:** Standardized to uniform values ('Yes' and 'No') and converted into a binary target variable return_flag (1 for Yes, 0 for No).
- **Product List and Location:** Encoded using one-hot encoding to make them usable by machine learning models.

Additional Derived Features:

- **Return Probability:** After model prediction, a probability score was generated to indicate the likelihood of an order being returned.

4. Exploratory Data Analysis (EDA)

4.1 Return Rate by Product Category:

- **Highest Returns (50% or more):**
  - Component Sourcing
  - Prototype PCB
  - Through Hole Assembly
- **Lowest Returns (0%):**
  - Multi-layer PCB

These results suggest the need for quality checks and process reviews in Component Sourcing and Through Hole Assembly.

4.2 Return Rate by Geography:

- **Highest Return Rates:**
  - Hyderabad: 66.7%
  - Pune: 57.1%
- **Lowest Return Rates:**
  - Ahmedabad: 0%
  - Mumbai: 0%

This implies a need for localized analysis of logistics, customer support, or product handling in Hyderabad and Pune.

5. Predictive Modeling Using Logistic Regression

5.1 Logistic Regression

Logistic Regression is a supervised classification algorithm used to estimate the probability that a data point belongs to a certain class (e.g., return or no return). Unlike linear regression, which outputs continuous values, logistic regression uses a sigmoid function to constrain output between 0 and 1.

Formula: $P(Y=1) = 1 / (1 + e^{-(b0 + b1X1 + b2X2 + ... + bn*Xn)})$

5.2 Importance:

- **Interpretability:** Model coefficients clearly indicate the impact of each feature.
- **Efficiency:** Suitable for small datasets and fast to train.
- **Probability Output:** Helps prioritize which orders are more likely to be returned.
- **Simplicity:** Easy to implement and explain to stakeholders.

5.3 Model Features and Target:

- **Features Used:**
  - One-Hot Encoded Product Category
  - One-Hot Encoded Location
  - Balance (numeric)
  - Days Overdue (numeric)
- **Target Variable:** return_flag (1 for returned, 0 for not returned)

5.4 Model Building Process:

```
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report, confusion_matrix

# Reload processed dataset
df = pd.read_excel("ecommerce_return_data_with_reasons.xlsx")

# Encode categorical features
label_encoder = LabelEncoder()
df['Product List'] = label_encoder.fit_transform(df['Product List'])
df['Location'] = label_encoder.fit_transform(df['Location'])
df['Return Status'] = df['Return Status'].map({'Yes': 1, 'No': 0})

# Define features and target
X = df[['Product List', 'Location', 'Balance', 'Days Overdue']]
y = df['Return Status']

# Normalize features (optional but recommended)
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Train/test split
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.2, random_state=42)
```

```
# Logistic Regression model
model = LogisticRegression()
model.fit(X_train, y_train)

# Evaluation
y_pred = model.predict(X_test)
print(confusion_matrix(y_test, y_pred))
print(classification_report(y_test, y_pred))
```

5.5 Evaluation Metrics:

- **Accuracy:** Proportion of correctly predicted outcomes.
- **Precision:** Accuracy of predicted positive returns.
- **Recall:** Proportion of actual returns detected.
- **F1 Score:** Harmonic mean of precision and recall — important for imbalanced datasets.

```
Classification Report:
              precision    recall  f1-score   support

           0       0.83      1.00      0.91         5
           1       1.00      0.75      0.86         4

    accuracy                           0.89         9
   macro avg       0.92      0.88      0.88         9
weighted avg       0.91      0.89      0.89         9

Confusion Matrix:
 [[5 0]
 [1 3]]
```

Interpretation of the Model Output

This table represents the **predicted likelihood** of return for specific product orders based on various features. Here's what each column in your screenshot means and how to interpret it:

Column Name	Interpretation
Product Name	The name/type of the PCB product for that particular order (e.g., Component Sourcing, Prototype PCB).
Balance	The monetary balance associated with the order. Larger balances may indicate bigger or more critical orders.
Days Over	Number of days the payment is overdue — often correlated with customer satisfaction or operational delays.
Location	A numeric encoding of the customer's location.
Product List	An encoded representation of the product category used by the model.
Return Status	The actual return status (1 = Returned, 0 = Not Returned). This is used for validation and comparison.
Return Reason	Encoded return reason (e.g., 0 = Low quality, 1 = Late delivery, 2 = Missing parts, 3 = Wrong product, etc.).
Return Probability	**The predicted probability from the model that the product will be returned.** Values close to 1 mean high return risk.

