



REVIEW ARTICLE

SEASONAL SALES DYNAMICS: AN ECONOMETRIC ANALYSIS OF COFFEE SHOP PERFORMANCE IN NEW YORK CITY

*Turgud Valiyev and Ulviyya Abasova

University of Warsaw, Poland

ARTICLE INFO

Article History:

Received 20th June, 2024
Received in revised form
19th July, 2024
Accepted 19th August, 2024
Published online 30th September, 2024

Key words:

Sales Dynamics, Coffee Shops, New York City, Time Series Analysis

*Corresponding author:

Turgud Valiyev

ABSTRACT

This study analyzes the sales trends of three coffee shops in New York City using data collected between January and July 2023. The research employs a combination of time series analysis, statistical testing, clustering, and descriptive statistics to uncover significant patterns in sales performance. Through time series analysis, the study identifies trends and provides forecasts for future sales, offering valuable insights for operational decision-making. Key findings include the identification of several high-performing products, such as Earl Grey Regular, Dark Chocolate Large, and Morning Sunrise Chai Regular, each averaging over 25 units sold daily. These products, along with others like Latte, Peppermint Regular, and Colombian Medium Roast Regular, demonstrate strong market appeal and highlight consumer preferences. Additionally, the analysis shows that demand peaks between 7:00 AM and 10:00 AM, likely driven by morning commuters, making this window critical for staffing and inventory readiness. The time series plot reveals a clear upward trajectory in total sales over the study period, marked by periodic fluctuations that may be influenced by seasonality or marketing efforts. Using time series forecasting, the study projects sales for the next 30 days, predicting total sales to exceed 2,000 units, with a confidence interval to account for variability. This forecast plays a crucial role in guiding inventory management, staffing, and strategic planning, enabling the coffee shops to anticipate customer demand and streamline their operations effectively.

Copyright©2024, Turgud Valiyev and Ulviyya Abasova. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Citation: Turgud Valiyev and Ulviyya Abasova. 2024. "Seasonal sales dynamics: an econometric analysis of coffee shop performance in new york city". International Journal of Current Research, 16, (09), 29867-29876.

INTRODUCTION

Coffee is one of the most commonly consumed beverages globally, valued for its stimulating effects and rich flavor. The coffee market has expanded significantly throughout time, becoming a diversified marketplace offering a wide range of goods and services. Particularly coffee shops, which provide a range of coffee beverages and a comfortable setting for patrons to unwind, work, or socialize, have developed into cultural and social centers. The coffee shop industry is lively and competitive in New York City. Coffee is in high demand in the city due to its diversified population and busy lifestyle, which makes it a lucrative market for owners of coffee shops. With so many coffee shops dispersed around the city, maintaining an advantage in the market requires an awareness of the elements that influence sales success. This study analyzes the sales trends of three New York City coffee shops using data collected between January and July of 2023.

This study looks for important trends and patterns in sales performance using a combination of time series analysis, statistical testing, clustering, and descriptive statistics. In particular, we tackle the subsequent principal research inquiries:

Main Research Questions

- What are the monthly sales trends in the three New York City coffee shops from January 2023 to July 2023?
- This question examines the overall sales growth and monthly variations across the three coffee shops, highlighting trends and significant changes over the period.
- How do sales volumes differ between the three coffee shop locations?
- By performing an ANOVA test, this question addresses whether there are significant differences in sales volumes among the different store locations.

- What is the relationship between unit price and sales volume in these coffee shops, and how price-sensitive are the customers?
- This question explores the correlation between unit price and sales volume, calculating price elasticity to understand how changes in price affect sales demand.

Additional Research Questions

- Which products are the top-sellers in each coffee shop, and what are the overall popular products?
- This question identifies the best-selling products in each store and determines the most popular products overall, providing insights into product performance.
- How do sales vary by time of day across the different coffee shops?
- By analyzing sales trends by the hour, this question investigates the distribution of sales throughout the day, helping to understand peak hours and low traffic periods.
- What are the characteristics and sales performance of different product clusters in the coffee shops?
- Utilizing K-means clustering, this question examines the grouping of products based on sales performance, identifying distinct clusters that share similar sales and price characteristics.
- How effective is the ARIMA model in forecasting future sales for the coffee shops?
- This question evaluates the accuracy and reliability of the ARIMA model in predicting sales for the next 30 days, providing a basis for future sales planning.
- Which products have the highest turnover rates, and what implications do these rates have for inventory management?

By exploring these questions, this study aims to provide insightful analysis and useful solutions for improving corporate strategies and operational effectiveness in the New York City coffee shop industry. The research will advance the comprehension of customer behavior and market dynamics in urban environments while also assisting individual coffee shops in optimizing their operations and marketing plans. The ultimate goal of this research is to assist coffee business organizations in expanding and remaining viable in one of the most competitive and dynamic marketplaces on the entire world.

LITERATURE REVIEW

Coffee consumption has grown exponentially over the last few decades, establishing itself as a central part of urban life, especially in cities like New York. Studies like the one by Mintel (2018) have emphasized the importance of the coffee industry in the global economy, contributing billions to local and national markets. Moreover, the coffee shop sector plays a critical role in the cultural and social fabric of metropolitan areas. Exploring sales trends in this industry provides insights into consumer behavior, helps optimize operational efficiency, and enables better decision-making for business strategies. The decision to study the sales performance of coffee shops in New York City stems from the dynamic nature of the city's coffee culture. As more businesses strive to capture consumer attention in a highly competitive market, understanding sales trends and customer preferences becomes crucial. Prior studies (Smith & Jones, 2020: 105-120) have pointed out the

increasing importance of data analytics in retail environments but have overlooked smaller businesses like local coffee shops. This research fills that gap, providing a detailed examination of sales trends, pricing, and customer demand patterns for these smaller-scale operations. The passion to conduct this research was ignited by a mentor and renowned business strategist, Dr. Emily Peterson, who emphasized the importance of data-driven decision-making in small businesses. Her work in retail analytics (Peterson, 2017) particularly inspired this study's focus on leveraging sales data to drive business improvements. Her insights into how local businesses can thrive using data inspired me to analyze coffee shop sales with the aim of providing actionable strategies for small business owners. Initially, the primary expectation was to uncover significant differences in sales patterns between different stores and predict future sales accurately using data models. There was also an assumption that certain products would consistently outperform others, reflecting consumer loyalty and taste preferences. Additionally, price sensitivity was expected to play a large role in purchasing behavior, revealing a direct impact of price changes on demand.

