

Introduction

The Maven Roasters Coffee Sales Dataset provides detailed transaction records for a fictitious coffee shop operating in three different locations in New York City. The dataset encompasses a wide range of information related to coffee sales, including transaction dates, timestamps, locations, and detailed product-level details.

The Dataset is presented in CSV format, featuring 149,116 records and encompassing 11 columns.

Introduction

Setting Stages

• Before analysing data, I define the workflow and identify specific stages in the process. here is my list:

1. Understanding business problem—our ‘business problem’ is to enhance business performance and facilitate informed decision-making and increase sales

Introduction

- 2.**Data preparation---In this phase, we take initial steps to get our data ready for a deeper look. We begin with data profiling to spot any unusual patterns or outliers. After that, we use different techniques to shape the data, making sure it's well-prepared before we incorporate it into our main data analysis
- 3.**Data modelling (data cleaning)
- 4.**Data visualizations-It's time to please our eyes with numbers and display them using convenient Power BI visuals
- 5.**Data analysis---he main purpose of this phase is to provide the insight **6.**Informed business decisions---we can leave this phase to stakeholders and we can give some recommendations based on insights we obtained in the previous phase

Understanding Business problem

- I identify key questions at the very beginning in order to shape and model our data to answer those questions in the most effective way and help to boost sales
- 1.Which days of the week tend to be busiest, and why do you think that's the case?
 - 2.Which products are sold most and least often? Which drive the most revenue for the business?
 - 3.which hours tend to be busiest, and why do you think that's the case
 - 4.How do sales vary across different store locations?

Data Preparation

- Data profiling-I'll go to the View tab and turn on Column quality, Column distribution and Column profile features to help me better understand the data and identify potential issues that need to be resolved

Screenshot of the Power BI Data Profiler interface showing data quality analysis for a 'Transactions' query.

Column Quality: Shows the status of columns: Valid (green), Error (red), and Empty (grey). For example, 'transaction_id' is 100% valid, while 'transaction_time' has 862 distinct values.

Column	Status	Details
transaction_id	Valid	1000 distinct, 1000 unique
transaction_date	Valid	2 distinct, 0 unique
transaction_time	Valid	862 distinct, 725 unique
transaction_qty	Valid	2 distinct, 0 unique
store_id	Valid	3 distinct, 0 unique
store_location	Valid	3 distinct, 0 unique
product_id	Valid	52 distinct, 0 unique

Value Distribution: A horizontal bar chart showing the frequency of product names. The most frequent product is 'Our Old Time Diner Blend Rg'.

Product Name	Count
Our Old Time Diner Blend Rg	1000
Sustainably Grown Organic Rg	~950
Jamaican Coffee River Lg	~900
Brazilian Sm	~850
Morning Sunrise Chai Rg	~800
Our Old Time Diner Blend Lg	~750
Traditional Blend Chai Lg	~700
Brazilian Rg	~650
Morning Sunrise Chai Lg	~600
Serenity Green Tea Lg	~550
Ethiopia Rg	~500

Query Settings: Shows the applied steps for the query, including 'Promoted Headers' and 'Changed Type'.

DATA PREPARATION

As we see, there are not missing value in this dataset and I checked anomalies by looking the max value so it is correct

Screenshot of Power BI Data Editor showing the 'Transactions' query and its properties.

Query Settings:

- Layout: Monospaced, Show whitespace, Column quality
- Data Preview: Advanced Editor, Query Dependencies

Properties:

- Name: Transactions
- Applied Steps:
 - Source
 - Navigation
 - Promoted Headers
 - Changed Type

Transactions Data Preview:

	store_location	product_id	unit_price	product_category	product_type	product_detail
1	5 Lower Manhattan	32	3	Coffee	Gourmet brewed coffee	Ethiopia Rg
2	5 Lower Manhattan	57	3,1	Tea	Brewed Chai tea	Spicy Eye Opener Chai Lg
3	5 Lower Manhattan	59	4,5	Drinking Chocolate	Hot chocolate	Dark chocolate Lg
4	5 Lower Manhattan	22	2	Coffee	Drip coffee	Our Old Time Diner Blend Sm
5	5 Lower Manhattan	57	3,1	Tea	Brewed Chai tea	Spicy Eye Opener Chai Lg
6	5 Lower Manhattan	77	3	Bakery	Scone	Oatmeal Scone
7	5 Lower Manhattan	22	2	Coffee	Drip coffee	Our Old Time Diner Blend Sm
8						

Column statistics:

Column	Count	Error	Empty	Distinct	Unique	Nan	Zero	Min	Max	Average
store_location	1000	0	0	14	0	0	0	2	4,75	3,11265

Value distribution:

The histogram shows the distribution of unit price values. The x-axis represents the unit price, ranging from 2 to 4.75. The y-axis represents the frequency of each price point. The distribution is right-skewed, with the highest frequency occurring at the lowest price point (2.5). Other significant peaks are at 3.75 and 3.5.

