IBM Workload 1
Data-Driven Optimization in Coffee Shops: A Business Canvas Approach

1. Introduction

The coffee shop industry is a highly competitive and rapidly evolving market. With rising customer expectations, fluctuating demand patterns, and the increasing dominance of large chains, independent coffee shop owners face significant challenges. Key obstacles include optimizing staffing levels to balance service quality with labor costs, managing fluctuating customer demand during peak and off-peak hours, and making data-driven decisions to enhance operational efficiency. Additionally, external factors such as weather conditions, local events, and seasonal trends further complicate demand forecasting and resource planning.

To address these challenges, this report adopts the Business Model Canvas (BMC) framework as the foundation for analysis. The BMC offers a structured method for assessing nine key business dimensions: customer segments, value propositions, channels, customer relationships, revenue streams, key resources, key activities, key partnerships, and cost structure. By applying this framework, the report provides a holistic view of how coffee shops create and deliver value, while identifying areas for optimization.

Central to this analysis is the application of data-driven insights. By leveraging predictive analytics, including time-series forecasting and multi-objective optimization models, coffee shop owners can make more informed decisions. For example:

- Forecasting models predict customer footfall based on historical sales data and external variables, helping managers anticipate demand fluctuations.
- Optimization models recommend ideal staffing levels to reduce labor costs while maintaining acceptable service quality, striking a balance between profitability and customer satisfaction.

By integrating the BMC framework with data analytics, this report provides actionable insights to help coffee shop owners make evidence-based decisions, improve resource allocation, and achieve a sustainable balance between operational efficiency and customer service quality.

2. Literature Review

2.1 Theoretical background on Business Canvas in retail

The concept of the Business Model Canvas (BMC) was developed by Osterwalder and Pigneur (2010), who began by clearly defining what a business model is. They described a business model as the rationale of how an organization creates, delivers, and captures value.

Rahardjo, Hasbullah and Taqi (2019) emphasized that the BMC is particularly useful for transforming abstract business ideas into visual representations, making it easier for entrepreneurs and decision-makers to communicate and refine strategic plans. In today's fast-paced and innovation-driven markets, the BMC effectively illustrates how enterprises capture value and drive revenue through targeted, customer-focused strategies.

According to the research by Chummee et al. (2022), the BMC method was successfully implemented in the retail sector. It addressed operational challenges—such as reduced sales and inefficiencies—by adjusting the business model to enhance customer engagement and value delivery.

2.2 Forecasting & stochastic modeling in retail

Sales forecasting is a crucial objective for enterprises aiming to manage their operations effectively. Traditional pricing and planning methods often rely heavily on experience and historical data, which limits their ability to respond to real-time market fluctuations and product variability (Li and Xin, 2024). In contrast, modern forecasting techniques frequently utilize time-series datasets, which contain time-related information that supports accurate predictions and statistical analysis (Kumar Jha and Pande, 2021).

In parallel, stochastic modeling offers a way to address uncertainty in resource planning and demand forecasting. Stochastic problems with recourse aim to allocate an optimal number of resources to maximize profit or minimize costs. This is particularly relevant in managing demand fluctuations, where the goal is to avoid incurring additional costs—such as expensive overtime during peak periods or excess staffing during times of low demand (Bisset and Terblanche, 2021).

These methods are especially valuable in the retail sector, where both short-term sales forecasting and long-term resource planning are necessary to maintain efficiency and customer satisfaction in dynamic environments.

2.3 Supporting rationale for chosen methodologies

This report adopts a real-world scenario from an IBM case study—the IBM Coffee Shop Project. Based on actual business data, the project utilizes two predictive approaches: forecasting and stochastic modeling. These AI-driven methods are applied to optimize operational workload and forecast revenue, enabling coffee shop owners to make data-informed cost adjustments and enhance overall efficiency.

In addition, the project uses the IBM Watson Cognos Analytics, specifically the IBM Analytics dashboard, to provide real-time insights into workload dynamics. This allows coffee shop owners to better visualize and understand their operations, supporting more precise, timely decision-making in response to fast-changing business conditions.

3. Data and Methodology

3.1 Data Summary

The dataset used in this project was originally developed by IBM to simulate the operations of a coffee shop chain and was made publicly available via Kaggle for business analytics and educational purposes. It includes synthetic records of customer transactions, product information, and inventory movement. Weather data was gathered through the weather data of the website and the U.S. holiday data was obtained from Kaggle to catch special events.

For this study, Store 3 during April 2019 was selected, which had consistent and complete entries across all relevant tables. Sales receipts provided detailed transaction-level information including date, time, product ID, quantity sold, unit price, staff ID, and customer ID. These were combined with product-level pricing and inventory data to construct variables such as actual sales (adjusted for stock availability), daily customer counts, and estimated employee efficiency. The resulting dataset was used as the foundation for demand forecasting and workforce optimisation modelling.

3.2 Forecasting Model

This project uses the Prophet model to forecast the customer number entering the coffee shop, based on time series aggregated at the daily level. For each day t, the target variable $y(t)$ is defined as the number of unique customers visiting the store.

