Sales Forecasting Systems for Large-Scale Retail Enterprises

Edoardo Santucci

Università Cattolica del Sacro Cuore, Brescia

April 29, 2025

Introduction

The retail sector faces constant demand fluctuations driven by seasonality, promotions, and economic shifts. Even minor forecast errors can lead to stockouts, excess inventory, or missed revenue each with substantial costs.

Accurate **sales predictions** enable retailers to:

- Balance inventory and avoid lost sales;
- Align staffing and logistics with demand;
- Time promotions for maximum impact.

Metro Italia S.P.A.

Metro Italia S.p.A.: one of Italy's foremost B2B wholesalers, supplying a broad assortment of food and non-food products to hospitality professionals, independent retailers, and institutional clients.

With over 50 strategically positioned locations nationwide, Metro Italia operates a dual-channel model:

- **Cash & Carry (C&C)**, for immediate and on-site purchases;
- **Full Service Delivery (FSD)**, for end-to-end logistics and home delivery.

Objective

- **Simultaneously generate robust 30-day sales forecasts for Metro Italy's entire network of Cash & Carry and Full-Service Delivery sites with maximum efficiency.**

Objective

- **Simultaneously generate robust 30-day sales forecasts for Metro Italy's entire network of Cash & Carry and Full-Service Delivery sites with maximum efficiency.**
- Benefits for Metro Italia S.P.A.:
  - Accurately plan inventory levels
  - Allocate staffing and logistics precisely
  - Launch promotional campaigns at the most strategic times

Data Collection & Preprocessing

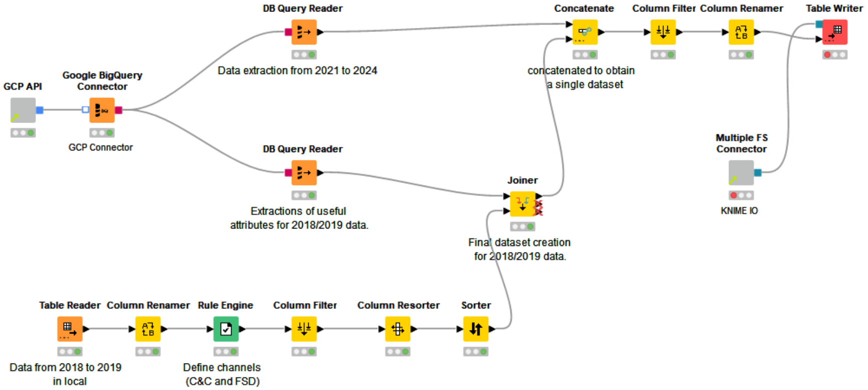


Figure: KNIME workflow extracting and merging C&C and FSD sales data from GCP and local files into a single dataset.

Exploratory Data Analysis

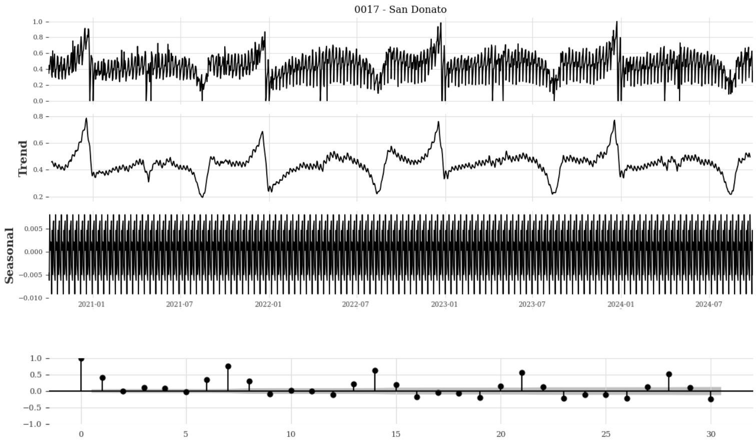


Figure: Time series decomposition of San Donato store data and corresponding Autocorrelation function (ACF) plot.