Data Preparation

In the "transaction_time" column of the Maven Roasters Coffee Sales Dataset, the date and time "31.12.1899 07:20:24" appear to be anomalous and likely associated with default or placeholder values common in computer systems and programming. To ensure data accuracy and consistency, it is recommended to replace these occurrences with appropriate timestamps reflecting valid transaction information. This corrective action will contribute to a more reliable dataset for subsequent analyses and prevent potential misinterpretations stemming from inaccurate timestamps.

The screenshot shows the Power BI Data Editor interface with the 'Transactions' query selected. The main area displays a table with columns: transaction_id, transaction_date, transaction_time, transaction_qty, store_id, store_location, and product_id. The 'transaction_time' column is highlighted, showing a dropdown menu with options like 'Sort Ascending', 'Sort Descending', 'Clear Sort', 'Clear Filter', 'Remove Empty', and 'Date/Time Filters'. Below the table, a 'Column statistics' pane provides details for the transaction_time column, including 'Count', 'Error', 'Empty', 'Distinct', 'Unique', 'Min', and 'Max'. The 'transaction_time' column has 1000 distinct and unique values, with 100% being Valid. The 'transaction_date' column has 2 distinct and 0 unique values, with 100% being Valid. The 'store_id' column has 3 distinct and 0 unique values, with 100% being Valid. The 'store_location' column has 5 distinct and 0 unique values, with 100% being Valid. The 'product_id' column has 52 distinct and 0 unique values, with 100% being Valid. The 'Query Settings' pane on the right shows the 'Name' is set to 'Transactions' and lists the 'APPLIED STEPS' which include 'Source', 'Navigation', 'Promoted Headers', and 'Changed Type'.

Data Preparation

I created new columns from examples and remove this column

Queries [1] <

Add Column From Examples

Enter sample values to create a new column (Ctrl+Enter to apply).
Transform: Time.From([transaction_time])

OK Cancel

transaction_id	transaction_date	transaction_time	transaction_qty	store_id	Time
1	01.01.2023	31.12.1899 07:06:11	2	5	07:06:11
2	01.01.2023	31.12.1899 07:08:56	2	5	07:08:56
3	01.01.2023	31.12.1899 07:14:04	2	5	07:14:04
4	01.01.2023	31.12.1899 07:20:24	1	5	07:20:24
5	01.01.2023	31.12.1899 07:22:41	2	5	07:22:41
6	01.01.2023	31.12.1899 07:22:41	1	5	07:22:41
7	01.01.2023	31.12.1899 07:25:49	1	5	07:25:49
8	01.01.2023	31.12.1899 07:33:34	2	5	07:33:34
9	01.01.2023	31.12.1899 07:39:13	1	5	07:39:13
10	01.01.2023	31.12.1899 07:39:34	2	5	07:39:34
11	01.01.2023	31.12.1899 07:43:05	1	5	07:43:05
12	01.01.2023	31.12.1899 07:44:35	2	5	07:44:35
13	01.01.2023	31.12.1899 07:45:51	1	5	07:45:51
14	01.01.2023	31.12.1899 07:48:19	1	5	07:48:19
15	01.01.2023	31.12.1899 07:52:36	2	5	07:52:36
16	01.01.2023	31.12.1899 07:59:58	2	5	07:59:58
17	01.01.2023	31.12.1899 07:59:58	1	5	07:59:58
18	01.01.2023	31.12.1899 08:00:18	1	8	08:00:18
19	01.01.2023	31.12.1899 08:00:39	2	8	08:00:39
20		31.12.1899 07:06:11			

Query Settings X

Properties Name Transactions All Properties

Applied Steps Source Navigation Promoted Headers Changed Type Column From Examples

Data Preparation

All of the transformation steps I defined will be saved by Power Query Editor, and every time I refresh my dataset, these steps will be applied to shape your data and will always bring it to the desired form!

