Test Project Session 1

IT Software Solution for Business

Independent Test Project Designer: Ramin Mohammaddoust

Independent Test Project Validator: Afshin Dehghani

Introduction

Belle Croissant Lyonnais, a renowned bakery/patisserie in Lyon, has accumulated a wealth of data related to its sales, customers, and products. This data holds valuable insights that can drive informed decision-making and contribute to the bakery's continued success. In this session, you will leverage the tools available in your production environment to analyze this data, uncover meaningful patterns and trends, and present your findings in a clear and concise manner.

Your analysis will focus on several key areas:

- **Sales Performance:** Evaluate the bakery's overall sales performance, identifying top-performing products, peak sales periods, and potential areas for improvement.
- **Customer Behavior:** Understand customer preferences, purchasing habits, and demographics to tailor marketing strategies and enhance customer satisfaction.
- **Product Trends:** Analyze product popularity, identify seasonal fluctuations, and explore potential new product offerings.
- **Operational Efficiency:** Assess operational efficiency by examining factors such as order fulfillment time, inventory turnover, and staffing levels.

Your findings will be presented to the Belle Croissant Lyonnais management team, providing them with actionable insights to optimize operations, increase sales, and enhance customer experiences.

Contents

This session package contains the following materials:

1. **Session Instructions (PDF):** Detailed instructions outlining the tasks to be completed and deliverables expected for this session.
2. **Sales Data (CSV Files):**

- sales_transactions.csv: Contains transaction-level data, including transaction ID, customer ID, date, product ID, quantity, and price.
- products.csv: Contains product information, including product ID, name, category, ingredients, price, cost, seasonal indicator, active status, and introduction date.
- customers.csv: Contains customer information, including customer ID, name, age, gender, postal code, email, phone number, membership status, join date, last purchase date, total spending, average order value, frequency, preferred category, and churn status.

3. **Data Dictionary (PDF):** Detailed descriptions of the data fields and their meanings in each CSV file are provided within the session instructions document.
4. **Common Folder:** This folder contains additional resources such as the Belle Croissant Lyonnais logo, icons, style guide, and other design assets that can be used throughout the development of the application.
5. **ARIMA Models (PDF):** A reference guide explaining the ARIMA (AutoRegressive Integrated Moving Average) model, its implementation, and evaluation for time series forecasting.

These materials provide all the necessary resources for competitors to successfully complete the Data Analysis and Reporting session.

Description of Project and Tasks

In this session, you will analyze data for Belle Croissant Lyonnais to gain insights into their operations, customers, and product performance.

Guidelines

1. **Easy to Use:** Present data and insights in a clear, understandable format.
2. **Looks Good:** Follow the Belle Croissant Lyonnais Style Guide for all visualizations and reports.
3. **Works Well:** Ensure all analyses and calculations are accurate and error-free.
4. **Secure:** Handle customer data confidentially and adhere to data privacy regulations.
5. **On Time:** Complete all tasks within the specified time limit.

Technical Considerations

1. **Data Cleaning:** Address missing values, inconsistencies, and formatting issues in the provided datasets.
2. **Data Analysis:** Apply appropriate statistical techniques for trend analysis, forecasting, and segmentation.
3. **Data Visualization:** Create clear and informative charts and tables to present findings.
4. **Modeling:** Implement time series forecasting, clustering, and other relevant algorithms.

Additional Considerations

- The analysis should be reproducible and well-documented.
- Use clear labels and explanations for all visualizations and tables.
- Organize information logically to facilitate easy understanding by stakeholders.

Instructions to the Competitor

1.1 Data Loading and Exploration

Objective

Demonstrate your ability to load, inspect, and understand the provided datasets, identifying potential data quality issues and preparing the data for further analysis.

Tasks

1. Load Data:

- Import the provided CSV files (sales_transactions.csv, products.csv, and customers.csv) into your chosen data analysis environment.

2. Initial Exploration:

- Display the first 5 rows of each DataFrame to showcase the structure and content.
- Determine the data types of each column and identify non-numeric columns.
- Check for missing values and inconsistencies in the data.

