

# DATA SCIENCE TOOLBOX: PYTHON PROGRAMMING

## PROJECT REPORT

(Project Semester: January- April 2025)

# DECODING RETAIL DYNAMICS: DEMOGRAPHICS, TRENDS & RETURNS

Submitted by

Gaurav Tiwari

Registration No.12315283

Programme and Section K23GD

Course Code INT 375

Under the Guidance of

Baljinder Kaur (UID:28968)

Assistant Professor

Discipline of CSE/IT

Lovely School of Computer Science

Lovely Professional University, Phagwara

## CERTIFICATE

This is to certify that Gaurav Tiwari, bearing Registration no. 12315283 has completed the INT375 project titled “DECODING RETAIL DYNAMICS: DEMOGRAPHICS, TRENDS & RETURNS” under my guidance and supervision. To the best of my knowledge, the present work is the result of his original development, effort, and study.

Baljinder Kaur (UID:28968)

Assistant Professor

School of Computer Science and Engineering

Lovely Professional University Phagwara,

Punjab.

Date: 11-04-2025

## DECLARATION

I, Gaurav Tiwari, student of BTech under CSE Discipline at Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 11-04-2025

Signature:

Registration No. 12315283

Gaurav Tiwari

## ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to my guide, Baljinder Kaur mam, for their invaluable guidance, support, and encouragement throughout the completion of this project. His expertise and constructive feedback have greatly contributed to the success of this work.

I would also like to extend my sincere thanks to Lovely Professional University for providing me such a wonderful opportunity to work on this project in the subject DATA SCIENCE TOOLBOX: PYTHON PROGRAMMING with subject code INT 375, helping with the necessary resources and skills that laid the foundation for my research.

This project, titled “DECODING RETAIL DYNAMICS: DEMOGRAPHICS, TRENDS & RETURNS”, has been a learning experience, and I would like to acknowledge the support of my peers, family, and all others who helped me in any manner.

Thank you all for your continuous support and motivation.

Name: Gaurav Tiwari

Section: K23GD

Roll No.: 59

Reg. No.: 12315283

Date: 11-04-2025

## CONTENTS

S No.	Title	Page No.
1.)	Introduction	6
2.)	Source of dataset	7
3.)	Dataset Preprocessing	8-9
4.)	Analysis on dataset (for each objective) i. General Description ii. Specific Requirements iii. Analysis results iv. Visualization	10-18
5.)	Conclusion	19
6.)	Future Scope	20
7.)	References	21

## INTRODUCTION

The retail sector has experienced a seismic transformation in recent years, driven by shifting customer preferences, economic changes, and the digital revolution. Understanding these evolving dynamics is critical for businesses aiming to stay competitive and customer-centric. The rise of e-commerce, omnichannel experiences, and hyper-personalization has reshaped how customers interact with brands, requiring new approaches to analyze purchasing behavior, trends, and returns.

This project, titled "Decoding Retail Dynamics: Demographics, Trends & Returns," presents a data-driven exploration of customer behavior and sales patterns using the Online Retail II dataset. Leveraging the power of Python libraries such as Pandas, Plotly, and Dash (with Dash Mantine Components), the project aims to build a futuristic and interactive analytics dashboard that visualizes key retail KPIs. Through customer segmentation, seasonality detection, purchase trend analysis, and return insights, the dashboard provides comprehensive visibility into the state of retail.

By processing, cleaning, and enriching the dataset, we extract meaningful insights such as best-selling products, top regions by revenue, frequent buyers, and seasonal sales trends. Visualizations such as bar charts, pie charts, scatter plots, line graphs, and heatmaps offer intuitive views of complex metrics. This helps retailers understand customer demographics, optimize inventory, forecast sales, and reduce return rates. Ultimately, this project emphasizes the role of data science in transforming retail intelligence into actionable strategy.

## SOURCE OF DATASET

The dataset used for this project is the publicly available "Online Retail II" dataset from the UCI Machine Learning Repository. It includes transactional data for a UK-based online retail store, covering invoice-level records of sales between 2009 and 2011. The data contains information such as invoice number, product description, quantity, invoice date, unit price, customer ID, and country.

This dataset is ideal for conducting customer and product-level analysis. It allows temporal exploration of sales trends, customer segmentation based on purchase behavior, and geographic distribution of revenue. The dataset was imported in CSV format and analyzed using Python tools for comprehensive data wrangling, enrichment, and visualization.

# DATASET PREPROCESSING

## 1. Data Loading

- Loaded the CSV file using Pandas' `pd.read_csv()`.
- Used `.head()`, `.info()`, `.describe()`, and `.isnull().sum()` to assess data quality.

## 2. Handling Missing Data

- Rows with missing Customer IDs were dropped to ensure valid customer-level analysis.
- Removed rows with zero or negative quantities and prices to maintain logical consistency.

## 3. Data Type Conversion

- Converted InvoiceDate to datetime using `pd.to_datetime()`.
- Extracted additional features: Year, Month, Day, Hour for trend and seasonality analysis.

