Project Report: Coffee Shop Analysis

1. Title

Coffee Shop Analysis Using MYSQL

2. Introduction

This project aims to analyze the sales performance and customer transaction patterns at a coffee shop using MySQL for data management and analysis. The data includes transactions, sales, and customer behavior, which can provide valuable insights into factors such as the time of transactions, the total bill amount, and monthly sales trends. This documentation outlines the methodology used for the analysis, covering how data was collected, processed, and analyzed**.**

3. Objectives

-To identify the most profitable product categories and stores.

-To understand customer preferences and tailor product offerings accordingly.

-To optimize inventory management and pricing strategies.

-To identify opportunities for growth and expansion.

4. Methodology

The project will follow a structured approach:

1. **Data Collection and Understanding** : The data is extracted from the MySQL database containing products table, stores table and transactions table. SQL queries will be used to extract, aggregate, and analyze the data.
2. **Data Cleaning and Preprocessing** : Cleaning the data and Check for missing values Identify any missing or null values in important fields such as store_id, product_id, transaction_qty, unit_price, and total_bill. Ensure that numerical fields like transaction_qty, unit_price, and total_bill are in numeric formats. Convert date and time fields (transaction_date, transaction_time) into appropriate datetime formats.
3. **Exploratory Data Analysis (EDA)** :
    - **Data Visualization:**

      Bar Chart showing total sales per store

Bar Chart for the most popular products based on quantity sold.

Pie Chart for showing product categories sales.

Stacked Bar Chart showing sales by product category.

- **Correlation Analysis:** Correlation analysis explores relationships between numerical variables in the coffee shop dataset, such as transaction_qty (items sold), unit_price (price per item), total_bill (total transaction value), and transaction_hour.

4. Insights and Recommondation:

- Interpret the results from visualizations and statistical analysis to derive actionable insights.
- Top-performing stores: Recommend strategies for improving sales at underperforming locations.
- Popular products: Focus marketing efforts on best-selling products and product categories.
- Customer behavior insights: Optimize staff scheduling and inventory management based on peak transaction times.

5. Reporting and Documentation:

- Summarize findings, visualizations, and key insights into a comprehensive report.
- Provide suggestions for future analysis or areas of improvement.

5. Tools and Technologies

The project will utilize the following tools and technologies:

- **Programming Language:** Python,MYSQL
- **Libraries:** Pandas, Matplotlib,MYSQL connector
- **IDE:** Jupyter Notebook
- **Data Source:** MYSQL Database

6. Expected Outcomes

Expected Outcomes for Coffee Shop Dataset Documentation (Including Stores, Products, and Transactions)

- **Sales Performance Insights**:

  - Identify best-selling products and categories to inform inventory management and marketing strategies.
  - Analyze seasonal trends to optimize product offerings based on customer preferences during different times of the year.

Customer Behavior Analysis:

- Understand purchasing patterns (e.g., peak transaction times, popular product combinations) to improve customer experience and increase sales.
- Segment customers based on purchase behavior to tailor promotions and loyalty programs.

Cross-Store Comparison:

- **Performance Benchmarking**: Compare performance metrics across different stores to identify best practices and areas for improvement.
- **Geographical Insights**: Understand how store location influences product sales, helping to tailor offerings to local preferences.

Inventory Management:

- **Stock Optimization**: Identify which products have high sales velocity and should be restocked more frequently, ensuring that popular items are always available.
- **Product Life Cycle**: Assess the sales trends of individual products over time to determine when products should be introduced, promoted, or discontinued.

7. Timeline

The project is expected to be completed within a [specific timeframe, e.g., 3 weeks], with the following milestones:

- Week 1: Data Collection and Preprocessing
- Week 2: Exploratory Data Analysis
- Week 3: Visualization, Reporting, and Final Submission

8. Conclusion

- Coffee is the dominant product category driving the majority of sales, followed by tea and coffee beans.
- Astoria is the top-performing store, while Los Angeles has the lowest sales.
- Premium Brewed Coffee is the most popular product, followed by tea-based drinks.
- Coffee Beans are less popular than brewed coffee, suggesting a preference for ready-to-drink or specific brewed coffee types.
- Product category performance varies across stores. For example, Gourmet Brewed Coffee is popular in Los Angeles and New York, while Brewed Herbal Tea is favored in Astoria and Chicago.