Deliverables:

- **File Name:** Session1_DataExploration.txt
- Provide the following information for each of the three CSV files:

- Data types of each column
- Inconsistencies and Anomalies:
  - **Invalid Dates:** Number of rows with dates outside the expected range (e.g., "2023-14-01").
  - **Negative Values: Number of rows with negative quantities or prices.**
  - **Invalid IDs: Number of rows with product IDs or customer IDs that don't exist in the corresponding files.**
  - **Unexpected Values: Number of rows with unexpected values in categorical columns in relation to the data dictionary provided.**
  - **Formatting Issues: Number of rows with extra spaces or inconsistent formatting in relevant columns in re**lation to the data dictionary provided.

1.2 Data Cleaning and Transformation

Objective

Demonstrate your ability to clean, transform, and standardize data to ensure accuracy, consistency, and suitability for analysis.

Tasks

1. Missing Values:

- Fill in missing values in the age column of the customers.csv file with the median age.
- Fill in missing values in the phone_number column of the customers.csv file with '0'.
- Fill in missing values in the promotion_id column of the sales_transactions.csv file with '0'.

2. Data Type Conversion:

- Convert the date columns in both sales_transactions.csv and customers.csv to datetime data type. For the time please add a random time between 9am to 5pm.

3. Data Standardization:

- Standardize phone numbers in the customers.csv file by removing all non-numeric characters except for + (spaces, dashes, parentheses).

Deliverables

1. **File Name:** customers_cleaned.csv

- File Type: CSV file (.csv)

2. **File Name:** sales_transactions_cleaned.csv

- **File** Type**:** CSV file (.csv)

1.3 Sales Trend Analysis

Objective

Calculate and visualize Belle Croissant Lyonnais's sales trends over time.

Tasks

1. Calculate total sales revenue, number of transactions, and average order value per month.
2. Create line charts for each of the three metrics over time (monthly).

3. Identify the top 3 months with the highest sales revenue and display them in a table.

Deliverables:

- **File Name:** Session1_SalesTrends.pdf
- **File Type:** PDF report containing:
  - Line chart: Total sales revenue per month
  - Line chart: Number of transactions per month
  - Line chart: Average order value per month
  - Table: Top 3 months by sales revenue (Month, Total Revenue)

1.4 Product Performance Analysis

Objective

Analyze and visualize the sales performance of Belle Croissant Lyonnais's products.

Tasks

1. Calculate total quantity sold and total revenue for each product.
2. Calculate profit margin for each product (Price - Cost).
3. Create a bar chart showing the total revenue for each product category ("Pastries", "Bread", "Tarte").
4. Create a table showing the top 3 best-selling products by quantity sold, including their names, total quantity, and total revenue.

Deliverables

- **File Name:** Session1_ProductPerformance.pdf
- **File Type:** PDF report containing:
  - Bar chart: Total revenue by product category
  - Table: Top 3 best-selling products (Product Name, Total Quantity Sold, Total Revenue)

1.5 Customer Analysis

Objective

Analyze and visualize customer demographics and loyalty program participation.

Tasks

1. Calculate and visualize the distribution of customer age groups (18-24, 25-34, 35-44, 45+) using a bar chart.
2. Calculate and display the distribution of customer gender ("M", "F") as percentages in a table.
3. Calculate and display the average spending per customer for each loyalty tier ("Basic", "Silver", "Gold") in a table.

Deliverables

- **File Name:** Session1_CustomerAnalysis.pdf
- **File Type:** PDF report containing:
  - Bar chart: Distribution of customer age groups
  - Table: Percentage distribution of customer gender

- Table: Average spending per loyalty tier

1.6 Time Series Forecasting

Objective

Predict Belle Croissant Lyonnais's daily total sales for the next 30 days using a time series forecasting model.

Tasks

1. Choose and implement an ARIMA model using the daily total sales data from sales_transactions_cleaned.csv.
2. Generate sales forecasts for the next 30 days.
3. Calculate the Mean Absolute Error (MAE) of the model.

Deliverables

- **File Name:** Session1_SalesForecast.csv
- **File Type:** CSV file (.csv)
- **Format:**
  - **Column 1:** Date (YYYY-MM-DD)
  - **Column 2:** Predicted_Sales (float)

1.7 Customer Segmentation and Recommendation

Objective

Demonstrate your ability to segment customers based on their purchasing behavior and develop a basic product recommendation system for Belle Croissant Lyonnais.

Tasks

1. Customer Segmentation:

- **Feature Engineering:** Create two new columns in the customers.csv file:
  - total_purchases: Calculate the total number of transactions for each customer.
  - avg_purchase_value: Calculate the average transaction value for each customer.
- **Clustering:** Using the total_purchases and avg_purchase_value columns, apply K-means clustering with 3 clusters to segment customers.

