September 2024

Data Science Lab: Ristoranti

Bentley, Caliaro, Rabie, Simeoni

ABSTRACT The ability to accurately predict a restaurant's annual revenue is an enormously valuable strategic tool. This project focuses on analyzing the time series of five restaurants in order to develop accurate predictive models that can provide reliable estimates of future revenue. The analysis includes a process of data cleaning, exploratory analysis, information enrichment, and advanced modeling, with the goal of refining and improving the accuracy of the predictions.

Table of contents

INTRODUCTION

The main objective of the project is to apply and compare several advanced time series forecasting models, including SARIMAX, Prophet, TBATS, and Holt-Winters, to accurately forecast restaurant invoices. The importance of obtaining accurate forecasts cannot be underestimated, as this information is critical for restaurant managers.

Forecasting annual invoices can be extremely beneficial to a restaurant for several reasons. Knowing the projected invoices for the coming year helps the restaurant to plan its finances, establishing budgets for operating expenses, investments, and tax payments; it allows optimal inventory management, reducing waste and shortages of food and beverages; it allows the restaurant to organize its staff efficiently, hiring more during busy periods and reducing them during off-peak periods; and finally, it helps to plan marketing and promotions, offering discounts or promotions during off-peak periods to attract more customers.

Each model selected for this project was chosen for its ability to capture specific time series characteristics, such as seasonality, trend, and the presence of exogenous regressors that may affect invoices. For example, SARIMAX is particularly useful for including seasonal effects and external variables, while Prophet is known for its flexibility in adapting to different frequencies.

DATASET

The “restaurants” dataset contains daily data of invoices and number of receipts from 5 restaurants, from 01/01/2018 to 05/31/2024. To this data were added 2 “weather” datasets, derived from data provided by ARPA (Regional Agency for Environmental Protection), which include daily average rainfall and temperatures for the cities of Piacenza and Pavia. Also included were the “SerieA” and “Champions” datasets provided by https://www.football-data.co.uk/, containing the number of daily Serie A and Champions League matches. These additional datasets were used as external regressors to better rare the prediction

METHODOLOGICAL ASPECTS

SARIMAX

The SARIMAX (Seasonal Auto Regressive Integrated Moving Average with eXogenous regressors) model is an extension of the ARIMA model that includes seasonality and exogenous regressors. An ARIMA model is a time series model that uses the autoregressive integrated moving average (ARIMA) to make forecasts. A SARIMAX model has three main components: the autoregressive (AR) component, the integration (I) component, and the moving average (MA) component.

The AR component models the correlation between an observation and a number of previous observations, capturing the effect that past observations have on the current one. The order of the AR component, denoted by the p parameter, represents the number of past observations used to predict the current observation.

The I component differentiates the data to make the time series stationary. Stationarity is an important property since many prediction models assume it. The order of differentiation, denoted by the parameter d, represents the number of times the data are differentiated.

The MA component models the forecast error as a linear combination of past forecast errors, trying to capture the effect these have on the current forecast error. The order of the MA component, denoted by the parameter q, represents the number of past forecast errors used to predict the current error.

The SARIMAX model extends ARIMA by including a seasonal component and exogenous regressors. The seasonal component models the correlation between an observation and observations at regular intervals (e.g., every year or month), while the exogenous regressors include external variables, such as weather conditions or holidays, that may influence forecasts. A SARIMAX model has seven parameters: p, d, q, P, D, Q and m, where the first three refer to the ARIMA component and the other three to the seasonal component. The m parameter represents the seasonality period (e.g., 12 for monthly data, 4 for quarterly).

Prophet

The Prophet model, developed by Facebook, is a time series model for time series forecasting with strong seasonal components. It is an open-source library for univariate time series forecasting that implements an additive model with trends, seasonality, and holidays. Prophet can be considered a nonlinear regression model, consisting of four main components: a linear or logistic trend curve, a Fourier-modeled annual seasonal component, a weekly seasonal component, and a component for holidays.

Prophet automatically detects changes in the trend from the data. The annual and weekly seasonal components model annual and weekly seasonal changes, respectively, while the holiday component captures the effects of holidays on the time series. Prophet is designed to be easy to use, automating many model parameters and enabling accurate predictions without manual intervention.

TBATS

The TBATS model is an advanced forecasting technique for particularly complex time series. The acronym TBATS stands for “Trigonometric, Box-Cox, ARMA Errors, Trend and Seasonal components,” reflecting its main features.

The trigonometric component is used to model non-regular seasonality by capturing complex cycles with trigonometric functions.The Box-Cox transformation stabilizes the variance, which is useful in the presence of heteroschedasticity, where the variability of the data changes over time.