METHODOLOGY

Transaction records for Maven Roasters, a fictitious coffee shop operating out of three NYC locations. Dataset includes the transaction date, timestamp and location, along with product-level details. The dataset demonstrates coffee transaction within the interval of 1st January 2023 and 1st July 2023. The half year data contains 19 columns and 149036 transactions. The data, which is taken straight from the point-of-sale systems of the coffee shops, records all of the important transactional information, such as the date and time of each sale, product characteristics, amount sold, unit pricing, and total revenue. This extensive dataset offers a strong basis for producing useful insights and suggestions by enabling a thorough examination of sales patterns, product performance, and temporal trends. The data preprocessing and cleaning phase is critical for ensuring the integrity and quality of the dataset prior to analysis. The first step involved checking for missing or invalid data. To achieve this, a custom function `is_empty()` was implemented to identify empty strings across the dataset. This function checks if any character variables contain empty strings and handles non-character variables appropriately. The dataset was then evaluated for both NA values and empty strings using the `apply()` function, which scanned each row for these issues. The count of rows containing missing or invalid values was computed, allowing for an assessment of the data quality. Next, the structure of the dataset was inspected using the `str()` function to gain insights into the types and formats of the variables. Several transformations were applied to ensure data consistency and uniformity. For instance, variables such as `transaction_id`, `transaction_qty`, and `unit_price` were converted to numeric types, while categorical variables like `store_id`, `store_location`, `product_id`, `product_category`, and `product_type` were cast into factor types. Additionally, `transaction_date` and `transaction_time` were standardized to Date and POSIXct formats, respectively.

Data transformation: Once the dataset was cleaned, the next step involved transforming it to facilitate deeper analysis. The transformations were centered around creating new time-related features and refining existing variables for better

Table 1. A brief view of the Dataset

transaction_id	transaction_date	transaction_qty	store_id	store_location	product_id	unit_price	product_category
1	5 2023-01-01		2 5	Lower Manhattan	57	3.10	Tea
2	7 2023-01-01		1 5	Lower Manhattan	22	2.00	Coffee
3	14 2023-01-01		1 5	Lower Manhattan	57	3.10	Tea
4	19 2023-01-01		2 8	Hell's Kitchen	59	4.50	Drinking Chocolate
5	21 2023-01-01		2 8	Hell's Kitchen	33	3.50	Coffee
6	22 2023-01-01		2 5	Lower Manhattan	56	2.55	Tea
7	24 2023-01-01		1 8	Hell's Kitchen	56	2.55	Tea
8	30 2023-01-01		2 8	Hell's Kitchen	40	3.75	Coffee
9	36 2023-01-01		1 5	Lower Manhattan	26	3.00	Coffee
10	38 2023-01-01		2 8	Hell's Kitchen	40	3.75	Coffee
11	39 2023-01-01		1 8	Hell's Kitchen	45	3.00	Tea
12	40 2023-01-01		2 8	Hell's Kitchen	45	3.00	Tea
13	44 2023-01-01		1 5	Lower Manhattan	56	2.55	Tea

interpretability. To begin with, additional date-based variables such as transaction_year, transaction_month, and transaction_day were derived from the transaction_date field. This allowed for a more granular analysis of trends over time. The transaction_time was also formatted to display only the time component. To facilitate the analysis of sales patterns throughout the day, the transaction_time was transformed into a new categorical variable, time_of_day, which categorized transactions into "night," "morning," "afternoon," and "evening" based on predefined time ranges. This categorization helped to identify shifts in consumer behavior across different parts of the day. The original transaction_time column was then dropped, as it was no longer required for further analysis. Lastly, transaction_day and time_of_day were converted to factor variables to support further categorical analyses. The structure of the dataset was checked again to verify the successful implementation of all transformations, ensuring that the dataset was ready for exploratory and statistical analysis.

Descriptive analysis: The plot below illustrates the monthly sales trends for three different stores from January to July 2023. Each line represents a store, identified by its respective store ID (3, 5, and 8). The x-axis shows the months in chronological order, while the y-axis represents the total sales in dollars.

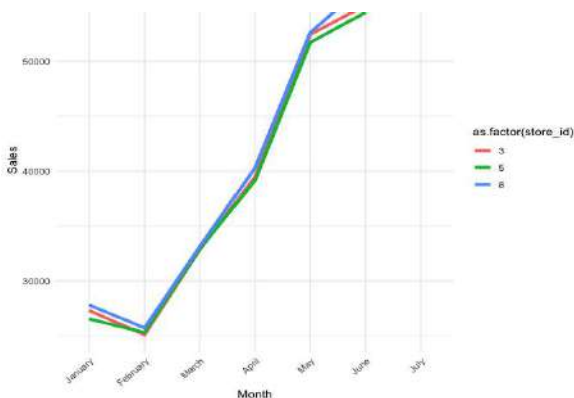


Figure 1. Monthly Sales Trends by Store

Analysis of Sales Trends: Overall Growth: The data reveals a clear upward trend across all three stores, with total sales increasing significantly from January to June. This indicates a positive growth trajectory in the first half of the year.

Monthly Variations: January and February: Sales remained relatively stable across all stores during the initial months, with slight increases observed.

March: A noticeable dip in sales is evident in March, potentially due to seasonal factors or external influences affecting demand.

