Lisbon School of Economics & Management
Universidade de Lisboa

MASTER
DATA ANALYTICS FOR BUSINESS

MASTER´S FINAL WORK
INTERNSHIP REPORT

AN ADAPTIVE FORECASTING FRAMEWORK FOR MULTI-UNIT RETAIL

NAYYERA YAHYA FATHY OSMAN ABOELNAGA

SUPERVISION:
ARTUR CUNHA

JANUARY - 2026

Lisbon School of Economics & Management
Universidade de Lisboa

MASTER
DATA ANALYTICS FOR BUSINESS

MASTER´S FINAL WORK
INTERNSHIP REPORT

AN ADAPTIVE FORECASTING FRAMEWORK FOR MULTI-UNIT RETAIL

NAYYERA YAHYA FATHY OSMAN ABOELNAGA

SUPERVISION:
ARTUR CUNHA

JANUARY - 2026

GLOSSARY

ETS – Error, Trend, Seasonality exponential smoothing model.

Fiscal Period – Standardized accounting period used for internal financial reporting, distinct from calendar months.

Global (Pooled Model) – Forecasting model trained jointly on data from multiple stores to share information across units.

JEL – Journal of Economic Literature classification system.

MAE – Mean Absolute Error

MAPE – Mean Absolute Percentage Error

MFW – Master's Final Work.

OLS – Ordinary Least Squares.

PCA – Principal Component Analysis.

Peer-Based Forecasting – Forecasting approach where predictions for data-sparse units are derived from similar data-rich units

QSR – Quick Service Restaurant

RMSE – Root Mean Squared Error

WAPE – Weighted Absolute Percentage error

ABSTRACT AND KEYWORDS

This Master's Final Work develops and evaluates an adaptive sales forecasting framework for a multi-unit fast food retail network characterized by heterogeneous historical data availability across stores. The work is based on a curricular internship conducted at LGG Advisors, where the client has multiple portfolios, one of which is a U.S.-based quick-service restaurant managing 61 stores across different states. The client's existing regression-based forecasting consistently exhibited an optimistic bias, leading to inefficient budgeting and planning.

The forecasting challenge arises from structural differences in historical length between stores, as the operator was constantly acquiring new stores, but their records were fragmented due to data migration issues. To address this challenge, the proposed framework combines pooled ridge regression for sufficient history stores, and a peer-based forecasting for data-sparse stores.

Principal component analysis and clustering are used to construct a store representation to identify matching peers to allow sharing information from similar long history stores. Model selection for short history stores is determined through performance-based selection using weighted absolute percentage error and will be continuously re-evaluated as additional data becomes available.

Empirical results show that the adaptive model improves forecast accuracy compared to the client's benchmark. Across the evaluation window, the hybrid model achieves 59% reduction in WAPE and reduces aggregate forecast deviation from $1.3 million to $97k. These improvements directly reflect better budgeting and estimation for the business.

Keywords: Sales forecasting, Multi-Unit Retail, Pooled regression, Data sparsity

TABLE OF CONTENTS

TABLE OF FIGURES

ACKNOWLEDGMENTS

I would like to begin by dedicating this work to my mother, who is no longer physically present, yet her soul, words, and light continue to guide me every day. Her strength, encouragement and resilience remain with me through every challenge, and her voice continues to push me forward even in her absence.

I would like to acknowledge my family for the role they played in shaping my resilience. I am especially grateful to my father, whose continuous support made this journey possible. I also thank my siblings, Ahmed, Lobna & Youssef for their encouragement, presence and moments of reassurance they offered during demanding times.

I would also like to express my heartfelt gratitude to my partner, whose support throughout this entire process invaluable. His belief in me, encouragement during difficult moments and constant presence gave me strength when I needed it most.

My sincere thanks go to the team at LGG where this internship was conducted, for welcoming me, exposing me to real business challenges and allowing me to grow both professionally and personally. I am also grateful for ISEG facilitating the connection with the host company enabling this internship opportunity.

I would also like to thank my academic supervisor, Professor Artur Cunha for his guidance and supervision.

Finally, I would like to thank my friends whose encouragement and support accompanied me throughout this journey.

1. INTRODUCTION

1.1 Background & Context

The rapid revolution of retail and food & beverage sectors necessitates highly accurate sales forecasting to support operational efficiency and strategic decision-making. The Master's Final Work (MFW) is based on a curricular internship conducted at LGG Advisors, a global service provider consultancy specializing in Finance & strategy and digital transformation. The work focuses on the development of a machine learning sales forecasting model for a major client, referred to as "The Client Retail Group". (TCGR) for confidentiality purposes. TCGR is a multi-unit franchise group that has a diverse portfolio. They manage multiple F&B chains across the United States, the chain under focus in the pizza industry with approximately 61 stores across different states.

1.2 Problem Motivation

The motivation of this MFW arises from the client's challenge & difficulty in meeting sales forecast demand using a traditional regression approach. During the internship, these forecasts were observed to exhibit a consistent optimistic bias, which led to inflated and avoidable costs. In fact, internal analysis indicated an over-forecasting of approximately $1.3 million in total sales over the final six fiscal periods of 2025 alone. Such forecast bias directly reflects on operational aspects, including inefficient demand planning and avoidable costs driven by excess inventory and over-staffing.

1.3 Structural Challenge

A central challenge underlying the client's forecasting problem is the heterogeneity of historical data across all stores. Although the stores operate under the same business model, they do not share uniform historical length. The portfolio consists of a mix of mature stores with long stable sales histories alongside recently acquired stores, which have limited and fragmented records.

This imbalance in data availability presents a structural constraint for demand forecasting. Forecasting methods rely heavily on long and consistent time series to capture patterns; otherwise, the model tends to perform poorly. As a result, applying a single model uniformly across the whole network will perform poorly for a significant portion of the stores.

In this context, the forecasting challenge extends beyond improving forecasts; it requires a system capable of adapting to different levels of data availability while maintaining consistency across the network.

1.4 Research Question

Based on the challenge outlined above, the research question guiding the Master's Final Work is: How can a deployable, high-accuracy sales forecasting model be designed to effectively manage heterogeneous historical depth data within a multi-unit retail franchise network?

To address this question, the primary objective of this work is the development and evaluation of a forecasting framework that accommodates differences in historical depth across stores while improving overall forecast accuracy.

Specifically, the objective of this work is to:

i. Improve sales forecast accuracy relative to the client's existing regression-based approach.
ii. Reduce systematic forecast bias that affects budgeting and operational planning.
iii. Demonstrate how improved sales forecasts translate into more reliable planning signals, including labour and budget estimates.

2. LITERATURE REVIEW

2.1 Demand Forecasting in Retail & Multi-Unit Environments

Demand forecasting plays a crucial role in many operational contexts especially the retail & food and beverage industries, as forecasting is needed for optimized operations. Accurate forecasts inform budgeting and financial planning to utilize resources and workforce as required. They also expand to other in-depth fields, such as inventory control from ordering ingredients in restaurants to managing stock levels in retail stores to avoid bottlenecks of stock outs or waste from overordering (Park et.al.,2022). This challenge is consistent with the case examined under study, inaccurate & overly ambitious forecasts that led to excess stock outs that are not needed, draining financials and not being able to match or come close achieving forecasts, which also encouraged operators of the business to seek a solution.

The operational costs of forecast errors can be significant, especially when forecasts are too high (over-forecasting). Sanders & Graman (2009) find in a warehouse case study that forecast errors can increase operational costs10-30%, depending on the environment and sector. Over-forecasting tends to incur excess labour hours & markdowns of unsold stock. Surprisingly, research on bias in forecasts shows that deliberately over-forecasting, which is introducing a positive bias was found to be more damaging cost-wise than not biasing the forecasts at all. This is because over forecasting often leads to waste. These findings indicate that improving forecast accuracy and avoiding excessive optimism in forecasts can have a tangible cost benefit for the retail and F&B sectors. (Sanders &Garman, 2009).

Another important aspect when forecasting for multi-store or multi- unit environments. Large retail chains and franchised brands must forecast demand hierarchically across many locations while accounting for heterogeneity in consumer behaviour, local events and external conditions. In the case examined in this report, the franchise operated across multiple U.S states where substantial variation was observed in how individual stores responded to events and seasonal effects. Such heterogeneity makes centralized forecasting particularly challenging, as aggregated forecasts can obscure meaningful local demand patterns as documented in the retail forecasting literature.

This challenge is consistent with the findings of Fildes (2019), who highlights that retailers face forecasting problems across hierarchical levels ranging from individual stores to the company overall. The paper emphasizes that while aggregation can reduce volatility, it may also mask important store-level differences that are critical for operational decision making (Fildes, 2019). Consequently, forecasting in multi-unit environments requires approaches that balance aggregation for stability with disaggregation for local relevance.