2. Recommendation Engine:

- **Product Affinity:** For each product, determine the top 3 other products most frequently purchased together in the same transaction.
- **Recommendations:** For each customer, recommend the top 3 products they haven't purchased yet, based on the products frequently purchased by other customers in their segment.

Deliverables

- **File Name:** Session5_Segmentation_and_Recommendations.csv
- **File Type:** CSV file (.csv)
- **Format:**

- **Column 1:** customer_id
- **Column 2:** cluster_label (1, 2, or 3)
- **Column 3:** recommended_product_1 (product_id)
- **Column 4:** recommended_product_2 (product_id)
- **Column 5:** recommended_product_3 (product_id)

1.8 Product Performance Analysis and Price Optimization

Objective

Demonstrate your ability to analyze product performance, identify pricing trends, and suggest data-driven price adjustments for Belle Croissant Lyonnais' products.

Tasks

1. Product Performance Analysis:

- **Sales Volume:** Calculate the total quantity sold and total revenue generated for each product. Sort the products by total revenue in descending order.
- **Profitability:** Calculate the profit margin (profit/revenue) for each product and sort the products by profit margin in descending order.
- **Sales Trends:** Analyze the sales trends of each product over time (monthly). Identify any seasonality or patterns in sales.

2. Price Analysis:

- **Price Sensitivity:** For each product, calculate the price elasticity of demand (PED). PED measures how sensitive the quantity demanded of a product is to changes in its price. Use the following formula to calculate PED:
- PED = (% change in quantity demanded) / (% change in price)
  You can use a simple percentage change calculation or a more sophisticated method like log-log regression.
- **Price Optimization:** Based on the calculated PED values and profit margins, suggest optimal price adjustments for each product. Consider the following guidelines:
  - If a product has high PED (elastic demand), a small price decrease could lead to a significant increase in sales volume and potentially higher overall revenue.
  - If a product has low PED (inelastic demand), a small price increase might not significantly impact sales volume and could lead to higher revenue.
  - Consider the product's profit margin when suggesting price adjustments. Aim to maximize profit while maintaining or increasing sales volume.

Deliverables

1. **File Name:** Session5_Product_Performance.csv

- **File Type:** CSV file (.csv)
- **Format:**
  - **Column 1:** product_id
  - **Column 2:** total_quantity_sold
  - **Column 3:** total_revenue
  - **Column 4: profit_margin (calculated as (total_revenue - total_cost) / total_revenue)**

2. **File Name:** Session5_Price_Analysis.csv

- **File Type:** CSV file (.csv)
- **Format:**
  - **Column 1:** product_id
  - **Column 2:** price_elasticity_of_demand
  - **Column 3:** suggested_price_change (in percentage, e.g., 5% increase or -3% decrease)

Additional Notes

- The price elasticity of demand can be calculated using various methods. Choose the method that you believe is most appropriate for the given data.
- Remember that price optimization is a complex process that involves multiple factors. Your suggestions should be based on the available data and your best judgment.

1.9 Customer Lifetime Value (CLTV) Calculation

Objective

Calculate CLTV for each customer.

Tasks

1. Calculate average purchase value per customer from sales_transactions_cleaned.csv.
2. Calculate purchase frequency (transactions per month) for each customer.
3. Calculate CLTV using the formula:

   CLTV = (Average Purchase Value) * (Purchase Frequency) * 36

Deliverable:

- **File Name:** Session1_CLTV.csv
- **File Type:** CSV file (.csv)
- **Format:**
  - **Column 1:** customer_id (int)
  - **Column 2:** cltv (float, rounded to 2 decimal places)

1.10 Churn Analysis

Objective

Analyze and compare CLTV for churned and active customers.

Tasks

1. Identify churned customers from customers_cleaned.csv.
2. Calculate the overall churn rate (percentage of churned customers).
3. Calculate average CLTV for churned and active customers separately.

Deliverable:

- **File Name:** Session1_Churn_Analysis.csv
- **File Type:** CSV file (.csv)
- **Format:**
  - **Column 1:** churn_rate (float, percentage rounded to 2 decimal places)

- **Column 2:** avg_cltv_churned (float, rounded to 2 decimal places)
- **Column 3:** avg_cltv_active (float, rounded to 2 decimal places)