

# Memo

To: Management

From: Data Science Department

RE: Coffee Shop Sales Report, July 2023

Transactions from our company's three New York coffee shops were examined to derive operational insights. A total of 149,116 transactions spanning January to June 2023 were analyzed. This report is organized into five sections, each addressing one aspect of the company's performance or operations. Concluding thoughts are offered in Section 6. An accompanying technical appendix details the processes of normalizing and querying the data.

## 1. How does the performance of the stores compare?

We are currently doing business out of Astoria, Hell's Kitchen, and Lower Manhattan. It is prudent to compare the sales numbers of our different locations to ensure that they are performing as expected. If one location were to perform much worse than the others, perhaps we might consider channeling more resources toward it to help boost sales, or look into closing the shop.

All three of our locations have very similar monthly sales from January to June. They all follow the same path starting at about \$26,000-\$27,000 in January, then holding roughly steady in February before boasting rapid growth through the spring months. In June, all locations achieved more than \$54,000 in sales. We will continue to monitor monthly numbers going forward, but for now, our high-level business strategies appear to be driving growth at all locations. The overarching takeaway here is that a business-as-usual approach is likely to yield continued success, though there may be opportunities to further optimize operations, as will be discussed in the next sections.

## 2. How do sales look across our different product categories?

Our prepared offerings (coffee, tea, bakery items, and drinking chocolate) are our bestsellers, both in terms of quantity and sales numbers. We should continue to prioritize the quality of these products to ensure a great customer experience every time. We might also consider expanding our tea offerings, which already set us apart from some other coffee chains, since we carry a high-quality assortment of herbal, black, green, and spiced tea blends. Increased growth with tea sales may help drive further overall growth.

Branded items which include "I Need My Bean!" t-shirts, diner mugs, and latte cups, are sold in smaller quantities. However, given their higher price point, these items generate a decent amount of revenue. It may be worth expanding our selection of branded goods to benefit from their higher margins. This may be something to look into with guidance from our product management, marketing, and operations experts. With that said, we should retain most of our focus on coffee, tea, and baked goods even if we do decide to pursue additional product opportunities in the branded space.

Packaged chocolate has not sold as well as our other offerings. Across all shops, total revenue from packaged chocolate sales is below \$4,500. This issue will be further examined in Section 4.

### 3. Within the top selling categories, what are the best-selling products?

Building on our discussion of categories from Section 2, this section takes a more detailed look at some of our best-selling products. It can be advantageous to understand our high sellers so that we can continue to support those offerings through purchasing, operational decisions, and strategic advertising.

Lattes dominate our coffee sales, both in terms of volume and revenue. As such, it is recommended that we stay on top of maintenance for our espresso machines so that we can continue to guarantee the consistent quality and availability of our lattes. We might find further success by offering to make lattes with different plant-based milks, as well. The different blends and roasts of brewed coffee are somewhat well-distributed – there is no clear best- or worst-selling product. Larger sizes are generally less popular, but they generate more revenue. It is recommended that we keep our coffee lineup as is while continuing to monitor sales and market trends.

Similar to coffee blends, sales of tea products are variable and size preference appears to play a large role. Peppermint tea and the chai variations generally sell well. Again, it is recommended to maintain the current tea selection and re-evaluate in a few months in case new sales trends emerge.

Chocolate croissants are far and away the most popular bakery item. All of the scones sell well, except for the oatmeal scone. This item lags the second least popular baked good by nearly \$1000 in total sales. It may not be worthwhile to continue carrying the oatmeal scone. Or, if we get a deal from our scone supplier, perhaps we can feature oatmeal scones to heighten their appeal.

### 4. What insights can be gleaned from the worst-selling product category?

As mentioned previously in Section 2, packaged chocolate has not sold as well as some of our other products. We carry three types: Chili Mayan, Dark Chocolate, and Sustainably Grown Organic chocolate. The quantities of each variety sold over time resemble our shops' overall sales. In other words, we have seen growth in chocolate sales from January to June. Quantities sold have roughly doubled over that time period, which is consistent with the growth observed for our most popular coffee product, a regular latte. Given this growth coupled with the convenience provided by the packaged aspect of the chocolates, it seems reasonable to retain this product category for the foreseeable future.

Essentially, we are still seeing growth across all product categories, from the best sellers to the worst. It may take some time to find more clear trends to better fine-tune our product offerings.

### 5. Are there any times of particularly high or low sales volume?

Pivoting away from products and toward operations, sales were examined from the perspective of transaction time. Are there certain times of the day where we might consider adding staff to the shops? On the flipside, do our sales justify the full range of our current hours of operation?