The ARMA component models forecast errors as a combination of past errors, improving the accuracy of forecasts.The trend component can be linear or nonlinear and captures changes in the direction of the time series, adjusting for ascending or descending trends.The seasonal component is flexible and handles multiple periodicities, capturing seasonality that repeats over different intervals (e.g., annual and weekly).

Holt-Winters

The Holt-Winters model, also known as the exponential damping method with trend and seasonality, is an established technique for forecasting time series with trend and seasonality. There are two main variants: additive and multiplicative.

In the additive model, seasonality is a variation that adds to the level and trend of the series, fit when the seasonal fluctuations have constant amplitude.In the multiplicative model, seasonality is a variation that multiplies the level and trend, fit when the amplitude of the fluctuations changes in proportion to the level of the time series.

Holt-Winters includes three main components: level, which represents the average value of the series; trend, which captures the general direction (increasing or decreasing); and seasonality, which models regular periodic variations.

Autocorrelation Function (ACF)

Autocorrelation Function (ACF) measures the degree to which values in a time series are correlated with its past values (lag). Specifically, ACF tells us how well a value in a series is correlated with previous values in that same series, over various time intervals (lag). The ACF helps to identify the order of the autoregressive element in an ARIMA model and to determine seasonality in the data.

Interpretation of the ACF:A positive ACF value for a certain lag indicates a positive correlation between the values in the time series at that time interval.A negative value indicates a negative correlation.ACF values close to zero suggest the absence of correlation for that lag.The ACF is particularly useful for identifying repeated patterns (such as seasonal cycles) and for assessing whether the time series is stationary or exhibits trends.

Partial Autocorrelation Function (PACF)

Partial Autocorrelation Function (PACF) measures the correlation between a value in the time series and its past values, eliminating the effect of intermediate values. In other words, PACF considers the direct correlation between the current value and its past lags, ignoring the effect of other nearest lags.

Interpretation of PACF:If the PACF shows significant correlations up to a certain lag but not beyond, it indicates that only a limited number of lags directly affect the present values.This is useful for determining autoregressive (AR) order in an ARIMA model, since the PACF truncates at the optimal autoregressive lag. Use of ACF and PACF in ARIMA modelingACF and PACF are crucial tools for determining the “p” (autoregressive) and “q” (moving average) parameters of an ARIMA model:The ACF can indicate the length of the moving average (MA) component of an ARIMA model, due to the decay of the autocorrelations at given lags.The PACF, on the other hand, is useful for determining the autoregressive (AR) component, by identifying the point at which the correlations truncate

AIC

The Akaike Information Criterion (AIC) is a metric used to compare statistical models, such as SARIMAX models, and evaluate which one fits the data best. The AIC balances the quality of a model's fit to the data with its complexity.

$$AIC = 2k - 2\ln(L)$$

k: The number of parameters in the model (more parameters, greater complexity).

L: The maximum likelihood (log-likelihood) of the model to the data (how well the model fits the data).

Interpretation: Models with a lower AIC are preferred. A low AIC means the model fits the data well with minimal complexity. AIC balances the trade-off between overfitting (model too complex) and underfitting (model too simple). A model with many parameters might fit the data better, but risks losing generalization.Why choose parameters that reduce AIC: A lower AIC indicates that the model is better able to capture patterns in the data without becoming overly complex. Choosing parameters that minimize AIC helps you get a model that is accurate and general, reducing the chance of overfitting the training data (overfitting) and maintaining a good level of accuracy on new predictions.In short, reducing AIC helps you choose a model that is overall better at balancing accuracy and complexity.

MAPE & RMSE

MAPE is a metric used to measure the accuracy of a forecasting model. It represents the average of the absolute percentage differences between the predicted values and the actual values. MAPE is expressed as a percentage, making it easy to interpret. A lower MAPE indicates better model performance.

RMSE is another commonly used metric to evaluate the accuracy of a model. It calculates the square root of the average of the squared differences between predicted and actual values. RMSE gives more weight to larger errors due to the squaring, making it sensitive to outliers. Like MAPE, a lower RMSE means better accuracy.

PRE-PROCESSING

Several datasets were worked on, including those covering weather data from the provinces of Piacenza and Pavia, restaurant invoicess located in these areas, and results of sporting events such as Serie A and Champions League matches.

The main objective was to merge and clean up these data to facilitate later analysis.The process began with importing the necessary libraries, including pandas for data management, matplotlib and seaborn for visualization, and os for file management.

For the weather data, monthly CSV files for the provinces of Piacenza and Pavia were loaded and combined into two separate DataFrames. Next, the most relevant columns were selected, such as date, temperature, and weather. Additional columns were added to quantify negative weather conditions, such as rain and wind.Weather data were combined with restaurant data, and missing values were handled by filling them with zero for invoicess or using average values for numeric variables.

