Coffee Shop Sales Analysis Project – Maven Analytics

I downloaded a coffee shop dataset from Maven's site. To begin the data analysis process, my first task was to familiarise myself with the dataset. This involved understanding the volume of transactions, the date range, and the products being sold.

__Objective 1 - Familiarization and Data Preparation__

To simplify handling the data and applying filters, I converted the dataset into a table. The dataset contains 149,456 rows of transactions, including timestamps for each transaction, as well as store locations and their corresponding IDs. This coffee shop operates across three store locations, each assigned a unique store ID.

The dataset also includes several unique products, which are categorised into nine distinct groups, each with their own product ID and associated unit price. Additionally, the dataset provides product type and detailed descriptions.

In a typical database or data model scenario, product-level and store-level details might be stored in separate tables. In such cases, tools like SQL or Power Query would be used to join these datasets. However, this dataset consolidates all relevant details into one table, simplifying the initial exploration.

The next step in the analysis is to calculate **Revenue** for each transaction. Revenue represents the total money earned from regular business operations, such as sales, before deducting any expenses. To calculate this, we will multiply the **transaction quantity** by the **unit price** for each transaction. Rather than integrating the Revenue column within the main dataset, it will be added to the right of the raw dataset for clarity and easy reference.

Next, we will create additional columns to derive **Month** and **Day of the Week,** based on the transaction date:

- **Month Name:** Using the MONTH function, we can extract the **month number** from the transaction date for each row in the dataset. To make it more readable, we will apply the TEXT function to convert the month number into a **three-letter abbreviation** (e.g., "Jan"). The formula will look like this:
  - =TEXT([@[transaction_date]],"mmm")

- **Day of the Week**: To identify the day of the week for each transaction, we will use the WEEKDAY function. This will return a number corresponding to the day, where Monday is 1 and Sunday is 7. To make it more user-friendly, we will create another column that reformats these numbers into **three-letter weekday abbreviations** (e.g., "Sun") using the TEXT function.
- **Hour:** Additionally, we will identify the **hour** when each transaction occurred. This can be done by applying the HOUR function, which extracts the hour component from the transaction timestamp. This column will give insight into the time of day of which transactions took place.

Objective 2 - Exploring the Dataset with Pivot Tables.

To analyse the dataset effectively, I created pivot tables in Excel to summarise key metrics. These pivot tables will focus on the following aspects:

- Revenue by Month
- Number of transactions by day of the week and the hour by day
- Number of transactions by product category
- Number of transactions and revenue by the product type

Pivot Table 1 – Revenue by Month

To generate this table, I used the **Pivot Table** function within Excel's Table feature and placed the result on a new sheet. On the first pivot table, I selected the calculated column Month Name to be set for the rows of the table, and the sum aggregate of the Revenue as the values. This table provides a clear view of how revenue trends vary across the months.

Row Labels	Sum of Revenue
Jan	$81,677.74
Feb	$76,145.19
Mar	$98,834.68
Apr	$118,941.08
May	$156,727.76
Jun	$166,485.88
Grand Total	**$698,812.33**

Pivot Table 2 – Number of Transactions by Day of the Week Number of transactions by the day of the week

To create this pivot table, I copied the first pivot table and pasted it so that an independent pivot table is generated. Then, the previously used Month name column was replaced with the calculated Weekday Name column. I then proceeded to replace the sum of revenue in values with a count aggregate of the Transaction Ids column, which uniquely identifies every transaction. This table shows the volume of transactions for each day of the week, highlighting peak activity periods.

Row Labels	Count of transaction_id
Mon	21643
Tue	21202
Wed	21310
Thu	21654
Fri	21701
Sat	20510
Sun	21096
Grand Total	**149116**

Pivot Table 3 – Number of Transactions by Hour of the Day

The 2nd pivot table was copied and pasted to be used as the base of this table. The Weekday name that was set as the rows of the table were replaced with the calculated column Hour, and the values were left to be the count aggregate of the transaction id.

Row Labels	Count of transaction_id
6	4594
7	13428
8	17654
9	17764
10	18545
11	9766
12	8708
13	8714
14	8933
15	8979
16	9093
17	8745
18	7498
19	6092
20	603
Grand Total	**149116**

Pivot Table 4 – Number of Transactions by Product Category

The 2nd pivot table was copied again, and the Weekday Name Column were replaced with the Product Category column. After this, the table was sorted by the number of transactions in descending order. This table will identify the most popular product categories, based on transaction volume.

