Online Retail Data Analytics Project

A summer internship project focused on leveraging data analytics to extract meaningful insights from online retail data.

Project Overview

This project applied Data Analytics techniques using SQL Server Management Studio (SSMS), Python, and Excel. The primary goal was to analyze the online_retail_II_org dataset, clean and preprocess the data, and extract meaningful insights.

Dataset Snapshot

536,642

Rows

Total number of transactions recorded.

8

Columns

Key attributes for each transaction.

This dataset provides a comprehensive view of online retail operations, enabling detailed analysis of sales patterns and customer behavior.

Key Data Attributes

Invoice

Unique transaction identifier.

StockCode

Unique product code.

Description

Product details.

Quantity

Units purchased/returned.

Price

Unit price in GBP.

Customer ID

Unique customer identifier.

Country

Customer's country.

Tools Utilized

Python

For advanced analysis, preprocessing, and visualization.

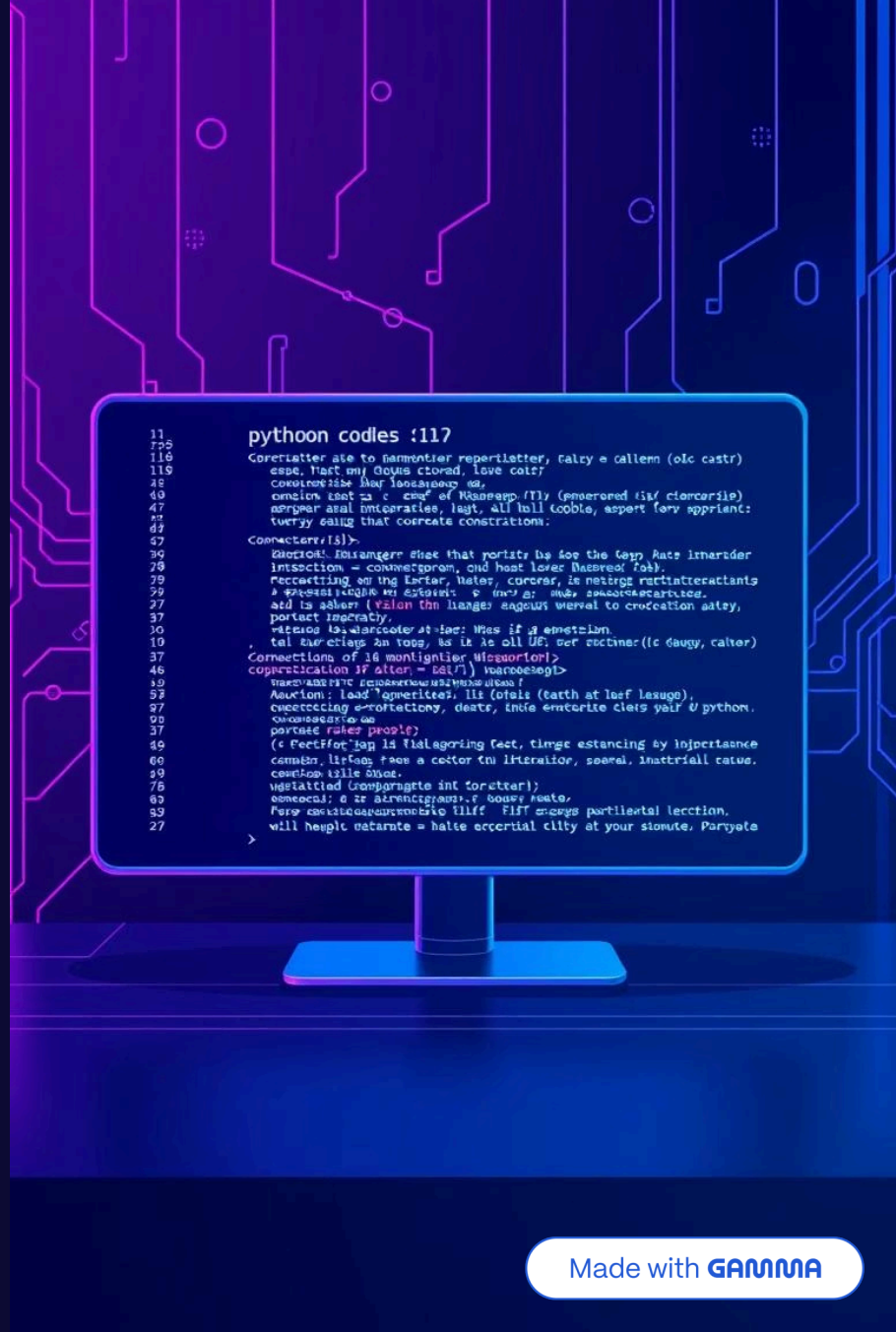
Python Analysis: Libraries & Loading

Imported Libraries

Pandas, NumPy, and Matplotlib were essential for data manipulation and visualization.

Dataset Loading

The dataset was efficiently loaded using `pd.read_excel()`.



Python Analysis: Data Cleaning

01

Missing Values

Cleaned using `dropna()` and `fillna()` to ensure data integrity.

02

Category Replacement

Missing categories/ingredients were replaced with 'Unknown' for consistent analysis.

Python Analysis: Grouping & Summarization

Grouping operations were performed to summarize sales data by various attributes.

Sales by Size

Aggregated sales data based on product size.

Sales by Category

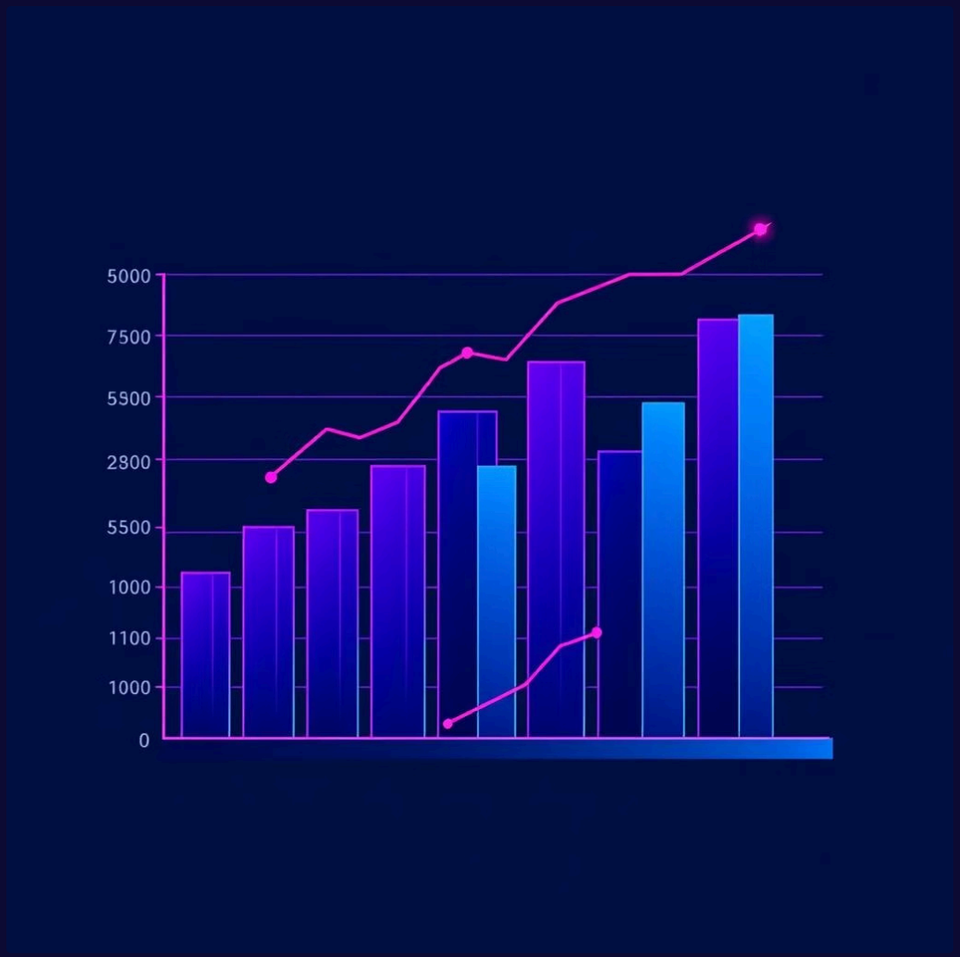
Summarized sales performance across different product categories.