Information on sporting events, such as results of Serie A and Champions League matches, was also added. Special attention was paid to filling in missing data during the Covid-19 period, using a multiple linear regression model to estimate missing values based on data available before and after the problem period.Next, monthly time series were created for each restaurant by aggregating daily data and calculating sums and averages.

This step was crucial to analyze the monthly trends and remove any unnecessary rows. Finally, pre-processing was completed by cleaning the specific data. Daily invoices data for each restaurant were extracted and saved, and a column was added for the day of the week.

Graphs and boxplots were created to visualize pre and post Covid changes in weekly invoices, and data were analyzed to identify any changes in weekly invoices cycles. All processed data were exported to separate CSV files for each restaurant and period, providing a solid basis for further analysis and reporting.

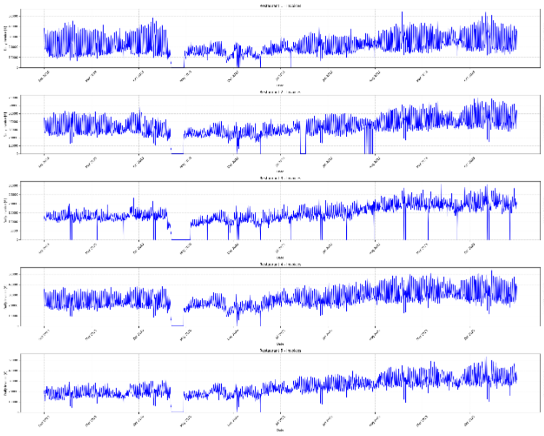
Multiple linear regression

The main objective is to ensure data continuity through the use of multiple linear regression over the period for restaurant transactions between March 1, 2020, and May 6, 2020, the period most plagued by the COVID problem.Initially, a series of preliminary operations are carried out to prepare the data.

Specifically, the “date” column is converted to datetime format, and missing values in the numeric columns, such as “TMEDIA °C,” “TMIN °C,” “TMAX °C,” and “atm cond,” are imputed using the average of the values. In addition, the categorical column “location” is completed using the most frequent value.

Data outside the COVID range are used to train a multiple linear regression model. Once the model is trained on the out-of-period data, it is used to predict the “total” values for the COVID period, thus replacing missing or outlier values with the estimates obtained through the regression.

Below we see the two graphs made: one represents the data before the correction and the other shows the data after the regression, allowing a clear and immediate comparison.



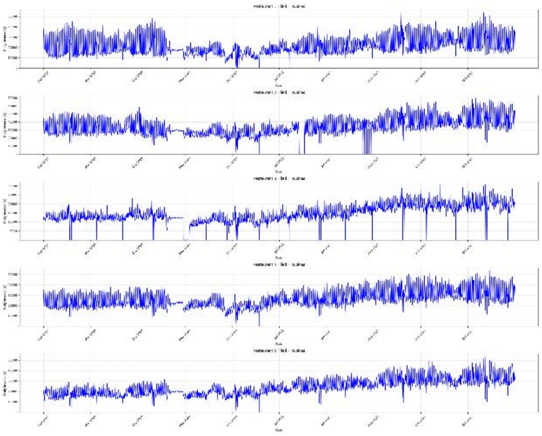


Figure 1: Plotting the original Daily Invoice for each restaurant

Figure 2: Plotting the Daily Invoice for each filled restaurants

DATA EXPLORATION

When exploring the data, it is essential to understand the variations and trends in the daily bill totals for each restaurant, broken down by day of the week. For this purpose, the distribution of daily bills for each day of the week was analysed using box plots.

The box plots provide an effective visual representation of the dispersion and trends in the data, highlighting medians, quartiles and outliers.

Variation in Daily Invoices: This analyses how the total daily invoices vary by day of the week for each restaurant. This helps identify the days with the highest or lowest expenses, providing an overview of weekly trends.

