



FACULTY OF MANAGEMENT

**SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
KATTANKULATHUR- 603 203**

APRIL 2025

**Mano Barathi.M
Reg. No.RA2352008010065**

AN ANALYSIS ON CUSTOMER RETURN BEHAVIOR AND PREDICTIVE MODELING



RESEARCH DESIGN

- 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.

Introduction:

- Product returns are a major concern in e-commerce due to cost, logistics, and customer dissatisfaction.
- The goal is to analyze return behavior and build a predictive model.

Objectives:

- Understand key drivers of returns
- Develop a logistic regression model to predict return probability
- Provide business insights and actionable recommendations.

TOOLS AND TECHNIQUES

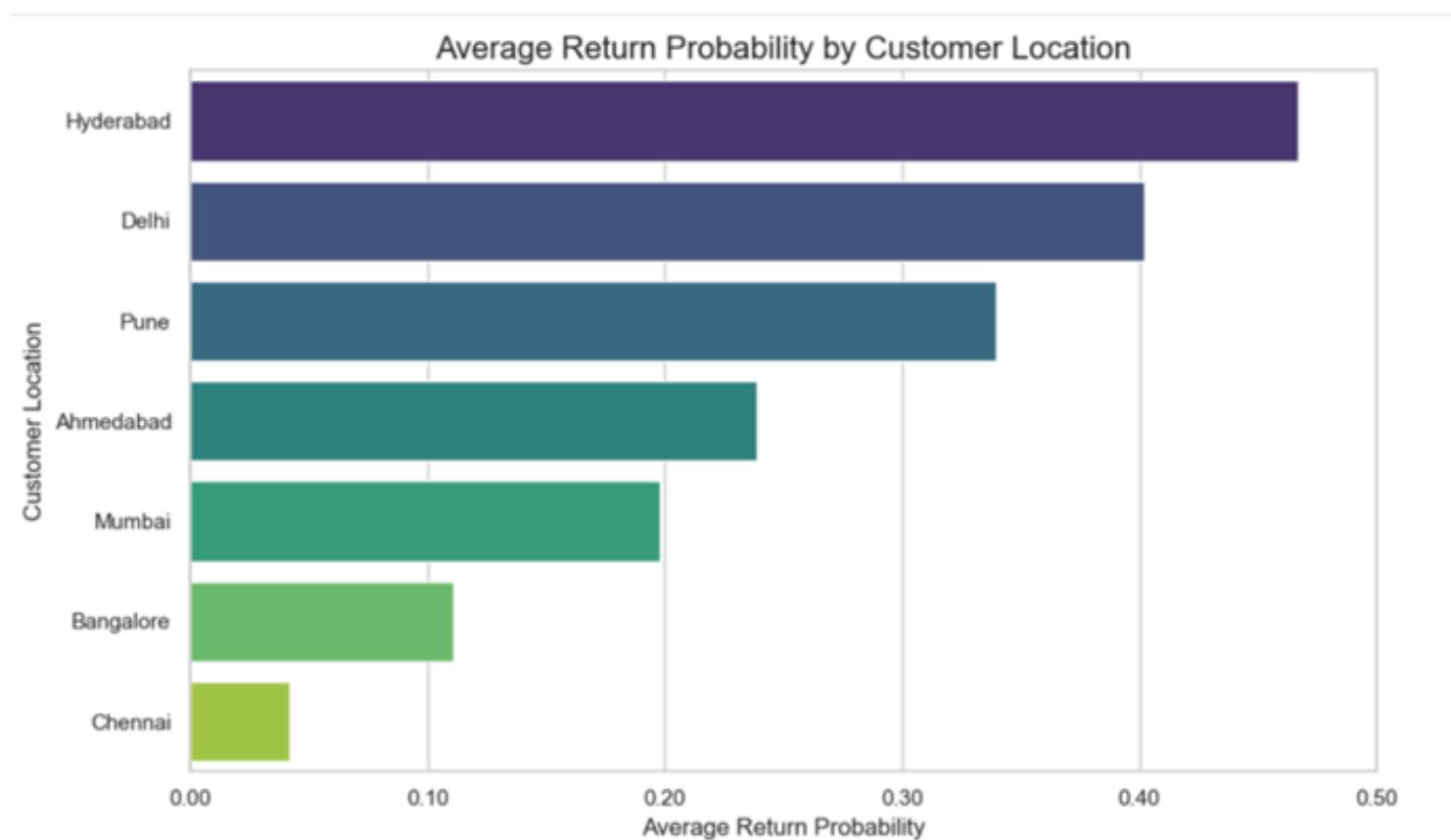
- **pandas:** Data manipulation and analysis
- **numpy:** Numerical computations
- **matplotlib & seaborn:** Data visualization
- **scikit-learn:** Machine learning modeling, preprocessing, and evaluation