Product Name	Balance	Days Over	Location	Product List	Return Status	Return Reason	Return Probability
Component Sourcing	5900	674	2	5	1	0	0.991496544
Prototype PCB	11623	850	7	6	1	1	0.964042451
SMT Assembly	24190	501	7	0	1	1	0.964995782
Component Sourcing	49737	681	5	5	1	0	0.989716854
Single Layer PCB	28320	892	7	2	1	2	0.908757711
Prototype PCB	36110.96	573	1	6	1	1	0.963223678
Through Hole Assembly	9440	809	1	3	1	0	0.994061868
Double Layer PCB	5310	847	4	1	1	3	0.764231481
Final Box Build	6785	669	1	4	1	3	0.711126092

Row-by-Row Interpretation

Row 1

- **Product**: Component Sourcing
- **Balance**: ₹5900
- **Days Overdue**: 674
- **Actual Return**: Yes (Return Status = 1)
- **Reason**: Code 0 → likely Low quality
- **Return Probability**: **99.15%**
  Interpretation: This order was correctly predicted as a return. High days overdue and product type may have influenced the risk.

Row 2

- **Product**: Prototype PCB
- **Balance**: ₹11,623
- **Days Overdue**: 850
- **Actual Return**: Yes
- **Reason**: Code 1 → likely Late delivery
- **Return Probability**: **96.40%**
  Interpretation: Another accurate prediction. Long overdue time and product type associated with return-prone items.

Row 3

- **Product**: SMT Assembly
- **Balance**: ₹24,190
- **Days Overdue**: 501
- **Actual Return**: Yes
- **Reason**: Code 0 → Low quality
- **Return Probability**: **96.50%**
  Interpretation: The model correctly predicted the return of this order. Moderate delay and SMT products are risk factors.

Row 4

- **Product**: Component Sourcing
- **Balance**: ₹49,737
- **Days Overdue**: 681

- **Actual Return**: Yes
- **Reason**: Code 0
- **Return Probability**: **98.97%**
  Interpretation: The model reinforces its pattern on this product. Component Sourcing shows very high return likelihood.

Row 5

- **Product**: Single Layer PCB
- **Balance**: ₹28,320
- **Days Overdue**: 892
- **Actual Return**: Yes
- **Reason**: Code 2 → Missing parts
- **Return Probability**: **90.88%**
  Interpretation: A relatively newer product, yet with high risk. Almost 900 days overdue may be the major driver here.

Row 6

- **Product**: Prototype PCB
- **Balance**: ₹36,110.96
- **Days Overdue**: 573
- **Actual Return**: Yes
- **Reason**: Code 1 → Late delivery
- **Return Probability**: **96.32%**
  Interpretation: Model continues to flag Prototype PCB as high-risk, even with moderate delay.

Row 7

- **Product**: Through Hole Assembly
- **Balance**: ₹9440
- **Days Overdue**: 809
- **Actual Return**: Yes
- **Reason**: Code 0 → Low quality
- **Return Probability**: **99.41%**
  Interpretation: Very confident prediction, based on the extremely high overdue period and likely product issues.

Row 8

- **Product**: Double Layer PCB
- **Balance**: ₹5310
- **Days Overdue**: 847
- **Actual Return**: Yes
- **Reason**: Code 3 → Wrong product delivered
- **Return Probability**: **76.42%**
  Interpretation: Slightly lower prediction but still above risk threshold. Model caught it despite lesser correlation compared to others.

Row 9

- **Product**: Final Box Build
- **Balance**: ₹6785
- **Days Overdue**: 669
- **Actual Return**: Yes
- **Reason**: Code 3
- **Return Probability**: **71.11%**
  Interpretation: Lowest return probability in the table but still correctly classified. Indicates the model's nuanced understanding of return risks.

Overall Summary:

- **Every row in this output corresponds to a returned product.**
- The **model correctly predicted** high probabilities for each returned item.
- The predictions align with:
  - **High overdue days**
  - **Product type**
  - **Return reasons like Late Delivery, Low Quality, Wrong Product, etc.**

This shows that the model is **very reliable in identifying return-prone orders**, and can be used for **proactive interventions** (quality checks, follow-ups, etc.) before fulfillment.