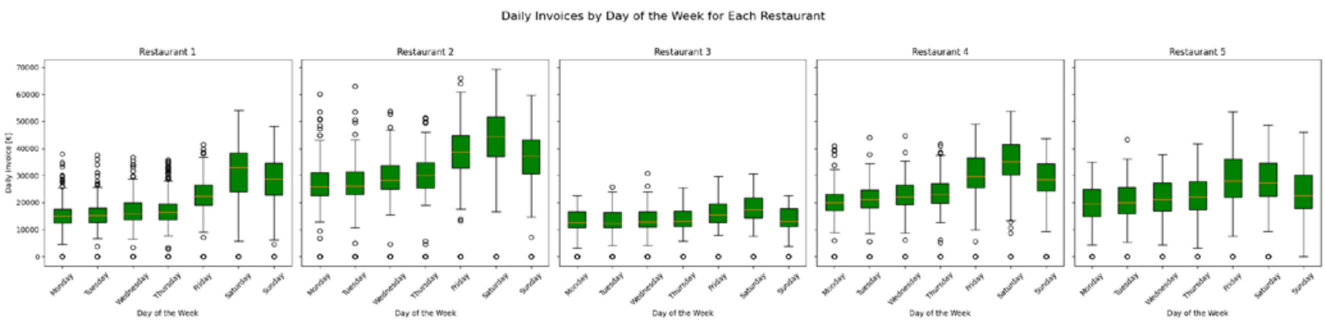


Figure 3: Plotting the weekly Daily invoice trend for each restaurant

It can be seen that, for all five restaurants, Saturday and Sunday bills are significantly higher than on the other days of the week.

Shown below is a scatterplot exploring the relationship between the number of daily batches and the average total of receipts for different restaurants. It is crucial to understand how specific variables influence total bills in restaurants. The graph displays the average total of bills (y-axis) as a function of the number of daily batches (x-axis), allowing trends or correlations to be observed.

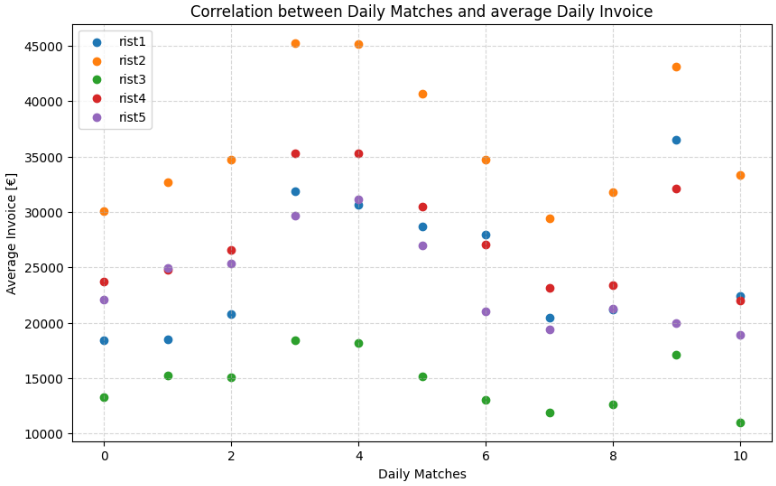
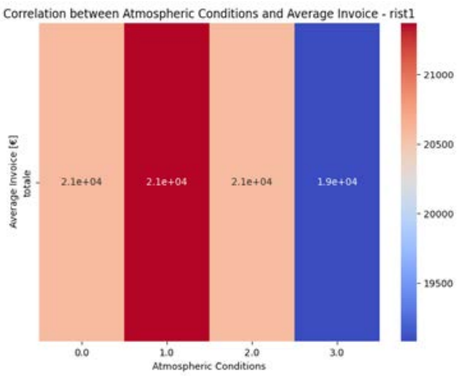


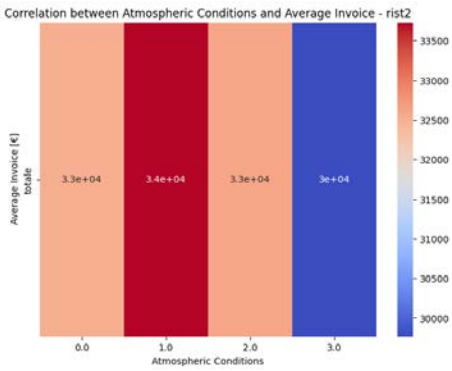
Figure 4: Scatterplot: Seeking for a correlation between Daily matches and average Daily Invoice

In order to explore the impact of weather conditions on restaurant sales, a heatmap was created to visualise the correlation between weather conditions and the average invoice total.

The analysis is based on the observation that the worse the weather, the lower the invoice total tends to be. The heatmap below represents the average invoice data in relation to weather conditions, where a colour scale is used to highlight the intensity of the variations.

This graph makes it possible to visually identify how changes in weather conditions affect sales and provides a clear indication of trends and correlations.