In applied retail forecasting, modelling choices are often constrained by the time stamps of available data whether, it is recorded as a normal calendar; monthly, weekly, daily, or following a fiscal calendar. In multi-unit retail and QSR organizations, store-level sales are commonly aggregated or recorded into a standardized fiscal period for financial operations. As a result, forecasting systems are typically built on fiscal-period sales. (Fildes, 2019)

Forecasting literature highlights that temporal aggregations alter the statistical properties of sales data by smoothing short-term variability and reducing apparent volatility. While this can improve forecast stability, it also limits the model's ability to capture short-duration effects. In our case Super Bowl as an event boosts sales of the chain. However, it's a one-day event. Hence, the effect is smoothed in a fiscal period. Consequently, event-driven sales fluctuations or localized shocks may appear weak or statistically insignificant when forecasts are produced at aggregated fiscal horizons (Hyndman & Athanasopoulos, 2018). In the context of this study, this limitation provides, important context for interpreting feature relevance and event effects in the proposed forecasting framework.

2.2 Classical Forecasting Models & Their Limitations

Classical univariate time-series models, most notably ARIMA (Auto-Regressive Integrated Moving Average) and exponential smoothing (ETS), have long been used as baseline approaches for sales forecasting (Fildes, 2019). ARIMA models capture temporal dependence through autoregressive and moving average as the model's parameters (p,d,q) and seasonal terms directly relate to the time series' memory and trend properties, on the other hand ETS frameworks decompose time series into level, trend & seasonal components (Cheng et al., 2025). These models rely heavily on steady and rich historical patterns, when these conditions are satisfied, these models can capture recurring trends and seasonal cycles effectively yielding reliable forecasts. (Fildes, 2019)

However, the literature also highlights several limitations of classical time-series models that become particularly pronounced in multi-store settings. First, these models require sufficiently long and consistent historical time series to reliably estimate parameters, especially seasonal components. When historical depth is limited, parameter estimates become unstable and forecast accuracy declines. This poses a challenge in retail environments when histories vary substantially and newer stores or recently acquired stores may have a limited number of observations, and this is the case of the client under study. Second, these models assume temporal patterns remain relatively consistent apart from noise, making them poor at handling structural breaks or regime shifts. Sudden changes in the underlying level of the series (due to one-off events, heatwaves, policy changes) can degrade forecast accuracy. Without explicit intervention or re-specification

ARIMA/ETS can struggle under structural (Liu, 2024). Even large transient shocks (e.g., pandemic lockdown) violate the model's assumptions of continuity and often lead to large errors if not manually adjusted. Furthermore, scalability is a challenge; classical approaches are inherently trained separately for each time series (Cheng et al., 2025). In a multi-store setting, this means building and maintaining hundreds of individual models. Aside from a heavy computational burden, a local modelling strategy cannot easily share information across series. Each store's model ignores data from other stores, even if there are common seasonal patterns or correlated demand drivers across locations. This lack of cross learning makes it difficult to forecast a heterogeneous setting where some stores have limited data or similar patterns. In practice, data pooling or hierarchical forecasting might be a solution for this condition (Fildes, 2019), but classical ARIMA/ETS frameworks are typically fit per series so cross-series information is not possible without additional frameworks (e.g., global models).

In summary, while ARIMA and ETS models provide well-structured and interpretable forecasts, they often prove insufficient for multi-store sales forecasting when faced with sparse histories, structural changes, and the need to scale across a large, diverse panel of time series. These limitations motivate the exploration of more flexible and data pooling approaches in the literature (Cheng et al., 2025).

2.3 Regression-Based & Machine Learning Forecasting Approaches

An alternative paradigm for forecasting is to use regression models that incorporate external predictors (beyond a time series' own lags) to predict future sales; this means leveraging rich features such as consumer price index & events as explanatory variables in a linear regression model. The appeal of the regression-based model is its ability to capture causal drivers and covariates that classical univariate methods cannot. For example, a spike in an event occurrence promotion can be directly modelled via indicator or price variables in a regression rather than hoping a time series model's seasonality will capture it. However, using many features introduces a different kind of challenge. One major issue is multicollinearity: retail features are often correlated with each other and with time. Traditional OLS regression becomes unstable when predictors are highly collinear, the coefficient estimates blow up in variance, making the model unreliable and prone to overfitting the noise in training data (Auguinis & Edwards, 2025). In forecasting

terms, an unstable regression might fit past sales well but yield out-of-sample predictions. Another issue is that many predictors given a short period of time can lead to a "high-dimensional" problem identified in retail forecasting, which is too many variables, too few observations (Fildes, 2019).

To address these issues, the report has turned to penalized regression (regularization) techniques that deliberately constrain or shrink coefficients to prevent overfitting. Two widely used methods are Lasso & Ridge regression, which add L1 or L2 penalties (respectively) to the ordinary least squares objective. Lasso (Least Absolute Shrinkage and Selection Operator) applies an L1 penalty, which tends to drive many coefficient estimates exactly to 0. This yields a sparse model that performs feature selection. Effectively choosing a subset of the most relevant predictors while eliminating the rest (Hastie, Tibshirani & Friedman, 2009/2017). Lasso is attractive in retail forecasting because it can explore a wide range of potential features. Ridge, on the other hand, uses an L2 penalty, which shrinks all coefficients toward zero but typically does not force any to zero. Ridge therefore retains all predictors but reduces their magnitudes, especially if they are only weakly informative, to prevent a single feature dominating the forecast. A key feature of Ridge is that it excels in handling multicollinearity. When features are highly correlated, Ridge will shrink their coefficients towards one another rather than arbitrarily dropping all, compared to Lasso's approach. This tends to distribute influence across collinear predictors, yielding more stable predictions (Auguinis & Edwards, 2025).

Regularization introduces a bias into the model, but it greatly reduces variance, which is the essence of the bias-variance trade-off that underpins its improved predictive performance, which is explicitly discussed here (Hastie, Tibshirani & Friedman, 2009/2017). By accepting a small bias, penalized models avoid overfitting the noise, yielding to lower error on new samples rather than an unbiased but high-variance OLS fit. This trade-off is particularly valuable in retail forecasting, where the goal is robust out-of-sample accuracy amid noisy sales data.

For multi-store forecasting with partially correlated predictors, the Ridge approach is often justified as the most stable choice. In addition, Ridge regression's shrinkage helps share information. This property is advantageous when predictors have shared effects across stores. By shrinking coefficients rather than selecting variables aggressively, Ridge

regression supports more stable and consistent forecasts across heterogeneous units (Hastie, Tibshirani & Friedman, 2009/2017).

Recent forecasting literature has examined the application of tree-based machine learning models, such as Random Forests and Gradient Boosting methods (including XGBoost), to retail and quick serve restaurant demand forecasting. These models are particularly attractive in retail settings due to their ability to capture non-linear relationships and complex interactions among explanatory variables. Unlike classical linear univariate time-series models, tree-based methods do not rely on strong distributional assumptions and can flexibly model heterogeneous demand responses (Montero-Manso & Hyndman, 2021), which has led to improved forecasting performance in some empirical retail applications (Makridakis et al.,2018). As a result, ensemble tree methods have been increasingly explored as alternatives to traditional forecasting approaches.

Despite these advantages, the academic evidence indicates that machine learning models are not universally superior to simpler statistical or regression-based approaches. Comparative studies show that the performance of tree-based models is highly dependent on data volume, signal-to-noise ratio, and temporal stability. In environments with limited historical data, short time series, or high variability across units, tree-based models are prone to overfitting & learning idiosyncratic patterns rather than persistent demand signals. This limitation is especially relevant in multi-store retail setting, where individual outlets often have uneven data histories differencing local demand drivers. Under such conditions, simpler models with fewer degrees of freedom frequently match or outperform complex tree-based models (Makridakis et al.,2018). These findings motivate the use of Ridge, which is a more constrained model that trades flexibility for stability.

Multi-unit retail networks generate panel data consisting of multiple store-level time series observed over common time periods. Forecasting in this setting is complicated by substantial heterogeneity across units, as individual stores differ in scale growth trajectories, seasonal intensity and responsiveness to external factors. Retail forecasting research emphasizes that this heterogeneity makes store-level prediction fundamentally different from single-series forecasting requiring methods that balance unit-specific behaviour with network wide behaviours (Fildes et. Al.,2022).

