E-Commerce Sales Forecasting Demo

By: Aidan Szwec

Table of Contents

[1. Executive Summary 3](#_Toc156916006)

[1.1 Objective of the Report 3](#_Toc156916007)

[1.2 Literature/ Sources 4](#_Toc156916008)

[1.3 Conclusion 4](#_Toc156916009)

[2. Data Preparation 5](#_Toc156916010)

[2.1 Data Preparation and Cleaning 5](#_Toc156916011)

[2.2 Data Sources and Overview 5](#_Toc156916012)

[2.3 Data Cleaning and Preprocessing: 6](#_Toc156916013)

[3. Data Exploration 7](#_Toc156916014)

[3.1 Objective of Data Exploration 7](#_Toc156916015)

[3.2 Objective of Data Exploration 7](#_Toc156916016)

[3.2.1 Missing Values: 7](#_Toc156916017)

[3.2.2 Time Period Analysis: 8](#_Toc156916018)

[3.2.3 Invoice Number Analysis: 8](#_Toc156916019)

[3.2.4 Stockcodes Analysis: 8](#_Toc156916020)

[3.2.5 Descriptions Analysis: 8](#_Toc156916021)

[3.2.6 Customers Analysis: 9](#_Toc156916022)

[3.2.7 Countries Analysis: 9](#_Toc156916023)

[3.2.8 Unit Price Analysis: 9](#_Toc156916024)

[3.2.9 Quantities Analysis 10](#_Toc156916025)

[3.2.10 Revenues Analysis: 10](#_Toc156916026)

[3.2.11 Conclusion of EDA: 10](#_Toc156916027)

[4. Data Forecast/Modeling 11](#_Toc156916028)

[4.1 Objective of Model Building: 11](#_Toc156916029)

[4.2 Forecasting Methodology and Approach: 11](#_Toc156916030)

[4.3 Model Evaluation: 11](#_Toc156916031)

[4.4 Interpretability and Feature Importance: 12](#_Toc156916032)

[4.5 Conclusion of Model Building and Forecasting: 12](#_Toc156916033)

[4.6 Conclusion Driving Business Impact through Informed Decision-Making 13](#_Toc156916034)

[5. Dashboard 14](#_Toc156916035)

[5.1 Dashboard Implementation 14](#_Toc156916036)

# Executive Summary

In this comprehensive e-commerce analytics and forecasting project, we delved into optimizing warehouse operations and predicting daily product sales for an online retailer. The analysis revolves around leveraging machine learning, exploring data insights, and delivering actionable results. Key findings and solutions include:

1. Warehouse Optimization:

- Identified opportunities to streamline storage and delivery processes.

- Proposed strategies for cost-effective warehouse management.

2. Sales Forecasting Model:

- Utilized CatBoost Regressor for accurate and granular daily sales predictions.

- Explored innovative features to enhance forecasting precision.

3. Data Exploration Highlights:

- Analysed customer behavior, product popularity, and country-wise transaction patterns.

- Uncovered insights into stock codes, descriptions, and unit prices.

4. Dashboard and Visualization:

- Demonstrated the power of analytics through an interactive dashboard.

- Visualized key metrics, trends, and predictions for user-friendly insights.

5. Business Impact:

- Translated analytics into actionable strategies for inventory planning and management.

- Showcased the potential for informed decision-making and competitive advantage.

### 1.1 Objective of the Report

The objective of this report is to delve into the sales data of an online retailer with the aim of optimizing warehouse operations and gaining actionable insights. The primary focus is on developing a predictive model to forecast daily product sales accurately, enabling effective inventory management. Through in-depth data exploration, the report aims to uncover patterns in customer behavior, product popularity, and transaction dynamics. Additionally, the goal is to showcase the analytical capabilities through the demonstration of machine learning techniques. The creation of a user-friendly dashboard serves to visually represent key metrics, trends, and predictions, providing stakeholders with an intuitive tool for decision-making. Ultimately, the report endeavors to translate analytical findings into practical recommendations for warehouse optimization and strategic planning, contributing to enhanced business efficiency and growth.