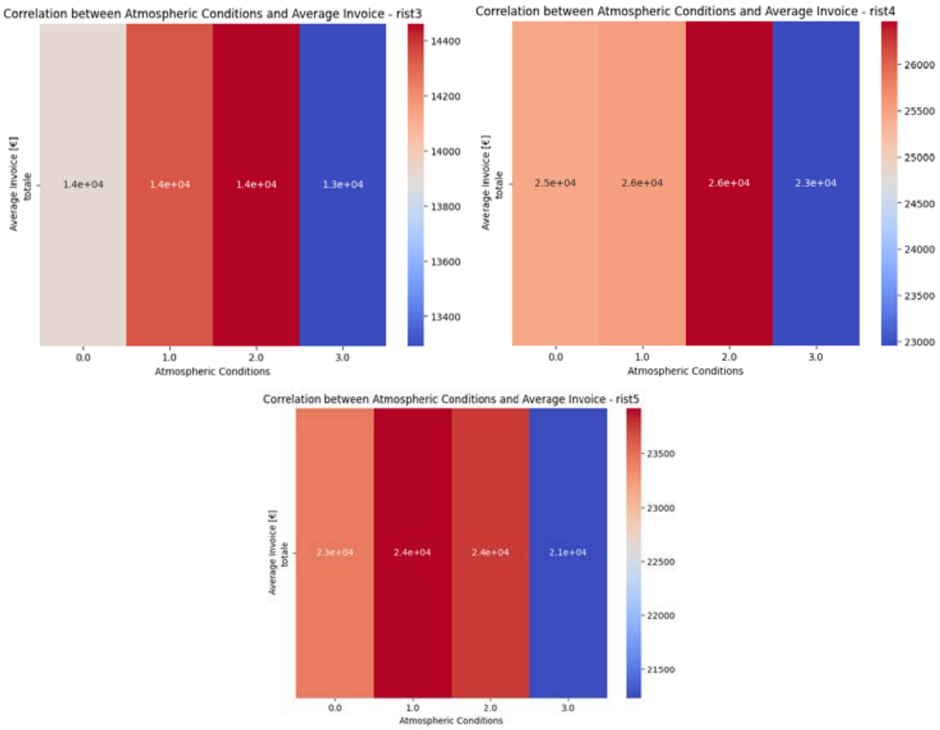


Figure 5: Heatmap: Seeking for a correlation between Atmosferic condition and average Invoice. I recall that the greater is 'atm_cond' the worst are the atmosferic condition. And it is clear from the graph below that the worst atmosphericcondition low the Invoice!

ANALYSIS

Analysis and Decomposition of Weekly and Daily Time Series

In order to better understand the characteristics of the weekly restaurant turnover time series, a decomposition into main components was performed: trend, seasonality and residuals. This process helps to identify and analyse the various components influencing the data, allowing a deeper understanding of the dynamics and patterns present.