Peak times are experienced at 10, 9, and 8 AM, in that order. Sales during these three hours comprise nearly 37% of all sales. It is recommended that we maintain clear lines of communication with each shop manager to ensure that they have the resources needed to provide a quality customer experience during these busy times.

As you know, the Astoria location has more limited hours (7 a.m. to 7 p.m.) compared to the other two locations that operate from 6 a.m. to 8 p.m. Astoria tends to be less busy than Hell's Kitchen and Lower Manhattan in the mornings, but has consistently stronger sales later in the day. There is a steep drop in

sales after 5 p.m. at Lower Manhattan, and after 6 p.m. at Hell's Kitchen. Based on observed sales, it is recommended to adjust the operating hours as shown in Table 1.

*Table 1*

<b>Location</b>	<b>Current Hours</b>	<b>Recommended Hours</b>
Astoria	7 a.m. - 7 p.m.	7 a.m. - 8 p.m.
Hell's Kitchen	6 a.m. - 8 p.m.	6 a.m. - 7 p.m.
Lower Manhattan	6 a.m. - 8 p.m.	6 a.m. - 6 p.m.

## 6. Conclusion

We have seen encouraging growth across all product offerings at our three coffee shop locations during this reporting period. The highest sales numbers have resulted from prioritizing the quality and variety of our coffee, tea, and bakery menu items. It is recommended to retain all current menu items, except for the oatmeal scone, and reassess the performance of these products after the next reporting period. Slight adjustments to the operating hours of each location are further recommended based on previous sales. This will allow the company to capitalize on the busiest times of each location while minimizing operational costs during less profitable hours.

# Technical Appendix

## Normalization

The data started off in first normal form (1NF). The total number of unique records matched the number of rows in the data frame, and therefore, each record was unique. It was determined that each cell contained a single value, as well. For some products, the size is included in the product detail, i.e., “English Breakfast Rg.” Consideration was given to separating the sizes (Sm/Rg/Lg) from the rest of the product information, but that maneuver would not make the most sense for this data set. For one, each product detail was already assigned a unique product ID, and it would be problematic to have identical descriptions corresponding to multiple different product IDs. Secondly, only a fraction of the items in the dataset include the size. If size was separated out as an additional column, a lot of records would have a null value in that column, which is undesirable.

Once the data was determined to be in 1NF, the normalization process progressed to consider second normal form (2NF). The goal of 2NF is to eliminate any partial dependencies such that non-key attributes depend on the entire primary key. To attain 2NF, the single large table needed to be split into multiple smaller tables. The information was initially categorized into three buckets: stores, transactions, and products. A “Store” table was created (Table A-1). Since store location depends solely on store ID, this table satisfies 2NF (and third normal form [3NF]).

Table A-1

Store ID (primary key)	Store Location
3	Astoria
5	Lower Manhattan
8	Hell’s Kitchen

A “Transaction” table was created as previewed in Table A-2. Date, time, quantity, and store ID are clearly dependent on transaction ID. Only one product is present per transaction, so transaction ID uniquely determines product ID. There is no discernable relationship between unit price and store or date. While product ID helps narrow down price, there is no reliable way to predict price given product ID and any column other than transaction ID. Therefore, unit price can be said to depend entirely on transaction ID.

Table A-2

Transaction ID (primary key)	Date	Time	Quantity	Store ID (foreign key)	Product ID (foreign key)	Unit Price
1	2023-01-01	07:06:11	2	5	32	3.00
2	2023-01-01	07:08:56	2	5	57	3.10
...	...	...	...	...	...	...

A “Product” table was conceptualized as shown in Table A-3, with product ID (primary key), product category, product type, and product detail. Unit price was discovered not to be a property of product but rather varied by transaction, so unit price was not included in the product table. Product category,

type, and detail depend on product ID, but transitive dependencies are present that will be dealt with when continuing on to 3NF.

Table A-3

Product ID (primary key)	Product Category	Product Type	Product Detail
1	Coffee Beans	Organic Beans	Brazilian - Organic
2	Coffee Beans	House blend Beans	Our Old Time Diner Blend
...	...	...	...

The Product table was further split to achieve 3NF. Product category depends on product type, which depends on product detail. These entities were split into separate tables starting from most general to most specific. A “Product Category” table was created first, as previewed in Table A-4.

Table A-4

Product Category ID (primary key)	Product Category
1	Coffee
2	Tea
...	...

The “Product Type” table was then created, incorporating the product category ID as a foreign key. The structure is previewed in Table A-5.

