Retail Stock Market Behavior

TEAM - DATA SCOUTS

TEAM DATA SCOUTS

1. MEESALA SREE SAI NATH

2. AKULA JITHENDRANATH

3. S ANSAR TEJMUL MOVIN

Introduction

Retail businesses generate enormous amounts of transactional data daily. Understanding purchasing patterns hidden in this data can help retailers optimize inventory, pricing, and customer engagement strategies.

This project analyzes the UCI Online Retail dataset (2010-2011) to uncover insights into purchasing behavior, product associations, and seasonal trends. Using data-mining and machine learning techniques, it explores how customer behavior shapes retail market dynamics and identifies actionable patterns for decision-making.

WORK PLANNING & DIVISION

DATE RANGE	PHASE	FOCUS AREA	KEY DELIVERABLES	LEAD
Nov 1 - Nov 5	Phase 1	Documentation & Planning	Research Objectives, Hypothesis Dataset Rationale, Methodology, Preprocessing plans	Tejmul Movin
Nov 6 - Nov 20	Phase 2	Data Exploration & Modeling	EDA Notebooks, Processed Dataset Trained Models	Akula Jithendranath
Nov 21 - Dec 3	Phase 3	Visualization & Presentation	Insights Report, Evaluation Metrics, Presentation Slides	M Sree Sai Nath

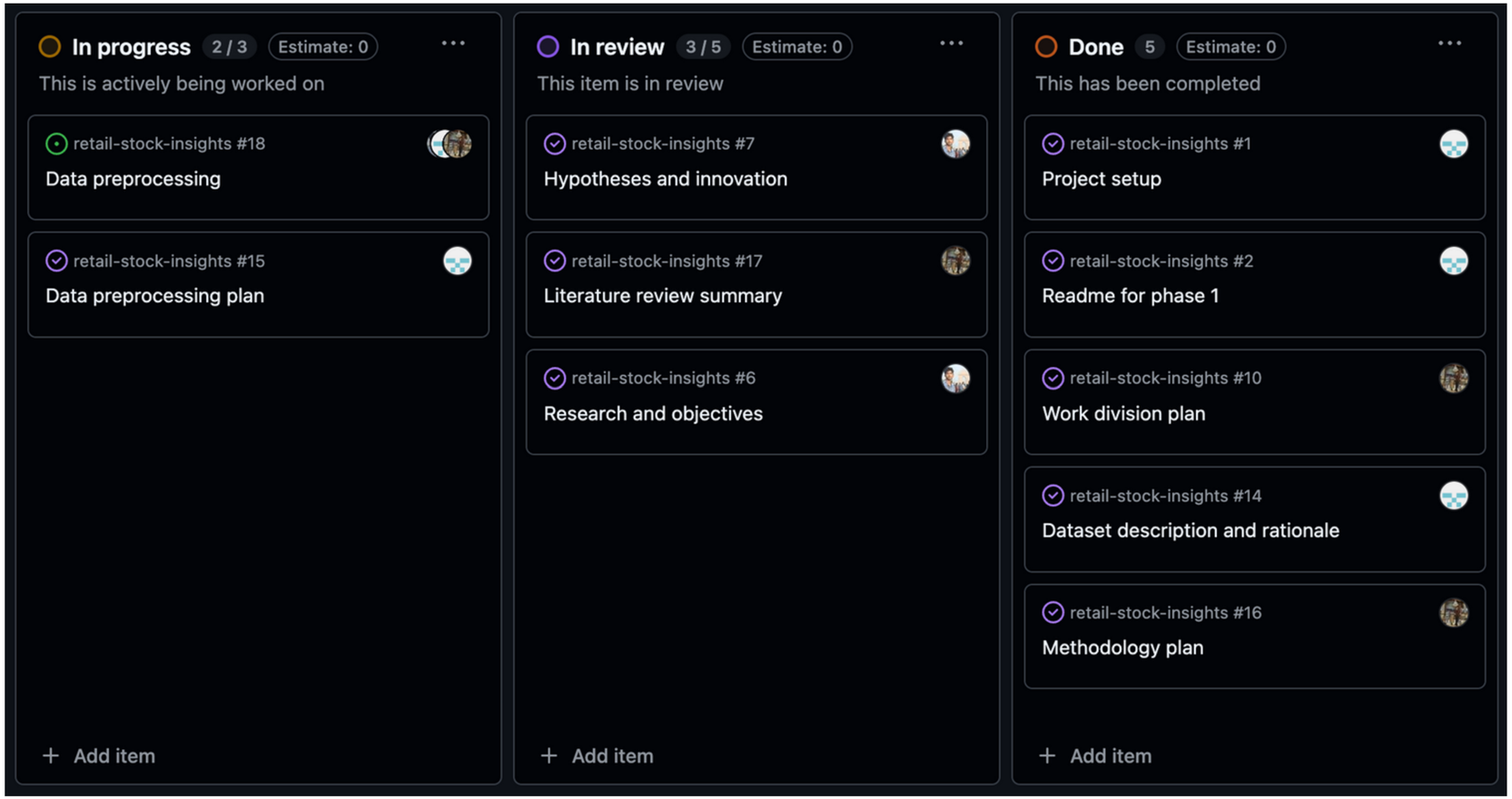
The primary objective of this plan is to ensure:

- Equal technical and documentation contributions across all members.
- Clear accountability through leadership rotation.
- Systematic workflow aligned with data-mining process stages.

