WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Sales Analysis and Forecasting for Coffee Sales Using SQL

Deriya Pritkumar Ramjibhai[1], Mr. Prashant Kothari[2]

[1] U.G. Student, Department of Computer Science and Engineering, Parul University, Vadodara, Gujarat, India
[2] Assistant Professor, Department of Computer Science and Engineering, Parul University, Vadodara, Gujarat, India
Corresponding Authors:
[1] 210305124023@paruluniversity.ac.in
[2] Prashant.kothari36174@paruluniversity.ac.in

ABSTRACT

This study analyzes six months of coffee sales data to uncover trends, predict future sales, and improve decision-making. The dataset includes transaction details such as date, product category, payment method, store location, and order quantity. SQL was used for data extraction, preprocessing, analysis, and visualization. The analysis highlights peak sales periods, customer buying patterns, and the impact of promotional campaigns. SQL-transformed data was used to forecast a 10% increase in sales over the next quarter. Findings support data-driven strategies to enhance revenue, improve customer retention, and optimize marketing efforts. Additionally, the study explores seasonal trends and identifies actionable insights for business growth.

Introduction

Background

Coffee shops and retail outlets generate a large amount of transaction data daily. Analyzing historical coffee sales data can help identify peak sales periods, customer preferences, and the impact of promotional activities. Data-driven insights can help optimize inventory management, forecast future demand, and enhance customer satisfaction.

Objective

The primary objective is to analyze past coffee sales data and forecast future trends using SQL. The study aims to:

- Identify peak sales periods and product categories contributing significantly to revenue.
- Analyze customer preferences based on payment methods and store locations.
- Build a forecasting model using SQL-based time series analysis.
- Evaluate the impact of seasonal variations on sales trends.

Literature Review

Previous Studies

- Previous studies focused on customer segmentation and predictive modeling to optimize sales.
- Time series models like ARIMA and Prophet have been widely used for sales forecasting.
- SQL-based techniques have been highlighted as efficient for data extraction, transformation, and querying of large datasets.

Theoretical Framework

- Exploratory Data Analysis (EDA): Uncovers hidden patterns, outliers, and correlations.
- Forecasting Models: SQL-based models analyze historical data to predict future trends.
- Data Cleaning Techniques: SQL ensures data quality by handling missing values, duplicates, and inconsistencies.

Research Gap

Most studies focus on predictive modeling using machine learning techniques or visualization tools. This study bridges the gap by leveraging only SQL for end-to-end data analysis and forecasting, demonstrating the potential of SQL in data-driven decision-making.

Data and Methodology

Dataset Description

- **Source:** Synthetic coffee sales data
- **Time Period:** Six months (January to June 2024)
- **Key Features:** Transaction ID, Date, Product Category, Price, Customer Demographics (age, gender, location), Payment Method, and Order Quantity

Data Preprocessing

- **Handling Missing Data:** Missing values were addressed using SQL queries to impute or remove incomplete records.

```
DELETE FROM COFFEE
WHERE PRODUCT_CATEGORY IS NULL OR PAYMENT_METHOD IS NULL;
```

- **Data Cleaning:** Duplicates were removed, and inconsistent formats were corrected using Excel.

```
-- REMOVING DUPLICATES BASED ON TRANSACTION ID
DELETE FROM COFFEE
WHERE TRANSACTION_ID NOT IN (
  SELECT MIN(TRANSACTION_ID)
  FROM COFFEE
  GROUP BY TRANSACTION_DATE, PRODUCT_CATEGORY, CUSTOMER_ID);
```

- **Data Transformation:** Aggregations, joins, and subqueries were used to prepare the data for analysis.

```
-- AGGREGATING SALES BY PRODUCT CATEGORY
SELECT PRODUCT_CATEGORY, SUM(UNIT_PRICE * TRANSACTION_QTY) AS TOTAL_REVENUE
FROM COFFEE
GROUP BY PRODUCT_CATEGORY;
```

Exploratory Data Analysis (EDA)

- **Descriptive Statistics:** Calculated mean, median, mode, and standard deviation to summarize data.

```
-- DESCRIPTIVE STATISTICS FOR SALES DATA
SELECT
  AVG(PRICE * ORDER_QUANTITY) AS AVG_REVENUE,
  MAX(PRICE * ORDER_QUANTITY) AS MAX_REVENUE,
  MIN(PRICE * ORDER_QUANTITY) AS MIN_REVENUE,
  STDDEV(PRICE * ORDER_QUANTITY) AS REVENUE_STDDEV
FROM COFFEE;
```

- **Trend Analysis:** Identified sales trends over time using SQL aggregate functions.

```
-- MONTHLY SALES TREND ANALYSIS
SELECT
  DATE_FORMAT(TRANSACTION_DATE, '%Y-%M') AS MONTH,
  SUM(PRICE * ORDER_QUANTITY) AS MONTHLY_SALES
FROM COFFEE
GROUP BY MONTH
ORDER BY MONTH;
```

- **Correlation Analysis:** Examined relationships between customer demographics and purchasing behavior.

```
-- CORRELATION BETWEEN STORE LOCATION AND REVENUE
SELECT
  STORE_LOCATION,
  SUM(UNIT_PRICE * TRANSACTION_QTY) AS TOTAL_REVENUE
FROM COFFEE
GROUP BY STORE_LOCATION
ORDER BY TOTAL_REVENUE DESC;
```

Results And Discussion

Key Insights

- Coffee and Tea contributed 67% of total revenue.
- Highest sales were observed during weekends, festival seasons, and promotional campaigns.

Implications

- Marketing Campaigns: Focus marketing campaigns on top-selling categories and leverage promotional periods to maximize revenue.
- Inventory Management: Optimize inventory levels by forecasting demand accurately to prevent stockouts and overstocking.
- Customer Retention: Enhance customer experience by providing personalized offers and loyalty rewards.

Conclusion

Summary of Key Findings

The analysis highlights key trends in customer purchasing behavior and predicts future sales with high accuracy. Insights gained from the study can guide inventory management, marketing, and operational strategies. Leveraging SQL ensures a comprehensive understanding of the data and provides actionable recommendations for business growth.

Future Scope

- Incorporate External Factors: Integrate economic indicators, competitor analysis, and market trends to improve forecasting models.
- Adopt Advanced Machine Learning Techniques: Explore algorithms such as XGBoost and LSTM for enhanced predictive capabilities.
- Expand Scope of Analysis: Include customer sentiment analysis and feedback to refine product offerings.

REFERENCES

1. Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice. https://otexts.com/fpp3/
2. Kaggle Dataset: E-commerce Sales Data (2024). https://www.kaggle.com/datasets
3. Microsoft Documentation: Power BI Best Practices (2024). https://learn.microsoft.com/en-us/power-bi/guidance/
4. SQL Data Analysis Guides – Online SQL Documentation. https://www.sqltutorial.org/
5. Industry Reports (2023). Digital Platform Retention Trends. Tech Insights. https://dev.mysql.com/doc/