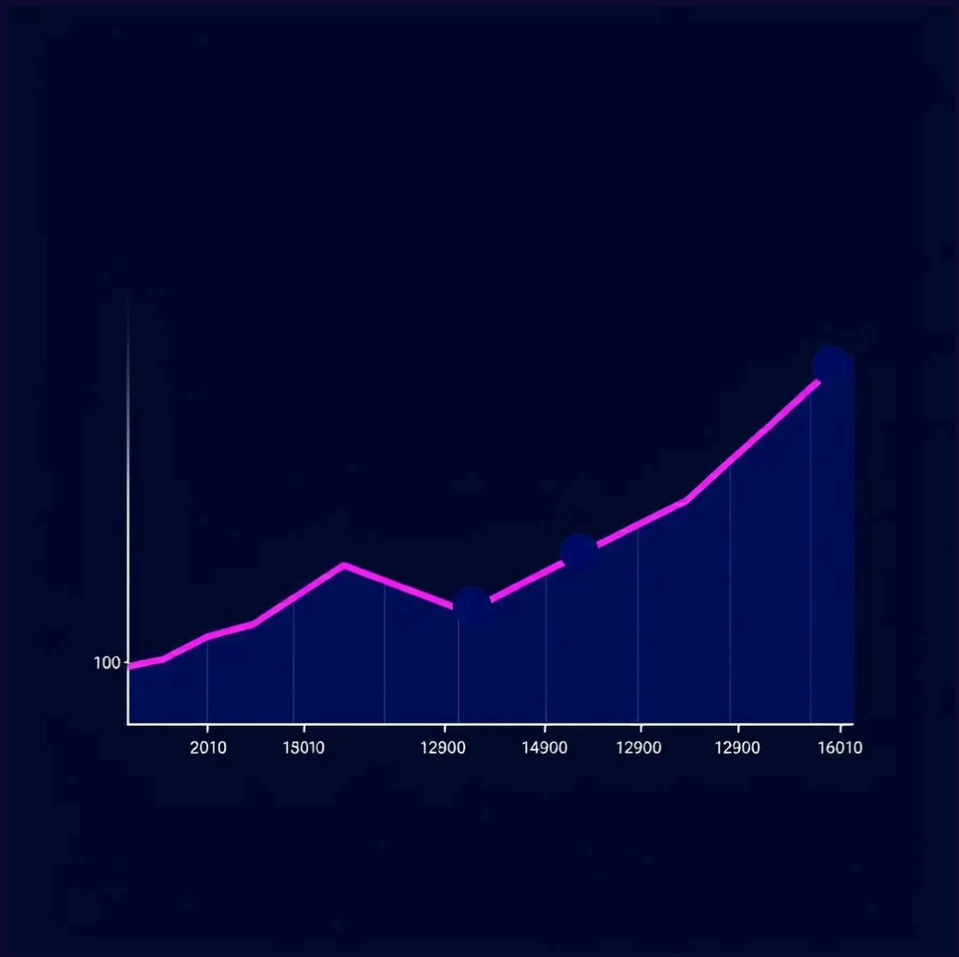
Sales by Ingredients

Analyzed sales patterns related to product ingredients.

Python Analysis: Visualizations

Bar and line charts were created to visualize sales patterns and trends.





Bar Charts

Effective for comparing sales across different groups.

Line Charts

Ideal for illustrating sales trends and changes over time.

Key Takeaways

Data-Driven Decisions

The project demonstrated how data analytics can inform strategic business decisions.

Comprehensive Toolset

Effective use of SQL, Python, and Excel for a holistic analysis approach.

Actionable Insights

Extracted meaningful patterns to optimize online retail operations.