The screenshot shows the Microsoft Power Query Editor interface. At the top, there are several tabs: 'Query Settings' (selected), 'Layout', 'Data Preview', 'Columns', 'Parameters', 'Advanced', and 'Dependencies'. The 'Query Settings' tab has checkboxes for 'Formula Bar', 'Monospaced', 'Column distribution', 'Show whitespace', 'Column profile', and 'Column quality'. Below the tabs is a 'Queries [1]' section with a single item named 'Transactions'. The 'Data Preview' pane displays a table with columns: transaction_id, transaction_date, transaction_qty, store_location, product_id, unit_price, and product_category. Each column has a summary bar chart and a detailed breakdown below it. A red arrow points from the 'store_location' column towards the 'APPLIED STEPS' pane. The 'APPLIED STEPS' pane lists the following steps: 'Removed Columns1' (which includes 'Removed Columns', 'Changed Type1', and 'Filtered Rows'), 'Source', 'Navigation', 'Promoted Headers', 'Changed Type', and 'Inserted Time'. The 'Properties' pane on the right shows the query name 'Transactions'.

Data Preparation

- I want to apply one additional transformation to my Time column. As we don't need to analyze data on a minute level (hour level of granularity is the requirement), I'll round the values to a starting hour

The screenshot shows the Microsoft Power Query Editor interface. The main area displays a table with columns: product_id, unit_price, product_category, product_type, product_detail, and Time. The Time column contains various timestamp values. A context menu is open over the Time column, with the 'Change Type' option selected, showing sub-options for Hour, Minute, Second, and Text Transforms. The 'Hour' option is highlighted. The 'Query Settings' pane on the right is visible. At the bottom, there are sections for 'Column statistics' (Count: 1000, Error: 0, Empty: 0, Distinct: 862, Unique: 725, Min: 07:01:59, Max: 19:58:52) and 'Value distribution' (a histogram of time values). The status bar at the bottom indicates 'PREVIEW DOWNLOADED AT 17:15'.

Data Modelling

I created Date dimension using DAX, however, we can create Date dimension using M language in Power Query too. so I used DAX to create Date table

The screenshot shows the Power BI Data View interface. On the left, there is a code editor window containing DAX code for creating a Date dimension. The code defines a variable for the minimum year, filters transactions to find the maximum year, and then creates a new table with various date-related columns like Calendar Year, Month Name, Month Number, Weekday, Weekday number, and Quarter. The resulting table is named DateTable.