To capture both temporal and external influences on customer demand, Prophet models the forecasted value as a sum of key components, along with two additional regressors:

$$y(t) = g(t) + s(t) + h(t) + \beta_1 \cdot t \cdot precip_t + \beta_2 \cdot severity_t + \xi(t)$$

Where:

- $g(t)$ models non-periodic trend changes over time
- $s(t)$ represents the periodic changes (e.g., hourly or daily seasonality)
- $h(t)$ represents holiday effects that occur on irregular dates
- $precip_t$ reflects the impact of total precipitation on customer footfall
- $severity_t$ captures the influence of external events such as promotions or service disruptions
- $\xi(t)$ is the error term accounting for random fluctuations not captured by other components

To stabilize the variance in customer counts, particularly where higher volumes tend to exhibit greater variability, a log transformation was applied:

$$y'_t = log(1 + y_t)$$

To represent capacity limits (for example, if the store can be only 1200 people maximum) and to avoid unrealistic forecasts like negative customer numbers or infinite growth, a logistic growth function was used, allowing forecasts to respect the store's capacity limits.

$$G(t) = \frac{C}{1 + exp(-k(t - m))}$$

Where:

- C is the maximum expected value
- k is the growth rate
- m midpoint or ***inflection*** point of growth curve

Further, regressors such as precipitation and Severity were normalised using Z Score, to make sure their approach to the model is unbiased and stop the dominance of high-magnitude features in the model and allow others to contribute as well.

3.3 Multi-Objective Optimisation Model

Simulation of Customer Demand

To account for uncertainty and fluctuations in customer traffic, customer demand was simulated using a log-normal distribution:

$$D \sim LogNormal\ (\mu, \sigma)$$

Where:

- μ = the average number of customers in a given time slot and scenario
- σ = the standard deviation controlling the variability of demand

Table 1 shows the mean demand (μ) used for each scenario and time slot:

Table 1. *Mean Customer Demand*

Scenario	Morning	Afternoon	Evening
Normal Day	62.26	33.58	27.83
Holiday	80	55	50
Promotion	90	70	60

The corresponding standard deviation parameters were set as follows:

$\sigma = 0.48\ (Morning), \sigma = 1.29\ (Afternoon), \sigma = 1.06\ (Evening)$

To avoid unrealistic extremes, simulated values were truncated between the 10th and 92nd percentiles. The resulting demand distributions formed the basis for the workforce optimisation model.

Optimisation Model Framework

The goal of the model is to determine the optimal number of staff per time slot to balance profitability and customer satisfaction. For each simulated demand value, the model selects a staff count $x \in [2, 15]$ that minimises the following objective:

$$Objective = \; -(\alpha \times Profit) + (\beta \times Penalty)$$

Where:

- $\alpha = 0.7, \beta = 0.3$
- Profit is calculated as:

  $$Profit = Total\,Revenue - Product\,Costs - Staff\,Costs$$

  $$Staff\,Costs = x \cdot h \cdot w$$

  With:

  $$x = number\ of\ employees$$

  $$h = shift\ duration\ in\ hours\ (e.g., Morning = 5, Afternoon = 5, Evening = 3)$$

  $$w = hourly\ wage\ (e.g., \$12)$$

- $Waiting\ Time\ (min) = max(0, \frac{((Demand\ per\ hour - Staff \times Service\ Rate))}{(Service\ Rate)}) \times 60$

  Penalty is calculated based on customer waiting time (Figure 1):

$$\text{Penalty}(W_q) = \begin{cases} (5 - W_q) \times 50 & \text{if } W_q < 5 \\ (W_q - 10) \times 50 & \text{if } W_q > 10 \\ 0 & \text{otherwise} \end{cases}$$

Figure 1. *Penalty Formulation*

4. Implementation and Insight

4.1 Result from forecasting model

The forecasting model demonstrated strong performance, with a Mean Absolute Error (MAE) of 49.85, a Mean Absolute Percentage Error (MAPE) of 10.25%, and an R^2 score of 0.94. These metrics indicate high predictive accuracy and a strong ability to explain the variation in customer demand.

As shown in Figure 2, the forecasted daily customer volume includes an 80% confidence interval. The blue line represents the median prediction, while the shaded area reflects model uncertainty. Customer traffic starts at a high level, peaking at around 1000 per day, and gradually declines to near-zero by late June. This downtrend may reflect seasonal closure or a significant business slowdown, potentially due to operational or structural factors. The forecast remains non-negative due to the log transformation and clipping applied, ensuring realistic forecast outputs under the Prophet framework.

