Executive Summary: Coffee Shop Sales

Project Overview:

This report presents the findings of the exploratory data analysis conducted on the transaction data of Maven Roasters, a coffee shop chain operating in three New York City locations. The analysis focuses on evaluating sales performance, understanding customer behaviour, and accessing product performance across store locations. The dataset captures transactional information, including dates, times, product details, store locations, quantities sold, and unit prices of the transactions made at the store.

Objective of the project:

A. **Dataset Understanding:** Assess structure, feature, and completeness of the data.

B. **Data quality checks:** Address missing data, outliers, and formatting issues.

C. **Descriptive analysis:** Summarize key performance indicators (KPIs) such as revenue, sales volume, and store performance.

D. **Visualization:** Present insights visually for clearer communication of trends and patterns.

E. **Relationships:** To identify correlations and patterns in customer purchases to inform business decisions.

F. **Advanced Data Analysis:** Explore advanced data analysis techniques for predictions.

G. **Recommendations:** Make actionable recommendations based on the insights of the analysis.

Recommendations:

To maximize profits for Maven Roasters, the following strategies and recommendations can be implemented, based on the analysis of the sales data:

Product Strategies:

- **Coffee** is a major revenue driver, contributing approximately 38% of total sales across all locations. Given this, focusing on maximizing coffee sales should be a priority.

- **Tea** is also a significant contributor, accounting for approximately 29% to 33% of sales. Like coffee, tea should be a key focus in maximizing sales.
- **Coffee Beans** and **Branded** products are premium items with high average unit prices and transaction values, and are essential for maximizing revenue. The top 20 highest-priced products are primarily from these categories, including items such as clothing and houseware. These products are also key for high-value transactions.
- **Flavors**, while having a low average unit price ($0.8), have the highest average quantity per transaction, indicating customers purchase them in bulk.
- **Bakery** and **Drinking Chocolate** products have moderate prices, and are purchased less frequently.

Location Strategies:

- While sales are evenly distributed across all three locations (Astoria, Hell's Kitchen, and Lower Manhattan), **Hell's Kitchen** slightly leads in overall sales.
- All three stores show steady growth in monthly sales from January to June with minimal differences in revenue, so all locations are valuable.
- **Hell's Kitchen** has a unique trend of leading in Coffee Bean sales (46% of total Coffee Bean sales) while still representing a minor portion of overall sales.
- **Lower Manhattan** has slightly stronger sales in the Bakery category.
- **Astoria** has a slightly higher share of Branded product sales.

High Value Transactions (HVT):

- High-value transactions make up only 2.19% of the total transaction volume, but contribute 9.08% of the total sales revenue. These transactions have an average value of $19.38, significantly higher than the overall average of $4.69.
- **Coffee Beans** and **Branded** categories are key for high-value transactions. All transactions in the Branded category are classified as high-value.
- Strategies focused on enhancing high-value sales in the **Tea, Bakery, and Flavors** categories could yield better returns, as these currently rely more on regular transactions.

Temporal Considerations:

- Sales revenue and transaction counts show consistent month-on-month growth from January to June. There was a 104% increase in transaction count from January to June and a 103.83% increase in sales amount during the same period.
- Both transaction frequency and sales revenue peak between 8 AM and 11 AM.

- There is a noticeable decline in transactions and revenue between 12 PM and 2 PM.
- Transaction frequency and revenue taper off progressively after 6 PM.
- Sales patterns are relatively stable across different days of the week.

Key Recommendations:

- **Focus on High-Value Products**: Strengthen the offerings in the **Coffee Beans** and **Branded** categories. Ensure that these products are well-stocked and actively promoted to capitalize on their high revenue potential.

- **Maximize Coffee and Tea Sales**: Given that Coffee and Tea sales make up two-thirds of revenue, maximizing sales of these products will likely be very impactful.

- **Optimize Staffing**: Ensure adequate staffing during peak hours (8 AM to 11 AM) to handle the high volume of transactions. Also, consider promotions during the midday dip to boost sales during slower times.

- **Enhance Sales in Other Categories**: Boost high-value sales in **Tea, Bakery, and Flavors** categories by bundling them with higher-value items or offering special promotions.

- **Location-Specific Promotions:** Leverage each store's unique sales trends, such as Hell's Kitchen for coffee beans, and offer tailored promotions accordingly.

- **Pricing Strategies**: Given the positive correlation between unit price and transaction value, consider strategic pricing adjustments on certain premium items while ensuring competitive pricing for frequently purchased items such as tea and coffee.

Results of the Analysis:

Data Overview:

The dataset comprises transaction records from three store locations of Maven Roasters, a coffee shop chain in New York City. Below are the key features of the dataset:

- **transaction_id**: A unique identifier for each transaction.
- **transaction_date**: The date of the transaction.
- **transaction_time**: The time of day when the transaction occurred.
- **transaction_qty**: The quantity of items sold in the transaction.
- **store_id**: Identifier of the store where the transaction occurred.
- **store_location**: The geographical location of the store.
- **product_id**: Unique identifier for each product.
- **unit_price**: Price of the individual product sold.
- **product_category**: The category to which the product belongs (e.g., Coffee, Tea, Bakery).
- **product_type**: A more specific classification of the product (e.g., Barista Espresso, Black Tea).
- **product_detail**: Final product name (e.g. Cappuccino, Cappuccino Large)

Data Highlights:

a) Dataset Size: The dataset contains 149116 rows and 11 columns; each row represents a transaction made at one of the three store locations in New York City.

b) Period: The dataset covers a period of 181 days, spanning from 1st January 2023 to 30th June 2023.

c) The dataset consists of 80 unique products, organized into 29 distinct product types, which are further grouped into 9 product categories.

d) Each product detail is associated with a unique product ID.

e) The product detail and product IDs can be used interchangeably for identification purposes.