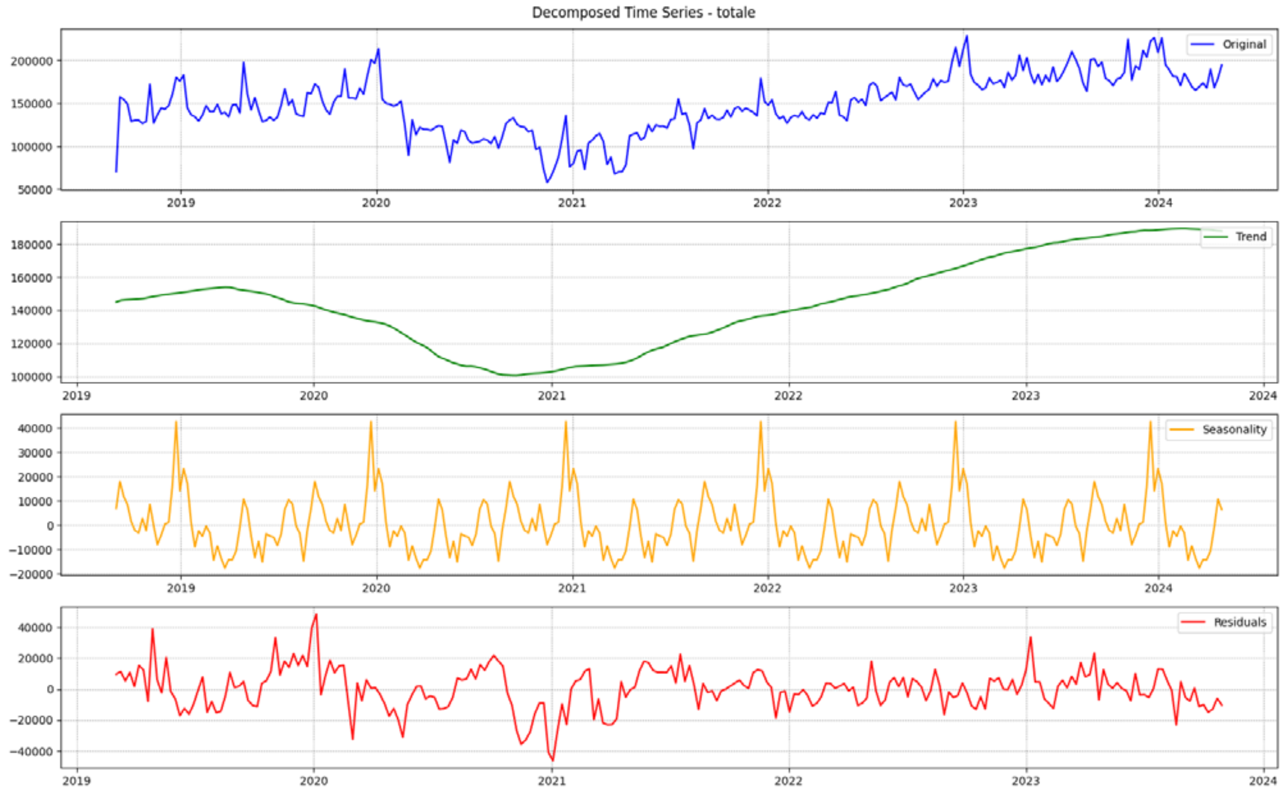


Figure 6: Decomposed time series – Weekly analysis

Original Series: The aggregated weekly time series shows significant variability, with clear peaks in 2019 and a decline in 2020. After 2021, steady growth is observed, with some fluctuations in values. This represents the overall trend of the data over time, including seasonal events and random variations.

Trend: The long-term trend shows a decline between 2019 and 2021, possibly due to an external event (such as a pandemic). After this period, the trend returns to stable growth, indicating a steady recovery until 2024. Removing seasonality, the general trend of the data can be observed.

Seasonality: Seasonal patterns show regular fluctuations on an annual basis, with repetitive peaks suggesting cyclical behaviour. These peaks could be related to periods of high demand or special events, such as holidays or tourist seasons. Seasonality is a recurring component that influences the data in a predictable manner.

Residuals: Residuals represent variations not explained by trend and seasonality. There are significant peaks in 2019 and 2020, but overall the residuals stabilise over time. This indicates that the decomposition model explains most of the variations well, leaving only a few unpredictable anomalies.

Figure 7: Decomposed time series - Daily Analysis

Original Series: The daily series shows strong variability with regular peaks and rapid fluctuations. The daily fluctuations are much more noticeable than in the weekly version. These rapid movements could be due to daily influences or short-lived events. In general, there is considerable volatility in the data.

Trend: The long-term trend remains similar to the weekly trend, with a noticeable decline between 2019 and 2021 followed by a recovery from 2022. However, more detailed and less linear variations can be seen in the daily trend. This suggests gradual changes in the phenomenon, highlighting small long-term fluctuations. Continued growth until 2024 is clear.

Seasonality: The daily seasonality shows regular and more frequent patterns than the weekly series. Repetitive cycles are observed almost daily, probably linked to weekly events or specific days. Seasonal fluctuations indicate the presence of strong short-term cyclic effects. The peaks repeat with regularity, showing a well-defined periodic behaviour.

Residuals: Residuals are much more volatile in daily data, with sudden peaks indicating anomalous events or random influences not captured by the model. This suggests the presence of unpredictable variables or daily noise. The high variability in the residuals makes it difficult to explain all fluctuations, indicating irregular events influencing the daily data.

For a computational issue we will use the weekly data: through satisficing tests we will determine the stationarity of the time series

ADF & KPSS test

To ensure that the forecasting models used are appropriate, it is essential to verify the stationarity of the time series. Stationarity is a necessary condition for many forecasting models, as these models assume that the statistical properties of the series remain constant over time. Two common tests to assess stationarity are the Augmented Dickey-Fuller Test (ADF) and the Kwiatkowski-Phillips-Schmidt-Shin Test (KPSS). Both tests are used to assess whether the time series has a unit root or is stationary.

- Augmented Dickey-Fuller (ADF) test:

Since the ADF statistic is -1.636 and the p-value is 0.464, we cannot reject the null hypothesis ($N0$), which states that there is a unit root in the time series (the series is not stationary). The critical value at the 5% level is -2.871, and since the value of the ADF statistic is not smaller than the critical value, it confirms that we cannot reject the null hypothesis.

Conclusion: According to the ADF test, the series is non-stationary (has a unit root).

- KPSS test (Kwiatkowski-Phillips-Schmidt-Shin): The KPSS statistic is 1.250766, which is much higher than the 5% critical value of 0.463. The p-value is less than 0.01, which indicates that we can reject the null hypothesis ($N0$), which states that the series is stationary.

Conclusion: According to the KPSS test, the series is not stationary. The fact that the value of the statistic is so high indicates that the series might be non-stationary with a trend component.