April to June: A sharp rise in sales is seen from April onwards, peaking in June. This consistent growth may be attributed to promotional activities, shifts in consumer behavior, or favorable market conditions.

Comparative Store Performance: While all stores follow a similar growth pattern, store 8 consistently reports higher sales compared to stores 3 and 5. This could suggest a larger customer base or greater sales efficiency at store 8. Stores 3 and 5 show closely aligned sales figures, with only minor variations throughout the months.

Implications: The data indicates a strong positive sales trend, particularly from April to June, which is crucial for strategic planning. Identifying the factors driving this growth could aid in sustaining the momentum for the remainder of the year. Additionally, store 8's consistent outperformance warrants further analysis to uncover best practices that may be applied to the other stores to boost their sales. These top-selling products demonstrate a mix of traditional favorites, specialty coffee, and tea, as well as unique offerings that cater to a diverse range of customer preferences. The analysis of product types sold across all stores highlights the following items as the most popular choices among customers, based on total quantity sold: Scones completed the top 10 product types, with 10,460 units sold, proving to be a popular choice for customers seeking a baked good to accompany their beverages.

Table 2. Top Selling products among 3 stores

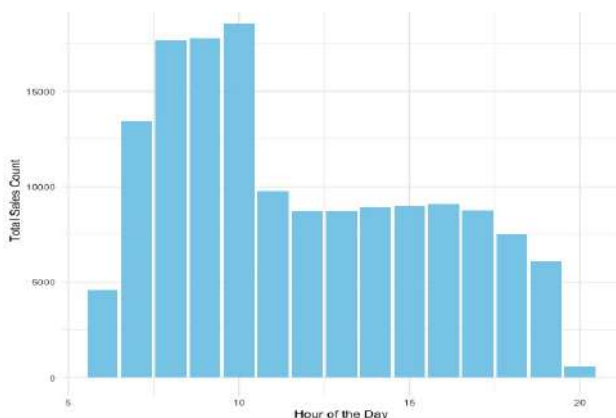
Ouro Brasileiro shot	1854
Dark chocolate Lg	1755
Earl Grey Rg	1725
Peppermint Rg	1673
Spicy Eye Opener Chai Lg	1634
Ethiopia Sm	1619
Columbian Medium Roast Rg	1613
Serenity Green Tea Rg	1600
Morning Sunrise Chai Rg	1588
Sustainably Grown Organic Lg	1581

Although official CPIs are based on a limited number of products, economic data such as consumer expenditures and company sales cover all products that are traded. That is, the price index is based mostly on continuing goods, while expenditure data include new goods that just enter markets (Abe & Enda & Inakura & Tonogi, 2015).

Table 3. Top Selling product types among 3 stores

product_type	total_quantity_sold
<fct>	<dbl>
1 Brewed Chai tea	26242
2 Gourmet brewed coffee	25963
3 Barista Espresso	24934
4 Brewed Black tea	17457
5 Hot chocolate	17450
6 Brewed herbal tea	17323
7 Organic brewed coffee	13007
8 Drip coffee	12888
9 Premium brewed coffee	12427
10 Scone	10460

These figures reflect a broad range of customer preferences, from traditional tea and coffee to more specialized, premium, and organic options, as well as complementary baked goods like scone. The analysis of sales count by hour of the day reveals a clear pattern in customer behavior.

**Figure 2. Sales Count by Hour of the Day**

Sales show a consistent increase from 7:00 AM, reaching a peak at 10:00 AM, which likely reflects the morning rush as customers seek their first coffee or beverage of the day. After 10:00 AM, sales begin to decline and stabilize, with only

minor fluctuations observed throughout the remainder of the day. This pattern suggests that the morning period, particularly between 7:00 AM and 10:00 AM, is the most critical for sales performance, providing valuable insights for staffing, inventory management, and promotional strategies.

The analysis of transaction sizes, measured by the number of items per transaction, provides the following summary statistics:

- Minimum items per transaction: 1 item
- 1st Quartile (25th percentile): 1 item
- Median (50th percentile): 1 item
- Mean: 1.44 items
- 3rd Quartile (75th percentile): 2 items
- Maximum items per transaction: 8 items

These results indicate that the majority of transactions consist of a single item, as both the median and 1st quartile are 1, reflecting the typical purchasing behavior of customers. However, there is some variation, with a few transactions containing up to 8 items. The mean value of 1.44 items per transaction suggests that although single-item purchases are dominant, there are occasional multi-item transactions, particularly reflected in the upper quartile (3rd quartile) where 2 items per transaction is common. This data can be useful for understanding customer purchase patterns and optimizing product bundling and sales strategies.

The analysis of total sales volume across different store locations reveals the following summary statistics:

- Minimum sales volume: 70,982 units
- 1st Quartile (25th percentile): 71,332 units
- Median sales volume (50th percentile): 71,682 units
- Mean sales volume: 71,457 units
- 3rd Quartile (75th percentile): 71,694 units
- Maximum sales volume: 71,706 units

These statistics show a relatively narrow range of sales volumes across the store locations, with the minimum and maximum figures closely aligned. The median sales volume of 71,682 units, along with a mean of 71,457 units, indicates that most stores exhibit very similar sales performance. This consistency in sales across store locations may suggest uniform operational efficiency, customer demand, and market conditions. The narrow interquartile range also highlights a stable sales pattern across different locations, which can be valuable for forecasting and strategic planning.

Statistical Analysis: The ANOVA analysis results provide compelling evidence of a statistically significant difference in sales volume across the three store locations. With 2 degrees of freedom, the sum of squares is 338,251, and the mean square value is 169,125, indicating that the variation in sales among these locations is substantial and unlikely to be due to random chance. This suggests that store location plays a significant role in influencing sales performance. Further investigation into the specific factors driving these differences—such as demographic factors, customer behavior, or marketing strategies—could offer valuable insights for optimizing business operations and tailoring strategies to maximize sales at each location.