Row Labels	Count of transaction_id
Coffee	58416
Tea	45449
Bakery	22796
Drinking Chocolate	11468
Flavours	6790
Coffee beans	1753
Loose Tea	1210
Branded	747
Packaged Chocolate	487
Grand Total	**149116**

Pivot Table 5 – Number of Transactions and Revenue by Product Type

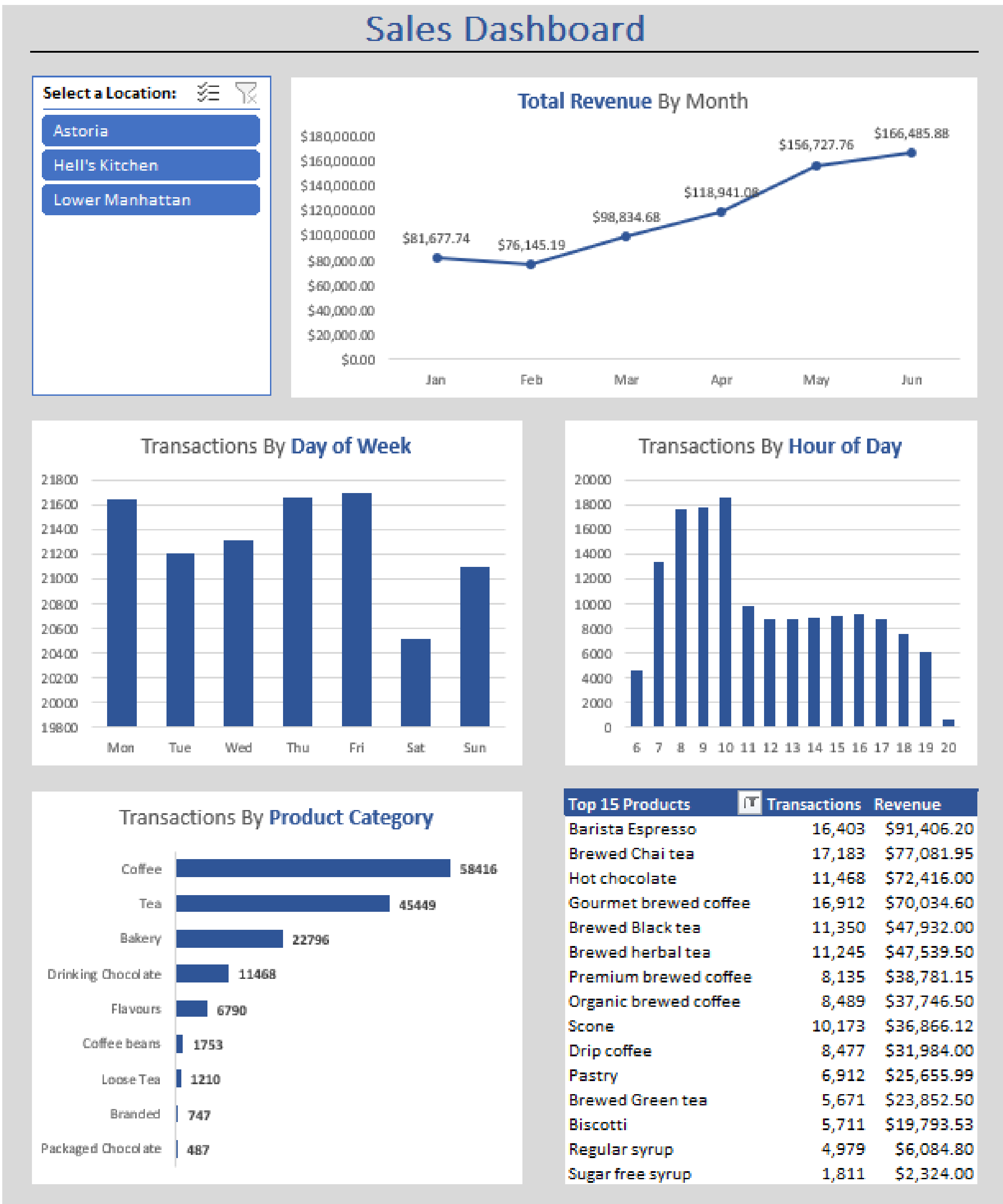
Using a copied pivot table, we set the rows of the table to be the Product Type column. The values of the table were set to be the count aggregate of Transaction Ids. This way, it can represent the number of transactions, and the sum aggregate of the revenue shows the revenue produced by all the transactions for each product. After setting the table with this, we sort the table by a descending order of revenue and then select the Top 15 rows. With this, the table will highlight the most profitable product types and their revenue.

