VISHAL CHAUDHARY · DATA ANALYST · DUBLIN, IRELAND

Retail & Customer Analytics · Personal Project

E-Commerce Sales Analytics

RFM customer segmentation, demand forecasting and churn detection on 541,909 real retail transactions — personal project.

541,909	R²=0.98	30%	5
Transactions Analysed	Forecast Accuracy	Reporting Time Saved	Customer Segments

TOOLS & TECHNOLOGIES

Python	SQL	Power BI	Random Forest	RFM Analysis	K-Means
Pandas	CRISP-DM				

Email vishal.ch1401@gmail.com	LinkedIn linkedin.com/in/vishal111	GitHub github.com/chaudhary521	Location Dublin, Ireland

PROBLEM STATEMENT

Online retailers often rely on aggregate sales figures without understanding which customers drive value, which are drifting away, and which products have predictable demand. This personal project set out to build a complete analytics pipeline that segments customers, forecasts demand, and surfaces churn risk using a publicly available but genuinely challenging real-world dataset.

DATASET

UCI Online Retail II dataset: 541,909 transactions from a UK-based e-commerce retailer covering 2009 to 2011. Features include invoice number, stock code, product description, quantity, unit price, customer ID, and country of purchase. Data spanned 38 countries and required significant cleaning — cancellations, negative quantities, and missing customer IDs were all present.

APPROACH & METHODOLOGY

Followed the full CRISP-DM methodology from business understanding to deployment. Built an RFM (Recency, Frequency, Monetary) scoring pipeline and applied K-Means clustering to identify five distinct customer segments. Developed a Random Forest demand forecasting model trained on seasonal and product-level features. Identified churn-prone customers using purchase gap analysis. Delivered findings through an interactive Power BI dashboard with country-level drill-down capability.

KEY TECHNICAL HIGHLIGHTS

› Built end-to-end RFM pipeline identifying five customer clusters: Champions, Loyal, At-Risk, New, and Dormant.

› Random Forest demand forecasting model achieved $R^2 = 0.98$ on holdout validation data.

› SQL-based ETL automation reduced manual reporting preparation time by 30%.

› Identified top 20 revenue-driving SKUs and their seasonal demand curves for inventory planning.

› Interactive Power BI dashboard with country and segment drill-down for regional sales analysis.

› Full CRISP-DM documentation ensures the pipeline is reproducible and auditable.

KEY INSIGHTS & RESULTS

The Champions segment — top 15% of customers by RFM score — generated 58% of total revenue. The At-Risk segment averaged a 45-day purchase gap, representing a significant re-engagement opportunity. UK transactions made up 89% of volume but international markets showed faster growth rates year-over-year.

BUSINESS IMPACT

This project demonstrates how a small analytics investment can reshape commercial strategy. The customer segmentation directly informs which groups should receive discounts versus loyalty rewards. The 30% reduction

in reporting time freed analyst capacity for insight generation rather than data preparation. The demand forecasting model supports inventory pre-positioning ahead of seasonal peaks.