A common approach in practice is to model each store independently using local forecasting models. While local models naturally capture store-specific characteristics, they perform poorly when historical data are limited or noisy. In retail environments, many stores have short or data availability constraints. Empirical evidence in retail forecasting shows that pooling information across related units can outperform purely individual estimation, particularly when data per unit are scarce (Montero-Manso & Hyndman, 2021).

While a wide range of regression and machine learning models have been proposed for retail forecasting, the literature consistently shows that no single approach dominates across all settings. In practice model choice is therefore confirmed by out-of-sample performance, particularly in such context where data is uniform regularized linear models including ridge are commonly found to offer a good balance between flexibility and stability.

2.4 Pooled, Information Sharing & Data Sparsity

To address these limitations, the approach of pooled modelling has been increasingly focused on in this literature. Global models are trained on data from all units simultaneously, allowing them to exploit shared patterns while accommodating heterogeneity across stores (Chen & Boylan, 2008). A central advantage of this approach is its ability to borrow strength across units. Common patterns such as seasonal effects of calendar driven demand shifts can be estimated more reliably using the combined data. Importantly, theoretical results show that global models are not inherently less expressive than local models and can approximate or exceed local forecast performance even in highly heterogeneous data (Montero-Manso & Hyndman, 2021).

The benefits of information sharing are particularly pronounced in scenarios where a store has insufficient historical data to support independent forecasting. In such cases, forecasts based solely on local information are unreliable (Hewamalage et al., 2022). By leveraging patterns learned from established stores, pooled approaches provide a stable baseline for data-sparse stores, with forecasts gradually adapting as more observations become available. Empirical studies confirm that in such cases global models tend to outperform local alternatives, highlighting the importance of shared learning (Hewamalage et al., 2022).

In summary, heterogeneous and sparse data present fundamental challenges for forecasting. The literature consistently supports the use of information sharing strategies that balance store-level differentiation with pooled learning. These insights provide a strong theoretical foundation for peer-based and pooled forecasting approaches.

2.5 Dimensionality Reduction, Clustering & Adaptive Forecasting Frameworks

Retail forecasting often involves high-dimensional features as sales that can be driven by numerous factors (operation, promotions, holidays, etc..). such abundance of predictors poses challenges for traditional modelling leading to overfitting and unstable estimates. Dimensionality reduction techniques like Principal Component Analysis (PCA) are therefore used to extract latent structures from these features while filtering out noise. PCA transforms original features into a small set of uncorrelated components that capture most of the information. Hence, increasing interpretability with minimal information loss (Ramos et al.,2022). In forecasting applications, this means key signals can be isolated from noisy covariates. For example, a recent study on supermarket sales found that incorporating PCA to summarise numerous promotional and price drivers yielded about a 10% improvement in forecast accuracy compared to using raw features. In summary, PCA is proven to improve levels of accuracy when there are redundant number of features by transforming features into a few informative factors, enabling focus on the underlying demand structure rather than spurious variation (Ramos et al.,2022).

Complementing dimensionality reduction, clustering methods address forecasting complexity by grouping similar stores or time series based on their patterns. Clustering aims to reduce heterogeneity in the data. By segmenting large retail network into more homogenous groups and applying shared parameters within each cluster will improve accuracy (Chen & Lu, 2021). This similarity-based approach also enables knowledge transfer or "peer" forecasting, which is especially valuable in cases where some stores have limited history. Instead of forecasting an under-resourced series in isolation, the series can borrow strength from its cluster peers. For instance, in a new product sales forecasting, researchers have shown that using historical sales data from similar products clustered by demand patterns, greatly improves predictive performance. In effect, clustering creates a framework for peer-based forecasting in which insight from data-rich units informs forecasts for conceptually similar but data-poor units. By first extracting

common factors via PCA and then pooling similar cases via clustering, a forecasting model can leverage cross-sectional information efficiently (Hwang et al.,2025).

Relying on a single forecasting model across heterogeneous units is increasingly recognized as limiting, given the diversity of demand patterns in retail settings. Different stores or products may exhibit unique seasonality, trends or volatility so it is very unlikely to find a model that fits all. Indeed, evidence from large-scale forecasting studies shows that no single model consistently dominates across all types of time series. For instance, the M4 competition that spanned thousands of series confirmed that methods that adapt to each series tend to outperform any fixed approach. These findings underscore the need for forecasting systems that are context sensitive. Developing one model for all units risks suboptimal accuracy in many cases, motivating the turn toward hybrid and adaptive frameworks (Ekiz Yılmaz & Yapar, 2025).

Two broad strategies have emerged to tackle heterogeneity; ensemble forecasting and performance-based model selection. Ensemble methods combine predictions from multiple models to exploit their complementary strengths. By averaging or otherwise aggregating forecast, ensembles can reduce error and model bias, guarding against the risk of picking a single wrong model (Kourentzes et al., 2018). There is substantial empirical evidence that forecasts combinations yield lower errors and more robust performance than even the best individual model. In contrast, model selection approaches maintain a library of candidate models but choose only the one that appears most suitable for a given period. This selection is typically based on observed performance (Ekiz Yılmaz & Yapar, 2025). For instance, the model with the lowest validation error or the model with the lowest validation error is chosen for each forecast cycle. Unlike static one-size fits all modelling. Such selection needs to be reevaluated, as the store can "graduate to a different model" given more data or different conditions (Ekiz Yılmaz & Yapar, 2025).

Beyond fixed ensembles or one-time selection, researchers are increasingly advocating adaptive forecasting systems that continually adjust model choices based on incoming data and performance feedback. In an adaptive framework, the process of choosing a model is not a one-off decision but an ongoing process responsive to new information (Ekiz Yılmaz & Yapar, 2025). For example, Hananya and Katz (2024) develop a dynamic selection scheme that automatically switches to the best machine

learning model in real time as a series of characteristics evolve. This kind of system monitors forecasting accuracy and can reassign the model when it detects regime changes or when an alternative model starts performing better. Adaptivity also involves considering data sufficiency, an adaptive system might employ simpler, more data efficient models for new or limited series and more complex models when ample data are available. The literature increasingly supports flexible, performance-driven frameworks. This dissertation aligns with that perspective, adapting a system-oriented view, rather than committing to a single model Ekiz Yılmaz & Yapar, 2025). It embraces an adaptive framework that assigns models according to what works best for each store and situation.

In summary, the literature highlights that sales forecasting in a multi-unit retail context is constrained by data heterogeneity and trade-off between model flexibility and stability. Classical time-series approaches struggle under sparse and uneven historical data, while highly flexible machine learning models may overfit in such settings. Research therefore increasingly supports pooled modelling, regularized regression & clustering as effective mechanisms for this business case.

3. CONTEXT

The dissertation was conducted during a four-month internship at LGG Advisors, a global consulting company specializing in finance, strategy, risk & Digital Transformation. The company focuses on building actual solutions; from automation tools & introducing AI to forecasting pipelines and Visualization dashboards with a strong emphasis on adaptability in a fast-paced AI landscape.

I was assigned to the digital transformation team during my internship, where I worked on both engineering and data focused projects. My work spanned multiple projects, a core assignment and the subject of this dissertation involved the design and deployment of a forecasting system for one of the internship organisation’s major clients (TCG). The objective was to build a scalable and adaptive demand forecasting pipeline to support budgeting and workforce planning.

Beyond the forecast initiative, my internship also included hands on exposure to automated bot developments, where I contributed to the design and deployment of intelligent assistants using frameworks such as Langchain & LangGraph. This involved

API Integration, prompt engineering and interface design. I was also involved in AWS-based ETL workflows as well as exploratory work in Power BI.

Together, these experiences provided me with a multi layered understanding how forecasting fits into a broader digital transformation strategy and how technical pipelines interface with decision making process across the organization.

The forecasting system was developed for a U.S- based fast food chain operator managing around 61 stores across six states: Alabama, Indiana, Kansas, Kentucky and New York. While the parent brand has a large footprint and thousands of stores across America, this forecasting solution was specifically built to support the client, which is has acquired 61 stores of the chain.

The store network is split into distinct cohorts:

- 31 long-history stores with over 18 months of consistent data
- 30 limited history stores which were most likely acquired and therefore have shorter & fragmented data. These stores are not necessarily new in operation but lack original data that the client could not acquire with the transition.

The heterogeneity in data depth across units created key modeling challenges particularly in balancing between data-rich solutions and those with insufficient historical data. However, this is a common problem in businesses specially during transitions or digital transformations, which enables loss of data.

The client's data infrastructure is hosted on AWS, where key operational datasets including sales, labor, budget are stored in structured Athena tables. The digital transformation at LGG Advisors managed the data ingestion and transformation workflows, ensuring that updated and standardized data is regularly available for analysis and modeling.