Row Labels	Count of transaction_id	Sum of Revenue
Barista Espresso	16403	$91,406.20
Brewed Chai tea	17183	$77,081.95
Hot chocolate	11468	$72,416.00
Gourmet brewed coffee	16912	$70,034.60
Brewed Black tea	11350	$47,932.00
Brewed herbal tea	11245	$47,539.50
Premium brewed coffee	8135	$38,781.15
Organic brewed coffee	8489	$37,746.50
Scone	10173	$36,866.12
Drip coffee	8477	$31,984.00
Pastry	6912	$25,655.99
Brewed Green tea	5671	$23,852.50
Biscotti	5711	$19,793.53
Regular syrup	4979	$6,084.80
Sugar free syrup	1811	$2,324.00
Grand Total	**144919**	**$629,498.84**

Objective 3 – Creating an Exploratory Dashboard

- Create Pivot Charts:
  - To show revenue by month as a line chart
  - Transactions by day of week and hour of day as column charts
  - Transactions by product category as a bar chart
- Assemble charts into a rough layout.
- Add a slicer for store locations and connect the slicer to all Pivot tables.

Dashboard

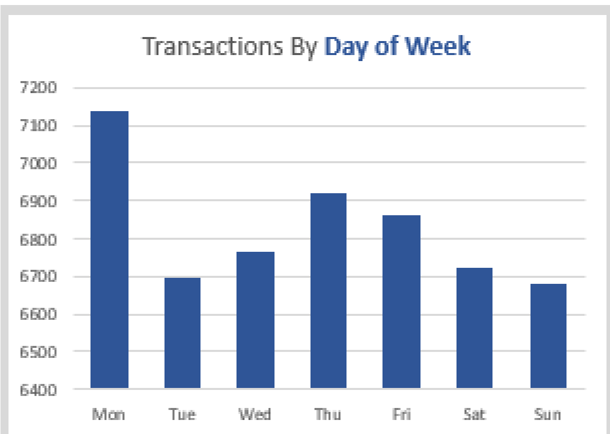


Objective 4 - Pattern and Insights

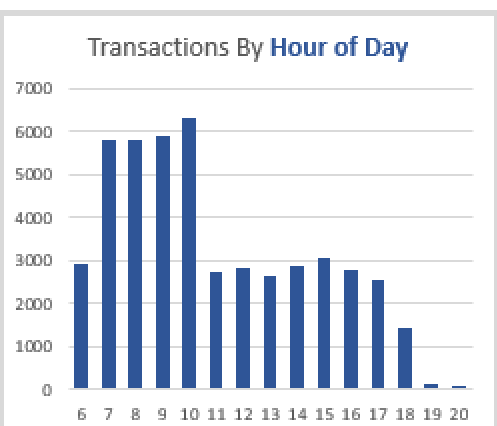
Analysing the visualisations reveals several interesting patterns, particularly in the **Day of the Week** and **Hour by Day** charts.

For instance, at the **Lower Manhattan Store location**, **Monday** consistently shows the highest transaction volume, indicating a significant spike in customer activity at the start of the week. This pattern suggests that many customers tend to make purchases at the beginning of the week, possibly due to routines, like starting their work week or replenishing supplies.

On the other hand, **Sunday** experiences the lowest transaction volume, as shown in the same charts. This creates an opportunity for the store owner to **strategically restock inventory on Sundays,** to prepare for the high demand on Mondays. By doing so, the store can maintain sufficient stock levels to meet customer needs and sustain high sales performance, throughout the year. This insight not only highlights customer behaviour, but also provides actionable recommendations to optimise inventory management and maximise revenue.