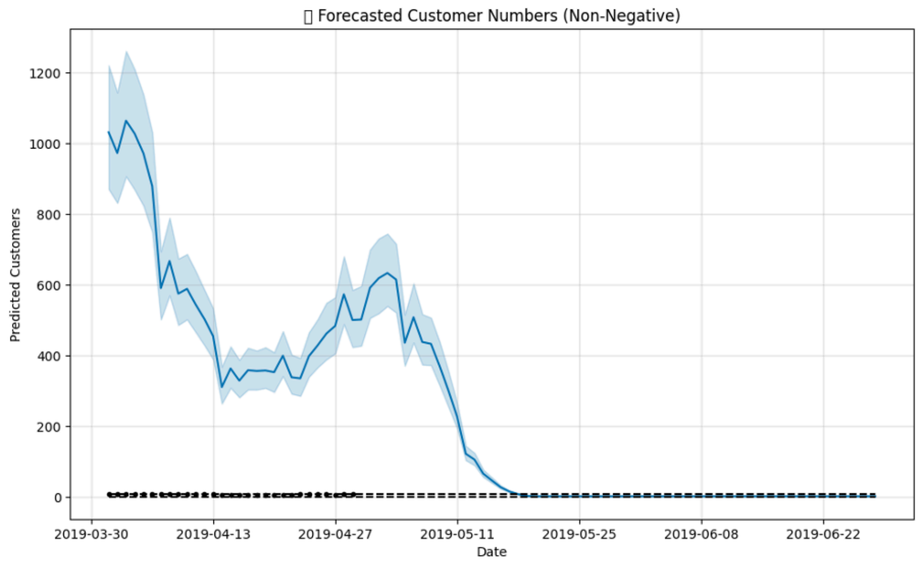

Figure 2. *Forecasting results*

Figure 3 compares the predicted and actual customer counts during the overlapping period in April. The blue solid line represents actual observed values, while the orange dashed line shows Prophet's forecasts. The two lines generally follow the same trajectory, indicating a good model fit. Minor deviations appear around April 6–10, but they remain within a reasonable margin of error and do not significantly affect overall forecast reliability.

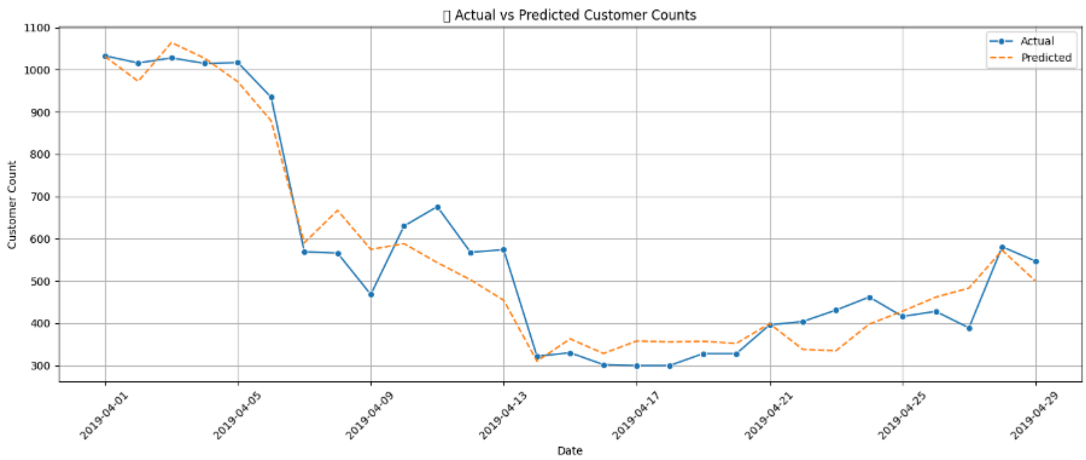

Figure 3. *Actual vs Predicted forecast*

4.2 Result from optimization model

The optimisation model was implemented in Python using a simulation-based search approach. For each scenario and time slot, customer demand was sampled from a truncated log-normal distribution. The model then iteratively evaluated staffing levels from 2 to 15 employees, selecting the number of staff that minimised the weighted objective function.

To generalise the simulation results into practical recommendations, customer demand values were grouped using Gaussian Mixture Model (GMM) clustering, forming low, medium, and high demand categories. For each cluster, the demand range, the most frequently optimal number of staff, and the corresponding average profit were calculated. This process produced a compact set of staffing recommendations that adjust to customer flow intensity. The results (Figure 4) demonstrate that:

- Recommended staff numbers increase with customer demand, ensuring service efficiency while avoiding overstaffing.
- High-demand scenarios such as holidays and promotions require more staff to maintain acceptable waiting times but still preserve profitability.
- The clustering framework supports flexible, scenario-specific scheduling strategies across time slots.

Scenario	Time Slot	Cluster	Demand Min	Demand Max	Recommended Staff	Average Profit
Holiday	Afternoon	Low Demand	11	89	2	110.82
Holiday	Afternoon	Medium Demand	90	224	3	281.51
Holiday	Afternoon	High Demand	227	355	7	516.67
Holiday	Evening	Low Demand	14	89	2	155.25
Holiday	Evening	Medium Demand	90	225	4	279.87
Holiday	Evening	High Demand	228	232	8	400.22
Holiday	Morning	Low Demand	45	89	2	155.94
Holiday	Morning	Medium Demand	90	160	3	269.07
Normal Day	Afternoon	Low Demand	7	89	2	105.05
Normal Day	Afternoon	Medium Demand	90	217	5	297.65
Normal Day	Evening	Low Demand	8	89	2	147.82
Normal Day	Evening	Medium Demand	90	129	5	250.88
Normal Day	Morning	Low Demand	35	89	2	140.44
Normal Day	Morning	Medium Demand	90	125	2	244.21
Promotion	Afternoon	Low Demand	15	89	2	114.67
Promotion	Afternoon	Medium Demand	90	221	3	285.77
Promotion	Afternoon	High Demand	227	452	9	575.25
Promotion	Evening	Low Demand	17	89	2	159.09
Promotion	Evening	Medium Demand	90	226	4	284.03
Promotion	Evening	High Demand	229	278	9	457.65
Promotion	Morning	Low Demand	50	89	2	163.42
Promotion	Morning	Medium Demand	90	180	3	268.58

Figure 4. *Recommendation Results*

5. Data-Driven Operational Insights

According to the Business Model Canvas framework (Osterwalder & Pigneur, 2010), nine key aspects should be considered when formulating strategy. For coffee shop owners, Leveraging big data tools, such as optimization modeling for profit management and predictive techniques for workload scheduling, enables more accurate measurement of operational conditions and data-driven decision-making（Tsiptsis, K., & Chorianopoulos, A., 2009).