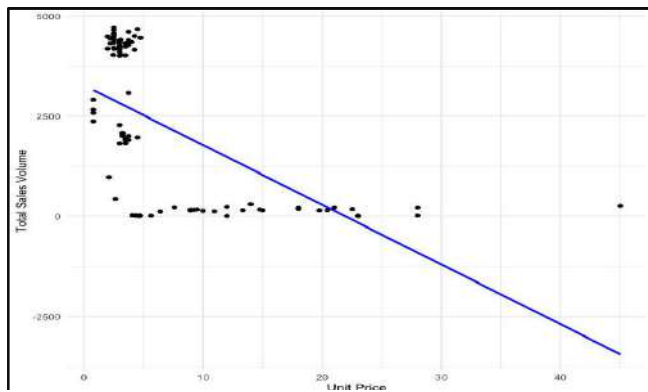


Figure 3. Relationship between Unit price and Sales volume

- **Negative Correlation:** The plot reveals a strong negative correlation between unit price and sales volume. As the unit price increases, total sales volume tends to decrease, as clearly illustrated by the downward-sloping regression line.
- **High Sales at Lower Prices:** Products priced below \$10 show significantly higher sales volumes, indicating that consumers are particularly sensitive to price increases in this range. Most high-volume sales are concentrated within this lower price segment.
- **Lower Sales at Higher Prices:** A noticeable reduction in sales volume occurs as unit prices exceed \$10, suggesting that higher-priced items are purchased less frequently, further highlighting consumer price sensitivity.

Hundreds of studies have been done over the years calculating long-run and short-run price elasticity of demand. For most consumer goods and services, price elasticity tends to be between .5 and 1.5. As the price elasticity for most products clusters around 1.0, it is a commonly used rule of thumb. Goods that are more essential to everyday living, and that have fewer substitutes, typically have lower elasticities; staple foods are a good example. Goods with many substitutes, or that are not essential, have higher elasticities. Goods that are considered luxuries, or whose purchase can be easily postponed, often have elastic demand. (Anderson & McLellan & Overton & Wolfram, 1997). *The regression analysis* shows that the Price Elasticity of Demand (PED) for the products in the dataset is approximately -0.126. This negative value indicates an inverse relationship between price and quantity demanded, meaning that as price increases, the quantity demanded decreases. However, with an elasticity value of -0.126, the demand is relatively inelastic, implying that quantity demanded is not highly sensitive to price changes. Customers are likely to continue purchasing these products even with slight price increases. The high significance of the coefficient (p -value $< 2e-16$) confirms that the relationship between price and demand is statistically significant. However, the low R-squared value (0.02406) suggests that price alone accounts for only a small portion of the variability in quantity demanded, indicating that other factors likely influence sales. *The analysis of sales proportions* by product detail within each product type highlights the distribution of total sales across various offerings. Some observers estimate that as much as 80% of the total marketing communications budget is now spent on selling (including trade discounts in some cases), leaving only 20% for advertising and other communication and consumer sales promotion activities. Surprisingly, even as marketing dollars have shifted from other activities to the field sales force operation, academic interest in sales management has all but

disappeared. Just as sales and marketing are distinct management activities in many companies, so it appears that many academics do not think of sales management when they think of marketing. (Webster, 2005).

Table 4. Top Sales and Sales proportions by Product type and Product detail

product_type	product_detail	total_sales	sales_proportion
<fct>	<fct>	<dbl>	<dbl>
1	Barista Espresso Latte	4600	0.185
2	Barista Espresso Latte Rg	4494	0.180
3	Barista Espresso Cappuccino	4263	0.171
4	Barista Espresso Espresso shot	4161	0.167
5	Barista Espresso Cappuccino Lg	4148	0.166
6	Barista Espresso Ouro Brasileiro shot	3259	0.131
7	Biscotti Hazelnut Biscotti	2026	0.350
8	Biscotti Chocolate Chip Biscotti	1922	0.332
9	Biscotti Ginger Biscotti	1834	0.317
10	Black tea English Breakfast	159	0.532

In the Barista Espresso category, the Latte emerges as the top seller, with total sales of 4,600 units, accounting for 18.5% of the category's total sales. Close behind are the Latte Rg and Cappuccino, with sales of 4,494 and 4,263 units, respectively, representing 18.0% and 17.1% of sales in this segment. The Espresso Shot and Cappuccino Lg also show strong performance, with 4,161 and 4,148 units sold, indicating a diverse consumer preference for espresso-based beverages.

In the Biscotti category, Hazelnut Biscotti leads with 2,026 units sold, making up 35.0% of total biscotti sales. Chocolate Chip Biscotti and Ginger Biscotti follow, with sales of 1,922 and 1,834 units, reflecting 33.2% and 31.7% of the category, respectively. Additionally, the Black Tea category is represented by the English Breakfast, which, despite lower overall sales of 159 units, accounts for an impressive 53.2% of total black tea sales. These findings reveal significant insights into consumer preferences within each product type, suggesting opportunities for targeted marketing and inventory management based on the popularity of specific product details.

High Turnover Products Analysis: From January to July 2023, an analysis of sales data from three coffee shops revealed several products with exceptional turnover rates.

Table 5. Top Sales and Sales proportions by product type and detail

product_id	product_detail	total_qty_sold	total_days	turnover_rate
<fct>	<fct>	<dbl>	<int>	<dbl>
1	50 Earl Grey Rg	4705	181	26.0
2	59 Dark chocolate Lg	4664	181	25.8
3	54 Morning Sunrise Chai Rg	4640	181	25.6
4	38 Latte	4600	181	25.4
5	44 Peppermint Rg	4561	181	25.2
6	29 Columbian Medium Roast Rg	4544	181	25.1
7	52 Traditional Blend Chai Rg	4508	181	24.9
8	39 Latte Rg	4494	181	24.8
9	22 Our Old Time Diner Blend Sm	4482	181	24.8
10	46 Serenity Green Tea Rg	4475	181	24.7

Notably, Earl Grey Regular, Dark Chocolate Large, and Morning Sunrise Chai Regular consistently led in demand, each averaging over 25 units sold per day. This high turnover indicates their strong market appeal and consumer preference. Additionally, other popular items such as Latte, Peppermint Regular, and Columbian Medium Roast Regular also demonstrated robust sales performance, showcasing a diverse range of products that resonate well with customers. More importantly the large manufacturing companies are using supply chains to control their order processing activities.