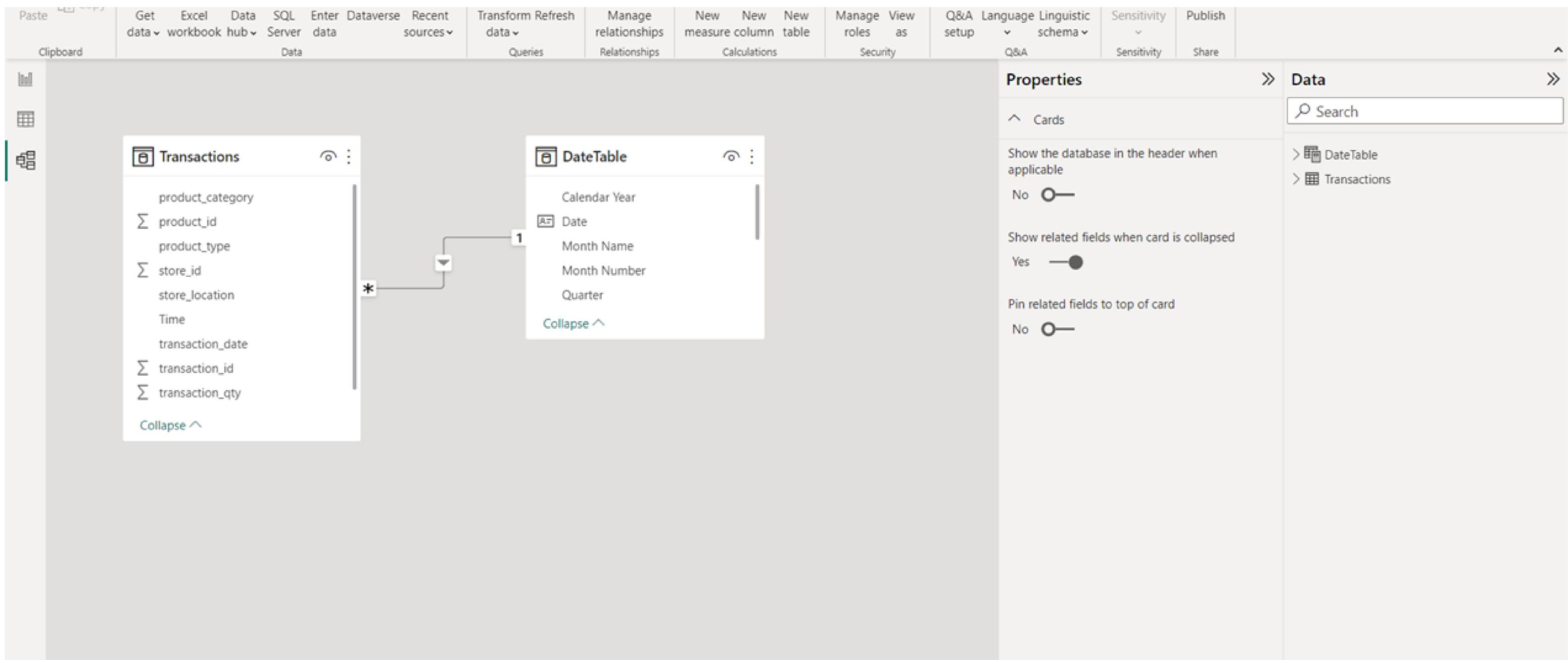
```
1 DateTable =  
2 VAR MinYear = YEAR ( MIN ( Transactions[transaction_date] ) )  
3 VAR MaxYear = YEAR ( MAX ( Transactions[transaction_date] ) )  
4 RETURN  
5 ADDCOLUMNS (  
6 FILTER (  
7 CALENDARAUTO ( ),  
8 AND ( YEAR ( [Date] ) >= MinYear, YEAR ( [Date] ) <= MaxYear )  
9 ),  
10 "Calendar Year", "CY " & YEAR ( [Date] ),  
11 "Month Name", FORMAT ( [Date], "mmmm" ),  
12 "Month Number", MONTH ( [Date] ),  
13 "Weekday", FORMAT ( [Date], "dddd" ),  
14 "Weekday number", WEEKDAY ( [Date] ),  
15 "Quarter", "Q" & TRUNC ( ( MONTH ( [Date] ) - 1 ) / 3 ) + 1)
```

On the right, the Data pane shows the structure of the DateTable. It includes columns for Date, Calendar Year, Month Name, Month Number, Weekday, Weekday number, and Quarter. Below this, the Data pane lists other tables and measures available in the model, such as Transactions, product_category, product_id, and product_type.

Date	Calendar Year	Month Name	Month Number	Weekday	Weekday number	Quarter
01.07.2023	CY 2023	July		7 Saturday	7	Q3
02.07.2023	CY 2023	July		7 Sunday	1	Q3
03.07.2023	CY 2023	July		7 Monday	2	Q3
04.07.2023	CY 2023	July		7 Tuesday	3	Q3
05.07.2023	CY 2023	July		7 Wednesday	4	Q3
06.07.2023	CY 2023	July		7 Thursday	5	Q3
07.07.2023	CY 2023	July		7 Friday	6	Q3
08.07.2023	CY 2023	July		7 Saturday	7	Q3
09.07.2023	CY 2023	July		7 Sunday	1	Q3

Data Modelling

I created relationship between tables.I used one-to-many relationship between Date table and Transactions table



Data Visualization

The dashboard's primary goal is analytical, emphasizing the identification of trends and patterns within historical data. Its purpose is to facilitate more informed mid to long-term decision-making by providing insights gleaned from past data analysis

I created some measures using DAX and then created folder to organize measures

Data Visualization

The screenshot shows the Power BI Data View interface. At the top, there's a navigation bar with options like Paste, Get data, Enter data, Transform data, Manage relationships, New measure, New column, New table, Manage roles, View as security, Q&A setup, Language schema, Sensitivity, Publish, and Share.