A central structural constraint in the project was the differentiation of the client's data as some tables followed a fiscal calendar, which comprises 13 fiscal periods per year, while other tables followed normal Gregorian calendars. Unlike standard monthly calendars, the client's fiscal year is divided into 13 equal-length periods, each corresponding to exactly four weeks (28 days), except for one designated period per year which spans five weeks (35 days). This five-week period accommodates the extra days

that accumulate due to the difference between a 52-week calendar (364 days) and the actual 365-day year (366 in a leap year). In the client's calendar, this extended period is Period 13 in a standard fiscal year. This structure is common in QSR and retail organizations as it ensures that each period contains the same number of weekends, making year-over-year comparisons more consistent than with calendar months. A key preprocessing step involved harmonizing these systems to align all external and internal data with the client's fiscal reporting periods. This alignment was critical for both training and evaluating forecasts within the same operational rhythm.

The fiscal harmonization process involved:

- Mapping daily and weekly records to their corresponding fiscal period and fiscal year.
- Engineering periodic components for instance sine/cosine transformations to model fiscal seasonality explicitly.
- Mapping external factors (CPI, event dates, etc..) to fiscal calendar.

This conversion ensured that all downstream modeling reflected the actual decision making of the business reducing risk of temporal misalignment between forecasts and operational execution.

4. METHODOLOGY

4.1 Problem formalization

The forecasting task addressed in this dissertation is defined at store-fiscal period level that focuses on predicting future sales for individual stores. Forecasts are generated at a fiscal period to align with the operational rhythm of the business. A central characteristic of the forecasting environment of this case, is the presence of substantial heterogeneity in the historical data availability across stores. While all stores operated under the same brand, they do not share uniform historical depth. Some stores have long, steady sales histories and others have fragmented records due to recent acquisitions, data migration issues or change in ownership. The uneven availability of historical data is not temporary or random but represents a structural constraint that must be explicitly addressed in the forecasting.

Stores with long histories allow for more reliable estimation of demand patterns and seasonal changes, while stores with limited data provide insufficient information to support fully independent accurate estimations. Applying a single forecasting strategy uniformly across stores would therefore be inappropriate, as it would either underperform for data-sparse stores or fail to fully exploit the information available in data-rich stores.

As a result, the problem can be described as follows; given historical store-level sales data and fiscal calendar structure, the objective is to generate one sales forecast per store for each future fiscal period. This must be achieved under the constraint of uneven historical data availability while maintaining forecast accuracy.

The remainder of this chapter outlines the methodological framework developed to address the problem. It discusses the feature engineering strategy, evaluation and benchmarking process and the adaptive mechanisms used to assign forecasting approaches to stores based on data availability and observed performance

4.2 Feature Engineering Strategy

4.2.1 Target Variable

The target variable, total net sales will be denoted as for store i observed in fiscal period t. The forecasting goal is to predict with a fixed forecast horizon h=1 fiscal period, using information available up to period t. Hence, all explanatory variables are constructed to ensure consistency across all variables and stores. Feature engineering was conducted under the constraint that all predictors must be historically available and aligned with the fiscal calendar.

Throughout the feature engineering process, explicit screening and validation were applied, where candidate variables were evaluated against target variable using correlation diagnostics and out-of-sample performance checks. Only check that demonstrated stable relationships with sales were kept.

4.2.2 Lagged Features

Lagged sales were included as predictors to capture temporal dependence and persistence in store-level demand. Lagged terms provide capture information on recent sales dynamics and momentum effects, which are central drivers in this business.

4.2.3 Fiscal Seasonality Representation

Seasonal effects were modeled explicitly at a fiscal period level to reflect the reporting structure that the business follows. Fiscal seasonality was encoded using a cyclic transformation of the fiscal period index. Let $pt \in \{1,\ldots,P\}$, seasonality was represented using sine and cosine transformations: ,

These transformations jointly encode the position of each fiscal period within the annual cycle. Preserving the circular structure of the fiscal calendar. This representation ensures continuity between the start and end of fiscal period. Compared to dummy-based seasonality, cyclical encoding reduces parameter dimensionality and allows more stability as variables.

4.2.4 External Covariates and Event Indicators

A wide range of event indicators was explored in the feature engineering process to capture irregular demand shocks. These included major U.S events and holidays such as Thanksgiving & sporting events. While Superbowl exhibited a consistent effect across stores, its magnitude at the aggregated level remained minimal

In practice, the explanatory power of most event variables was limited. A key reason is that many of these events are of short duration, often occurring on a very limited time or a fall on a single day while the target variable is aggregated over an entire period. As a result, event effects tend to be diluted when averaged across periods, reducing their signal to noise ratio. This issue was also persistent, when weather as a variable was tested, heatwaves or extreme cold days that might push sales of the pizza chain did not show a strong magnitude for the same issue.

4.2.5 Excluded Variables

Several categories of external variables were tested during feature exploration but ultimately excluded. Cost related variables, such as delivery cost or operational costs were found to be highly correlated with sales as they’re derived from total net sales hence, they were excluded as introducing this variable would introduce endogeneity and risk instability of the model.

Promotional features were initially considered and other features that might be considered as a contributing variable, but they were not present in the dataset provided by the client.

4.2.6 Feature selection Guardrails

The final feature set was determined through a combination of empirical testing and methodological constraints. Features were retained only if they were exogenous to sales and did not introduce multicollinearity. The guardrail approach ensured that final features offer interpretability over marginal gains from unstable or redundant predictors.

4.3 Model benchmarking & Evaluation

The objective of benchmarking framework is to evaluate forecasting approaches under consistent conditions to identify the most suitable modeling strategy, rather than assuming a single optimal model.

A set of candidate model families was considered, representing different levels of flexibility and complexity. These included regularized linear regression models and tree-based machine learning approaches. All models were trained using the same features described in section 4.2 to ensure that performance differences reflect modeling assumptions rather than variation in available information.

4.3.1 Evaluation Metric

Forecast accuracy was evaluated using weighted absolute percentage error (WAPE), defined as:(1)

WAPE was selected as the primary evaluation metric due to its stability for retail and QSR businesses. Unlike mean squared error, WAPE scales errors relative to observed sales, providing a size-adjusted measure of accuracy that better reflects operational impact of forecast deviations across heterogeneous stores. Concretely, WAPE sums the absolute errors across all store-period observations and divides by the total observed sales, which means large-volume stores contribute proportionally more to the overall score — preventing small stores from distorting the metric. A WAPE of 5% means that in aggregate the model's total absolute error equals 5% of total actual sales. The implementation used in this work calculates WAPE as follows:

```
# WAPE defined as a standalone function using NumPy
```

```
def wape(y_actual, y_forecast):                # y_actual, y_forecast: arrays
    numerator   = np.sum(np.abs(y_actual - y_forecast)) # total absolute error
     denominator = np.sum(np.abs(y_actual))             # total actual sales
    return (numerator / denominator) * 100         # expressed as a percentage
```

The numerator accumulates the absolute difference between forecast and actual sales across all stores and periods; the denominator is the sum of actual sales. This ratio is therefore weighted toward higher-volume stores, which is appropriate since forecast errors at large stores have greater operational impact than equivalent percentage errors at small stores. A full code listing of the feature engineering, model training and WAPE computation is provided in Appendix A.

4.3.2 Validation Strategy

Model evaluation was conducted using a fixed holdout strategy. For each store, the last 6 fiscal periods were reserved as an out-of-sample validation window k=6. To ensure the reliability of the evaluation, minimum historical depth constraints were enforced due to uneven historical depth across stores. Stores were required to have at least 12 historical fiscal periods to be eligible for model training. A stricter threshold of 18 historical fiscal periods was used for evaluation.

This evaluation approach preserves data quality and avoids information leakage, this approach reflects the practical constraints of multi-store retail forecasting where models are trained on historical data and assessed on recent, unseen periods.

4.4 Forecasting Architecture

4.4.1 System Overview

The forecasting framework is applied as a system with conditional logic instead of a single estimator across all stores to account for the differences between each store. The segmentation is enforced during execution and forms the essence of the model's architecture. The pipeline produces a single forecast per store-fiscal period, but the model family used can differ by store depending on historical depth and observed validation performance.

4.4.2 Long-History Stores

For stores with sufficient history (>18 fiscal periods), the forecasting is performed using pooled ridge for predictions. This approach allows information sharing while maintaining stability in coefficient estimates

As discussed in the literature, pooled is appropriate in multi-unit retail environments where stores share common patterns but differ in scale and noise characteristics. The pooled model leverages cross-sectional difference to reduce variance and improve generalization cross correlated predictors.