In this report, K-means clustering applied to real-world sales data from the coffee industry, analyzing over 140,000 transactions across three New York City locations. The results, visualized in figure 5, reveal four consumer groups with distinct behavioral patterns based on time, location, and purchasing activity.

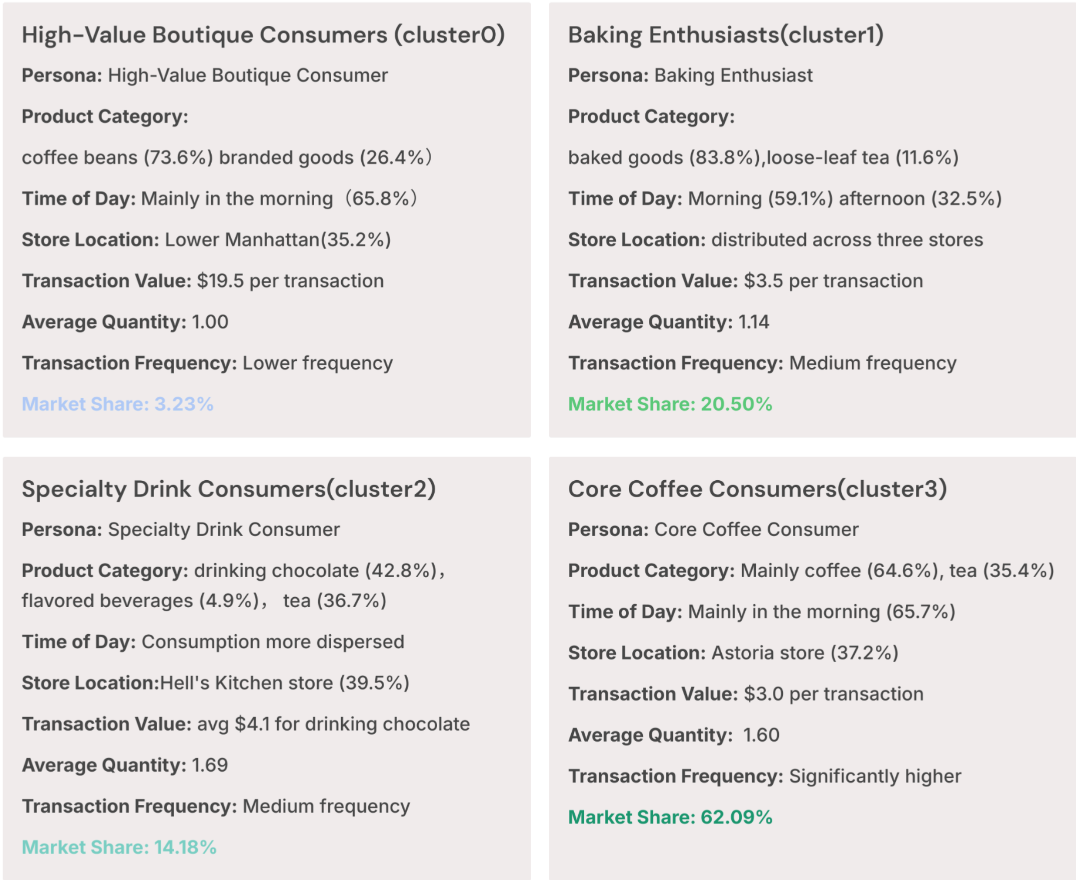

Figure 5. *Clusters Profile*

Notably, Cluster 3—Core Coffee Consumers—plays a central role in daily operations, contributing over 62% of total transactions and accounting for approximately 31–37% of customers across all locations. Their purchasing behavior is highly predictable: most orders (65.7%) occur in the morning, likely driven by commuting routines, with the highest concentration (37.2%) at the Astoria store, a residential area in Queens.

For managers, this insight highlights the operational pressure experienced during peak hours, and helps explain the prevalence of bottlenecks such as understaffing, long queues, and stockouts in the retail coffee sector.

6. Strategic Focus Areas

6.1 Workload planning strategy

Through regular input of historical data and forecasting techniques, customer demand could be quantified and revealed by dashboard. Coffee shop owners can furtherly refer to weather prediction and future holidays metrics to assist in the speculation for following time periods. The Multi-Objective Optimization Model developed a framework to connect profit and labor cost metrics.

Considering the customer satisfaction penalty and profit maximization, shop owners should arrange more staff to serve in those higher demand scenarios, such as holidays and promotions. By avoiding excessive waiting time, those profitability-service balances can be reached while remaining positive profit margins.

6.2 Operation strategy

Channel & Customer Relationships:

According to Reilly's Law of Retail Gravitation, customers are not drawn to large individual stores (Duddy, 1932). Coffee Empire can opt for store size with customer stream and location. Additionally, they can expand channels such as digital sales to complement physical store downsizing.