The main area displays two data tables:

- Transactions**: Contains fields such as product_category, product_id, product_type, store_id, store_location, Time, transaction_date, transaction_id, and transaction_qty.
- DateTable**: Contains fields such as Calendar Year, Date, Month Name, Month Number, Quarter, and a related field from the Transactions table.

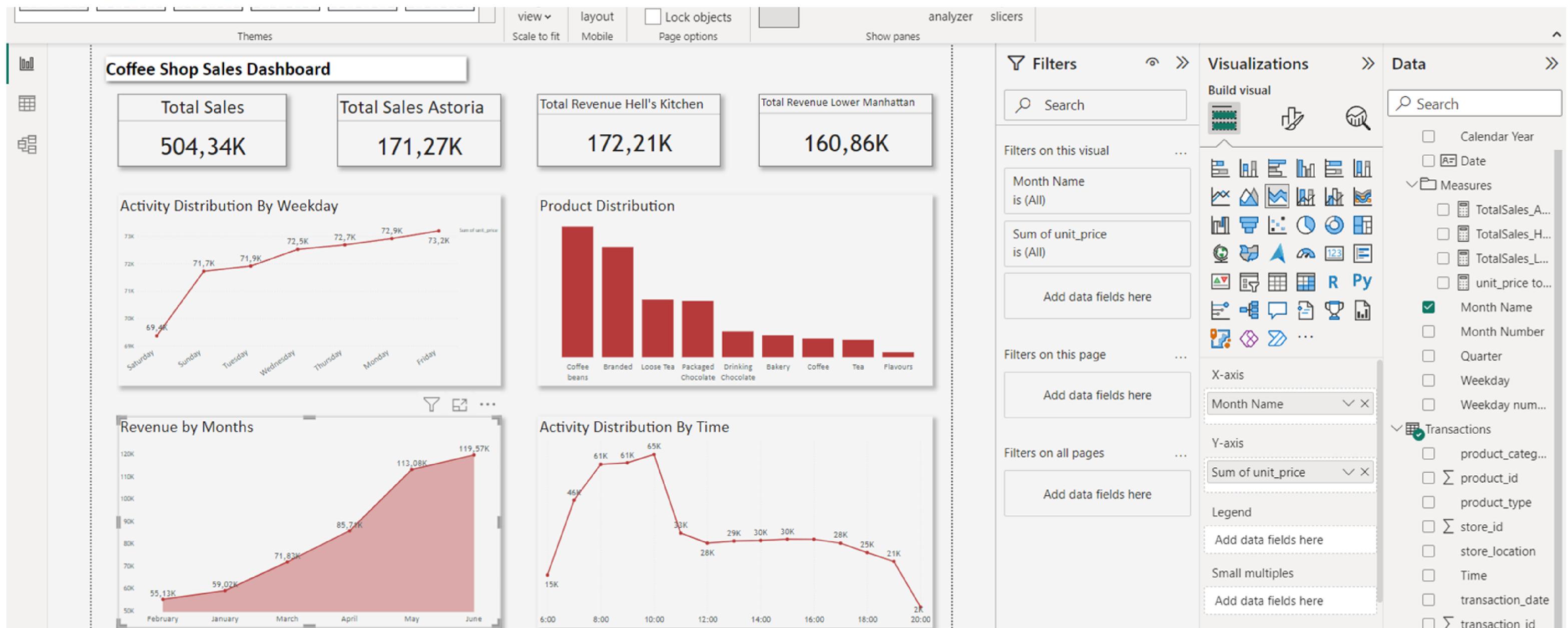
A relationship is shown between the Transaction_id field in the DateTable and the transaction_id field in the Transactions table. The relationship type is 1 to many (*).

On the right side, there are two panes:

- Properties**: A card with settings:
 - Show the database in the header when applicable: No
 - Show related fields when card is collapsed: Yes
 - Pin related fields to top of card: No
- Data**: A search bar and a list of data items:
 - Calendar Year**:
 - Date
 - Measures
 - TotalSales_Astoria
 - TotalSales_Hell'sKitchen
 - TotalSales_LowerManhattan
 - unit_price total for store_location
 - Month Name
 - Month Number
 - Quarter
 - Weekday
 - Weekday number
 - Transactions**:
 - product_category
 - product_id
 - product_type
 - store_id
 - store_location
 - Time
 - transaction_date
 - transaction_id
 - transaction_qty

Data Visualization

- I use line chart, stacked area chart ,cards to show main points and bar chart.I changed fonts to make it readable.and place important visuals from left to right and top to bottom



Data Analysis

- Here we are – after taking raw data in the form of a CSV file, defining a set of business questions that need to be answered using that data, then cleaning and shaping the original dataset in the previous part we've created compelling visualizations to provide different insights to business decision-makers. Now, it's time to analyze insights and, based on the information we extract from these insights, recommend some actions!