Table A-5

Product Type ID (primary key)	Product Type	Product Category ID (foreign key)
1	Gourmet brewed coffee	1
2	Brewed Chai tea	2
...	...	...

Finally, the “Product Detail” table was created, incorporating the product type ID as a foreign key. The structure is previewed in Table A-6.

Table A-6

Product Detail ID (primary key)	Product Type ID (foreign key)	Product Detail
1	17	Brazilian - Organic
2	28	Our Old Time Diner Blend
...	...	...

The resulting database is normalized up to 3NF. It contains five tables pertaining to store, transaction, product category, product type, and product details.

## SQL Queries

The data were queried using SQLITE3 and the results were corroborated using pandas. The queries leveraged to derive insights on the five questions explored in the body of the report are explained here.

### 1. How does the performance of the stores compare?

Transaction data were aggregated into monthly buckets to improve interpretability. Note that this method of aggregation is responsive to differences in number of days per month, so, February appears to have a slight dip but sales rates actually remained roughly consistent with January performance. Still, this method is useful for looking at overall growth. Sales were calculated by multiplying quantity by unit price. Sum of sales were grouped by store location to yield monthly sales for each store location. The query involved performing an inner join on the transaction table with the store table so that transaction month, store location, and calculated sales were returned, grouped by store location and month.

### 2. How do sales look across different product categories?

Three queries were carried out to investigate product category performance. The first was the most general, selecting product category, sum of quantities sold, and sum of sales amounts. This was achieved by a series of joins, starting with the transaction table with (1) the product detail table on the product ID column, then (2) joining the product type table on the common product type ID column as a bridge to finally get to (3) product category by joining on the product category ID column. All joins were default inner joins. The data were grouped by product category and ordered by sales so that best-selling categories were listed at the top.

Based on the output from the first query, two follow-up queries were conducted. Since the product category “Branded” showed a low quantity sold but decent sales numbers, a query was built to extract the detailed product names belonging to the branded category. The product detail table was joined with product type and product category, and a “where” clause was used to filter for only items corresponding to the branded category. This resulted in the identification of three products.

Since sales of items in the tea category are second only to coffee sales, further investigation of tea offerings was conducted. The same query as above was implemented, except that the “where” clause specified the tea category instead of the branded one. The list of tea product details helped to inform the general discussion of current product offerings and potential recommendations for future expansion.

### 3. Within the top selling categories, what are the best-selling products?

This set of queries builds on knowledge gained from previous queries, i.e., that coffee, tea, and bakery are the top categories. As before, a series of joins was performed to relate the transaction, product detail, product type, and product category tables. Sum of quantity and sum of calculated sales columns were selected along with product detail. Product category filters were applied to only select products within coffee, tea, and bakery categories. Records were grouped by product detail and ordered by quantity in descending order. Both quantity and sales were examined and discussed.

#### 4. What insights can be gleaned from the worst-selling product category?

The worst-selling product category was discovered to be packaged chocolate from previous queries, so, the first query of this section examined the quantity of packaged chocolate sold over time. The transaction table was joined with product detail, product type, and product category. Results were limited by specifying the packaged chocolate category in the “where” clause. The sums of quantities sold per product detail per month were returned. In this way, the performance of the three different types of packaged chocolate could be assessed.

A second query was written to create a basis of comparison for the packaged chocolate by looking at the top selling product: regular lattes. This query selected month, product detail name and sum of quantity from the transaction table joined with the product detail table. Results were limited to regular lattes and were grouped by month. It was observed that the overall pattern in quantity sold echoed that of packaged chocolate, though at a much larger scale.

Finally, a loop of queries was constructed to look at quantities of items sold in all product categories. A loop was used instead of one large query so that multiple smaller tables were returned for each different product category, which made the results easier to look at and interpret. Transaction date, product category, and quantity were returned from the table built from joining transactions, product detail, product type, and product category. Results were grouped by month.

#### 5. Are there any times of particularly high or low sales volume?

The first query under this topic sought to identify the busiest and quietest sales times without distinguishing between store locations. Transactions were grouped by hour, and the sum of quantity and sum of calculated sales were selected. Records were ordered by descending order of sales. Relatively low sales at certain hours prompted an inquiry into whether all store locations were open at all times. A query was written to return time by hour and store location from the transaction table joined with the store table. The results indicated that the Astoria location had more limited hours than the other two shops. Therefore, the potential for further discrepancies between the locations was investigated by way of a third query. The transaction table was again joined with the store table. Sum of quantities and sum of calculated sales were selected, grouped by hour and store location.