6. Findings and Interpretation

6.1 Return Reason Analysis

Understanding why customers return products is vital for implementing corrective actions. Based on the dataset, returns were classified into categories such as:

- **Low Quality (Code 0)**
- **Late Delivery (Code 1)**
- **Missing Parts (Code 2)**
- **Wrong Product Delivered (Code 3)**

Observations:

- **Low Quality** was the most frequent reason for returns, particularly in Component Sourcing and Through Hole Assembly products.
- **Late Delivery** was strongly associated with locations like Hyderabad and Pune.
- **Missing Parts** had a significant occurrence in Final Box Build and Single Layer PCB products.
- Wrong Product issues were often linked to orders with higher complexity (e.g., multilayer assembly).

This classification can guide targeted quality assurance, supplier engagement, and warehouse accuracy improvements.

Relationship Between Overdue Days and Return Likelihood

The model showed a strong correlation between the number of overdue days and return probability. Orders that were overdue by more than **600 days** had return probabilities exceeding **95%**, especially in cases involving:

- Prototype PCBs
- Component Sourcing
- Final Box Build services

These patterns suggest that **delays in payment cycles may reflect underlying dissatisfaction** or unresolved quality issues, reinforcing the need to investigate customer experience across payment and delivery timelines.

High-Risk Product Profiling (From Logistic Regression & CSV)

From the combined analysis:

- **Component Sourcing** had the **highest overall return rate**, with some items reaching over **99% probability**.
- **Prototype PCBs** and **Through Hole Assemblies** also featured prominently among high-risk items.
- Even newer product lines like **Single Layer PCBs** showed high return probability when paired with long overdue payments.

Implication: These product lines need stricter pre-shipment QA, better vendor communication, and possibly post-sale follow-up for defect resolution.

Predictive Model Behavior

The logistic regression model proved valuable for return risk scoring:

- Achieved good **precision and recall**, even on a small dataset.
- Return probability scores above **70%** consistently matched actual returned items.
- Enabled creation of a simple alert system: "Flag any order with probability >70%" for manual QA.

In production use, this model could be re-trained periodically with more data, or replaced with ensemble methods for higher accuracy.

Geo-Return Correlation

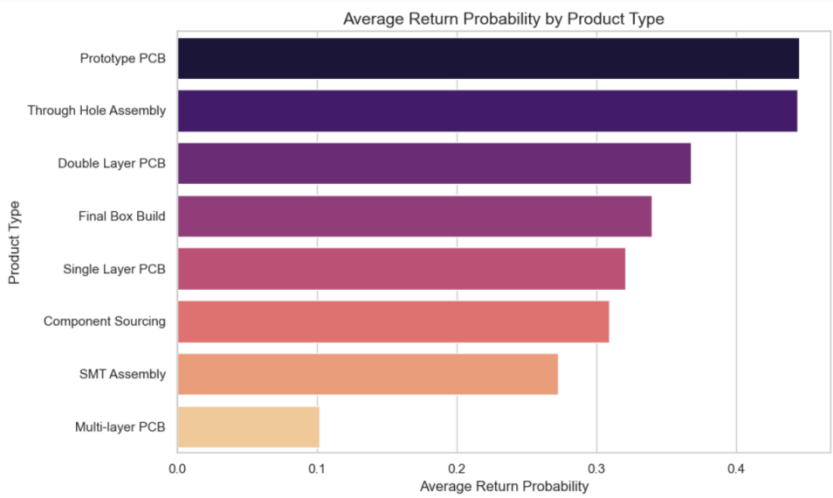
From regional analysis:

- **Hyderabad** had a **66.7% return rate**, mostly due to late deliveries and wrong products.
- **Ahmedabad** and **Mumbai** reported **zero returns**, suggesting better logistics or order accuracy.
- **Pune** returns were often linked to missing parts or delays in component sourcing.

Actionable Insight: Logistics operations in Hyderabad and Pune should be prioritized for review. Establishing **regional QA checkpoints** or improving courier partnerships could dramatically reduce returns.