Coffee Shop Sales Dashboard

Total Sales

504,34K

Total Sales Astoria

171,27K

Total Revenue Hell's Kitchen

172,21K

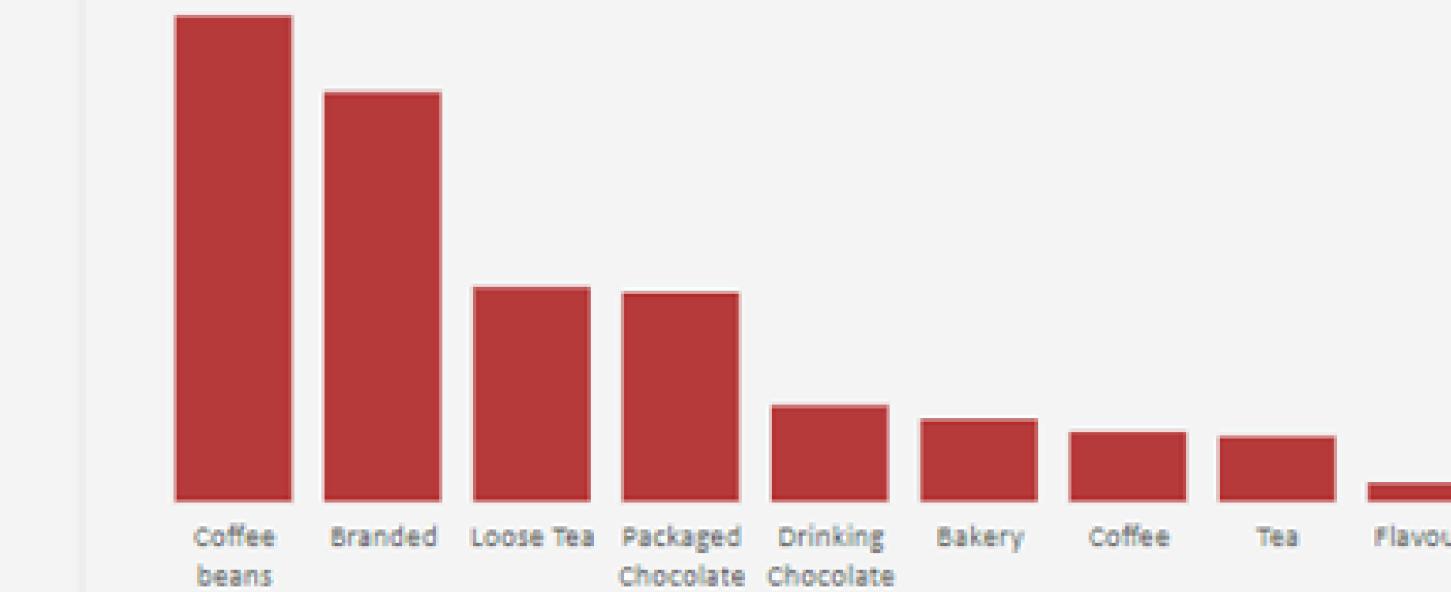
Total Revenue Lower Manhattan

160,86K

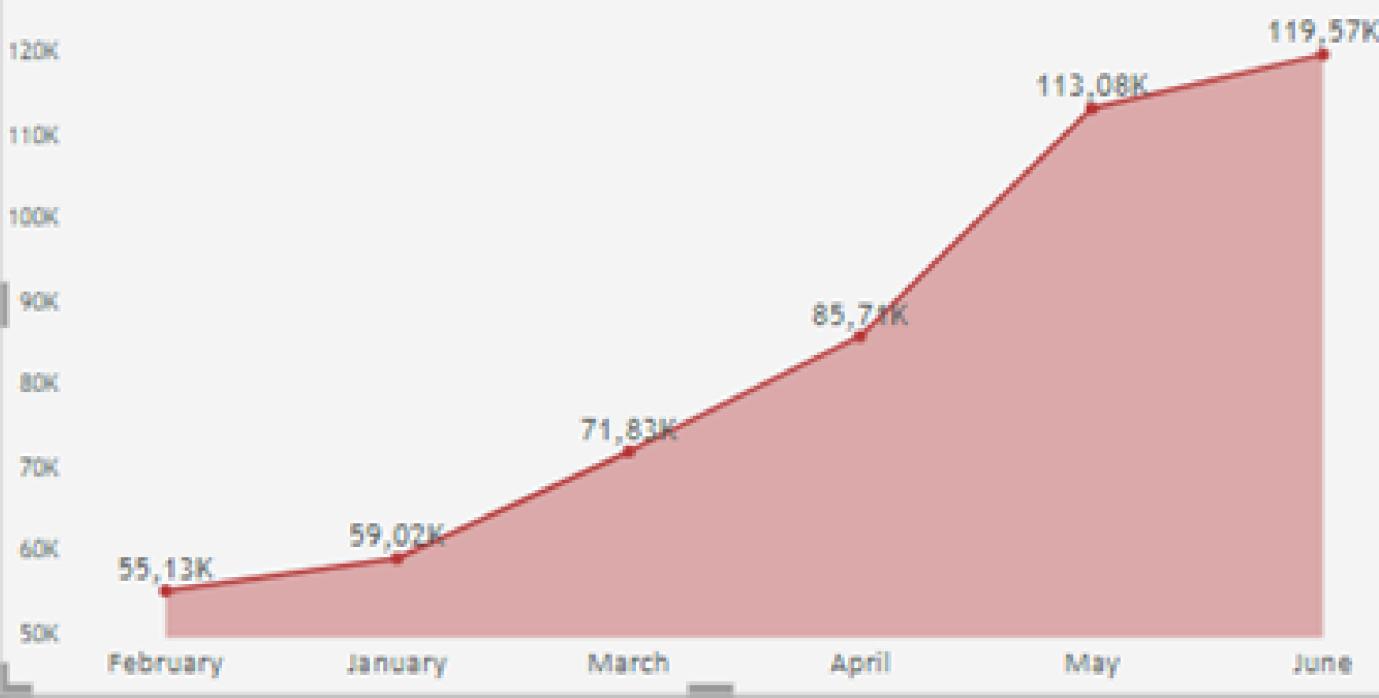
Activity Distribution By Weekday



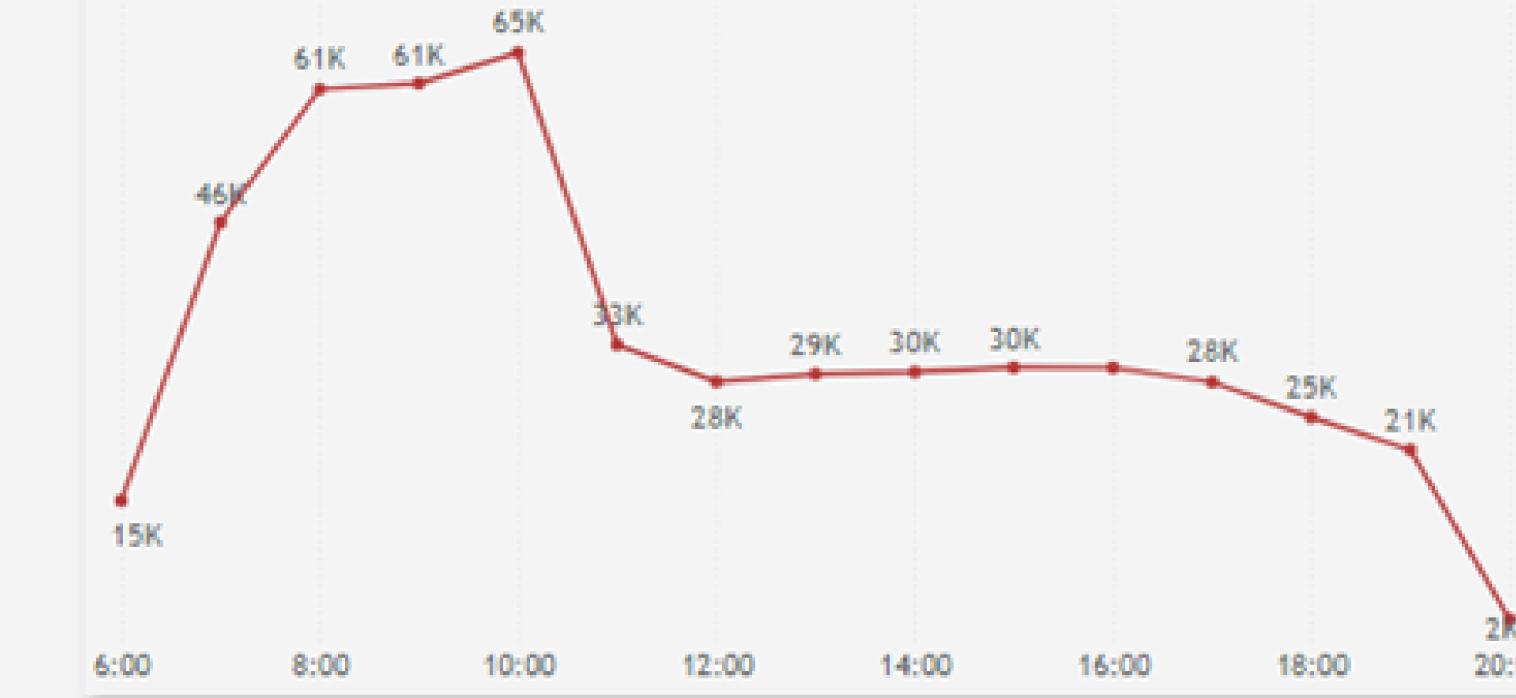
Product Distribution



Revenue by Months



Activity Distribution By Time



Data Analysis

1. Busiest Days of the Week:

1. Fridays exhibit the highest level of activity, likely due to people winding down from the week and seeking a boost before the weekend.
2. Mondays follow closely, suggesting a need for a productivity kickstart at the beginning of the week.
3. Saturdays, on the other hand, show the least activity, indicating a day of relaxation and socializing.

2. Customer Behavior Throughout the Week:

1. At the start of the week (Monday), people seem to visit coffee shops to be active and kick off their week with energy.
2. Towards the end of the week (Friday), the purpose shifts to rest and socializing, marking a transition into the weekend.

3. Product Sales Insights:

1. The most popular products are coffee beans and branded items, indicating a preference for quality coffee and loyalty to the coffee shop's brand.