Overall conclusion: ADF test: We cannot reject the hypothesis that the series has a unit root, so the series is non-stationary. KPSS test: The KPSS test indicates to us that the series is non-stationary, confirming the evidence of a trend.

ACF & PACF

To better understand the structure and temporal dependence of the data, it is useful to examine the autocorrelation (ACF) and partial autocorrelation (PACF) functions of the time series. These tools help identify the presence of seasonal patterns and determine the optimal order for models such as ARIMA and SARIMA.

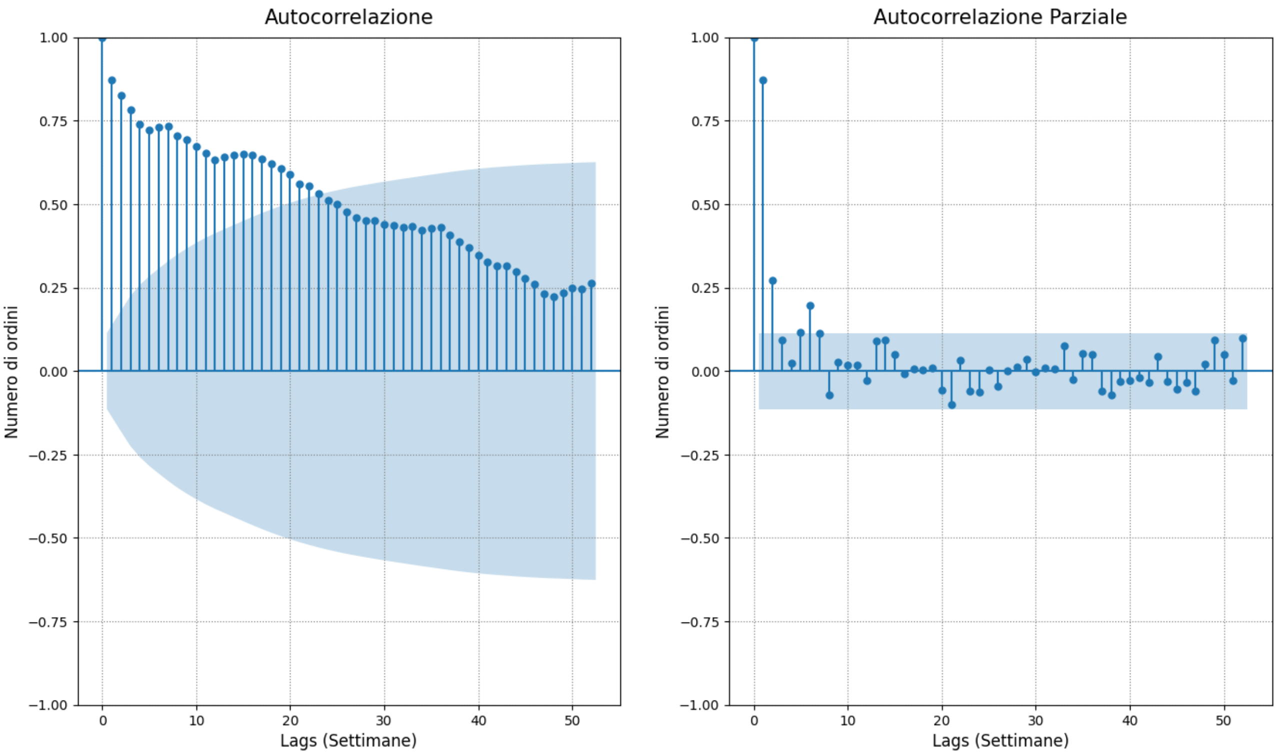


Figure 8: ACF & PACF

ACF (Autocorrelation) As we already knew, our series exhibits seasonality, and the plot of ACF confirms this pattern. As we increase the lags, we observe that the significance of the lags gradually decreases, which is consistent with the presence of long-term seasonality.

The first significant lag is lag 1, which is not surprising. This indicates that the number of orders in the previous week can directly influence the current week's orders. This suggests a strong short-term dependence. Another significant lag is observed around lag 7 and lag 14, which could indicate a weekly pattern, being that 7 represents a whole week. This suggests that one week's orders influence the following week's orders, confirming a seasonal weekly cycle. In general, the ACF does not show a sharp “cut-off” but decreases slowly, suggesting the presence of an autoregressive component and the need to differentiate the data to make them stationary.

PACF (Partial Autocorrelation) In the PACF plot, we observe that lag 1 has the most significant peak, followed by a smaller peak at lag 2. This means that the orders of the immediately preceding week have the strongest direct correlation with those of the current week.