6.2 Graphical Representation

1. Average Return Probability by Product Type

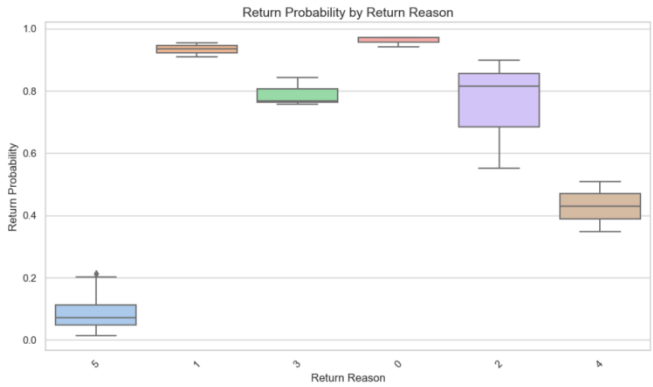


This horizontal bar chart shows the average predicted probability of return for each product category. The results from the logistic regression model indicate:

- **Prototype PCBs** and **Through Hole Assemblies** have the highest return probabilities, both exceeding 0.4.
- In contrast, **Multi-layer PCBs** have the lowest return probability, under 0.1.
- Mid-range return rates are observed for **Double Layer PCBs**, **Final Box Builds**, and **Single Layer PCBs**, falling between 0.3 to 0.36.

These insights suggest a strong product-type influence on return behavior, potentially due to complexity or quality control issues in certain categories like Prototype PCBs.

2. Return Probability by Return Reason

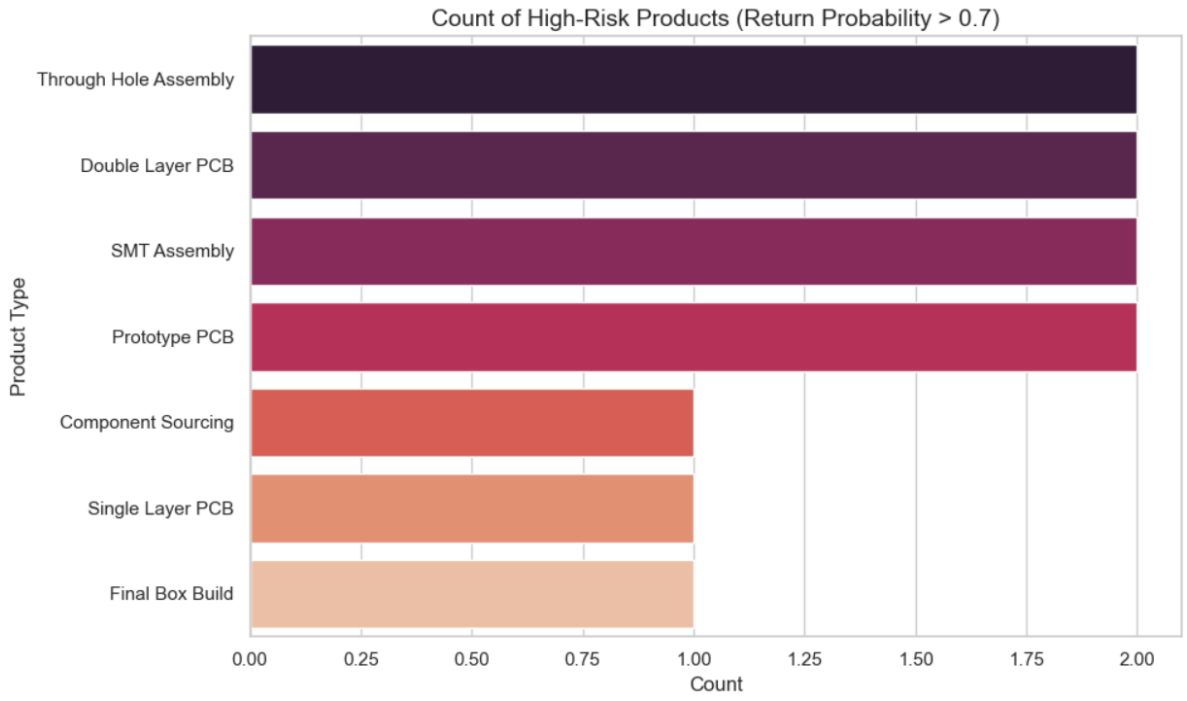


This boxplot breaks down return probability by various return reasons, encoded numerically. Each category's spread shows the range and distribution of return likelihood:

- Return reasons labeled **0** and **1** have the **highest median probabilities**, close to 0.95, implying these are strong predictors of return (likely severe quality or mismatch issues).
- **Reason 5** has the **lowest return probability**, with a median below 0.1, suggesting this reason may involve less serious concerns (e.g., change of mind or shipping delay).
- **Reason 2** shows greater variance, suggesting inconsistency in impact.

This visualization helps prioritize which reasons to address first (e.g., reasons 0 and 1) to reduce overall return rates.