### 1.2 Literature/ Sources

Kaggle Link/ Credit: <https://www.kaggle.com/code/allunia/e-commerce-sales-forecast>  
Data Link: <https://www.kaggle.com/datasets/carrie1/ecommerce-data>

Snapshot:

A table with numbers and a list of items

Description automatically generated

### 1.3 Conclusion

In conclusion, this report extensively explores the analysis of an online retailer's sales data with a focus on the UK market. The objective is to optimize warehouse operations and enhance predictive capabilities using regression tools in the realm of machine learning and AI modeling. By harnessing advanced techniques, the report demonstrates the importance of predictive analytics in understanding and forecasting sales trends. The insights derived from this analysis empower businesses to make informed decisions, improve operational efficiency, and strategically plan for future success in a dynamic retail landscape.

# Data Preparation

The introduction to the Jupyter Notebook forecast provides a detailed exploration of sales data from a UK online retailer. This comprehensive analysis aims to uncover valuable insights, optimize warehouse operations, and predict daily quantities of sold products. The forecast utilizes advanced machine learning and AI modeling techniques, with a primary focus on regression tools. Through this introduction, readers will gain a clear understanding of the objectives and significance of the Jupyter Notebook forecast in enhancing business decision-making and strategic planning in the competitive retail landscape.

### 2.1 Data Preparation and Cleaning

This section focuses on the meticulous process of preparing and cleaning the dataset for effective analysis. Key steps include:

- Initial Assessment: Understand the dataset's structure and variables.

- Handling Missing Values: Implement strategies to address incomplete data.

- Time Period Analysis: Identify patterns, trends, and seasonality in sales data.

- Invoice Number Examination: Ensure accuracy and consistency.

- Stockcodes and Descriptions Validation: Confirm accuracy for reliable data.

- Customer and Country Insights: Explore transaction patterns and identify key markets.

- Unit Price and Quantities Analysis: Understand patterns and anomalies for accurate forecasting.

- Revenue Examination: Analyze factors influencing revenue.

-Conclusion of Data Exploration: Summarize findings, setting the stage for further analysis.

### 2.2 Data Sources and Overview

This section provides a detailed exposition of the data sources and offers a comprehensive overview of the dataset. The primary focus is on establishing the provenance and reliability of the data used in the analysis. The dataset, obtained from Kaggle (with credit to the contributor [insert Kaggle link]), serves as the foundational cornerstone for the subsequent analytical processes.

Dataset Origin: The dataset originates from [Kaggle](<https://www.kaggle.com>) a reputable platform for sharing and discovering datasets. Acknowledgment and credit are extended to the original contributor, fostering transparency and credibility.

Kaggle Link/ Credit: <https://www.kaggle.com/code/allunia/e-commerce-sales-forecast>  
Data Link: <https://www.kaggle.com/datasets/carrie1/ecommerce-data>

Data Integrity and Relevance: A critical assessment of the dataset's integrity is conducted to ensure the reliability of the information. This involves verifying the consistency, accuracy, and completeness of the data, laying the groundwork for robust analytical outcomes.

Scope and Granularity: The section outlines the scope and granularity of the dataset, delineating the specific variables and attributes encompassed. This includes but is not limited to InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, and Country.

Data Exploration Significance: The overview sets the stage for subsequent data exploration endeavors, emphasizing the strategic importance of the dataset in unraveling patterns, trends, and insights crucial for warehouse optimization and sales forecasting.

This detailed exploration of data sources and overview ensures a transparent and informed foundation for the subsequent phases of the analysis, contributing to the credibility and efficacy of the overall project.

### 2.3 Data Cleaning and Preprocessing:

Ensuring the cleanliness and reliability of our dataset is paramount. The initial phase of our data preparation involves meticulous cleaning and preprocessing steps.

Missing values can be stumbling blocks in our quest for accurate insights. How would you like us to handle missing data? Are there specific variables where missing values might carry more significance, guiding us to adopt tailored strategies?

The cleanliness of the dataset hinges on addressing outliers and anomalies. What criteria should we use to identify and handle outliers, and do you have any particular concerns or expectations regarding their impact on our analysis?

Standardizing and formatting data is another crucial step. Are there specific data formats or standards we should adhere to, and how might these considerations align with the overarching objectives of our analysis?

Moreover, dealing with duplicates and ensuring data consistency are pivotal tasks. Do you have any specific preferences or guidelines for handling duplicate entries and maintaining data consistency throughout the dataset?

Understanding your preferences and priorities in these cleaning and preprocessing steps will guide us in creating a robust foundation for subsequent analyses. Your insights into these aspects will contribute significantly to the reliability and integrity of our findings.