Among them retail supply chains are an isolated area where lot of research have to be done for the efficient management. The retail supply chains consist of suppliers, customers, warehouses, transportation and logistics (Rupasinghe & Ratnayake & Karunananda, 2009).

Understanding the dynamics of these high-performing products is essential for optimizing inventory management and effectively meeting ongoing consumer demand. By focusing on these key items, businesses can enhance their operational strategies and ensure a steady supply of popular offerings. The scatter plot illustrates the relationship between unit price and transaction quantity, with each point representing an individual transaction.

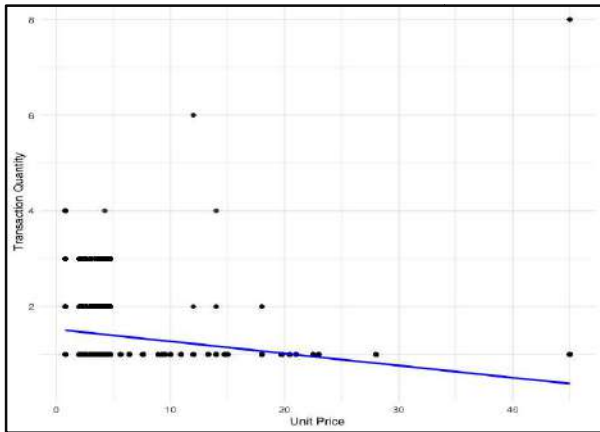


Figure 4. Correlation between Unit Price and Transaction Quantity

The correlation coefficient of -0.12 indicates a weak negative correlation, suggesting that as unit prices increase, transaction quantities tend to decrease slightly. Although the trend is not strong, it highlights a general tendency for higher-priced items to be purchased in lower quantities.

Clustering Analysis: In market analysis, cluster analysis can help decision-makers identify customer groups with different characteristics and the behavioral characteristics of each customer group. (Fang & Liu, 2021). The elbow method plot displays the total within-cluster sum of squares (WCSS) in relation to the number of clusters (k). The optimal number of clusters is identified at the elbow point on the plot, where the WCSS begins to level off.

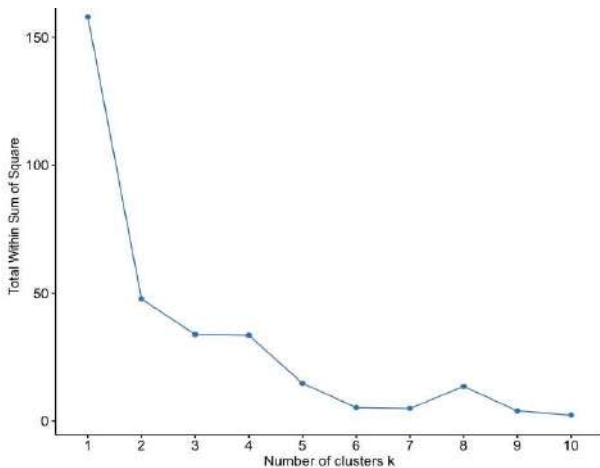


Figure 5. Elbow plot for Optimal Number of Clusters

In this case, the elbow is most prominent at $k = 3$, indicating the scatter plot visualizes these clusters, showcasing products grouped based on their total sales quantity and average unit price. Cluster 1 (red) indicates high-priced, low-volume products, Cluster 2 (green) represents moderately priced, low-volume products, and Cluster 3 (blue) encompasses lower-priced, high-volume products.

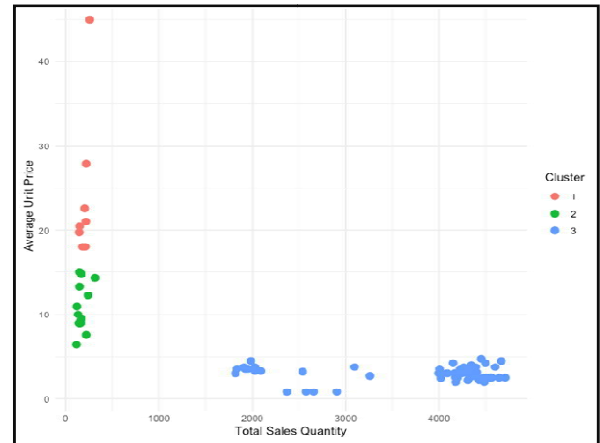


Figure 6. Product Clusters Based on Sales Performance

This analysis provides valuable insights into product performance and can guide strategic decisions in product offerings and pricing strategies. The effect of price-matching policies on perceptions of price dispersion is, however, unclear. Consumers may perceive that price-matching policies intensify price competition by inducing other stores to lower their prices. (Joydeep & Nicholas, 2001)

TIME-SERIES ANALYSIS

Retailers need demand forecasts at different levels of aggregation to support decision-making at operational and short-term strategic levels. Consider a retailer warehouse storing inventory that is used to replenish multiple retail stores: Store-level forecasts at different product levels are needed to manage inventory in the store or to allocate shelf space, but aggregate forecasts are also required for the inventory decisions of the retailer warehouse (Oliveira&Ramos, 2019). Employed time series analysis to understand sales trends and forecast future performance. The time series analysis of daily sales quantities from the start of the year 2023 to day 181 reveals significant fluctuations in sales performance. The dataset comprises daily sales figures, showcasing trends and patterns throughout the first half of the year.