Statistical Technique:

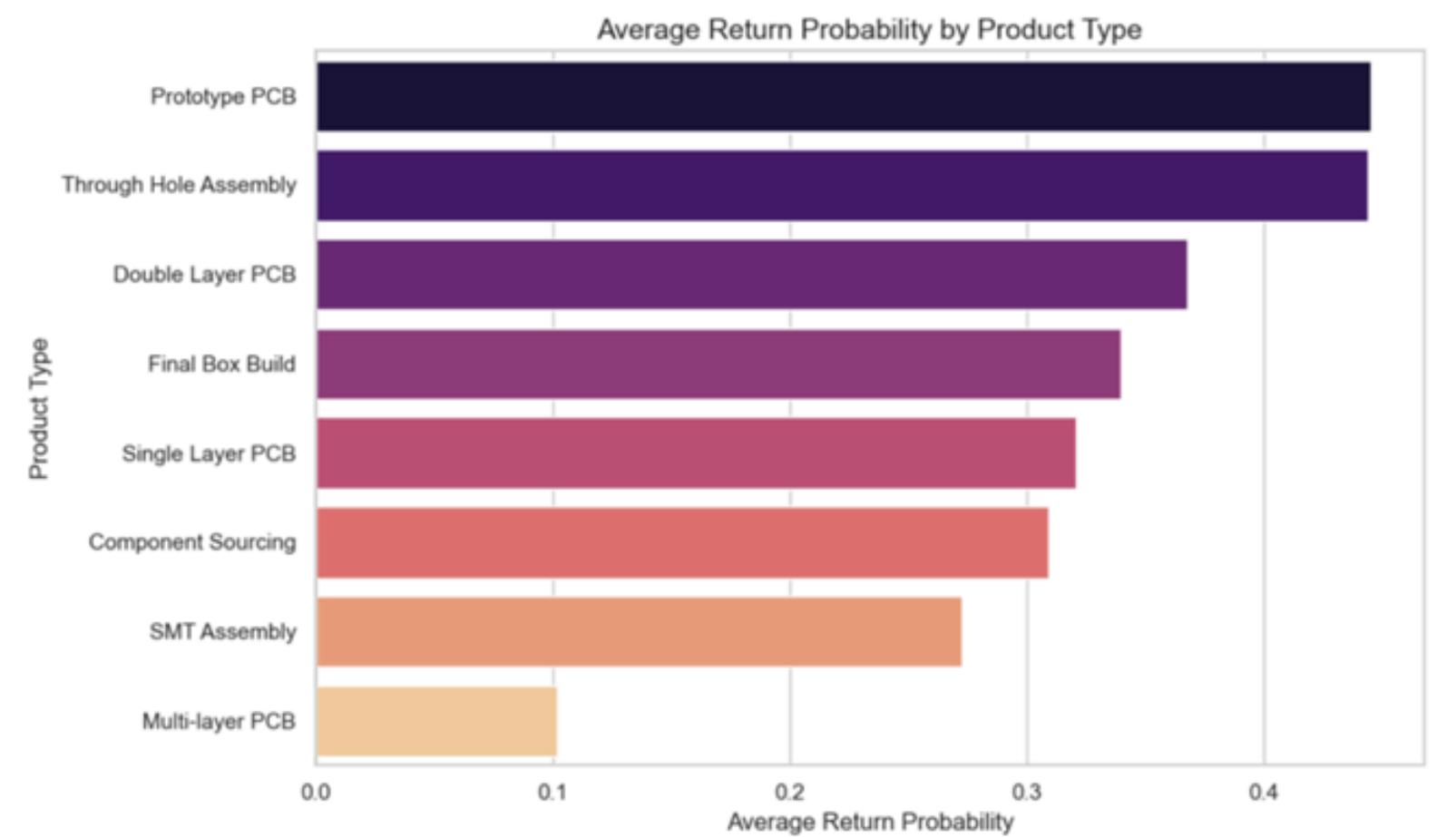
- **Logistic Regression:** A supervised machine learning algorithm used to estimate the probability of a binary outcome—in this case, whether a product will be returned or not.