The framework is implemented in Python, with each component relying on a specific library. The Ridge regression model is trained using sklearn.linear_model.Ridge from scikit-learn (Pedregosa et al., 2011), which handles coefficient estimation with L2 regularization. Store dimensionality reduction is performed using sklearn.decomposition.PCA, which compresses correlated store-level features into principal components for clustering. Store segmentation then uses sklearn.cluster.KMeans to group stores into behavioral clusters, and sklearn.preprocessing.StandardScaler is applied to standardize features before PCA. All data manipulation, lagged feature construction, and fiscal period alignment are handled with pandas, while NumPy is used for array operations including the sine/cosine seasonality encoding (np.sin, np.cos) and WAPE computation. CPI data is fetched live from FRED via the fredapi library. On the infrastructure side, pyathena is used to query store sales and labor data from AWS Athena, and boto3 manages reading and writing Parquet files to and from S3. The pooled ridge model is initialized with the following parameters:

- Regularization: alpha =5.0
- Holdout window size: k = 6
- Forecast Horizon: H=6
- Minimum eligibility for history training: 12 periods
- Long-history cohort threshold: 18 periods

c) Features used by pooled model

The execution layer of pooled ridge used a fixed and explicitly defined features which are the following:

- Lagged sale: total_net_sales_m1
- Lagged advertising: adv_m1
- Lagged CPI: CPI_m1
- First cyclic seasonality: p13_sin, p13_cos
- Irregular fiscal period is_5w
- Store fixed effect (dummy store identifier)
- Event indicator Super Bowl

The featured list is fixed at runtime to ensure comparisons across model families reflect modeling assumptions rather than differences in available information. The pooled-ridge model produces store-level sales forecasts for each fiscal period in the forecast horizon along with evaluation metrics computed over the holdout window.

4.4.3 Short History stores & Peering Architecture

Short-history stores are (<18 fiscal periods) provide insufficient temporal depth to support stable estimation of the pooled regression framework at store-level. Key issues include limited observation relative to number of predictors, unstable coefficient estimates, sensitivity to shocks, which dominate when historical depth is short.

Although Ridge regularization reduces variance, it cannot compensate for a lack of historical data. As a result, the framework does not attempt to apply the long history pooled ridge uniformly across all stores and instead introduces peer modeling approach.

To address this limitation, the framework constructs a store-level representation using a broader set of features compared to the used selected features in pooled modeling and applies principal component analysis. PCA is used to compress correlated store characterizes and reduce noise and multicollinearity.

Based on resulting PCA scoring, stores are assigned to 3 clusters. Cluster locking ensures stores are compared to peers within same behavioral group, reducing risk of transferring coefficient across groups that have different patterns. Clusters are also used to provide the client the difference in behavior for each cluster

Peers are identified for short-history stores based on clustering and store histories. Peer construction enforced 3 peers for each short-history store & for peers to have >18

periods. Candidate peers are ranked according to Euclidian distance in the PCA, and the 3 closest eligible stores are selected.

for each data-sparse store, forecasts are created by transferring information from its matched peer stores. This process does not involve training separate forecasting models for peers. Instead, the pipeline retrieves already generated pooled model forecasts of the normalized forecasting table.

Let denote the selected peer for short history store . The final forecast for each short store is obtained by first averaging the pooled-model forecasts of the peer stores, which represents the expected sales learned from similar long-history stores. Because the peer stores may operate at a different scale, this peer forecast is then adjusted by a scaling factor that accounts for the difference in magnitudes between the short history stores and its peers. The scaling factor is defined as the ratio between the average observed sales of short history store and the corresponding average of observed sales of peer stores as shown below.

4.4.4 Adaptive Model Selection & Store Graduation Approach

For each store, forecast accuracy is calculated per store using WAPE, computed over a rolling out-of-sample evaluation window. WAPE is used as an evaluation metric to prevent the dominance of large stores and ensure model selection is personalized and reflects each store.

For each store, the framework compares the WAPE achieved by the pooled ridge and the peer-based forecasting. The model yielding lower error is selective as the forecasting model for this store until evaluation takes place again.

To ensure stability and that the decision behind switching to models is not based on noise, a tolerance margin is applied. Model reassignment only takes place, when improvement in WAPE exceeds the predefined threshold to reflect meaningful performance.

While peer-based approach is provided as a solution for short history stores, it is noteworthy to elaborate that it is not enforced as a mandatory solution for all short history stores as model selection remains strictly performance driven.

Over time, as additional observations become available, relative performance of peer-based and pooled may change and that reflects more accuracy. In this sense, graduation from peer-based forecasting is not restricted but decided through continuous evaluation. However, the framework supports the graduation of short-history stores from peer to pooled modeling only if WAPE improves sufficiently.

5. Empirical Results

5.1 Store Representation via PCA

Unlike the pooled forecasting model, which was built on a selective set of features that range from internal to external features. PCA, on the other hand, is constructed using a much broader feature set. These features cover various aspects such as sales scale, labor structure, demand mix, and event-related uplifts. The purpose of PCA is not prediction, but representation to capture differences between stores.

Table I: PCA dominant positive & negative drivers

Component	**Explained Variance**	**Positive Drivers**	**Negative Drivers**
PC1	30.4%	Log_avg_net_sales, avg_net_sales, halloween_uplift	Avg_phone_pct, avg_other_pct, avg_occupancy_pct
PC2	17.6%	Sales_roll3_std_mean, super_bowl_uplift, sales_cv	Avg_other_pct, avg_labor_pct, avg_occupancy_pct
PC3	12.0%	Back_to_school_uplift, trend_recent_growth, halloween_uplift	avg_food_pct, avg_ad_pct, avg_net_sales

The 1st component (PC1), explaining 30.4% of total variance, represents differences in overall store scale. Positive drivers are associated with higher sales intensity and stronger events uplifts, while negative drivers relate to more service-heavy operations. As a result, PC1 measures the difference in overall sales intensity across stores.

The 2nd component, PC2, explains 17.6% of the variance; it reflects more demand stability against volatility, while negative drivers' contributions are from labor and occupancy-related features. Higher PC2 values correspond to stronger sales variability and event

sensitivity, while lower values reflect more stable demand patterns. These components captured how predictable a store can be.

The 3rd component, PC3, explains 12% of the variance, and captures recent growth dynamics. It is driven by trend and seasonal responsiveness and negative contributions from average sales and advertising intensity.

Accounting for all 3, these components summarize 3 key dimensions of store behavior: size, stability, and growth. The results of PCA scores are used as inputs to the clustering stage, which will be discussed in the following section 5.3.

5.2 Store Behavior Clusters

Following the behavioral and a representation defined by PCA, stores are clustered using PCA scores. Clustering is performed on derived components to group stores that exhibit similar combinations of scale, growth, and volatility rather than similar levels of historical length.

The 2-D projection shown in Figure 1 serves solely as a visualization aid and does not reflect the full dimensionality of the clustering process, as all 3 PCs that were used for clustering.

The 1st cluster corresponds to lower volume, operationally stable stores. These stores have a lower range of PC1 values and exhibit relatively low PC2 scores, indicating smaller average sales with limited short-term volatility and service-heavy operations.

The 2nd cluster corresponds to medium-volume stores characterized by higher responsiveness to calendar-driven events. While average sales are moderate, these stores exhibit greater volatility across periods, making them less predictable.

Finally, the 3rd cluster corresponds to higher-volume stores with strong growth. These stores combine larger scales with lower volatility, reflecting more efficient operations and mature stores.

These clusters are not meant to act as performance rankings, but a structural classification to identify how each group behaves and assign suitable models. Stores within the same cluster may differ in state or size, yet exhibit similar behavior to demand drivers. These clusters are utilized in the following section through model assignment.

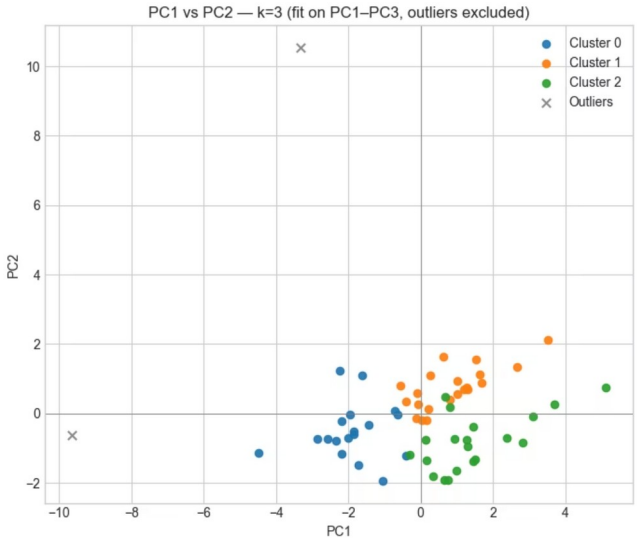

Figure 1: Store Behavior Clusters

5.3 Model Selection

Model selection takes place after clustering. This rule determines the suitable forecasting approach for a store depending on historical length and testing metrics. Initially, each store is assessed based on historical length; stores with at least 18 fiscal periods are directly assigned to the pooled model, while stores with fewer than 18 fiscal periods are flagged as short history stores and go through additional evaluation.