A screenshot of a computer

Description automatically generated

# Data Exploration

This section delves into the depths of the sales dataset, aiming to uncover valuable insights that lay beneath the surface. Through exploratory data analysis (EDA), we navigate the intricacies of customer behavior, product dynamics, and transaction patterns. The primary goal is to unearth patterns that will serve as a cornerstone for building a precise sales forecasting model. Each exploration is driven by a quest for actionable insights, contributing to a comprehensive understanding of the dataset and facilitating strategic decision-making.

### 3.1 Objective of Data Exploration

The objective of data exploration is to unravel the hidden patterns, trends, and insights within the sales data. This phase goes beyond the surface-level understanding, aiming to provide a deeper comprehension of the dataset's characteristics. Through exploratory data analysis (EDA), our goal is to extract meaningful information that will inform strategic decision-making.

As we embark on the journey of data exploration, we seek to uncover patterns in customer behavior, product popularity, and transaction dynamics. The insights gained will not only enhance our understanding of the dataset but will also serve as a foundation for building a robust sales forecasting model.

Your guidance on specific areas of interest or hypotheses to investigate during the EDA process will be instrumental. Are there particular aspects of customer behavior, product preferences, or regional variations that you believe could significantly impact sales? Your insights will steer the direction of our exploratory analysis, ensuring that it aligns closely with your business objectives.

Ultimately, the objective of data exploration is to transform raw data into actionable insights. By identifying key trends and patterns, we aim to provide a solid foundation for the subsequent phases of the project, leading to more informed decision-making and strategic planning.

### 3.2 Objective of Data Exploration

The Exploratory Data Analysis (EDA) phase is a pivotal step in deciphering the nuances of the sales dataset, providing essential insights for strategic decision-making. Each sub-section within EDA delves into specific aspects, unraveling key patterns and trends that offer valuable business intelligence.

### 3.2.1 Missing Values:

Understanding and addressing missing values is crucial for data completeness. This analysis identifies which fields lack data, enabling informed decisions on imputation or exclusion and ensuring the integrity of subsequent analyses.

### 3.2.2 Time Period Analysis:

Analysing temporal trends helps identify seasonality, peak sales periods, and overall business cycles. This insight is instrumental in planning marketing strategies, stock management, and resource allocation.

### 3.2.3 Invoice Number Analysis:

Examining invoice numbers provides a sequential understanding of transactions. This aids in uncovering purchasing patterns, order frequency, and customer loyalty, offering valuable information for targeted marketing.

### 3.2.4 Stockcodes Analysis:

Understanding the popularity of stockcodes sheds light on customer preferences and trending products. This information aids inventory management, ensuring that popular items are well-stocked to meet demand.

A graph with a number of data

Description automatically generatedA graph of a number of bars

Description automatically generated with medium confidence

### 3.2.5 Descriptions Analysis:

Analysing product descriptions provides insights into the characteristics and appeal of items. This knowledge can guide marketing efforts, helping to highlight product features that resonate with customers.

A graph of blue and white bars

Description automatically generatedA graph of a bar graph

Description automatically generated with medium confidence

### 3.2.6 Customers Analysis:

A comprehensive examination of customer-related patterns includes customer behavior, preferences, and loyalty. This information is essential for personalized marketing, customer retention, and enhancing the overall shopping experience.

A graph of a bar graph

Description automatically generated with medium confidence

### 3.2.7 Countries Analysis:

Analysing transactional patterns across different countries identifies global market dynamics. This insight informs international expansion strategies, tailored marketing campaigns, and localization efforts.

A graph of green squares

Description automatically generated

### 3.2.8 Unit Price Analysis:

Scrutinizing unit prices reveals pricing structures and variations. This information is critical for pricing strategies, identifying premium products, and understanding customer sensitivity to price changes.

A graph of numbers and lines

Description automatically generated with medium confidence

### 3.2.9 Quantities Analysis

Examining quantities sold provides insights into product demand and popularity. This knowledge guides inventory management, ensuring adequate stock levels for high-demand items.

A graph of a graph of a graph

Description automatically generated with medium confidence

A graph of a line

Description automatically generated with medium confidence

### 3.2.10 Revenues Analysis:

In-depth revenue analysis uncovers patterns and fluctuations in financial metrics. This understanding is essential for financial planning, identifying revenue drivers, and optimizing sales strategies.