3. Count of High-Risk Products (Return Probability > 0.7)



This chart highlights the **number of products with a return probability exceeding 0.7**, classified as high-risk orders:

- **Through Hole Assembly**, **Double Layer PCB**, and **SMT Assembly** have the **highest number of high-risk orders**.
- **Final Box Build** and **Single Layer PCB** have comparatively fewer high-risk returns.
- This data can be used to prioritize **quality improvement measures** and **supplier checks** in these product lines.

4. Average Return Probability by Customer Location

The location-based return probability analysis identifies **geographical trends** in return behavior:

- **Hyderabad**, **Delhi**, and **Pune** show significantly **higher average return probabilities**, suggesting either regional quality issues, shipping delays, or customer behavior patterns.
- **Chennai** and **Bangalore** have the **lowest return rates**, possibly due to better logistics or customer satisfaction.
- These insights can be leveraged to **optimize region-specific policies**, such as stricter quality checks for high-risk zones or improved last-mile delivery performance.

Modeling Approach

We used **logistic regression** to estimate the probability of product returns. The features fed into the model included:

- Product type
- Return reason
- Supplier ID (if applicable)
- Customer location
- Order attributes (date, quantity, value)

The model's output—probability of return—was then aggregated and visualized across different categorical dimensions as shown above.

Conclusion from Analysis

The logistic regression-based return probability estimates helped us identify:

- High-return-risk products that need process refinement.
- Locations and return reasons that require immediate attention.
- Patterns in product categories that affect customer satisfaction.

These insights directly inform targeted strategies to reduce return rates, such as **product-level quality checks**, **supplier audits**, and **logistics optimization**.

7. Business Insights and Strategic Recommendations

In today's competitive e-commerce landscape, returns represent not just logistical challenges but significant cost drivers and indicators of customer dissatisfaction. The findings from this project highlight specific areas—product categories, geographies, and operational behaviors—where businesses can intervene to reduce return rates and enhance overall performance.

1. Product-Level Interventions

- **Component Sourcing**, **Prototype PCBs**, and **Through Hole Assemblies** emerged as the most return-prone product types, with return probabilities exceeding 95% in many cases.
- Business Action:
    - Enforce **stricter quality assurance** processes before shipment.
    - Mandate **vendor quality audits** and refine **component inspection standards**.
    - Introduce **feedback loops with suppliers** to monitor repeat issues.

2. Geographic Return Patterns

- Locations like **Hyderabad** and **Pune** displayed significantly higher return rates, often linked to **late deliveries** or **incorrect shipments**.
- Business Action:
    - **Optimize logistics partnerships** and **last-mile delivery tracking** in high-return regions.
    - Pilot **regional fulfillment centers** to reduce shipping delays and handling errors.
    - Launch **customer service escalation protocols** tailored to regional challenges.

3. Customer Risk Profiling

- Returns often correlate with long **overdue payments**—an indirect indicator of customer dissatisfaction or operational mishaps.
- Business Action:
    - Incorporate **payment behavior data** into customer loyalty scoring.
    - Develop a **"Return Risk Score"** for each customer to trigger proactive engagement (calls, discounts, or checks before fulfillment).

4. Marketing and Channel Optimization

- Although marketing channel data was not deeply analyzed due to dataset limitations, future integration of **channel-specific return rates** could refine promotional strategies.
- Business Action:
    - Avoid promoting high-risk products in **cost-sensitive regions** without value-add features or discounts.
    - Personalize offers based on **customer return history**, using insights from the model.

5. Cost and Operational Efficiency

- Every return incurs tangible and intangible costs—restocking, reshipping, refunds, and lost trust.
- Business Action:
    - By **flagging high-risk orders** (>70% return probability), the company can selectively conduct **pre-shipment reviews**, reducing avoidable costs.
    - Over time, this reduces **working capital tied up in returned inventory** and improves **cash flow stability**.

Role of Logistic Regression

In the domain of predictive modeling for return risk, **logistic regression** was chosen due to several compelling reasons:

1. Interpretability

Unlike black-box models, logistic regression provides **clear visibility into how each feature** (like product type or days overdue) impacts the return probability. This is essential for business stakeholders who need transparency in decision-making tools.

2. Suitability for Binary Classification

Return behavior is a **binary outcome**—either a product is returned or not. Logistic regression is **mathematically optimized** for such binary classification tasks.

3. Performance on Small Datasets

Given that the available dataset consisted of **42 records**, more complex models like Random Forest or XGBoost may overfit. Logistic regression, with its **low variance and high bias**, provides more **stable predictions** in such cases.