For Coffee Empire, predictive analytics can be used to foster customer loyalty to tailor promotions, product recommendations, and marketing messages based on individual customer preferences and behaviours.

Key Resources & Key Partnership:

In this project, staffing levels are the main goal to be optimized. However, the forecast model and optimization model show clear sales peaks and demand variability in different scenarios which may lead to understaffing during peak times or overstaffing during low demand periods. Businesses that adjust staffing levels based on dynamic demand forecasts experience higher operational efficiency and cost savings. Coffee Empire can implement part-time, on-call, and seasonal staff strategies to handle fluctuant situations.

The staffing optimization model does not mention coffee supply issues directly but optimizing staffing and maintaining steady sales require a reliable supply of high-quality coffee beans and materials. Compton emphasizes that strong partnerships with reliable suppliers ensure consistent product availability, mitigating risks associated with disruptions (2019). Coffee Empire can diversify supplier sources for coffee and materials to ensure supply chain resilience and prevent disruptions during peak sales periods.

7. Proposed Interventions - IBM Watson Cognos Analytics

Using IBM Cognos Analytics, a comprehensive business intelligence dashboard was developed to visualise key operational data and performance metrics for coffee shop management. The interactive panels display critical indicators including product category revenue distribution, time-slot sales patterns, and transaction trends. This solution enables managers to import historical data to identify underlying patterns in customer behavior and product performance. The dashboard reveals important insights such as the dominance of coffee and tea products in revenue generation, the significantly higher sales volumes during morning periods, and weekly transaction patterns. By transforming complex data into accessible visual formats, the dashboard empowers coffee shop owners to enhance both operational efficiency and customer satisfaction through informed business strategies.

8. Validation

The IBM Watson Cognos Analytics dashboard developed offers valuable business intelligence but faces several methodological and technical constraints that impact its generalizability and real-world application.

Dataset limitations: The primary limitation was the restricted time frame of our dataset, covering only April 2019 for Store 3, as visible in the transaction date visualizations. This short temporal window prevents the identification of longer-term trends, seasonal patterns, or year-over-year comparisons that would be crucial for comprehensive business planning. Additionally, the sales and revenue charts by week, especially the notable drop in week 5, suggest potential data completeness issues that could skew forecasting results. The dataset also relied solely on Kaggle's synthetic coffee shop data, which may not accurately reflect real-world variability in customer behavior or operational challenges faced by actual coffee shops.

Model validation challenges: While the forecasting model demonstrated promising patterns, several validation challenges emerged. The product category analysis (Figure 6) shows clear dominance of coffee and tea products, but our model had limited ability to account for complex interactions between product categories during different time slots. The time-based visualizations reveal significant variations in customer traffic patterns (particularly morning peak volumes), but validating these patterns would require more extended time-series data.

The dashboard effectively displays patterns such as the regular syrup product being the highest quantity item, but validating whether these product preferences remain consistent across seasons would require additional data collection periods.

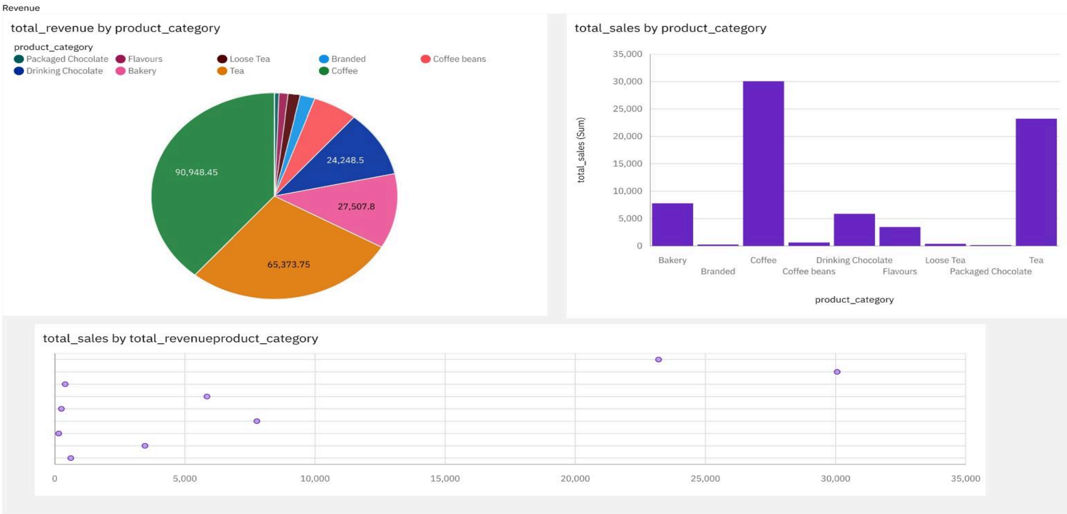

Figure 6. *Product Category Analysis*

Generalizability issues: The revenue distribution by product category and time slot (Figure 7) reveals specific patterns particular to Store 3's customer base and location, raising questions about the model's applicability to coffee shops in different markets, sizes, or service models. Without comparative data from multiple stores, it is difficult to assert that the staffing recommendations would optimise operations across diverse coffee shop contexts. The dashboard visualisations clearly show morning as the highest revenue period, but this pattern might vary significantly for shops in different locations or serving different customer segments.