## 4. Duplicate Removal

- ❑ Checked for duplicates using `.duplicated()` and removed them.

## 5. Feature Engineering

- Created TotalPrice column ( $\text{Quantity} * \text{UnitPrice}$ ).
- Grouped data by Country, Product, and Customer to derive new aggregated features.
- Applied RFM (Recency, Frequency, Monetary) scoring for customer segmentation.

## 6. Binning and Categorization

- Segmented customers based on total purchase value into Low, Medium, and High value.
- Binned products by return frequency to understand product-level return behavior.

## 7. Statistical Summarization

- ❑ Generated summaries for revenue, order quantity, average spend per customer.

## 8. Data Readiness

- ❑ The dataset was cleaned, enriched, and structured for visualization via Dash components.

# DETAILED ANALYSIS BASED ON PROJECT OBJECTIVES

## Objective 1: Understand Customer Demographics

- Chart Used: Donut Chart, Treemap
- Purpose: Analyse number of customers by country and average spend.
- Insight: Most revenue generated by UK customers; few high-value customers from EU.

## Objective 2: Analyse Purchase Trends

- Chart Used: Line Plot, Heatmap
- Purpose: Identify monthly revenue trends and peak shopping hours.
- Insight: Seasonal trends with spikes in Q4; major purchases during work hours.

## Objective 3: Customer Segmentation

- Chart Used: Scatter Plot (RFM segments)
- Purpose: Cluster customers into groups (e.g., loyal, new, lost).
- Insight: Helps prioritize customer retention strategies.

#### Objective 4: Detect Seasonality in Sales

- Chart Used: Time Series Decomposition
- Purpose: Detect cyclical patterns and trends.
- Insight: Holiday season impacts visible; steady baseline growth.

#### Objective 5: Cancellation and Returns Insight

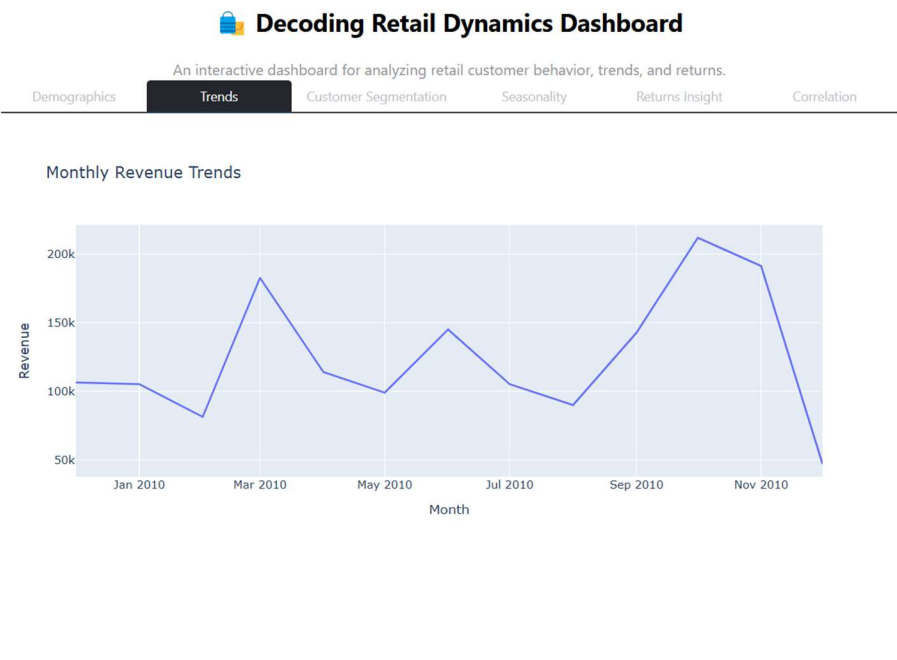
- Chart Used: Bar Chart, Pie Chart
- Purpose: Visualize frequency and monetary impact of returns.
- Insight: Certain products have high return rates, suggesting quality or delivery issues.