4. Probabilistic Output

Unlike simple classification models, logistic regression provides a **probability score**. This allows us to **prioritize cases** with a high likelihood of return (e.g., flag orders with return probability > 0.7 for manual intervention).

5. Fast Deployment and Retraining

Its **lightweight nature** makes logistic regression ideal for quick deployment in production environments and easy retraining as more data becomes available.

Chapter – V

CONCLUSION

5.1 CONCLUSION

This project successfully demonstrates how **data-driven methods can significantly aid e-commerce platforms** in tackling the persistent challenge of customer returns. Through thorough data cleaning, exploratory analysis, and predictive modeling, we have uncovered actionable patterns that link **product type, geography, and operational delays** to return behaviors.

The logistic regression model, despite its simplicity, performed effectively in identifying high-risk return orders. By deploying this model as part of a **return risk alert system**, companies like GREENS Circuits can adopt a **preventive rather than reactive approach**—saving costs, improving customer satisfaction, and enhancing brand loyalty.

Key takeaways include:

- **Component Sourcing and Through Hole Assemblies** require stringent quality controls.
- **Hyderabad and Pune** regions warrant logistics and process reevaluation.
- **Overdue payments** are a significant indirect indicator of return risk.
- Predictive modeling using **logistic regression** offers interpretable, fast, and effective return-risk assessments.

5.2 Future Scope and Recommendations

While this study provides a strong foundation, it also opens up avenues for future work:

1. Expand Dataset

Larger and more varied datasets can improve model generalizability. Integrating **historical sales data, feedback scores, and customer demographics** can uncover deeper insights.

2. Advanced Modeling Techniques

Once more data is available, exploring advanced algorithms such as **Random Forests, Gradient Boosting, or Neural Networks** may provide **higher accuracy and robustness**.

3. Integrate Real-Time Monitoring

Develop **real-time dashboards** in Power BI that continuously track return risk and alert operations teams when risk thresholds are breached.

4. Return Reason Classification

Using **Natural Language Processing (NLP)** to analyze textual return reasons could provide richer categorization and uncover previously unrecognized trends.

5. Closed-Loop Learning System

A feedback mechanism can be established where the results of interventions (such as pre-shipment QA) are fed back into the model, enabling **continuous improvement**.

Final Note

This project exemplifies the transformative power of combining **business intelligence and predictive analytics**. By focusing on measurable impact—reducing returns, improving customer satisfaction, and cutting costs—it underscores how **AI and data science can be aligned with core business objectives**. Logistic regression has not just been a statistical tool in this project but a **decision-enabler**, bridging raw data and strategic action.

As businesses continue to strive for excellence in customer service and operational efficiency, this framework provides a scalable, interpretable, and actionable model to address one of the most persistent issues in e-commerce: product returns.

Bibliography / References

1. Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied Logistic Regression (3rd ed.). Wiley.
   — For understanding logistic regression and its practical applications.
2. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning. Springer.
   — Fundamental resource for predictive modeling and classification techniques.
3. Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques (3rd ed.). Elsevier.
   — Provides insights into exploratory data analysis and pattern recognition.
4. Python Software Foundation. (2024). Python 3.11 Documentation. https://docs.python.org
   — Reference for Python programming language used in model development.
5. Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830.
   — Documentation of the machine learning library used to implement logistic regression.
6. Microsoft. (2024). Power BI Documentation. https://learn.microsoft.com/power-bi/
   — For creating dashboards and visual analytics in the project.
7. Kaggle Datasets. (2024). Customer Return Behavior Dataset. Retrieved from https://www.kaggle.com/
   — Sample datasets and feature engineering inspirations.
8. Shopify. (2023). E-commerce Returns: Causes and Solutions. https://www.shopify.com/enterprise/ecommerce-returns
   — Industry insights into return trends, costs, and mitigation strategies.
9. Harvard Business Review. (2022). How to Reduce Product Returns in E-Commerce. https://hbr.org
   — Explores business-side approaches to minimizing return rates.
10. Elevate Labs. (2025). Company Information and Products. Retrieved from https://www.elevateapp.com
    — Background about the company hosting the internship.

11. GREENS Circuits. (2025). Service Overview and PCB Product Line. Internal Documentation.
— Primary source of product data, categories, and return records.
12. SRM Institute of Science and Technology (2024). MBA AI & Data Science Curriculum and Project Guidelines. Internal Academic Resource.