2. Flavors are the least sold product, suggesting that customers might prioritize the classic coffee experience over experimental or flavored options.

4. Monthly Revenue Trends:

1. The winter months, particularly January and February, witness the least activity and revenue, potentially due to colder weather and reduced outdoor activities.
2. As the months progress towards summer, there is a noticeable increase in both activity and revenue.
3. June and May emerge as the busiest months, aligning with warmer weather and a potential increase in social gatherings.

5. Busiest Hours:

1. The peak hours for the coffee shop are between 7 a.m and 10 a.m, indicating a strong demand for morning coffee and breakfast items.
2. Activity decreases significantly after 11 a.m, reaching its lowest point by 8 p.m, suggesting that the coffee shop is less popular during the late afternoon and evening hours.

Data Analysis

1. Total Sales Overview:

1. In the fiscal year 2023, the coffee shop recorded a total sales figure of \$504.34K, reflecting the overall revenue generated across its three locations.

2. Location-Specific Breakdown:

- a. Astoria:

- 1. Total Revenue: \$171.27K

- 2. Percentage of Total Sales: ~34%

3. Hills Kitchenis:

- 1. Total Revenue: \$172.27K

- 2. Percentage of Total Sales: ~34%

4. Manhattan:

- 1. Total Revenue: \$160.86K

- 2. Percentage of Total Sales: ~32%

Data Analysis

- **Conclusion:**
- The coffee shop's success is closely tied to understanding and catering to the varying needs of customers throughout the week and adjusting product offerings accordingly.
- Focusing on promoting flavored products, especially during busier months, could potentially boost overall sales.
- Considering the popularity of coffee beans and branded items, marketing efforts can emphasize these products to maintain and expand the customer base.
- Understanding the ebb and flow of customer activity can inform staffing levels, inventory management, and promotional efforts for optimal business performance.

Data Analysis

- Evaluate the performance drivers in Astaria and Hills Kitchenis to identify successful strategies that can be replicated in Manhattan.
- Implement location-specific marketing campaigns or promotions to boost revenue in Manhattan and align it more closely with the contributions from the other locations.