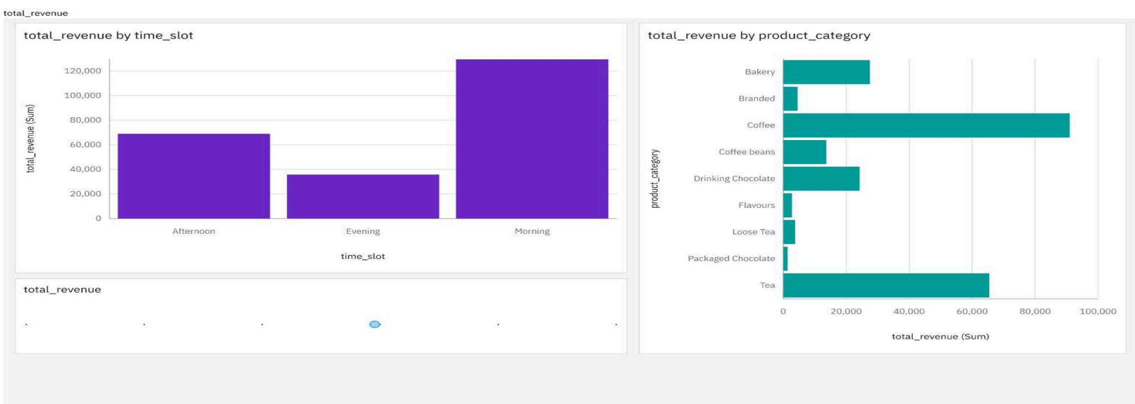

Figure 7. *Revenue Distribution*

9. Group Work

Despite the limitations of the project timeline, our group followed the initial project plan closely. Each member took on a defined role and contributed to completing the project deliverables. As our milestones were closely connected, team members provided dynamic support to one another when needed, ensuring progress remained consistent.

A record of our group collaboration, including meeting minutes, is provided in the appendix as a reference.

To better illustrate our workflow, Figure 8 presents our project timeline using the Agile Project Management (APM) method. APM is an iterative approach where teams and stakeholders collaborate to define requirements and continuously prioritize tasks (Hass, 2007).

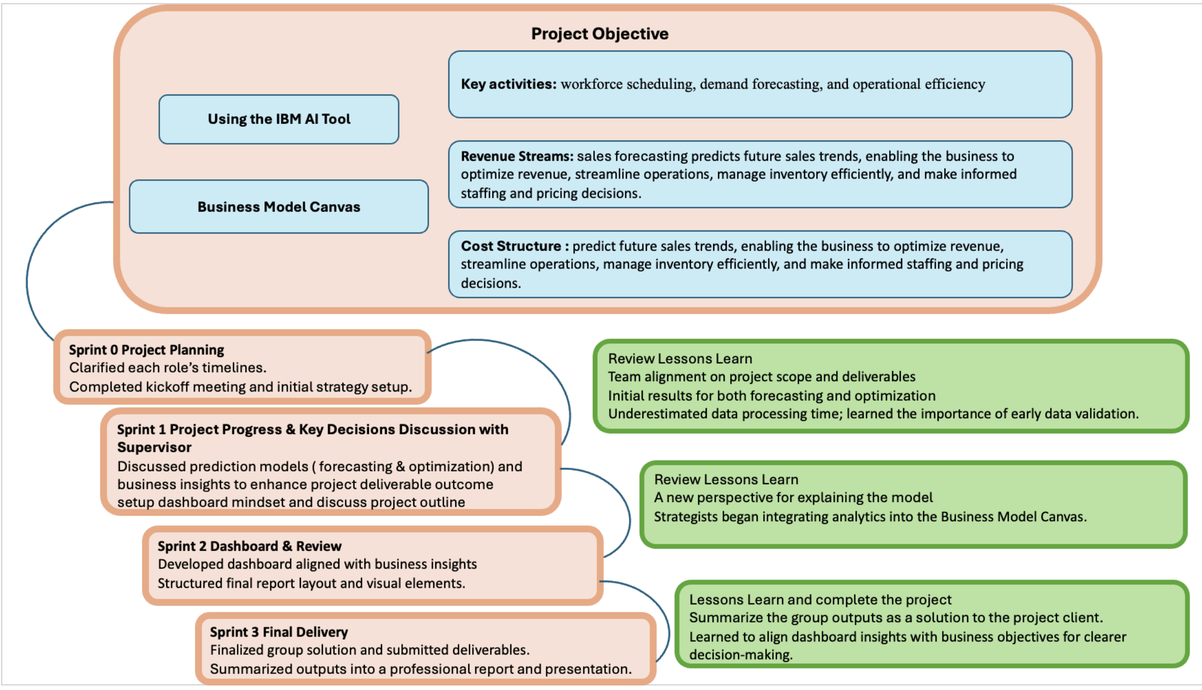

Figure 8. *The agile project life cycle model*

10. Conclusion

Data-driven strategies transform coffee shop management by enabling precise workload planning, optimized staffing, and targeted marketing. By implementing predictive analytics, dynamic staffing models, and customer-focused operational improvements, coffee shop owners can efficiently process information flows, anticipate demand fluctuations, and make evidence-based decisions.