## Demographics

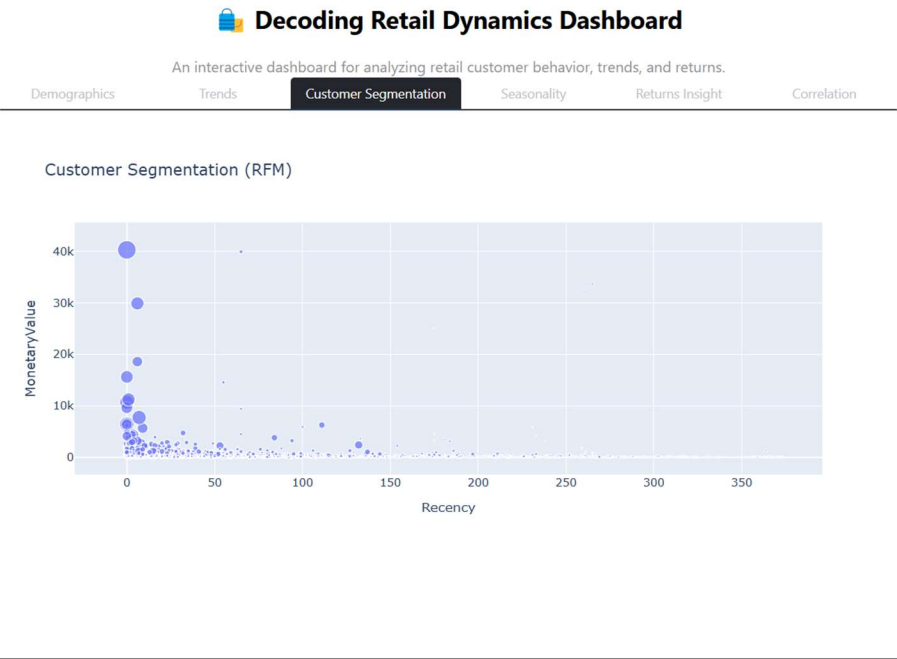




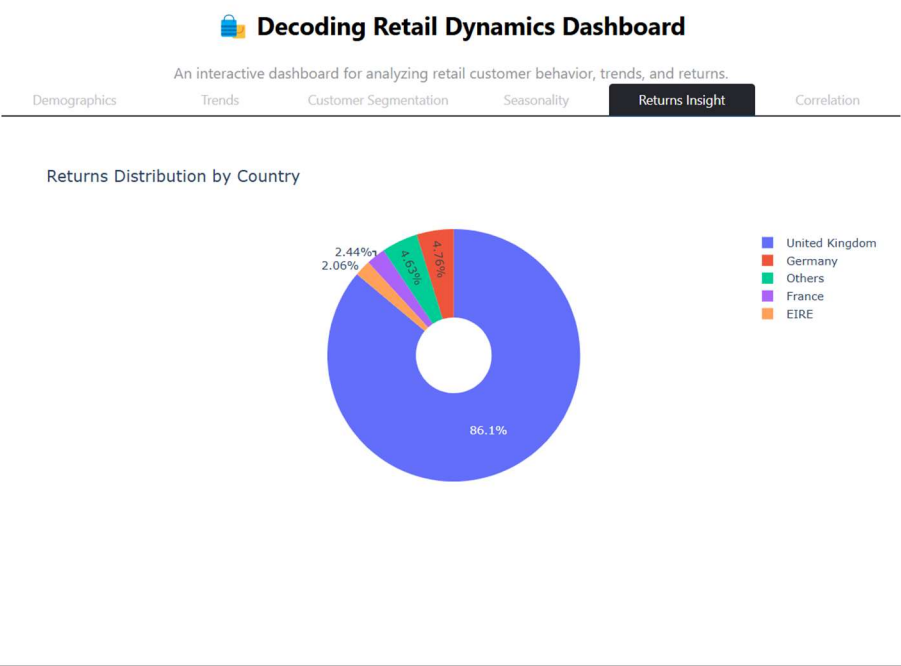
Trends



Customer Segmentation



Seasonality



Returns Insight



Correlation



CONCLUSION

The "Decoding Retail Dynamics" project successfully demonstrates how data analytics can be used to uncover customer behavior patterns, optimize product strategies, and improve operational decisions. By segmenting customers, visualizing trends, and analyzing return behavior, retailers can implement targeted marketing and inventory decisions.

Key findings:

- Majority of revenue comes from a small segment of loyal customers.
  - Sales exhibit strong seasonal trends, peaking during the holidays.
  - High return rates are concentrated around specific products and customers.
- These insights are crucial for businesses seeking to adapt in a competitive and fast-changing retail environment. With effective use of interactive visualizations and EDA, this project highlights how data-driven decision-making can significantly enhance retail performance.

## FUTURE SCOPE

1. Real-time Dashboard Integration:
  - o Connect live sales data APIs to create a real-time business intelligence system.
2. Predictive Modeling:
  - o Use machine learning models (e.g., XGBoost, LSTM) to forecast future sales and returns.
3. Product Recommendation Engine:
  - o Build collaborative filtering-based engines to recommend products to customers.
4. Churn Prediction:
  - o Predict which customers are at risk of leaving based on behavior patterns.
5. Inventory Optimization:
  - o Integrate stock levels and supplier lead times to recommend reorder points.
6. Enhanced UI/UX:
  - o Include more dashboard filters (price range, category, region) and theme toggles.
7. Multi-language Support:
  - o Enable dashboards in different languages to enhance accessibility for global users.

## REFERENCES

- UCI Machine Learning Repository - Online Retail II Dataset
- Dash Plotly Documentation
- Dash Mantine Components Documentation
- Python Pandas & Plotly Libraries
- Retail Analytics Blogs and Whitepapers