### 3.2.11 Conclusion of EDA:

Summarizing the EDA journey, this section highlights key findings and actionable insights. The information gained from EDA serves as the foundation for subsequent model building, empowering the business to make data-driven decisions for growth and efficiency.

# Data Forecast/Modeling

This section focuses on constructing a robust predictive model to forecast daily product sales, leveraging advanced machine learning techniques. The objective is to provide the business with a reliable tool for anticipating market demands, optimizing inventory management, and enhancing strategic decision-making.

### 4.1 Objective of Model Building:

The primary goal of model building is to develop an accurate and interpretable forecasting tool. This involves selecting suitable algorithms, defining relevant features, and implementing a robust validation strategy to ensure the model's efficacy.

### 4.2 Forecasting Methodology and Approach:

The forecasting methodology centers around predicting daily product sales based on historical data. The process involves Exploratory Data Analysis (EDA), data preprocessing, and feature engineering. The chosen model, CatBoost Regressor, is explained for its suitability in regression tasks. A sliding window time series approach is employed to address challenges posed by significant sales increases during specific periods.

### 4.3 Model Evaluation:

The performance of the model is assessed using metrics such as Root Mean Square Error (RMSE). This evaluation provides insights into the model's accuracy and precision in predicting sales quantities. Visualization of model predictions against actual sales further enhances the understanding of its efficacy.

### 4.4 Interpretability and Feature Importance:

Understanding the factors influencing the model's predictions is crucial for interpretability. Shapley values are explored to determine feature importance, offering insights into which variables have the most significant impact on sales forecasts. This transparency aids stakeholders in trusting and utilizing the model effectively.

A graph with numbers and a bar

Description automatically generated

A graph and chart with orange lines

Description automatically generated

### 4.5 Conclusion of Model Building and Forecasting:

This section concludes the model-building process, summarizing key findings, challenges, and successes. It emphasizes the practical application of the forecasting model in improving inventory management, strategic planning, and overall business efficiency. The insights derived from the model pave the way for data-driven decision-making, providing a competitive edge in the dynamic retail landscape.

A graph showing the growth of a graph

Description automatically generated with medium confidence

### 4.6 Conclusion Driving Business Impact through Informed Decision-Making

In this comprehensive endeavor, we embarked on a journey to optimize warehouse operations and enhance the predictive capabilities of an online retailer through advanced machine learning and AI modeling. The synergy between meticulous data preparation, insightful exploration, and robust model building culminated in actionable insights that can significantly benefit the business.

Key Takeaways:

1. Strategic Forecasting: The developed forecasting model, anchored by the CatBoost Regressor, provides a granular understanding of daily product sales. This empowers the business to anticipate demand patterns and align inventory management strategies accordingly.

2. Operational Efficiency: The insights derived from exploratory data analysis offer a holistic view of customer behavior, product popularity, and transaction dynamics. This newfound understanding serves as a compass for streamlining warehouse operations, optimizing storage, and enhancing overall operational efficiency.

3. Informed Decision-Making: The model's interpretability, backed by Shapley values, ensures transparency and instills confidence in the decision-making process. Stakeholders can easily grasp the factors influencing sales predictions, fostering a data-driven culture within the organization.

4. Competitive Advantage: By leveraging machine learning techniques, businesses gain a competitive edge in a dynamic retail landscape. Informed decisions based on accurate sales forecasts contribute to adaptive strategies, positioning the company ahead of market trends and competitors.

Business Impact:

- Inventory Planning: Accurate sales predictions enable proactive inventory planning, minimizing stockouts and overstock situations. This directly translates to cost savings and improved customer satisfaction.

- Cost-Effective Warehouse Management: Streamlined warehouse operations, guided by data-driven insights, pave the way for cost-effective management. From storage optimization to efficient order fulfillment, every aspect contributes to resource utilization.

- Enhanced Customer Experience: Understanding customer preferences and behaviors allows for personalized marketing strategies and a tailored product offering. This enhances the overall customer experience, fostering loyalty and repeat business.

In conclusion, this integrated approach to data analysis and forecasting not only unlocks the potential for operational excellence but also positions the business strategically in the market. The culmination of data-driven insights equips stakeholders with the tools needed to navigate uncertainties, adapt to changing demands, and ultimately thrive in the ever-evolving landscape of e-commerce.

# Dashboard

### 5.1 Dashboard Implementation

(Coming Soon)