As mentioned in the methodology chapter, forecast accuracy is evaluated over the most recent 6 fiscal periods using WAPE. The additional evaluation for flagged short history stores compares the pooled modeling WAPE against the peer modeling WAPE. If the pooled model outperforms peer modeling by a margin of three percentage points, the pooled model is assigned; otherwise, the peer modeling is assigned.

As a result of this evaluation, 36 stores have been assigned to pooled modeling, and 25 stores have been assigned to peer modeling. Model assignment will be reevaluated periodically, allowing short histories to be evaluated after acquiring more historical periods and graduating to pooled modeling if conditions are satisfied.

5.4 Store-Level Performance

This section presents illustrative store-levels using the final hybrid forecasting framework. Each store is forecasted using its assigned model, either pooled or peer-based, on the mentioned model selection rule in the previous section. The objective is to

showcase how the model behaves across different store profiles, as well as the difference between the model's performance and the client's performance.

All figures display actual sales, model forecasts, and client forecasts over the same historical window (P6-P11).

5.4.1 Long History Stores

Figures 2 and 3 present two representative long history stores forecasted using a pooled model, as these stores possess sufficient direct estimation within the pooled regression framework.

Across both examples, the pooled model exhibits close alignment with actual sales, capturing both directional and level changes in demand. Relative to the client forecast, the pooled model reduces optimism bias, particularly in periods where actual sales fall below budgeted expectations.

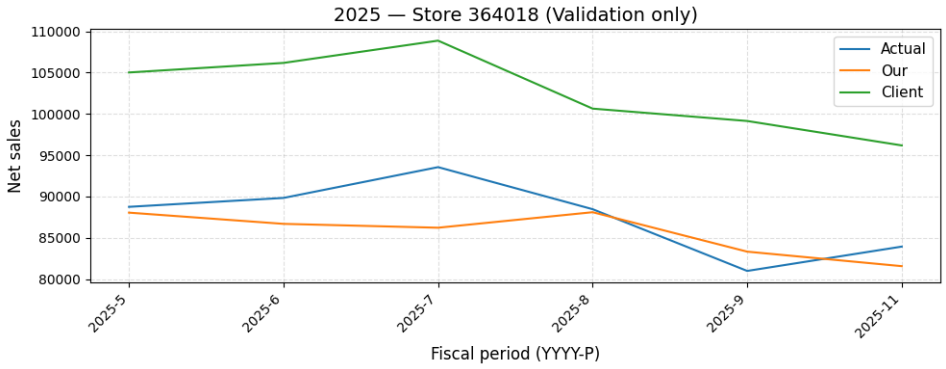

Figure 2:State: New York | No. of Periods: 44

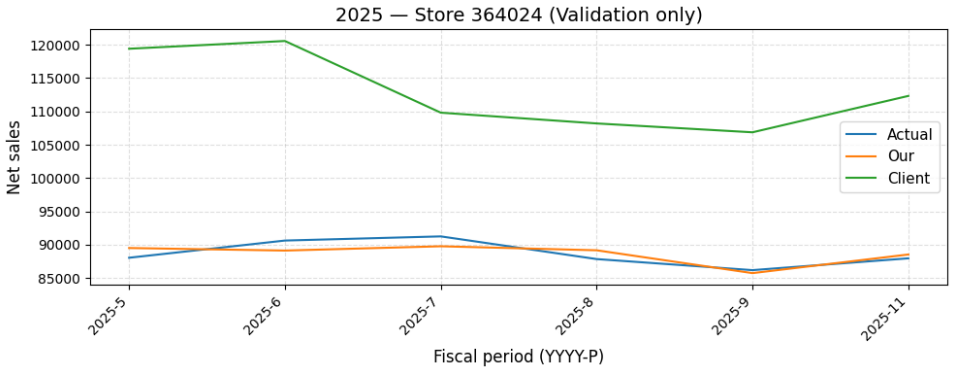

Figure 3:State: Michigan | No. of Periods: 33

5.4.2 Short History Stores

Figures 4 & 5 present two representative short history stores using a peer-based approach. Due to limited historical depth, these stores went through the model selection logic and were accordingly assigned to peer modeling.

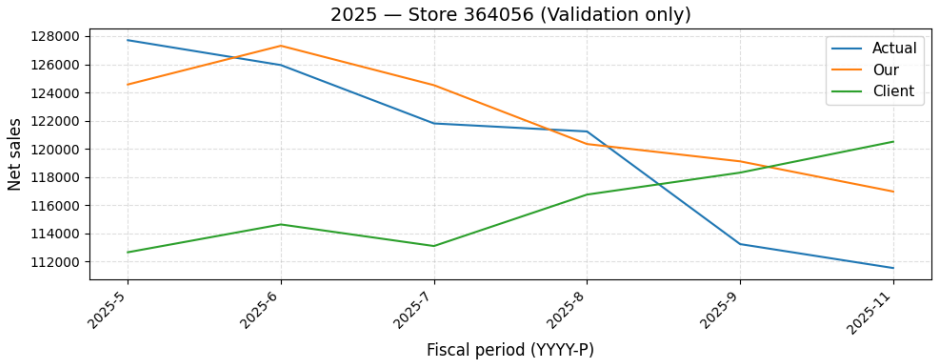

Figure 4: Michigan | No. of Periods: 9

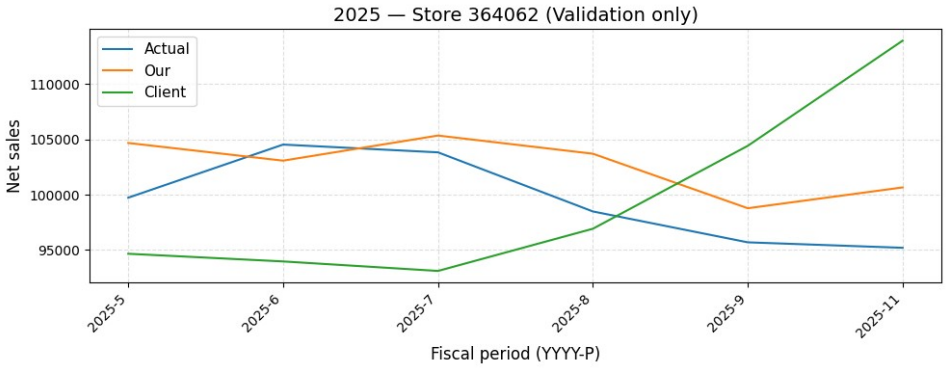

Figure 5: Indiana | No. of Periods: 9

In both examples, peer-based provides stable demand trajectories that avoid excessive volatility and optimism. However, short history stores still exhibit higher uncertainty.

5.5 Aggregate historical Alignment

Figure 6 presents aggregate net sales across all stores over the fiscal the last 6 fiscal periods. The figure compares average net sales with forecasts produced by the hybrid model and the client benchmark.

Across the evaluation window, the hybrid model remains closely aligned and tracks observed sales dynamics. However, a negative deviation in P10 was observed, where the hybrid forecast slightly overestimates actual sales. This reflects a temporary lag in

reacting to decline rather than a persistent forecasting bias, but the model realigns in the following period.

Table 2 quantifies these patterns over the full P6-P11 window. Total actual sales amount to $32.73 million compared to $32.82 million under the hybrid model and $34.03 million under the client forecast. This corresponds to only 0.3%, equivalent to $97k deviation from reality compared to the client's benchmark, which is 3.99%, equivalent to $1.3 million.

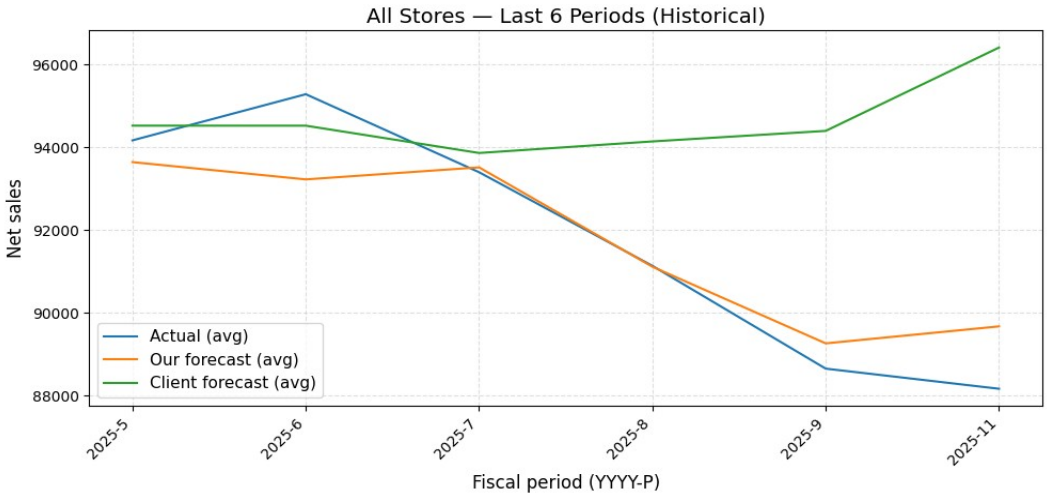

Figure 6: Aggregate Forecast Performance Across All Stores (Historical Periods)

Table II: Aggregate gap analysis

Fiscal Period	Actual Sum	Our Forecast Sum	Client Forecast Sum	Client Gap	Client Gap %	Our Gap	Our Gap %
2025-1	4,901,299	4,455,294	5,266,582	365,283	7.45%	-446,005	-9.10%
2025-2	5,093,144	5,150,133	5,260,981	167,837	3.30%	56,989	1.12%
2025-3	5,231,150	5,350,763	5,625,539	394,389	7.54%	119,613	2.29%
2025-4	5,457,598	5,485,464	5,743,634	286,036	5.24%	27,866	0.51%
2025-5	5,650,379	5,692,208	5,671,698	21,319	0.38%	41,829	0.74%
2025-6	5,717,317	5,746,781	5,671,650	-45,667	-0.80%	29,464	0.52%
2025-7	5,604,144	5,716,362	5,632,066	27,922	0.50%	112,218	2.00%
2025-8	5,468,372	5,534,019	5,648,812	180,440	3.30%	65,647	1.20%
2025-9	5,319,219	5,337,999	5,664,098	344,879	6.48%	18,780	0.35%
Totals	**48,442,622**	**48,469,023**	**50,185,060**	**1,742,438**	**3.60%**	**26,401**	**0.05%**

5.5.1 Forward Horizon

Figure 7 illustrates the behavior of the final hybrid model relative to the client's forecast over the forward period available in the dataset from the client's side (P12-P13). During the historical window, the hybrid model captures overall decline and subsequent stabilization across stores. In the forward horizon, both approached a recovery. However, the hybrid model exhibits a more restrained adjustment following the observed downturn.

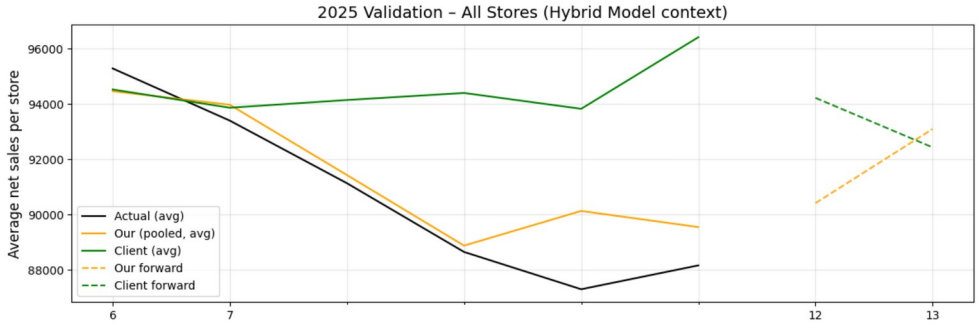

Figure 7: Forward Horizon

5.6 Forecast Accuracy Metrics

As summarized in Table 3, the hybrid model consistently outperforms the client's forecasts across all four-accuracy metrics. From a business perspective, improvements in forecast accuracy directly translate into more reliable planning and budgeting. Since operational and financial decisions are driven by forecasted demand. Hence, reducing errors helps limit unnecessary cost allocations and improves efficiency.

Table III: Performance metrics

Metric	Hybrid model	Client forecast	Improvement vs client
MAE	4,800	11,900	59% lower
RMSE	6,400	16,500	61% lower
MAPE	5.6%	13.2%	57% lower
WAPE	5.3%	13.1%	59% lower

5.7 Operational Implications

In addition to evaluating sales forecast accuracy, the final hybrid forecasting framework was used to derive operational implementations for labour planning and

budget allocation. These results are not produced through a separate predictive model but are derived mechanically from sales forecasts using historically observed relationships.

5.7.1 Labor Cost Estimation

Labor cost and hours were not modelled using an independent forecasting framework. Instead, it was derived directly from sales forecasts produced by the final hybrid model, using the historical labour-to-sales ratio at the store level. For each store s, the historical labour ratio was computed:where t denotes fiscal periods with available historical observations.

To ensure robustness, especially for stores with limited historical data, some safeguards were applied. For instance, in case labour ratio could not be derived, this ratio would be estimated using the median labour ratio across all stores.

The labour cost forecast for each store and period was obtained by multiplying the predicted sales by this ratio. Figure 8 presents an illustrative comparison that compares actual labour costs compared to hybrid model forecasts and client's benchmark.

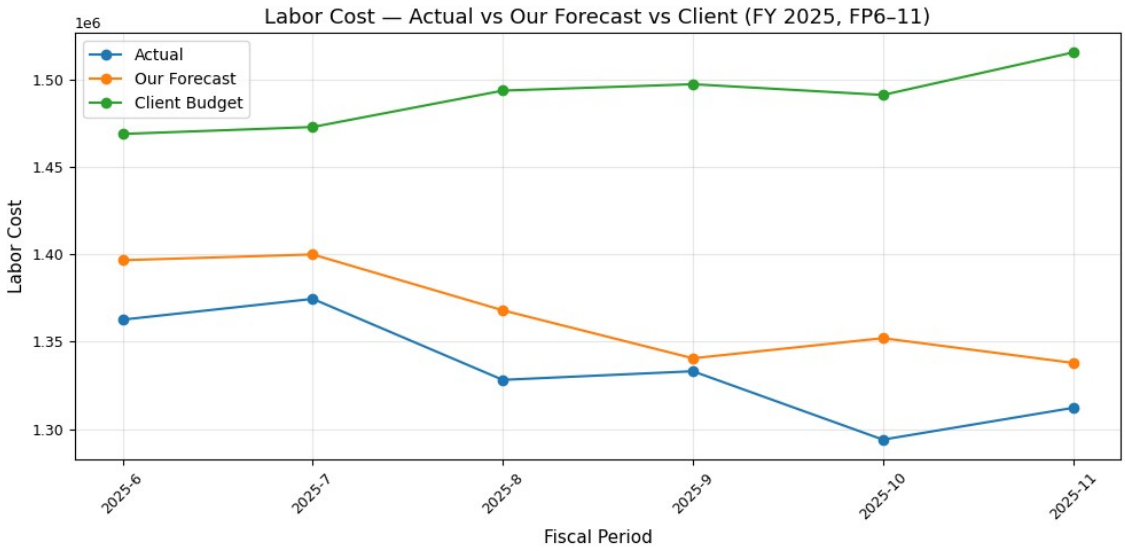

Figure 8: Cost comparison

5.7.2 Budget Gap Analysis

Figure 9 and table 4 present budget gap for P12 and P13, calculated as the difference between final sales forecast by the hybrid model and budget values provided by the client. This visualization highlights clear variation across states, with period 12 showing larger

negative gaps such as Indiana and Michigan, indicating budgets in these locations were set above the level implied by updated sales forecasts.

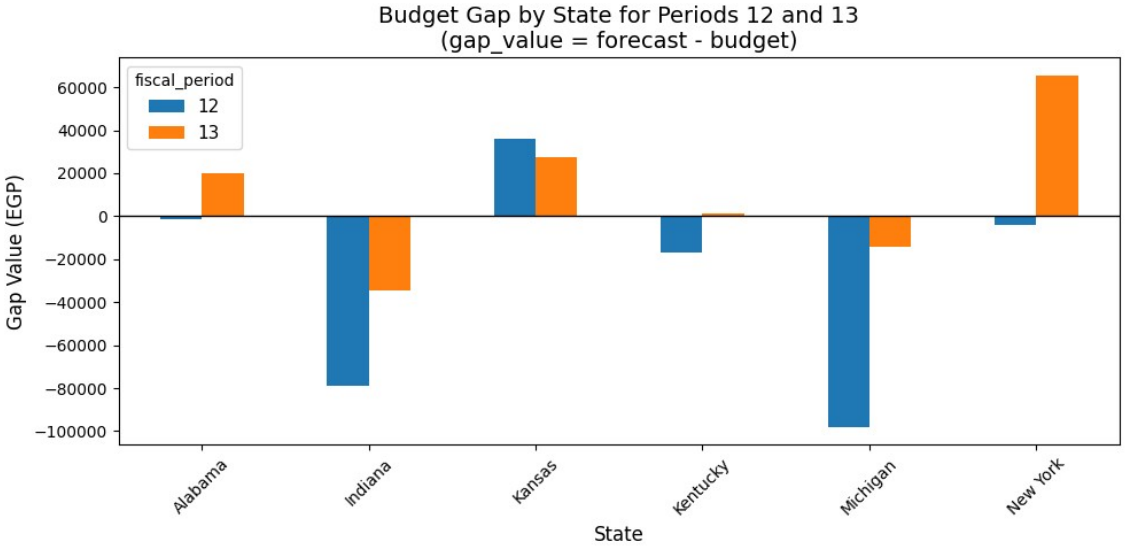

Figure 9: Budget gap comparison

Table IV: Budget gap values

Fiscal Year	Fiscal period	Total Forecast	Total Budget	Gap Value
2025	12	5,490,190	5,653,459	-163,269
2025	13	5,611,170	5,546,126	65,044

6. CONCLUSION

The master's final work addressed the challenges of forecasting sales in a multi-unit food & beverage network characterized by heterogeneous historical data availability. This

project was focused on US fast food chain operator managing 61 stores in different states, where traditional regression forecasts consistently overestimated demand and led to inefficient operational decisions and inefficient cost allocation.

To overcome these limitations, an adaptive forecasting framework was developed that assigns models based on data availability and observed performance rather than applying a single approach globally. Stores with sufficient historical data were automatically assigned pooled model & stores with limited historical length went through an evaluation process that either yields peer modelling or pool modelling based on WAPE performance.

Empirical results show that this hybrid framework substantially improved forecast accuracy achieving 59% reduction in WAPE compared to client's forecast. At an aggregate & store-level forecasts aligned with actual sales significantly and reduced optimism. These gains translated into a more reliable budgeting assumptions and improved operational planning.

The framework also highlighted important consideration when developing a forecasting model, effective forecasting required a deep understanding of how the business operates. It is mandatory to understand operational realities that are not captured in data. In addition, data completeness and coverage emerged as a key constraint, particularly when aggregating daily information to fiscal periods and realizing its incompleteness. Data quality checks are also mandatory to perform in depth before starting the modelling process.

The value of the proposed framework was recognized by the client, who requested its extension to a larger portfolio operating approximately 197 stores. Future work may therefore focus on scaling the system to larger networks and incorporating machine learning, such as SageMaker AutoML.

Overall, this work demonstrated that adaptive, performance-driven forecasting systems can effectively manage data heterogeneity in a multi-unit retail setting delivering both improved accuracy and business value.

REFERENCES

Chen, H., & Boylan, J. E. (2008). Empirical evidence on individual, group and shrinkage seasonal indices. *International Journal of Forecasting*, 24(3), 525–534.

Chen, I.-F., & Lu, C.-J. (2021). Demand forecasting for multichannel fashion retailers by integrating clustering and machine learning algorithms. *Processes*, 9(9), 1578.

Ekiz Yılmaz, T., & Yapar, G. (2025). A robust hybrid forecasting framework for the M3 and M4 competitions: Combining ARIMA and ATA models with performance-based model selection. *Applied Sciences*, 15(17), 9552.

Fildes, R., Ma, S., & Kolassa, S. (2019). Retail forecasting: Research and practice. *International Journal of Forecasting*, 35(1), 1–19.

Hewamalage, H., Bergmeir, C., & Bandara, K. (2022). Global models for time series forecasting: A simulation study. *Pattern Recognition*, 124, 108441.

Hwang, S., Lee, Y., Jeon, B.-K., & Oh, S. H. (2025). Sales forecasting for new products using homogeneity-based clustering and ensemble method. *Electronics*, 14(3), 520.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice* (2nd ed.). OTexts.

Liu, J. (2024). Navigating the financial landscape: The power and limitations of the ARIMA model. *Highlights in Science, Engineering and Technology*, IFMPT 2024, 88.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. *PLOS ONE*, 13(3), e0194889.

Montero-Manso, P., & Hyndman, R. J. (2021). Principles and algorithms for forecasting groups of time series. *International Journal of Forecasting*, 37(2), 553–570.

Montero-Manso, P., & Hyndman, R. J. (2021). Principles and algorithms for forecasting groups of time series: Locality and globality. *International Journal of Forecasting*, 37(4), 1632–1653.

Park, J., Lee, K., & Jang, S. (2022). A review of forecasting studies for the restaurant industry: Focusing on results, contributions and limitations. *Global Business & Finance Review*, 27(2), 61–77.

Sanders, N. R., & Graman, G. A. (2009). Quantifying costs of forecast errors: A case study of the warehouse environment. *Omega*, 37(1), 116–125.

Aguinis, H., & Edwards, J. R. (2014). Methodological wishes for the next decade and how to make wishes come true. *Journal of Management Studies*, 51(1), 143–174.

Cheng, T., Jiang, Z., Lim, A., & Yao, T. (2025). Forecasting sales of new products: A systematic literature review. *International Journal of Production Economics*, 279, 109453.

Fildes, R., Petropoulos, F., Kolassa, S., & Rostami-Tabar, B. (2022). Retail forecasting: Research and practice. *International Journal of Forecasting*, 38(4), 1283–1318.

Hananya, R., & Katz, R. (2024). Adaptive model selection in time series forecasting: A dynamic evaluation framework. *Expert Systems with Applications*, 238, 122038.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.

Kourentzes, N., Barrow, D., & Crone, S. (2018). Neural network ensemble operators for time series forecasting. *Expert Systems with Applications*, 41(9), 4235–4244.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

Ramos, P., Santos, N., & Rebelo, R. (2022). Performance of state space and ARIMA models for consumer retail sales forecasting. *Robotics and Computer-Integrated Manufacturing*, 34, 151–163.

APPENDIX A – KEY CODE SNIPPETS

A.1 Feature Engineering – Fiscal Period Mapping

The following function maps daily or monthly dates to the client's 13-period fiscal calendar. It expands the calendar into a day-level lookup table and merges on exact date, ensuring that CPI, event flags and other daily signals are correctly aligned with the fiscal period in which they fall.

```
def map_to_fiscal_period(df, date_col, calendar):
    # Expand calendar: one row per day per fiscal period
    cal_daily = (
        calendar.assign(_date=lambda d: d.apply(
            lambda r: pd.date_range(r["beginning_date"],
                                    r["ending_date"], freq="D"), axis=1))
        .explode("_date")[["_date", "fiscal_year", "fiscal_period"]]
    )
    out = df.copy()
    out["_date"] = pd.to_datetime(out[date_col]).dt.normalize()
                            return      out.merge(cal_daily,      on="_date",
how="inner").drop(columns="_date")
```

A.2 Pooled Ridge Training and WAPE Evaluation

The snippet below shows the core of the PooledRidgeForecaster.train() method. Store dummy variables (fixed effects) are added to the design matrix via pd.get_dummies. The model is first trained on a training split, evaluated on a per-store holdout (k=6 periods), and then re-fit on all eligible data. WAPE on the holdout is computed as the ratio of total absolute error to total actual sales.

```
# Build design matrix with store fixed effects
X  =  pd.get_dummies(df[feature_list  +  [store_col]],  columns=[store_col],
drop_first=True)
y = df[target]
model = Ridge(alpha=alpha).fit(X_train, y_train)
# Holdout WAPE
preds = model.predict(X_valid)
```

```
wape = np.sum(np.abs(y_valid - preds)) / np.sum(np.abs(y_valid))
```

A.3 Peer-Based Forecast Scaling

For short-history stores, the forecast is derived by scaling the average pooled-model forecasts of k=3 peer stores by a ratio that accounts for differences in store scale. The scaling factor is the ratio of the short store's average historical sales to the average of the peers' first 10 periods of observed sales. This ensures the forecast is anchored to the short store's own sales level while borrowing the temporal dynamics from data-rich peers.

```
scale = avg_short_store / mean(avg_peer_first10)

peer_forecast = scale * mean(pooled_forecast_peer1,
                             pooled_forecast_peer2,
                             pooled_forecast_peer3)
```