Modeling Framework

- **Global Approach**
  Global models were chosen to take advantage of shared temporal and cross-sectional patterns across stores, substantially reducing computational costs compared to training individual models per store

Modeling Framework

- **Global Approach**

  Global models were chosen to take advantage of shared temporal and cross-sectional patterns across stores, substantially reducing computational costs compared to training individual models per store

- **Forecasting Models**

  Linear Regression, Random Forest, LightGBM and LSTM

Modeling Framework

- **Global Approach**

  Global models were chosen to take advantage of shared temporal and cross-sectional patterns across stores, substantially reducing computational costs compared to training individual models per store
- **Forecasting Models**

  Linear Regression, Random Forest, LightGBM and LSTM
- **External Regressors**

  Incorporated month of year and day of week features to capture strong seasonal and weekly demand cycles

Modeling Framework

- **Global Approach**
  Global models were chosen to take advantage of shared temporal and cross-sectional patterns across stores, substantially reducing computational costs compared to training individual models per store
- **Forecasting Models**
  Linear Regression, Random Forest, LightGBM and LSTM
- **External Regressors**
  Incorporated month of year and day of week features to capture strong seasonal and weekly demand cycles
- **Efficient Implementation via Darts**
  Darts library provides a unified interface for data preparation, model fitting, and parallelized hyperparameter tuning

Evaluation Strategy

Data Partitioning

- Holdout Train–Test Split
- Rolling-Window Backtesting - *to emulate live forecasting and assess robustness as new data arrive*

Evaluation Strategy

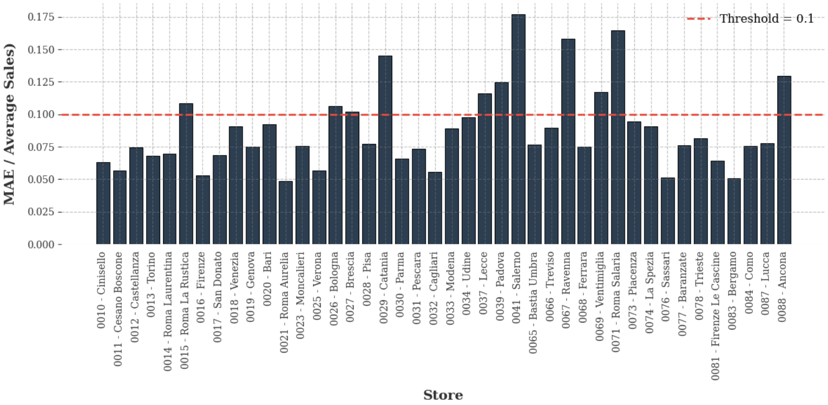
Data Partitioning

- Holdout Train–Test Split
- Rolling-Window Backtesting - *to emulate live forecasting and assess robustness as new data arrive*

Performance Metrics

- Mean Absolute Error (MAE)
- Root Mean Squared Error (RMSE)
- **Relative MAE** - *MAE normalized by each store's average sales, enabling fair cross-store comparison and alignment with business decision thresholds*

C&C Channel: Results Summary



Model	Relative MAE
Lin. Regr.	13.89%
RF	11.06%
LightGBM	11.37%
LSTM	**8.82%**

Figure: Relative MAE per store for the best-performing model in the Cash & Carry channel.