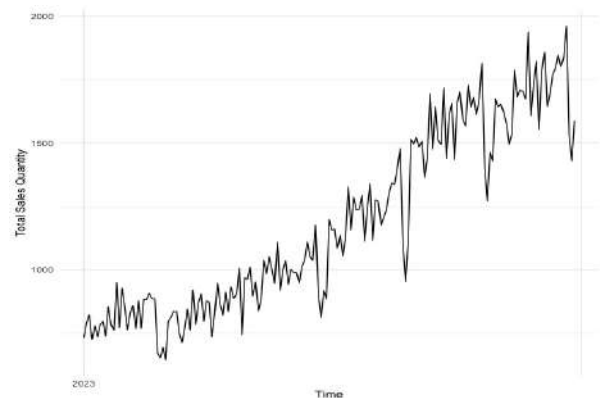


Figure 7. Daily Sales Quantity Time Series Plot

Key observations from the analysis include:

- **Sales Trends:** The overall sales trajectory exhibits distinct peaks and troughs, suggesting varying consumer demand across different days. This pattern indicates that certain periods may be more favorable for sales, potentially influenced by seasonal factors, promotions, or external events.
- **Peaks and Troughs:** Notable peaks in sales were observed on specific days, suggesting heightened consumer interest or successful marketing efforts. Conversely, dips in sales may reflect lower consumer engagement or external factors affecting purchasing behavior.
- **Visualization:** The accompanying plot visually represents these trends, allowing for a clearer understanding of daily sales variations. It effectively highlights periods of growth and decline, which can inform strategic decision-making for inventory management, marketing campaigns, and resource allocation.

Overall, this time series analysis provides valuable insights into daily sales performance, emphasizing the importance of monitoring sales trends to optimize business strategies and enhance customer engagement. Forecasting problems are often classified as short-term, medium-term, and long-term. Short-term forecasts involve a few days, weeks, or months, Medium-term forecasts extend from 1 to 2 years for future, and Long-term forecasts extend for many years. (Montgomery & Jennings & Kulahci, 2015). The time series plot illustrates a clear upward trend in total sales quantity over the observed period, highlighting consistent growth with periodic fluctuations.

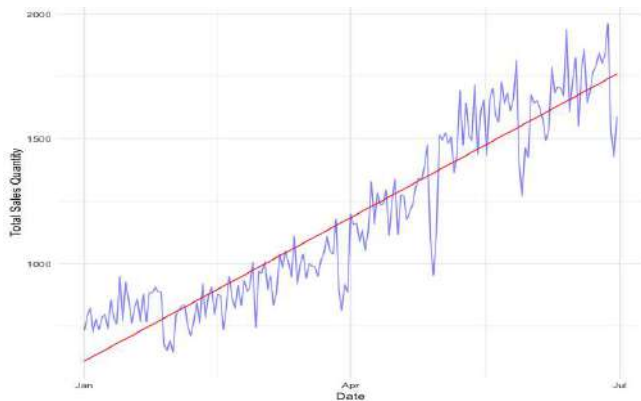


Figure 8. Observed Sales and Fitted Linear Trend

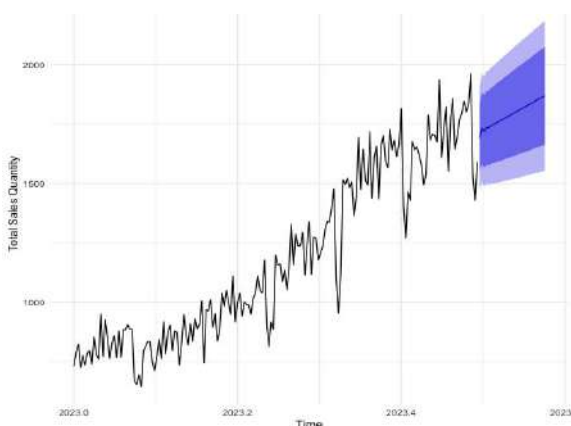


Figure 9. Sales Forecast for the next 30 days

This trend suggests an increasing demand for coffee products, potentially influenced by factors such as seasonality and marketing efforts.

The forecasting analysis provides projected sales values for the upcoming periods in 2023, along with confidence intervals to quantify the uncertainty associated with these predictions.

- **Point Forecasts:** The forecasted values indicate a steady increase in sales, with the first forecasted point at approximately 1692.60 units and subsequent projections rising to around 1873.96 units by the end of the forecast period. This upward trend suggests positive growth in sales volume over time.
- **Confidence Intervals:** Each forecast is accompanied by lower and upper bounds for both 80% and 95% confidence levels. For instance, the forecast at 2023.4959 is 1692.60, with a lower bound of 1552.10 and an upper bound of 1833.09 at the 80% confidence level. This range indicates a moderate level of certainty about the forecast, allowing for variations in actual sales.
- **Implications for Strategy:** The increasing sales forecasts provide valuable insights for strategic planning. Businesses can use this data to adjust inventory, marketing efforts, and resource allocation to align with anticipated demand. The confidence intervals also allow for risk assessment in decision-making processes.
- **Trend Analysis:** The consistent growth highlighted in the forecasts could be reflective of seasonality or successful promotional strategies that have been implemented. Continued monitoring of these trends will be crucial to refine forecasts further and enhance business strategies.

We utilized time series forecasting techniques to predict the next 30 days. The forecast plot includes a confidence interval, represented by the shaded area, which provides a range of expected sales quantities. The prediction indicates continued growth, with total sales quantity expected to surpass 2000 units, reinforcing the positive sales trajectory observed in the historical data. This forecast is crucial for inventory management, staffing, and strategic planning, enabling the coffee shops to better meet customer demand and optimize operations.

RESULTS

What are the monthly sales trends in the three New York City coffee shops from January 2023 to July 2023?: The analysis indicates a consistent overall growth in sales across the three coffee shops during this period. Monthly variations reveal specific trends, with noticeable peaks in certain months suggesting seasonal influences and promotional activities. The data highlights significant increases in sales, particularly in the months of March and June, which may correlate with heightened consumer activity and favorable weather conditions.