INTERPRETATIONS

- 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.

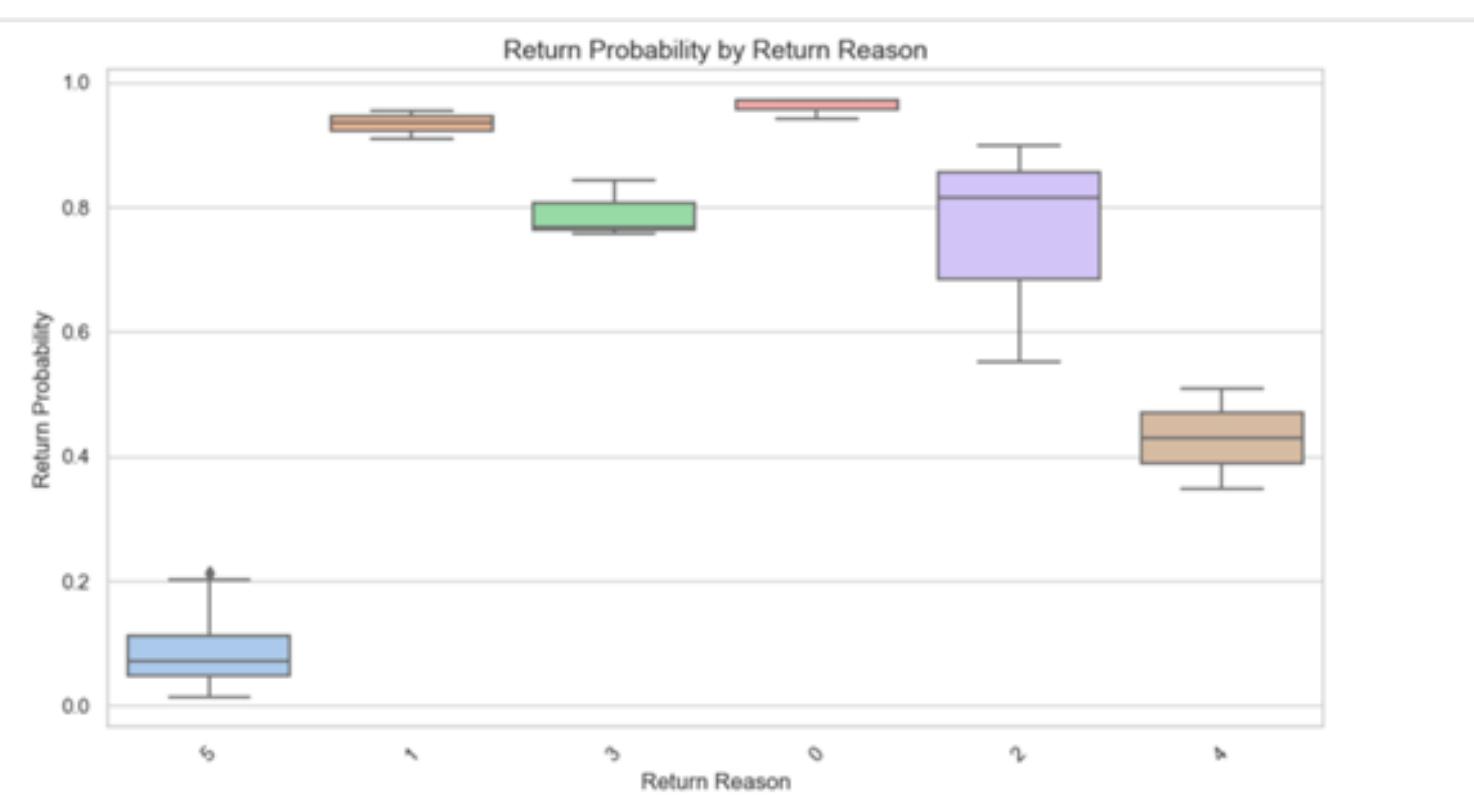


- Prototype PCBs and Through Hole Assemblies have the highest return probabilities, both exceeding 0.4.
- In contrast, Multi-layer PCBs have the lowest return probability, under 0.1.
- Mid-range return rates are observed for Double Layer PCBs, Final Box Builds, and Single Layer PCBs, falling between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