Data Preprocessing:

A. **Null Values and Duplicates:** The dataset was thoroughly checked for null values and duplicate records. No null values or duplicate entries were found. 100% data is available for further analysis.

B. **Data Type Adjustments:** Data types for certain features were updated to ensure consistency and accuracy. For example, categorical features like "transaction_date" and "product_category" were converted to appropriate formats.

C. **Feature Engineering:**

   a. Date-Based Features: Derived new variables such as:

      i. Day of the Month: Numerical values representing days (e.g., 1, 2, 3...).

      ii. Day of the Week: Categorical values (e.g., Monday, Tuesday, etc.).

      iii. Month: Extracted month information from the transaction date.

   b. **Time-Based Features:** Extracted "Hour of the Day" from the transaction time to analyse sales patterns by time.

D. **Transaction Value:** Added a new feature calculated as the product of "transaction_qty" and "unit_price" to measure the sales amount or revenue generated by each transaction.

Key Findings:

Summary Metrics:

a) **Total Transactions:** 149,116 transactions.

b) **Total Revenue:** $698,812.33.

c) **Units Sold:** 214,470 units.

d) **Average Daily Transactions:** 823.85 transactions per day.

e) **Average Daily Revenue:** $3,860.84.

f) **Revenue per Transaction:** $4.69.

g) **Revenue per Unit Sold:** $3.26.

h) **Increase in Monthly Revenue (From January to June):** 103.83 percent.

i) **Increase in Transaction Frequency (From January to June):** 104.18 percent.

Store Performance:

a) The Maven Roasters own three stores across New York, Astoria, Hell's Kitchen, and Lower Manhattan.

b) All stores perform competitively, with Hell's Kitchen slightly leading.

Product Analysis:

a) **Coffee dominates:** Across all locations, coffee contributes around **38% of the total sales.**

b) Coffee and Tea accounts for **two thirds of the revenue**.

c) **Coffee products** (Coffee and Coffee Beans), together generate **44% of the revenue**.

d) Branded, Loose Tea, Flavours and Packaged Chocolates, together account for 5.38% revenue.

e) **Premium products:** The high-priced products are from Coffee Beans and Branded categories, which include items such as Clothing and Houseware.

f) **Least-priced products:** These are from Tea, Coffee, and Flavours categories.

Temporal Trends:

a) **Upward Trend:** Sales revenue and transaction counts showed steady month-on-month growth from January to June.

b) **Day wise consistency:** Daily transaction counts and sales revenue remained steady across the week with minimal variation.

c) **Peak Hours:** Both transaction frequency and sales revenue peaked between 8 AM and 11 AM.

d) **Midday Dip:** A noticeable decline in transaction counts and revenue occurred between 12 PM and 2 PM.

e) **Afternoon Increase:** Transaction frequency and revenue increased slightly after 2 PM.

f) **Evening Decline:** Both metrics tapered off progressively after 6 PM.

g) **Stable Sales Proportions:** The proportion of each product category in the monthly sales remains stable across all months, with minimal variation observed.

h) **Consistent Category Contributions:** While Coffee Beans, Bakery, and other categories show some fluctuation, their overall contribution remains consistent month-to-month.

Customer Behaviour:

a) **58 percent** of transactions having a quantity of 1 unit, and 39% consisting of 2 units.

b) Fifty percent of the purchases have generated a revenue of $3.75 each. Additionally, fifty percent of the transactions involve items priced at $3 or less.

c) **Frequent Purchases:** Coffee, Tea, and Bakery categories consistently show the highest transaction frequencies and quantities, contributing significantly to Maven Roasters' revenue.

d) **Premium Purchases:** Coffee Beans and Branded product categories have high unit prices, resulting in higher sales amounts, highlighting their positions as premium products

e) **Low-cost purchases:** Flavour category have an average unit price of $0.8 and average transaction value of $1.24.

f) **Flavours** have the lowest average unit price but the highest average quantity per transaction, indicating customers purchase them in bulk during a single transaction.

g) **Bakery products** make an average transaction value of $3.6

h) **Price-Quantity Relationship:** Unit Price negatively affects the quantity purchased, However weak relationship suggests other factors also influence buying behaviour.

i) **Pricing-Revenue Relationship:** Unit Price positively affects transaction value, indicating that pricing plays a key role in driving revenue. HIgher the price, higher the revenue.

j) **Quantity-Revenue Relationship:** While transaction quantity also positively impacts transaction value, other factors have a stronger influence, as the relationship between quantity and transaction value is comparatively weaker.

High Value Transactions:

a) **Sales Contribution:** High-value transactions account for $63,437.99, contributing 9.08% of total sales.

b) **Volume Representation:** They make up only 2.19% of the total transaction volume, highlighting their exclusivity.

c) **Average Transaction Value (ATV):** The ATV for high-value transactions is $19.38, significantly higher than the overall average of $4.69.

d) **Key Categories:**

    i) Coffee Beans stands out with a higher average transaction value and substantial total high-value sales.

    ii) The Branded category has all its transactions classified as high-value.

e) **Disproportionate Impact:** High-value transactions, though fewer in number, contribute disproportionately to total sales, emphasizing their importance to revenue generation

Limitations:

a) The data provides transactional information, it does not include consumer profiles.

b) The information regarding the cost of the product to provide recommendations regarding cost and revenue.