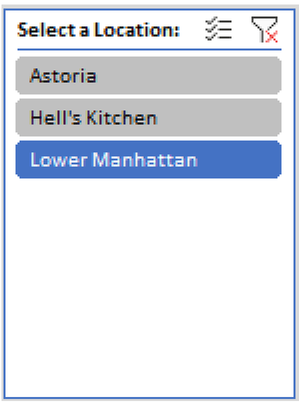
Examining the **Hour by Day** chart for the **Lower Manhattan Store**, we see that customer traffic peaks between **7:00 AM and 10:00 AM**, with minimal activity during the later hours of operation.



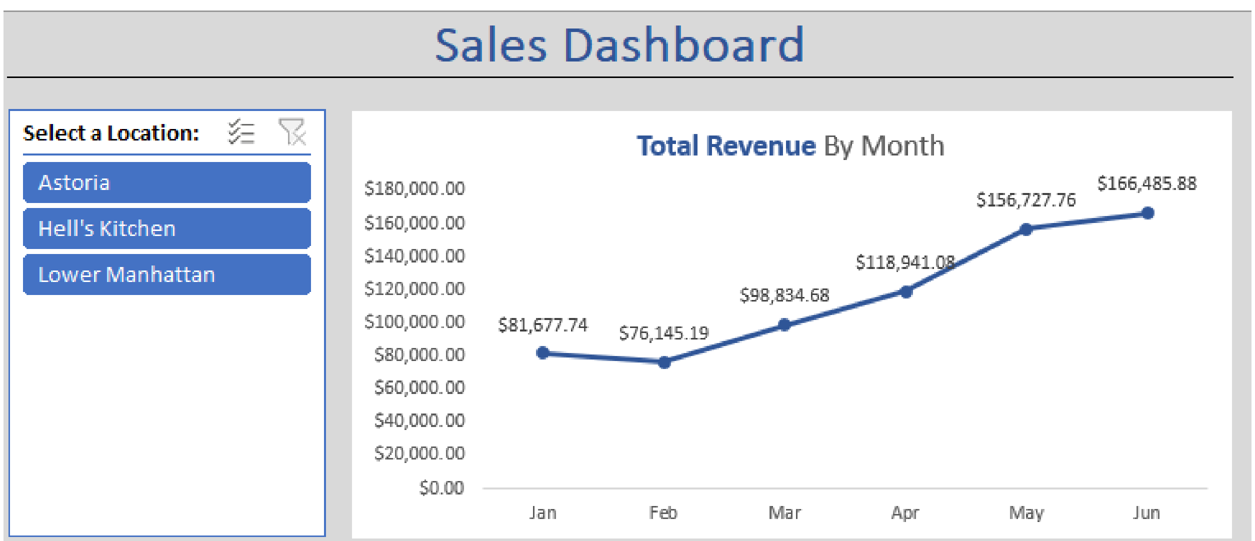
This pattern suggests that the store may not be making full use of its operating hours, particularly in the last few hours of the day. This insight indicates that the store owner could be **losing money by keeping the store open for extended hours** (beyond 18 hours of operation). A more financially sound approach might involve **closing the store earlier** to save on operational costs, such as electricity and gas. Instead of serving few or no customers during these hours, the store could allocate this time to **restocking and maintenance tasks**, ensuring that all store operations are completed by **8:00 PM**.

This strategy highlights opportunities for **improved time and shift management**, optimising both cost efficiency and productivity for the store.

These insights are easily observed by selecting **Lower Manhattan** as the store location within the slicer on the dashboard.



Overall, stores tend to get more transactions/business into the summer months of the year, this insight is provided by the line chart that displays **Total Revenue By Month**, when the slicer is cleared.



Analysing the **Product Popularity** pivot table for the **Lower Manhattan Store**, we find that the **Barista Espresso** is the most popular product, generating **$31,051 in revenue**. This insight allows the owner to make an informed decision to stock similar products that could attract more transactions and boost revenue further.

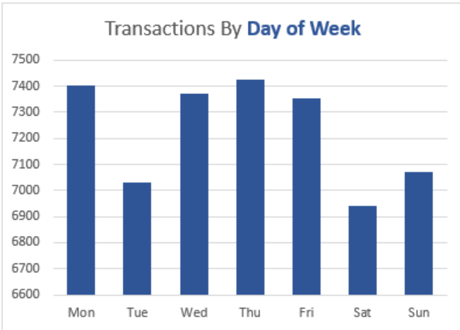
Additionally**, herbal and aromatic teas** rank highly among the Top 15 products sold at the store. This suggests an opportunity to explore other products with similar flavour profiles to appeal to existing customers and attract new ones. Conducting market research on complementary or trending tea offerings could help maximise this potential.