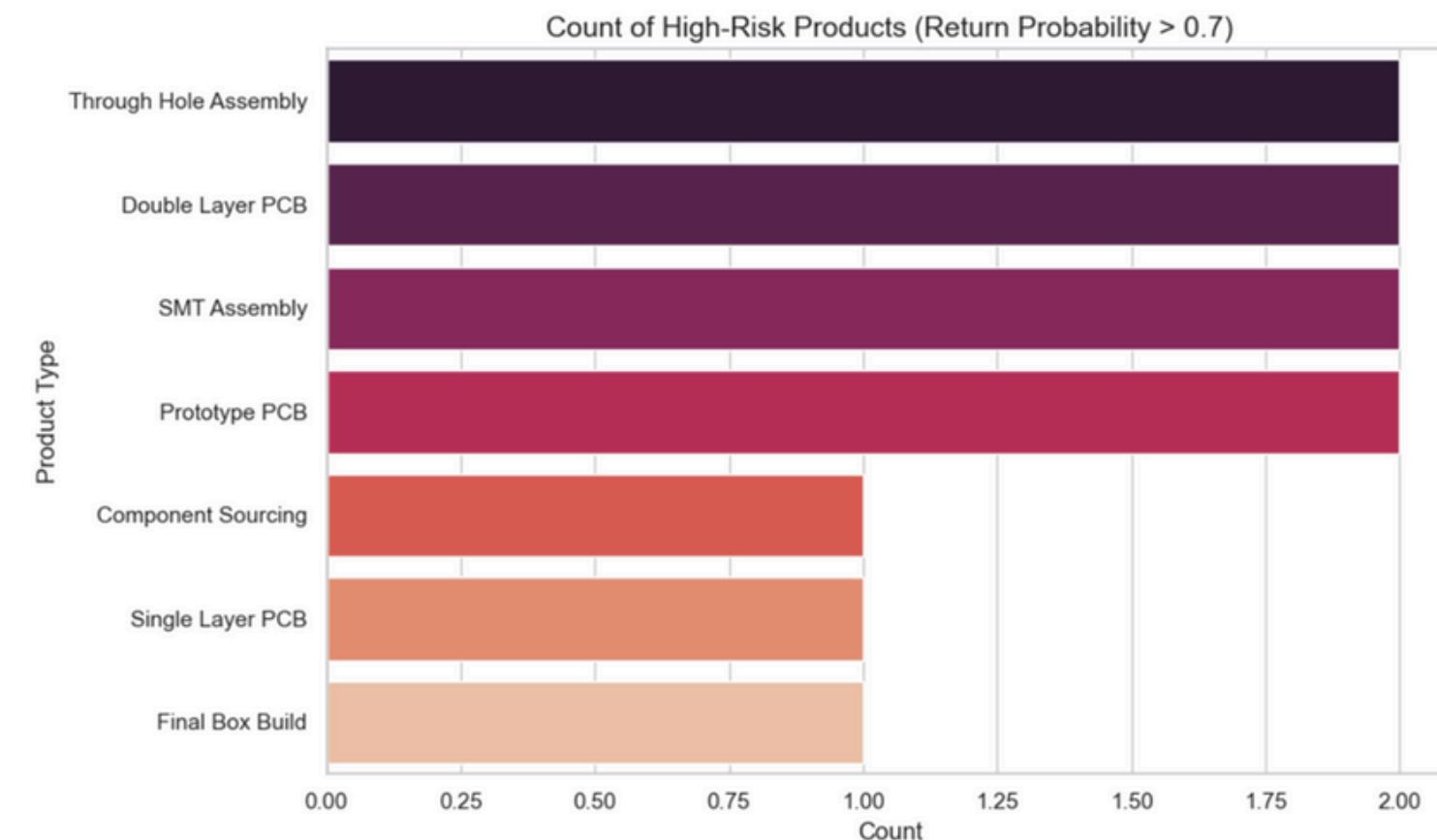


INTERPRETATIONS

- 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.



- 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.



LOGISTIC REGRESSION

Probabilities from Logistic Regression

- LOGISTIC REGRESSION DOESN'T JUST SPIT OUT "YES" OR "NO" (0 OR 1). IT ACTUALLY CALCULATES THE PROBABILITY THAT A GIVEN INPUT BELONGS TO CLASS 1 (E.G., "RETURNED").
- THE SIGMOID FUNCTION SQUASHES THE OUTPUT TO A RANGE BETWEEN 0 AND 1,

	precision	recall	f1-score	support
0	0.76	1.00	0.87	13
1	1.00	0.73	0.85	15
accuracy			0.86	28
macro avg	0.88	0.87	0.86	28
weighted avg	0.89	0.86	0.86	28

Accuracy: 0.8571428571428571

RETURN RISK ANALYSIS

Return Probability

The predicted probability from the model that the product will be returned. Values close to 1 mean high return risk.

	Location	
0.9967083048680316	Location	Bangalore
0.9955479807671309	Location	Hyderabad
0.5677781554301089	Location	Pune
0.5210181088488061	Location	Chennai
0.4005770831943784	Location	Coimbatore
0.2987296926839235	Location	Delhi
0.06427935570908927	Location	Ahmedabad
0.05484575320952578	Location	Mumbai

Double Layer PCB	0.912902911215864
Through Hole Assembly	0.7461992888169044
Prototype PCB	0.5032844260738052
Component Sourcing	0.45656159048500455
Single Layer PCB	0.3684239419446827
SMT Assembly	0.23544998147557403
Multi-layer PCB	0.08054987077470799
Final Box Build	0.07365991549653303

RETURN RISK ANALYSIS

Return Probability

The predicted probability from the model that the product will be returned. Values close to 1 mean high return risk.

0.9994074072940888	Return Reason	Defective item
0.9886177174611671	Return Reason	Late delivery
0.9694987566118242	Return Reason	Low quality
0.7798991495415236	Return Reason	Missing parts
0.35301572240701984	Return Reason	Wrong product delivered
0.05724490206451237	Return Reason	not returned

CONCLUSION

- 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.



A large, bold, white sans-serif text "THANK YOU" is centered in the upper half of the image. The background is a solid dark blue. Behind the text, there are several abstract graphic elements: a cluster of overlapping blue chevrons pointing right in the top left, another cluster in the top right, and a faint, dark silhouette of a city skyline with a prominent skyscraper on the right side.

**THANK
YOU**