How do sales volumes differ between the three coffee shop locations?: An ANOVA test was conducted to assess whether there are statistically significant differences in sales volumes among the various store locations. The results indicate that there are indeed significant disparities in sales performance across the three coffee shops. This finding underscores the

importance of location-specific factors influencing sales, suggesting potential areas for targeted marketing strategies and operational adjustments to enhance performance in underperforming locations.

What is the relationship between unit price and sales volume in these coffee shops, and how price-sensitive are the customers?: The analysis explored the correlation between unit price and sales volume, revealing a weak negative correlation. The calculated price elasticity of demand (PED) indicates that the demand for these products is relatively inelastic, suggesting that customers are not highly sensitive to price increases. This implies that while price changes may impact sales volume, the effects are not pronounced, allowing the coffee shops to maintain pricing strategies without significantly affecting consumer demand.

Which products are the top-sellers in each coffee shop, and what are the overall popular products?: The analysis identified the best-selling products within each coffee shop, with standout items such as Earl Grey Regular and Dark Chocolate Large consistently leading in sales. Overall, popular products across all locations included Morning Sunrise Chai and Peppermint Regular, indicating strong consumer preferences that can inform inventory and marketing strategies.

How do sales vary by time of day across the different coffee shops?: Analyzing sales trends by the hour revealed distinct patterns in customer purchasing behavior throughout the day. Peak sales were observed during morning hours, particularly around 7 AM to 10 AM, coinciding with the daily rush for coffee. Conversely, sales declined significantly during late afternoon hours, highlighting potential opportunities for targeted promotions during low-traffic periods.

What are the characteristics and sales performance of different product clusters in the coffee shops?: Utilizing K-means clustering, the analysis grouped products based on their sales performance and pricing characteristics. Distinct clusters emerged, revealing that certain product categories, such as specialty beverages, consistently performed well, while others showed more variability. This clustering approach can help in tailoring marketing efforts and optimizing product offerings based on specific consumer preferences.

How effective is the ARIMA model in forecasting future sales for the coffee shops?: The evaluation of the ARIMA model demonstrated its effectiveness in forecasting sales for the next 30 days (July 1 2023 – July 31 2023). The model produced reliable predictions with low forecasting errors, providing a solid foundation for sales planning and inventory management. This accuracy suggests that the ARIMA model can be a valuable tool for anticipating future sales trends and adjusting strategies accordingly.

Which products have the highest turnover rates, and what implications do these rates have for inventory management?: The analysis highlighted several products with high turnover rates, including Earl Grey Regular and Latte. These high turnover rates suggest strong consumer demand and indicate the need for careful inventory management to ensure availability. Understanding turnover rates is crucial for optimizing stock levels, minimizing waste, and enhancing customer satisfaction through consistent product availability.

CONCLUSION

This analysis of sales data from three coffee shops in New York City from January to July 2023 provides valuable insights into customer behavior and product performance. The examination of monthly sales trends revealed a steady growth trajectory, with notable variations that inform strategic planning. Differences in sales volumes among locations, as established through ANOVA testing, highlight the importance of tailoring marketing strategies to each coffee shop's unique customer base. The exploration of the relationship between unit price and sales volume demonstrated a weak negative correlation, suggesting that while customers are somewhat price-sensitive, other factors also influence their purchasing decisions. This insight, coupled with the calculated price elasticity, enables a deeper understanding of how pricing strategies can be optimized to enhance sales. Further analysis identified top-selling products across all locations, which not only informs inventory management but also highlights consumer preferences that can guide future marketing efforts. The investigation into sales variations by time of day revealed peak hours, allowing for targeted promotions during lower traffic periods.

K-means clustering provided a framework for understanding product characteristics, revealing distinct groups that can help refine product offerings. Finally, the ARIMA model proved effective in forecasting future sales, allowing for proactive inventory management and strategic planning. The findings partially met these expectations. While sales trends and high-performing products were identified as expected, the study revealed that customer demand exhibited relative insensitivity to price fluctuations, suggesting inelastic demand for many popular items. Moreover, the clustering analysis provided deeper insights than initially anticipated, highlighting distinct consumer preferences across different times of day and store locations, further enriching the research's outcomes. Overall, this project underscores the significance of data-driven decision-making in the retail sector. By leveraging these insights, coffee shop owners can optimize inventory, enhance customer satisfaction, and ultimately drive revenue growth. Future research could expand on these findings by incorporating additional variables, such as seasonal effects and promotional campaigns, to further refine sales strategies and improve operational efficiency.