When examining the **Bar Chart for Transactions by Product Category** at the Lower Manhattan location, it becomes clear that **Loose Teas and Coffee Beans** account for relatively few transactions. In contrast, categories like **Teas and Coffee** dominate in popularity. This information supports a strategic shift in purchasing decisions, emphasising products that align with customer preferences, and avoiding overstocking less popular items.

By aligning inventory and product selection with these insights, the store can drive higher sales and optimise resource allocation.

Top 15 Products	Transactions	Revenue
Barista Espresso	5,320	$31,051.00
Brewed Chai tea	5,066	$24,008.75
Gourmet brewed coffee	5,217	$23,201.20
Hot chocolate	3,405	$22,494.50
Brewed Black tea	3,611	$16,215.50
Brewed herbal tea	3,535	$15,994.50
Premium brewed coffee	2,617	$12,907.00
Scone	3,506	$12,660.67
Organic brewed coffee	2,575	$12,009.30
Drip coffee	2,475	$9,817.00
Pastry	2,409	$8,890.49
Brewed Green tea	1,700	$7,646.00
Biscotti	1,975	$6,777.78
Regular syrup	2,100	$2,614.40
Sugar free syrup	830	$1,152.80

When selecting the store location to be **Astoria**, we see a major difference in customer behaviour across the week and on an hourly basis of store operations. Customers perform most purchases in the **middle of the week** (Thursdays). From **Wednesday to Friday** the Store sees **high levels of transactions** and it also sees high transactions at the **start of the week** (Monday). Looking at the column chart for Transactions By Day of Week we can see the lowest customer activity occurring on Saturdays, this may be an ideal day for which the store owner can execute **restocking and administrative activities**.

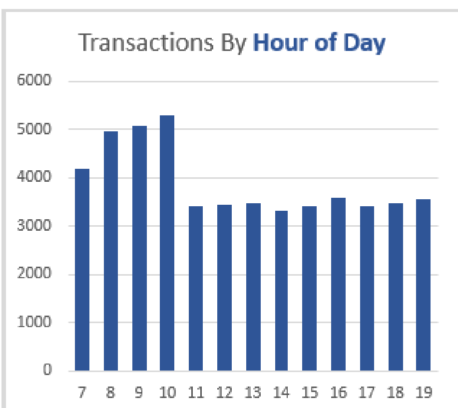


When examining **transactions by the hour of day**, the **Astoria Store** demonstrates marginally better performance throughout its operational hours compared to the **Lower Manhattan Store**. Notably, during the **last two hours** of operation, Astoria still records **3,000 to 4,000 transactions**, whereas Lower Manhattan sees fewer than **500 transactions** during the same time frame.

Additionally, the charts reveal a key difference:

- The **Astoria Store** ceases business operations at **7:00 PM**, while Lower Manhattan continues until **8:00 PM**.
  This discrepancy suggests that the Astoria store owner may have already recognised the **significant dip in business during later hours** and made an informed decision to **close earlier**, optimising operational costs such as labour, electricity, and other expenses.

This comparison highlights the importance of tailoring store hours to customer behaviour patterns, allowing for more efficient operations and potentially higher profitability.



To make informed decisions for improving business at the store, the owner can examine the **Top 15 Products** chart to identify which items are the most popular. At this store, **Barista Espresso** and **Brewed Chai Tea** are the favourites among customers, but **Hot Chocolate** also ranks third in terms of revenue.

Given the success of Hot Chocolate, the owner could explore related products, such as offering **different brands** or **flavours of hot chocolate**, to see if this can drive additional revenue. Introducing variations may cater to customer preferences and encourage repeat purchases.

Additionally, by analysing the **Top 15 Products**, the owner can make more strategic decisions on **discontinuing less popular items**. Dropping slow-selling products can help save on purchasing costs, reducing waste and freeing up space for more profitable options.

This approach ensures that the store's product lineup is continuously optimised to meet customer demand and maximise profitability.