However, the project also faced several challenges. The dataset's limited timeframe (one month of data from a single store) restricted the identification of seasonal patterns or long-term trends. Integration of external variables such as weather and holidays required careful alignment and normalization to ensure model accuracy. Additionally, platform constraints within the IBM Cognos Analytics trial version limited the interactivity and real-time adaptability of the dashboard, making it less practical for daily operational use.

Despite these limitations, the proposed interventions offer a robust decision-support system that balances service quality with profitability. With further development, including extended data collection and more intuitive dashboard platforms, these solutions have the potential to deliver significant operational improvements and long-term competitive advantages in an increasingly complex retail environment.

References

Alarcón, G.S., 2024. *Marketing communication strategies in maintaining customer loyalty*. *Join: Journal of Social Science*, 1, pp.32–51. Available at: https://doi.org/10.59613/1p9r6968

Almazrouee, A.I., Almeshal, A.M., Almutairi, A.S., Alenezi, M.R. and Alhajeri, S.N. (2020) 'Long-term forecasting of electrical loads in Kuwait using Prophet and Holt–Winters models', *Applied Sciences*, 10(16), p. 5627. Available at: https://doi.org/10.3390/app10165627 (Accessed: 27 March 2025).

Bisset, C. and Terblanche, S.E., 2021. A systematic literature review on stochastic programming models in the retail industry. *South African Journal of Industrial Engineering*, 32(3), pp.97–110. Available at: https://doi.org/10.7166/32-3-2611

Chang, Y.-T., Wang, I.-C., Li, C.-H. and Wen, S.-P., 2010. Application of K-means clustering in customer segmentation: A case study in retail industry. *Journal of Business and Retail Management Research*, 5(1), pp.92–107.

Chummee, P., Chanprasert, S., Nakhon, S. and Ayutthaya, S., 2022. Applying Business Model Canvas (BMC) for planning marketing strategies and strategies for the customers of the consumer and consumer retail business. Case Study: Kong Khong Market Business, Ban Sang Som. *International Journal of Early Childhood Special Education*, 14, pp.2282–2286. https://doi.org/10.9756/INT-JECSE/V14I1.267

Compton, N., 2019. Strong partnerships and European collaboration. *Angewandte Chemie International Edition*, 58, pp.12722–12724. https://doi.org/10.1002/anie.201909779

Duddy, E.A., 1932. The Law of Retail Gravitation. William J. Reilly. *American Journal of Sociology*, 37, pp.826–828. https://doi.org/10.1086/215884

Hass, K.B., 2007. The blending of traditional and agile project management. *PM World Today*, 9(5), pp.1–8.

Journal of Supply Chain Management, n.d. *Wiley Online Library*. Available at: https://onlinelibrary.wiley.com/journal/1745493x [Accessed 20 Mar. 2025].

Kumar Jha, B. and Pande, S., 2021. Time series forecasting model for supermarket sales using FB-Prophet. In: *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE. https://doi.org/10.1109/iccmc51019.2021.9418033

Kumar, A., Maheshwari, P. and Raj, P.C., 2022. A hybrid Prophet–LSTM model for PM2.5 forecasting: An empirical study on Indian cities. *Frontiers in Environmental Science*, 10, p.945628. Available at: https://www.frontiersin.org/articles/10.3389/fenvs.2022.945628/full

Li, D. and Xin, J., 2024. Deep learning-driven intelligent pricing model in retail: From sales forecasting to dynamic price optimization. *Soft Computing*. https://doi.org/10.1007/s00500-024-09937-z

Lloyd, S., 1982. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2), pp.129–137.

MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pp.281–297. Berkeley: University of California Press.

McKinsey & Company, 2023. *What is personalization?* McKinsey & Company. Available at: https://www.mckinsey.com

Osterwalder, A. and Pigneur, Y., 2010. *Business model generation: A handbook for visionaries, game changers, and challengers*. Hoboken, NJ: John Wiley & Sons.

Rahardjo, B., Hasbullah, R. and Taqi, F.M., 2019. Coffee shop business model analysis. *Integrated Journal of Business and Economics*, 3(2), pp.140–152.

Raji, I.D., Oyelade, O.N. and Oyelade, O.O., 2021. A comparative analysis of ARIMA and Prophet models for forecasting electricity consumption. In: *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*. IEEE, pp.204–209. Available at: https://ieeexplore.ieee.org/document/9418033

Taylor, S.J. and Letham, B., 2018. *Forecasting at scale*. Google Books. Available at: https://books.google.co.uk/books?hl=en&id=x54cEAAAQBAJ

Tian, Y., Zhao, Y., Lu, R. and Xu, J., 2022. Short-term power load forecasting based on Prophet model. *Energies*, 15(6), p.2158. https://www.mdpi.com/1996-1073/15/6/2158

Tsiptsis, K. and Chorianopoulos, A., 2009. *Data mining techniques in CRM: Inside customer segmentation*. Wiley.