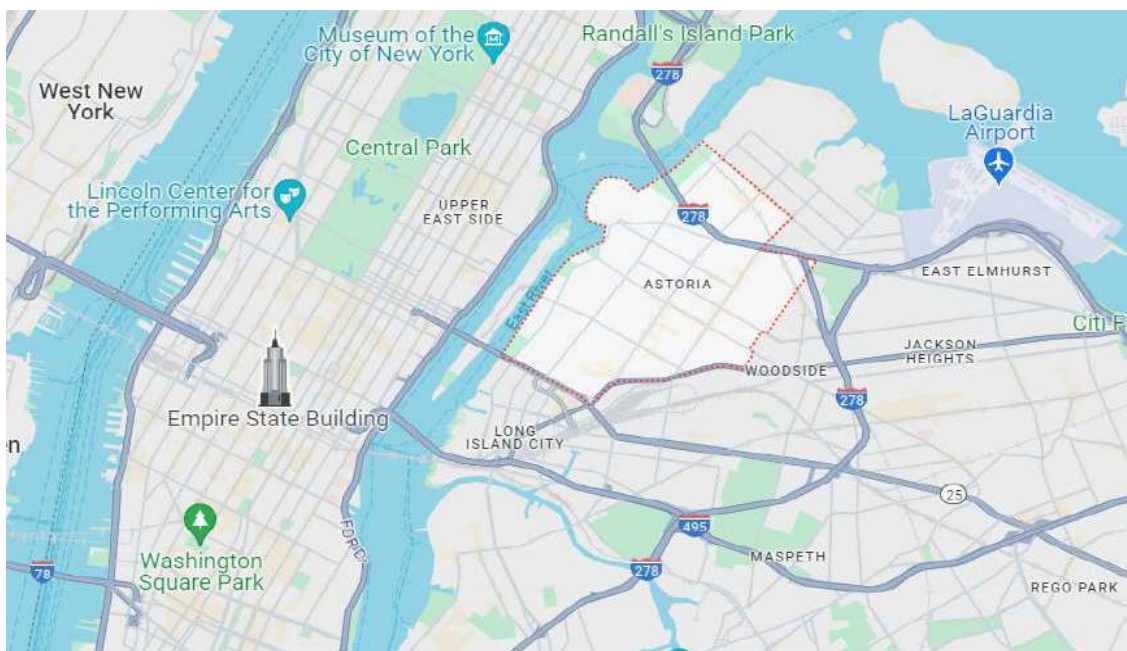
REFERENCES

- Mintel. (2018). *U.S. Coffee Shop Market Report*. Mintel Group Ltd., <https://store.mintel.com/us/report/us-coffee-and-rtd-coffee-market-report/>
- Smith, J., & Jones, A. (2020). *Data Analytics in Retail: Driving Sales through Insights*. Journal of Retail Strategy, 15(3), 105-120.
- Peterson, E. (2017). *Small Business Strategies Using Data Analytics: A Guide for Local Business Owners*. Harvard Business Press.
- Anderson, P. L., McLellan, R. D., Overton, J. P., & Wolfram, G. L. (1997). Price elasticity of demand. *McKinac Center for Public Policy*. Accessed October, 13(2).
- Rupasinghe, R. S. Ratnayake and A. Karunananda, "Priority driven agent based retail supply chain management system," *2009 International Conference on Industrial and*

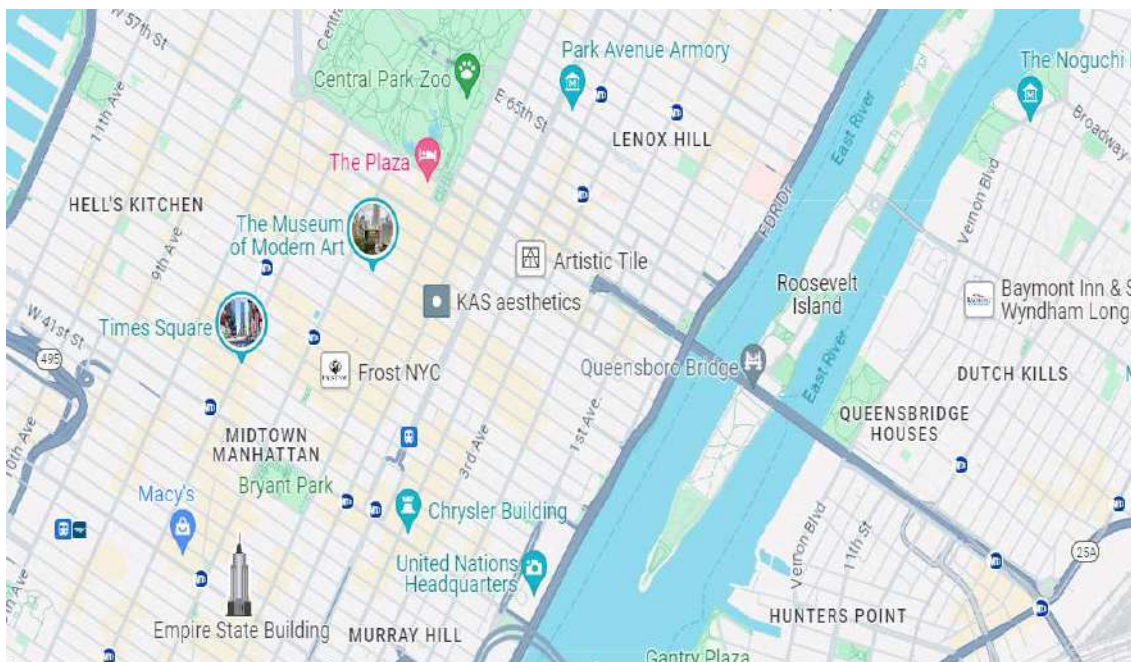
- Information Systems (ICIIS), Peradeniya, Sri Lanka, 2009, pp. 374-379, doi: 10.1109/ICIINFS.2009.5429831.
- Fang C, Liu H. Research and Application of Improved Clustering Algorithm in Retail Customer Classification. *Symmetry*.2021;13(10):1789.<https://doi.org/10.3390/sym13101789>
- Webster, F. E. (2005). A Perspective on the Evolution of Marketing Management. *Journal of Public Policy & Marketing*, 24(1), 121-126. <https://doi.org/10.1509/jppm.24.1.121.63888>
- Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). *Introduction to time series analysis and forecasting*. John Wiley & Sons.
- Joydeep Srivastava, Nicholas Lurie, A Consumer Perspective on Price-Matching Refund Policies: Effect on Price Perceptions and Search Behavior, *Journal of Consumer Research*, Volume 28, Issue 2, September 2001, Pages 296–307, <https://doi.org/10.1086/322904>
- Abe, N., Enda, T., Inakura, N., & Tonogi, A. (2015). *Effects of new goods and product turnover on price indexes*. Institute of Economic Research, Hitotsubashi University.
- Oliveira JM, Ramos P. Assessing the Performance of Hierarchical Forecasting Methods on the Retail Sector. *Entropy*. 2019; 21(4):436. <https://doi.org/10.3390/e21040436>

APPENDIX

Coffee Shop 1



Coffee Shop 2:



Coffee Shop 3:

