Weather- and Event-Aware Demand Forecasting for Ordering Decisions in Small Convenience Retail

Fouad Bouabdi-Benaatou

Personal Information

Student ID: 2242740
Student Name: Fouad Bouabdi-Benaatou
Course: Research Methods and Professional Development ACFA7004
Teacher: Dr Aliyu Sambo
Assessment: Research Proposal
Address: University of Wales Trinity Saint David
Birmingham

Publication Date: October 21, 2026
Email: 2242740@student.uwtsd.ac.uk

Contents

1 Introduction **3**
1.1 Context and Background . . . 3
1.2 Research Problem . . . 3
1.3 Research Aim and Objectives . . . 4
1.4 Significance of the Research . . . 4

2 Literature Review **4**
2.1 Machine Learning for Retail Demand Forecasting . . . 4
2.2 External Drivers: Calendar Events and Weather . . . 5
2.3 From Forecasts to Ordering Decisions . . . 5
2.4 Theoretical and Technical Frameworks . . . 6
2.5 Relation to This Study . . . 6
2.6 Summary . . . 6

3 Research Design and Methodology **6**
3.1 Research Design . . . 6
3.2 Research Methods . . . 7
3.3 Materials . . . 8

4 Expected Outcome **9**

5 Timeline **9**

List of Figures **11**

List of Tables **11**

List of Abbreviations **11**

1 | Introduction

1.1 Context and Background

Demand forecasting is the task of estimating how much of each product customers will buy in a future period. In retail, these forecasts are used for ordering, stock control and staff planning, so they affect both costs and customer service [1]. The field sits between data mining, machine learning and operations research, and it is a difficult one. A single store may sell thousands of products, many of them sell only a few units per day, and demand changes with prices, promotions, holidays and the weather. Fildes et al. describe this as a dimensionality problem: there are too many possible variables and too little data to estimate their effects [1].

For a long time, statistical methods were the standard. Exponential smoothing models are still widely used because they are simple and work well for many regular series [2]. For products with many zero-sales days, Croston's method and the bias-corrected version proposed by Syntetos and Boylan are common choices [3].

The main recent development is the move to machine learning (ML). The M5 competition used Walmart sales data for 3,049 products in ten stores across three US states, covering more than five years. All of the top-performing methods were based on ML, and most used gradient-boosted trees such as LightGBM [4, 5]. An important reason for their success was cross-learning, where one model is trained on many related series at the same time [4]. Some ML methods have also given better forecasts than Croston-type methods on daily SKU data with intermittent demand [6]. In a study of a bakery chain that focused on calendar special days, ML methods were more accurate than established forecasting methods [7].

A second trend is the use of external data. Badorf and Hoberg studied 673 stores and found that the effect of weather on daily sales can reach 23.1% depending on the store location. They also showed that weather forecasts improved sales forecasts for up to about seven days ahead [8]. A study of 146 convenience stores in China found that higher temperature was linked to higher sales, but the effect depended on the product category [9].

A third trend is more attention to the decision that follows the forecast. Work on prescriptive analytics and on the "predict, then optimise" approach argues that prediction models should be designed and assessed with the later decision in mind [10, 11]. In inventory control, this links forecasting with the newsvendor problem, which balances the risk of ordering too much against the risk of ordering too little [12]. New public datasets with stockout labels and weather data, such as FreshRetailNet-50K, now make it possible to study these questions on realistic data [13].

This study is placed in the UK convenience sector. According to the Association of Convenience Stores, there are 50,486 convenience stores in mainland UK, and forecourt shops are counted within this sector [14].

1.2 Research Problem

The specific problem addressed in this study is how a small convenience store, such as a forecourt shop or an off-licence, should decide how much of each product to order when demand changes with events and weather. If the store orders too much, short-life products expire and become waste. If it orders too little, shelves are empty and sales are lost. In the author's own experience of stock control on a motorway forecourt, these decisions are often made by staff using judgement and simple rules, for example reordering what was sold in the previous week. This experience motivates the study, but it is not used as evidence.

Four challenges make this problem hard. First, the strongest methods have been shown at scale. The M5 results depended on cross-learning across thousands of series [4], and in a data-driven newsvendor study ML methods clearly outperformed traditional methods only when the dataset was large enough [12]. A single small store has a much shorter and narrower sales history. Whether the gains of ML carry over to this setting is therefore an open question, not an established result.

Second, most studies judge forecasts only by accuracy. Kolassa shows that the "best" point forecast depends on which error measure is used [15], so an accurate forecast is not automatically a good order. A store needs an order quantity that balances waste and stockouts, and this depends on costs as well as on the forecast [12, 10].

Third, recorded sales are not the same as demand. When a product is out of stock, sales understate what customers wanted, and models trained on these sales give biased forecasts [13, 16]. Recent work has linked censored-demand forecasts to ordering policies, but it used data from a large fresh-grocery platform with 898 stores [16].

Fourth, small businesses face practical barriers. A recent systematic review reports that SMEs struggle with access to data, skills, finance and infrastructure when adopting AI, and that they prefer tools which are easy to use [17]. For this reason, a design requirement of this study is that the solution must be light and cheap to run.

1.3 Research Aim and Objectives

The main research question is: *Can a weather- and event-aware ML approach, trained on the limited data of a single small store, lower the combined cost of waste and stockouts compared with simple ordering rules?*

Aim: To design and evaluate a weather- and event-aware forecasting and ordering prototype for small convenience stores, and to measure whether it reduces the combined cost of waste and stockouts compared with simple ordering rules when sales history is limited.

Objectives:

O1. To critically review the literature on retail demand forecasting, external demand drivers and data-driven ordering, and to select suitable models and baselines.

O2. To prepare public retail sales data for single-store experiments, enriched with calendar events and daily weather, and with stockout information where it is available.

O3. To develop and compare forecasting models: simple statistical baselines and machine learning models, each with and without weather and event information.

O4. To develop an ordering module that turns forecasts into order quantities with a cost-based ordering rule, with stated assumptions for shelf life and delivery days.

O5. To evaluate all approaches on past data with both forecast accuracy measures and decision measures (waste, stockouts and total cost), and to repeat the tests for sales histories from 3 to 24 months.

O6. To build a simple prototype interface and check that it runs on a standard laptop without special hardware.

1.4 Significance of the Research

For the research community, this study gives evidence on a question that large benchmarks do not answer: how much of the advantage of ML remains when a store has little data, and at what point simple rules are good enough. It also judges forecasts by the cost of the orders they produce, which follows the decision-focused view of recent work [10, 11].

In practice, the main benefit is at store level: fewer expired products, fewer empty shelves on hot days or event days, and less time spent on manual ordering. The effect on national food waste should not be overstated, because retail produces only about 2% of UK food waste by weight [18]. Possible applications include forecourt shops, off-licences and small franchise stores. If the approach works well, it could later be extended to larger stores, but this is outside the scope of the present project.

2 | Literature Review

The purpose of this review is to examine the existing literature on retail demand forecasting and to identify the main trends, arguments and gaps that are relevant to ordering in small stores. The review covers three themes: machine learning for retail demand forecasting, external drivers of demand such as calendar events and weather, and methods that turn forecasts into ordering decisions. It focuses mainly on peer-reviewed work published from 2019 onwards, with a small number of older studies that are still the basis of current methods. Table 1 summarises the key studies.

2.1 Machine Learning for Retail Demand Forecasting

Fildes et al. give a broad review of retail forecasting, from strategic decisions down to daily store-level forecasts [1]. They note that causal models usually beat simple benchmarks, but that there was not yet enough evidence on

how well ML performs in retail. Their review first appeared online in 2019, before the M5 competition, so this view reflects the evidence available at that time.

The M5 competition changed this picture. It used Walmart data for 3,049 products in ten stores, with calendar events, prices and SNAP days as explanatory variables but no weather data [4]. All of the winning methods were ML-based, and the winning solution combined many LightGBM models trained at different levels of aggregation. The organisers also found that the many zero values at product–store level made forecasting hard, and that training one model across many related series worked better than forecasting each series alone [4]. LightGBM itself is a gradient boosting method designed to train quickly on large tabular data [5], which helps to explain its popularity in this setting.

Spiliotis et al. reached a similar but more careful conclusion on daily SKU data from a single company. Some ML methods gave better accuracy and less bias than Croston's method and its variants, especially when trained across many SKUs [6]. Croston-type methods, including the bias correction of Syntetos and Boylan [3], and exponential smoothing [2] therefore remain the natural baselines.

Taken together, the evidence that ML improves retail forecasts is now fairly strong. However, it comes from large datasets with many related series. This is an important condition, and it is examined in Section 2.5.

2.2 External Drivers: Calendar Events and Weather

Huber and Stuckenschmidt studied daily demand in a bakery chain with special attention to holidays and other calendar special days [7]. They found that ML methods were more accurate than established methods, and that treating the problem as classification rather than regression gave better results. Their study shows that event information can be learned from data, but it also used data from many stores of one chain.

Weather has been studied in more detail. Badorf and Hoberg analysed 673 brick-and-mortar stores using a random coefficient model that allows non-linear effects [8]. The effect of weather on daily sales reached 23.1% depending on the store location and 40.7% depending on the sales theme, and weather forecasts improved sales forecasts for up to about seven days. Tian et al. studied more than six million transactions in 146 convenience stores in China [9]. Higher temperature was linked to higher sales, while rainy days led customers to buy more items with higher prices in one order. Again, the results depended on the product category.

These two studies agree on one point that matters for this research: weather effects are local and product-specific. A model for one small store therefore has to learn its own response to weather and events, but it has much less data to learn from than the chains in these studies. This tension has not been tested directly in the studies reviewed.

2.3 From Forecasts to Ordering Decisions

The theoretical basis for ordering under uncertain demand is the newsvendor model. It chooses the order quantity that balances the expected cost of overstock against the expected cost of a stockout, so the best order is a quantile of the demand distribution rather than its mean [12]. Huber et al. applied a data-driven version of this model to perishable products in a German bakery chain, using ML and quantile regression [12]. They found that ML approaches clearly beat traditional methods if the dataset was large enough, and that most of the benefit came from better forecasting. Bertsimas and Kallus developed a wider framework for going from data to decisions, including auxiliary data and censored observations, and tested it on inventory for a large media distributor [10]. Elmachtoub and Grigas went further and proposed training prediction models with a loss based directly on the decision cost [11]. Their tests used shortest-path and portfolio problems, not retail ordering, so their value for small-store replenishment is still unclear.

Two further points support a decision-based evaluation. First, the "best" point forecast depends on the chosen error measure [15], so accuracy alone cannot show whether a forecast leads to good orders. Second, sales data are censored when products run out. The FreshRetailNet-50K dataset provides hourly stockout labels, promotions and weather for 898 stores so that this problem can be studied [13]. Azar used it to link censored-demand forecasts with twenty ordering policies and reported fill rate and profit, using a pipeline that runs on a normal CPU [16]. This is the closest existing work to the present study, but it again uses data from a large platform rather than a single small store.

Table 1: Summary of key studies

Study	Setting	Main contribution	Limitation for this study
Makridakis et al. [4]	Walmart, 10 stores	ML (LightGBM) won M5; cross-learning	Large chain; no weather
Spiliotis et al. [6]	Daily SKU data	Some ML beats Croston-type methods	Accuracy only
Huber & Stuckenschmidt [7]	Bakery chain	ML more accurate, with focus on special days	Many stores, one chain
Badorf & Hoberg [8]	673 stores	Weather effects are non-linear and local	Not convenience retail
Tian et al. [9]	146 convenience stores	Weather effects depend on category	No ordering model
Huber et al. [12]	Bakery chain	Data-driven newsvendor; ML needs enough data	Large dataset
Azar [16]	Fresh-grocery platform	Forecasts linked to order policies	90 days; large platform

2.4 Theoretical and Technical Frameworks

The study builds on the newsvendor model as its theoretical framework, because it directly expresses the trade-off between waste and stockouts. On the technical side, the reviewed studies rely mainly on gradient-boosted trees, especially LightGBM [4, 5], quantile regression [12], and statistical baselines such as exponential smoothing and Croston-type methods [2, 3]. Forecasts are normally evaluated on hold-out periods with standard error measures [2], while decision studies add cost, fill rate or profit [12, 16].

2.5 Relation to This Study

Most of the reviewed studies share three features: they use data from large chains or platforms, they evaluate mainly forecast accuracy, and they treat events and weather separately from the ordering decision. This study is different in three ways. First, it tests the ML approach under the data limits of a single small store and measures how the benefit changes as the sales history becomes shorter. Second, it judges each method by the waste and stockouts produced by its orders, not only by accuracy. Third, it combines calendar events, weather and a newsvendor ordering rule in one light pipeline that can run without special hardware, which fits the needs of small businesses [17]. The expected contribution is therefore not a new algorithm, but new evidence on when existing methods are worth using in a setting that the literature has largely ignored.

2.6 Summary

The literature shows that ML, especially gradient-boosted trees, can improve retail forecasts, that calendar events and weather have real but local effects, and that forecasts should be judged by the decisions they support. The main gaps are that this evidence comes almost entirely from large datasets, that few studies evaluate ordering outcomes, and that the small-store setting has not been tested directly in the studies reviewed.

3 | Research Design and Methodology

3.1 Research Design

This study follows Design Science Research (DSR). In DSR, the researcher builds an artefact that addresses a real problem and then evaluates how well it works. Hevner et al. state that the artefact must address a relevant problem and must be evaluated rigorously [19]. In this study, the problem is the daily ordering decision in small convenience stores described in Section 1.2. The artefact is a forecasting and ordering tool. It takes a store's sales history, calendar events and weather, forecasts the demand for each product, and turns the forecast into an order quantity. Its relevance comes from the waste and empty shelves that small shops face. Its rigour comes from testing it against the simple rules that these shops use today.

The work follows the six activities of the Design Science Research Methodology (DSRM) of Peffers et al. [20]. Fig. 1 shows how they map to this project. *Problem identification* is the ordering problem and the literature gap set out in Sections 1.2 and 2. The *objectives of a solution* come from the aim: a lower combined cost of waste and stockouts than simple rules, with limited data, on ordinary hardware. *Design and development* covers the data pipeline, the forecasting models and the ordering module (O2–O4). *Demonstration* means running the tool on single-store data and in a simple prototype interface (O6). *Evaluation* is the backtesting of all methods for different lengths of sales history (O5). *Communication* is the MSc dissertation and the documented code.

DSR fits this study because the research question is practical: it asks whether a designed tool performs better than what small shops do now. A purely theoretical analysis could not answer this, and a survey or case study would describe current ordering without testing an alternative. A pure simulation would test methods but would not produce a usable tool. DSR also allows the tool to be improved between rounds. For example, if the first full evaluation in week 10 shows poor results for slow-selling products, the features or the ordering rule for these products will be changed and tested again in a second round.

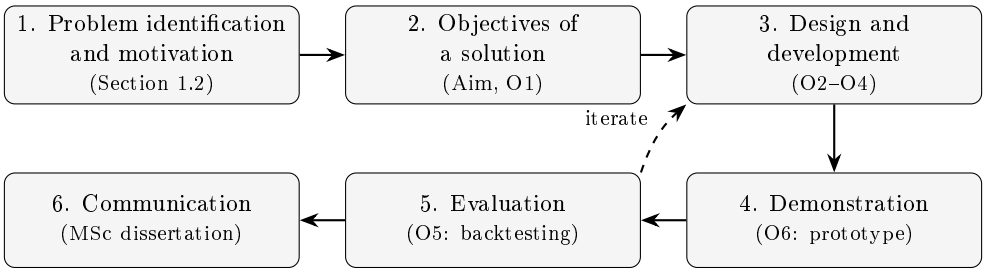

Figure 1: DSRM activities [20] mapped to this study

3.2 Research Methods

Overview. The study uses a quantitative method. Every outcome of interest can be measured as a number: forecast errors, units wasted, units short, cost and running time. This allows a fair and repeatable comparison between methods. Qualitative feedback from store staff would be useful, but it is left for the MSc project because the main question is about performance.

Strategy. The research strategy is software development combined with a controlled computational experiment. The pipeline is developed in Python and then tested under different conditions, changing one factor at a time: the forecasting model (baselines or LightGBM), the features used (none, calendar events, or calendar events and weather), the length of sales history (3, 6, 12 or 24 months), the ordering rule (order equal to the point forecast, or a newsvendor quantile), and the ratio between waste cost and stockout cost. An experiment is suitable because all methods can be tested on exactly the same periods and products, so differences in results can be linked to the factor being changed. To imitate a small store, the M5 experiments use one store and a subset of food products, with the training history cut to the lengths above. Two simple rules represent current practice: ordering the quantity sold on the same weekday of the previous week, and ordering up to a moving average of recent sales. These are compared with the statistical baselines and the LightGBM models. In the FreshRetailNet experiments, periods labelled as out of stock are treated as censored, so that low sales caused by empty shelves are not learned as low demand [13].

System design. Fig. 2 shows the architecture. Sales history, calendar events and weather are merged and turned into features such as lagged sales, day of week, event flags and temperature. The forecasting module produces point forecasts and quantile forecasts. The ordering module then sets the order quantity to the forecast quantile at the critical ratio

$$\tau = \frac{c_u}{c_u + c_o}, \tag{1}$$

where c_u is the cost of a lost sale and c_o is the cost of an unsold unit, which follows the newsvendor logic used by Huber et al. [12]. A simple inventory simulation applies each order, removes products that pass an assumed shelf life, and records waste and stockouts. Shelf life and delivery days are stated assumptions for each product category.

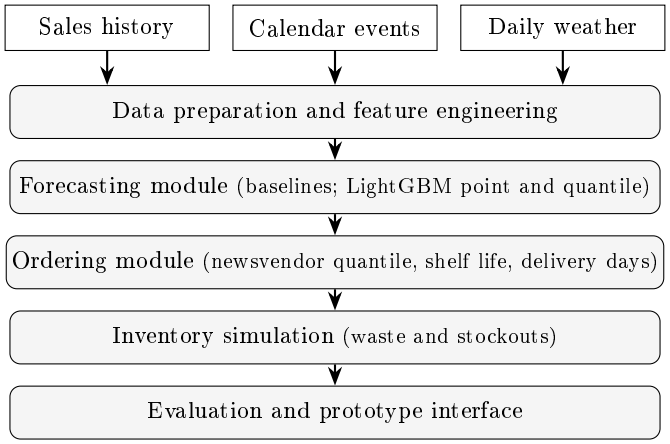

Figure 2: Proposed system architecture

Data analysis and evaluation. All methods are evaluated with rolling-origin backtesting: the model is trained on data up to a forecast origin, tested on the following days, and the origin is then moved forward. Forecast quality is measured with WAPE, RMSSE and bias. Decision quality is measured with waste units, stockout units, fill rate and total cost,

$$C = c_o \times \text{waste} + c_u \times \text{stockouts}. \tag{2}$$

Results are averaged across products, and bootstrap confidence intervals are used to check whether differences between methods are larger than random variation. Training time and memory use are also recorded for objective O6.

3.3 Materials

Research data. Two public datasets are used. The M5 data contains daily Walmart sales for 3,049 products in ten stores across three US states over more than five years, together with calendar events and prices, but no weather [4]. Its long history makes it suitable for testing different history lengths and event effects. Because M5 gives only the state of each store, daily temperature and rainfall for a representative location in each state will be taken from a public historical weather service, such as Open-Meteo, after checking its terms of use. This weather match is coarse, and it is treated as a limitation.

The second dataset, FreshRetailNet-50K, contains 90-day series for perishable products in 898 stores, with hourly stock status, promotions and weather, and it is released under a CC-BY-4.0 licence [13]. It is used for the weather and stockout parts of the study. Its short period means it cannot be used to study yearly events. The two datasets therefore complement each other. Neither of them comes from UK convenience stores, so results will be presented as evidence about the method rather than about UK shops directly. If written permission is obtained, anonymised sales data from a UK forecourt may be added in the MSc project.

Ethical considerations. The public datasets contain product-level sales and no personal data. Licence conditions will be followed, including attribution for CC-BY data and the competition terms for M5. If workplace data is used, it will only be collected with written permission from the company and ethical approval from the university. It will contain aggregated product sales only, with no customer or staff information, and it will be stored securely and not published.

Software. Python 3 with pandas and NumPy for data handling, LightGBM for gradient boosting [5], statsmodels for exponential smoothing, a custom implementation of Croston and SBA, scikit-learn for metrics, Matplotlib for charts, Streamlit for the prototype interface, Git for version control, and Overleaf for writing.

Hardware. A standard laptop with a multi-core CPU and 16 GB of RAM, without a GPU. This is deliberate, because the solution must be cheap enough for a small business, and recent work suggests that a LightGBM ordering pipeline can run on a CPU [16]. A free cloud notebook service may be used as a backup for the larger FreshRetailNet experiments.

4 | Expected Outcome

Anticipated results. With 12 or 24 months of history, LightGBM models with calendar and weather features are expected to give lower total cost than the simple ordering rules and the statistical baselines. This expectation follows the M5 results and the data-driven newsvendor study [4, 12]. With only 3 or 6 months of history, the advantage is expected to become smaller, and the statistical baselines may perform as well as ML, since Huber et al. found that ML performed clearly better only with enough data [12]. Weather features are expected to help for some product groups much more than for others, because earlier studies found that weather effects depend on the product category [8, 9]. When the cost of a lost sale is higher than the cost of an unsold unit, the newsvendor rule is expected to reduce stockouts compared with ordering to the point forecast, with some extra waste. Finally, the whole pipeline is expected to run on an ordinary laptop.

Implications. For research, the main result will be an estimate of how much sales history a store needs before ML ordering becomes worth using, which large benchmarks do not provide. For practice, the result can be turned into simple guidance for small retailers, and the prototype shows how such a tool could look. The code and experimental setup will be documented so that other researchers can repeat or extend the tests.

Limitations. Neither dataset comes from UK convenience stores, so the results show how the method behaves rather than proving its value for UK shops. The weather match for M5 is coarse because only the state of each store is known, and FreshRetailNet covers only 90 days. In M5, recorded sales are used as demand, so hidden stockouts may bias the models. Shelf lives, delivery days and costs are assumptions, and real supplier rules such as case sizes and minimum orders are simplified. The prototype will not be tested with real store staff at this stage. Each of these limitations will be reported clearly, and some of them can be reduced in the MSc project if workplace data becomes available.

5 | Timeline

The work runs from October 2026 to May 2027 in two stages, as shown in Fig. 3. Before the project starts, the proposal is submitted, the ethics form is sent in early November 2026, and the topic is agreed with the supervisor. Small tasks, such as checking the dataset licences and setting up the software, are also done before January. The project itself runs for 16 weeks, from 11 January to 30 April 2027. May is kept as extra time in case it is needed.

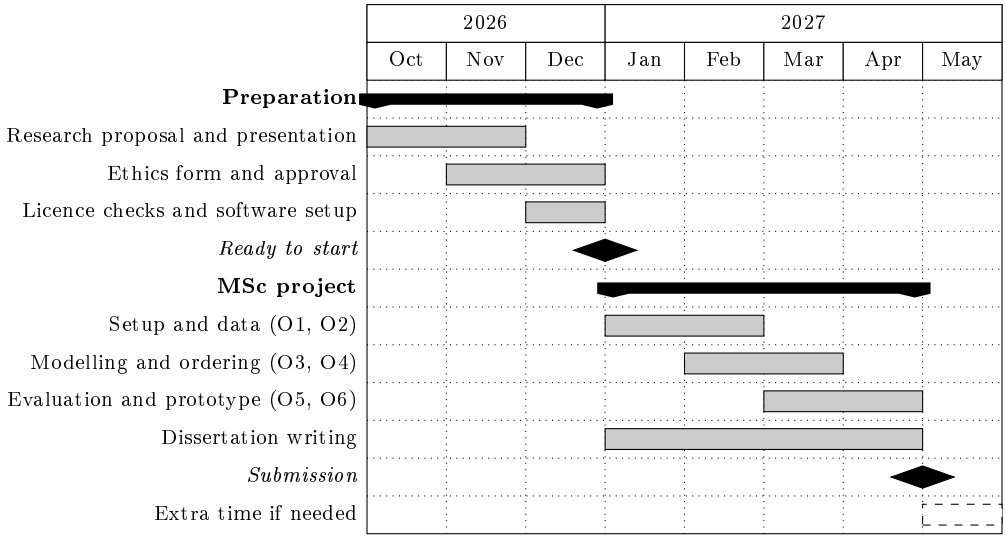

Figure 3: Overview of the work, October 2026 to May 2027

Fig. 4 shows the detailed weekly plan. The tasks are grouped into five phases: setup and data, modelling, evaluation, prototype, and writing. Supervisor meetings are planned every two weeks. Three of them are formal review points at the milestones: the data pipeline ready (week 4), the first full evaluation (week 10) and

submission (week 16). Writing starts early. The introduction and literature review chapters are drafted in the first weeks, and each later chapter is drafted soon after the related work is done. Two weeks of buffer before submission give time to fix problems if data preparation or model tuning takes longer than planned.

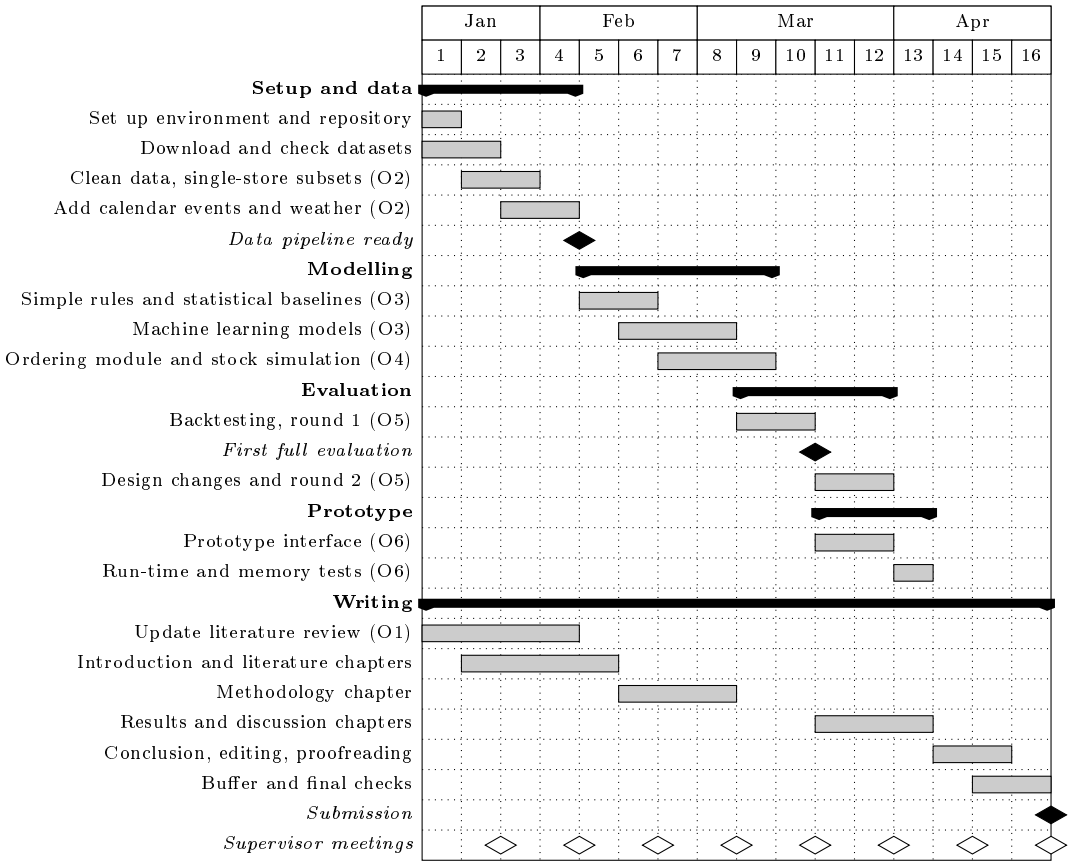

Figure 4: Detailed weekly plan for the MSc project (week 1 starts on 11 January 2027)

List of Figures

1 DSRM activities [20] mapped to this study . . . 7
2 Proposed system architecture . . . 8
3 Overview of the work, October 2026 to May 2027 . . . 9
4 Detailed weekly plan for the MSc project . . . 10

List of Tables

1 Summary of key studies . . . 6

List of Abbreviations

AI	Artificial Intelligence
CC-BY	Creative Commons Attribution (licence)
CPU	Central Processing Unit
DSR	Design Science Research
DSRM	Design Science Research Methodology
GB	Gigabyte
GPU	Graphics Processing Unit
LightGBM	Light Gradient Boosting Machine
ML	Machine Learning
MSc	Master of Science
O1–O6	Research objectives 1 to 6
RAM	Random Access Memory
RMSSE	Root Mean Squared Scaled Error
SBA	Syntetos–Boylan Approximation
SKU	Stock Keeping Unit
SME	Small and Medium-sized Enterprise
SNAP	Supplemental Nutrition Assistance Program
UK	United Kingdom
US	United States
WAPE	Weighted Absolute Percentage Error

References

[1] R. Fildes, S. Ma, and S. Kolassa, "Retail forecasting: Research and practice," *International Journal of Forecasting*, vol. 38, no. 4, pp. 1283–1318, 2022.

[2] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*. Melbourne, Australia: OTexts, 3rd ed., 2021.

[3] A. A. Syntetos and J. E. Boylan, "The accuracy of intermittent demand estimates," *International Journal of Forecasting*, vol. 21, no. 2, pp. 303–314, 2005.

[4] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "M5 accuracy competition: Results, findings, and conclusions," *International Journal of Forecasting*, vol. 38, no. 4, pp. 1346–1364, 2022.

[5] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)*, pp. 3146–3154, 2017.

[6] E. Spiliotis, S. Makridakis, A.-A. Semenoglou, and V. Assimakopoulos, "Comparison of statistical and machine learning methods for daily SKU demand forecasting," *Operational Research*, vol. 22, pp. 3037–3061, 2022.

[7] J. Huber and H. Stuckenschmidt, "Daily retail demand forecasting using machine learning with emphasis on calendric special days," *International Journal of Forecasting*, vol. 36, no. 4, pp. 1420–1438, 2020.

[8] F. Badorf and K. Hoberg, "The impact of daily weather on retail sales: An empirical study in brick-and-mortar stores," *Journal of Retailing and Consumer Services*, vol. 52, p. 101921, 2020.

[9] X. Tian, S. Cao, and Y. Song, "The impact of weather on consumer behavior and retail performance: Evidence from a convenience store chain in China," *Journal of Retailing and Consumer Services*, vol. 62, p. 102583, 2021.

[10] D. Bertsimas and N. Kallus, "From predictive to prescriptive analytics," *Management Science*, vol. 66, no. 3, pp. 1025–1044, 2020.

[11] A. N. Elmachtoub and P. Grigas, "Smart "predict, then optimize"," *Management Science*, vol. 68, no. 1, pp. 9–26, 2022.

[12] J. Huber, S. Müller, M. Fleischmann, and H. Stuckenschmidt, "A data-driven newsvendor problem: From data to decision," *European Journal of Operational Research*, vol. 278, no. 3, pp. 904–915, 2019.

[13] Y. Wang, J. Gu, L. Long, X. Li, L. Shen, Z. Fu, X. Zhou, and X. Jiang, "FreshRetailNet-50K: A stockout-annotated censored demand dataset for latent demand recovery and forecasting in fresh retail." arXiv preprint arXiv:2505.16319, 2025.

[14] Association of Convenience Stores, "The Local Shop Report 2025," tech. rep., Association of Convenience Stores, 2025.

[15] S. Kolassa, "Why the "best" point forecast depends on the error or accuracy measure," *International Journal of Forecasting*, vol. 36, no. 1, pp. 208–211, 2020.

[16] J. Azar, "A reproducible, leakage-free pipeline for censored demand forecasting and inventory optimization on FreshRetailNet-50K," *Inventions*, vol. 11, no. 5, p. 89, 2026.

[17] S. G. Ayinaddis, "Artificial intelligence adoption dynamics and knowledge in SMEs and large firms: A systematic review and bibliometric analysis," *Journal of Innovation & Knowledge*, vol. 10, no. 3, p. 100682, 2025.

[18] X. Malik, L. Smith, I. Stewart, and N. Burnett, "Food waste in the UK," Research Briefing CBP-7552, House of Commons Library, Apr. 2024.

[19] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design science in information systems research," *MIS Quarterly*, vol. 28, no. 1, pp. 75–106, 2004.

[20] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A design science research methodology for information systems research," *Journal of Management Information Systems*, vol. 24, no. 3, pp. 45–77, 2007.