WORK PLANNING & DIVISION

Phase Leadership and Workflow Management:

In each phase, the designated Phase Lead will be responsible for reviewing the Pull Requests (PRs) raised by team members. Task assignments, progress tracking, and updates will be managed using the Kanban board, ensuring organized workflow and clear accountability throughout the project lifecycle.



RESEARCH QUESTIONS & OBJECTIVES

Product Popularity & Revenue

- Which categories and prices lead to higher sales?
- What factors predict high-value orders?

Temporal Pattens & Seasonability

- How do sales vary by season, day, or time?
- Which periods show peak or low customer activity?

Customer Segmentation & Behaviour

- Can we cluster buyers (frequent, seasonal, bargain)?
- What predicts customer lifetime value or churn?

Association & Cross-Selling

- Which products are frequently bought together?
- What bundling strategies increase sales?

Predictive Modeling

- Which features affect transaction value most?
- Which model performs best — Random Forest, XGBoost, etc.?

Geographic & Market Growth

- Which countries contribute most to revenue?
- Can we group regions with similar buying patterns?

KEY HYPOTHESES

Area	Hypothesis	Testing Approach
Product Performance	**H1:** Higher prices lead to fewer purchases.	Compare price vs. quantity sold across products.
	H2: Product features can predict high-value (£500+) orders.	Use ML classification to identify large orders.
	H3: 20% of products drive 80% of revenue.	Rank products by sales; test Pareto pattern.
Temporal Patterns	**H4:** Sales peak during Q4 (holiday season).	Analyze monthly revenue trends.
	H5: Weekday vs. weekend sales differ.	Compare average order values by day type.
	H6: Most purchases occur in daytime hours.	Plot hourly order counts from timestamps.

ADVANCED HYPOTHESES

Area	Hypothesis	Testing Approach
Customer Segmentation	**H7:** Customers can be grouped by RFM traits.	Cluster using Recency, Frequency, and Monetary scores.
	H8: Buying behavior varies by country.	Compare order value and frequency across top countries.
Association & Cross-Selling	**H9:** Some products are often bought together.	Apply Apriori or FP-Growth to find product pairs.
Predictive Modeling	**H10:** Price & quantity drive order value.	Train regression/tree models; check feature importance.
	H11: Customer purchase trends can predict future sales.	Use past transaction data to forecast upcoming demand.
Market Expansion	**H12:** Few countries contribute most revenue.	Rank revenue by country to find key markets.

RESEARCH METHODOLOGY PLAN

PHASE	FOCUS AREA	KEY METHODS / TOOLS	EXPECTED OUTCOME
Phase 1	Literature Review & Data Preprocessing	Review retail analytics studies; handle missing/outlier data; feature engineering (RFM, TotalSales, time features).	Clean, structured dataset and theoretical foundation.
Phase 2	Exploratory Analysis & Modeling	Descriptive statistics, trend & pattern detection; K-Means (Segmentation), Apriori/FP-Growth (Association), XGBoost & Random Forest (Prediction).	Early data insights and preliminary predictive models.
Phase 3	Visualization & Evaluation	Plotly, Dash/Streamlit dashboards; SHAP & feature-importance explainability; model validation & performance review.	Interactive visuals, validated results, and business insights.

DATA PREPROCESSING SUMMARY

Step	Action Taken	Purpose / Outcome
Data Import & Inspection	Loaded dataset (541,909 rows, 8 columns)	Verified structure and missing values
Duplicate Removal	Removed 5,268 duplicates	Ensured unique, clean records
Missing Value Handling	Dropped rows with missing CustomerID (135,037)	Prepared for customer-level analysis
Invalid Entries	Removed negative / zero Quantity and UnitPrice (8,912 rows)	Ensured valid transactions
Feature Engineering	Added TotalPrice, Year, Month, DayOfWeek, Hour	Enabled trend & revenue analysis
Outlier Removal (IQR)	Filtered extreme values (≈ 59k rows removed)	Improved data consistency
Standardization	Cleaned Country names (37 unique)	Ensured uniform grouping and analysis

LINK : Data Preprocessing Notebook

EXPLORATION & NEXT PREPROCESSING STEPS

Early Insights from Exploration	Next Steps in Preprocessing
Country Trends: UK dominates sales volume	**Feature Aggregation (RFM):** Build Recency, Frequency, Monetary features per customer
Top Products: Few items generate most revenue	**Categorical Encoding:** Convert Country and top StockCode values to numerical form
Top Customers: Identify high-value buyers for targeting	**Scaling & Normalization:** Apply StandardScaler / MinMaxScaler for model readiness
Monthly & Hourly Patterns: Peak in Q4 and office hours	**Final Data Cleanup:** Remove non-product entries (e.g., ‘POSTAGE’, ‘DOT’)

Thankyou

~ Team DATA SCOUTS