

Project Summary – EDA Report

Dataset Overview

- The dataset appears to be a **retail transaction dataset**, containing fields such as InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, and Country.
 - Primary focus is on customer transactions across different regions (e.g., United Kingdom, Iceland, Finland, Italy).
-

Data Cleaning

- **Missing Values:**
 - CustomerID: ~135,000 missing values
 - Description: ~1,450 missing entries
 - **Action Taken:** These were likely removed or imputed depending on relevance.
 - **Duplicates** and **null entries** were assessed and handled.
 - Columns were also standardized and cleaned (date formats, text casing, etc.).
-

Key Analyses and Visuals

- **Country-wise Customer Analysis:**
 - Breakdown of invoice counts by customer and country using groupby aggregations.
 - **Sales Activity by Day:**
 - Bar plot showing the number of transactions by day of the week.
 - **Monthly Frequency:**
 - Visualization of transaction frequency across different months from Dec 2010 to Dec 2011.
-

Insights

- The **United Kingdom** is the dominant source of transactions.

- **Most invoices** are generated on **Wednesdays and Thursdays**, indicating mid-week sales peaks.
- **December 2010 to December 2011** was the analysis period, with notable spikes during certain promotional months.