Appendices

Appendix A. Data Source

- **Coffee shop sale data from Kaggle:**

  https://www.kaggle.com/datasets/keremkarayaz/coffee-shop-sales

- **Holiday data:**

  https://www.kaggle.com/datasets/donnetew/us-holiday-dates-2004-2021

- **Weather data - extracted using API:**

  https://openweathermap.org/api

Appendix B. Dashboard Results and Insights

Coffee_empire_dashboard.ipynb

Appendix C. K-means Cluster Details of Customer Segmentation

Figure 1. *Clustering Scatters*

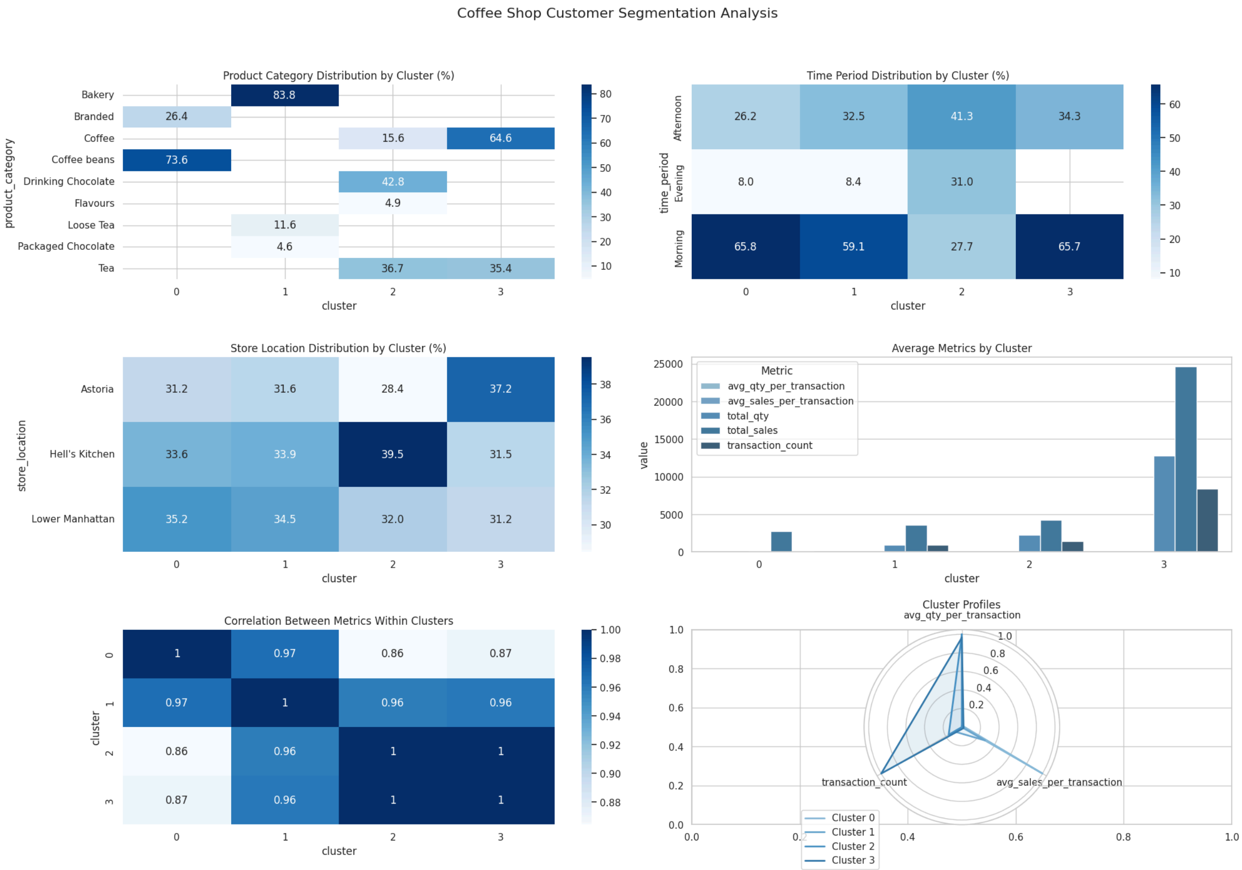

Figure 2. Clusters Features Distribution

Appendix D. Other Figures of Forecasting

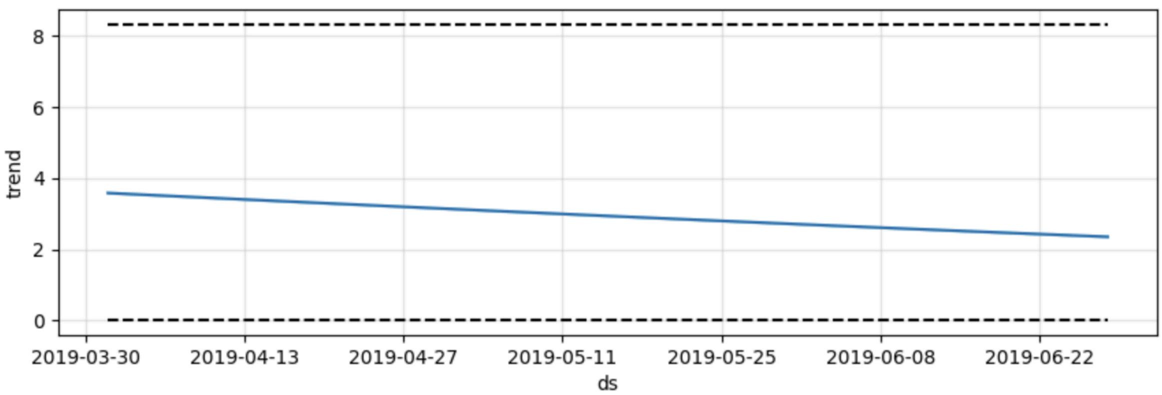

Figure 3. Trend analysis