

# Product Return Rate Analysis Report

## 1. Introduction

Returns are a critical metric in e-commerce and retail analytics, impacting customer satisfaction, revenue, and operational efficiency. This project aims to analyze return trends using a simulated dataset of product orders. The objective is to identify high-return categories, suppliers, and regions, and uncover actionable insights using data analysis and visualization techniques.

## 2. Abstract

This project focuses on return rate analytics using a synthetic order dataset that includes product details, marketing channels, return flags, and reasons. The analysis pipeline covers data generation using Python, transformation using SQL and Pandas, and dashboard creation using Power BI. The core goal is to determine return percentages by category, supplier, and region, empowering decision-makers to optimize product offerings, logistics, and marketing strategies.

## 3. Tools and Technologies Used

Tool	Purpose
Python	Data generation, transformation, analysis
Pandas	Grouping, aggregation, and calculation
SQLite / SQL	Structured query-based analysis
Power BI	Visualization and dashboard creation
Excel (CSV)	Initial dataset format for portability

## 4. Steps Involved

- Step 1: Dataset Creation (Python)

A dataset of 500 synthetic orders was created using Python and the Faker library. Columns included: order\_id, product\_name, category, order\_date, quantity, price, region, country, marketing\_channel, is\_returned, and return\_reason.

- Step 2: Data Enrichment

Added supplier field (randomly assigned from a fixed list). Cleaned and validated data types for price, date, and categorical fields.

- Step 3: SQL Analysis (Optional)

Imported data into SQLite for relational queries. Example query:

```
SELECT category, region,  
       COUNT(*) AS total_orders,  
       SUM(is_returned) AS total_returns,  
       ROUND(SUM(is_returned) * 100.0 / COUNT(*), 2) AS return_percentage  
FROM orders  
GROUP BY category, region;
```

- Step 4: Return Rate Analysis (Python)

Used Pandas to compute return percentage:

```
df.groupby(["category", "region"]).agg(total_orders=('order_id', 'count'),  
total_returns=('is_returned', 'sum')).assign(return_percentage=lambda x:  
(x.total_returns / x.total_orders * 100).round(2))
```

- Step 5: Power BI Dashboard

Loaded CSV into Power BI Desktop. Created interactive visuals: KPI cards (overall return rate, total returns), bar charts (return % by category, region, supplier), slicers (date, region, marketing channel), and matrix views.

## 5. Conclusion

The project successfully demonstrated a complete workflow for analyzing product returns using modern data tools. Key insights such as high-return categories or problematic regions can drive improvements in supply chain, vendor selection, and product quality. Integrating Python, SQL, and Power BI provided both flexibility and depth in analysis. The framework is easily extendable to real datasets and scalable for enterprise applications.