C&C Channel: Results Summary

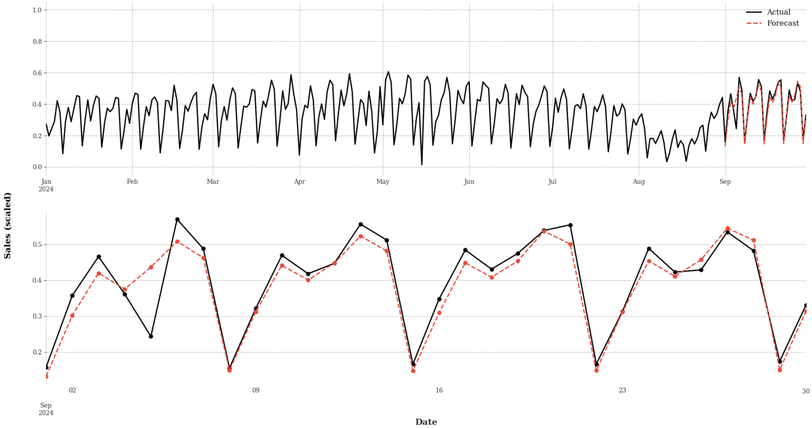
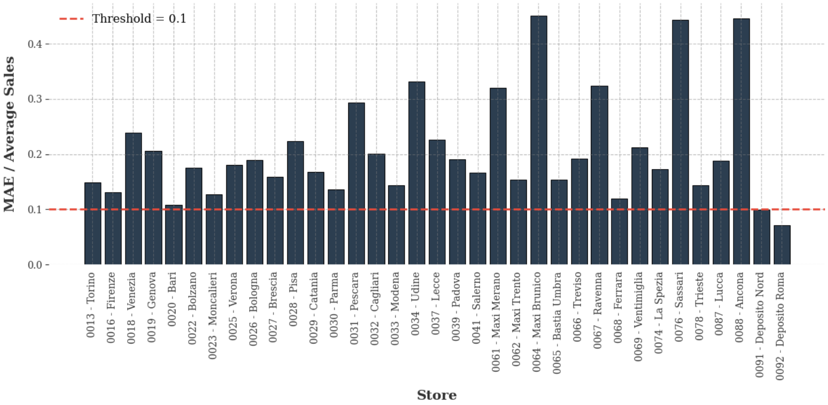


Figure: Actual vs Forecasted daily sales for **San Donato** using the best-performing model (LSTM)



Model	Rel. MAE
Lin. Regr.	23.5%
RF	24.3%
LightGBM	**20.7%**
LSTM	31.0%

Figure: Relative MAE per store for the best-performing model in the Full-Service Delivery channel

FSD Channel: Results Summary

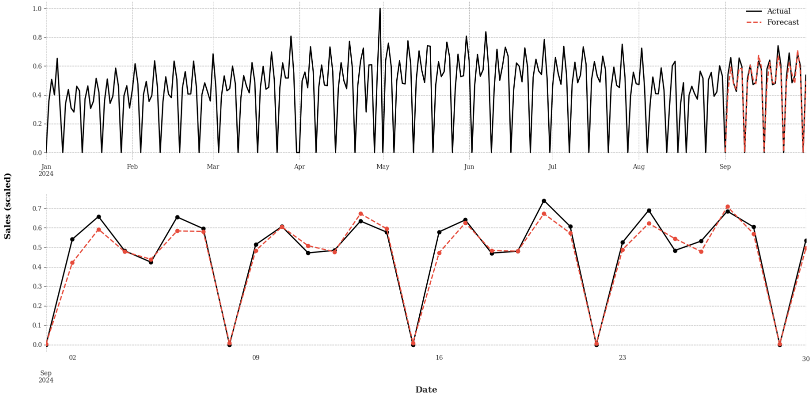


Figure: Actual vs Forecasted daily sales for **Deposito Roma** using the best-performing model (LightGBM)

Conclusions

- **LSTM** *model demonstrates a superior capacity to capture the inherent cyclicality of the Cash & Carry channel*

Conclusions

- **LSTM** *model demonstrates a superior capacity to capture the inherent cyclicality of the Cash & Carry channel*

- **LightGBM** *delivers robust performance when addressing the intricate and rapidly expanding requirements of the Full-Service Delivery channel, a domain that remains particularly challenging to forecast given its relative novelty and ongoing growth*

Conclusions

- **LSTM** *model demonstrates a superior capacity to capture the inherent cyclicality of the Cash & Carry channel*
- **LightGBM** *delivers robust performance when addressing the intricate and rapidly expanding requirements of the Full-Service Delivery channel, a domain that remains particularly challenging to forecast given its relative novelty and ongoing growth*
- *Future work will focus on incorporating exogenous drivers such as promotions, holidays, and weather, and on deploying continuous retraining pipelines to ensure real-time adaptability*

Thank you for your attention!