Lag 7 in the PACF shows another significant peak, which suggests that the weekly patterns are directly correlated, which supports the idea of weekly seasonality. The other lags do not show particularly significant peaks, suggesting that autoregression is only important in the short term and the more distant lags do not have much direct impact on the series. Conclusions The ACF plot shows a clear weekly seasonality, with significant correlations at lags 1, 7 and 14, while the PACF confirms that autocorrelation is strong in the short term, with a direct impact on the closest lags, especially lag 1.

Lag plots

are useful tools for visualizing the relationship between time series values and their lags. In this case, the lag_plots function explores how current values in the total series are correlated with values at specific lags (1, 7, and 14 days). This analysis can reveal recurrent patterns and help identify seasonality and periodicity in the data

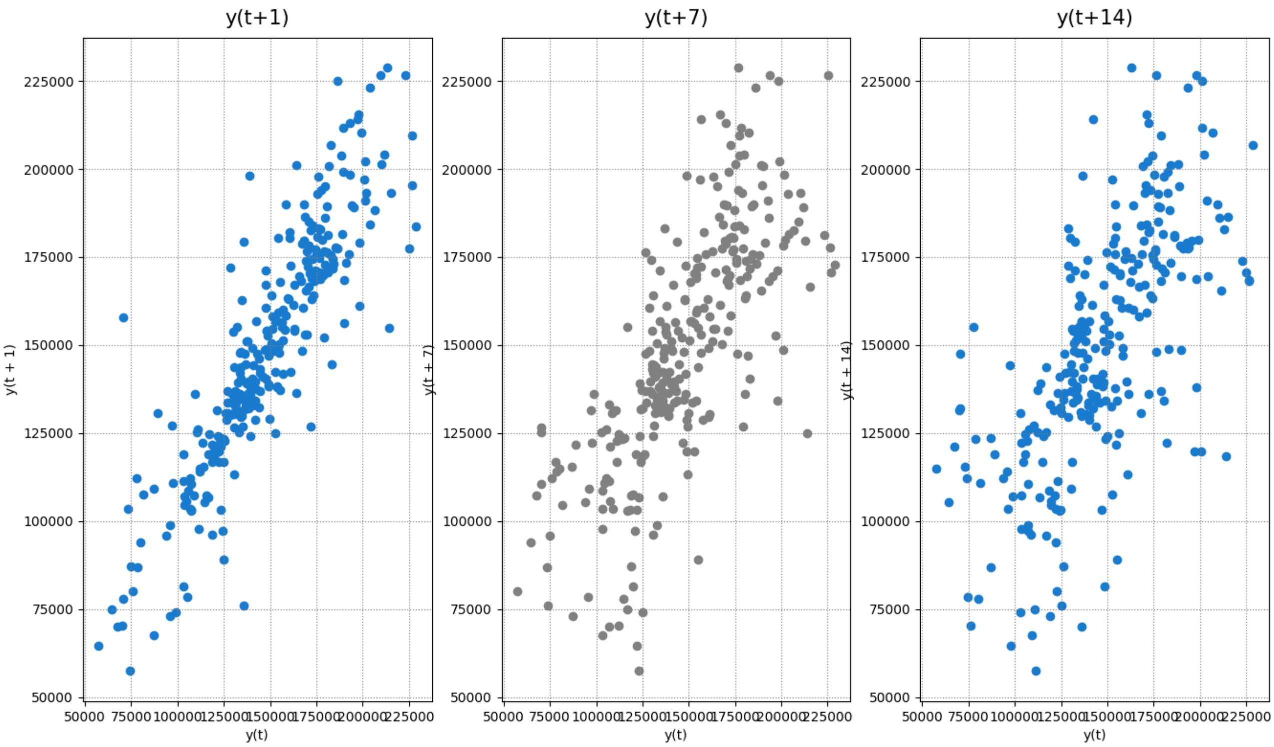


Figure 9: three scatter plots to visualize the relationship between the time series and its lags (lags) at specific intervals. The plots show the correlation between the time series y(t) values and its lags values of 1, 7 and 14 days

Remarks: Lag 1: The lag 1 plot shows a clear positive linear correlation between

$y(t)$ *y(t)* and $y(t+1)$ *y(t + 1).*

This confirms that the previous week's values have a strong impact on the current week's values, as observed in the ACF graph. The linear relationship indicates that there is a strong short-term dependence between consecutive values in the time series.

Lag 7: The lag 7 graph shows a more diffuse but still positive correlation. This is consistent with the idea of weekly seasonality: values from the same previous week (7 weeks earlier) influence current values,

but with a less rigid correlation than in lag 1. The point spread indicates that there are other variables or noises that influence this relationship.

Lag 14: A positive correlation is also observed in the plot of lag 14, similar to that of lag 7, but with a slightly larger scatter. This confirms that there is a biweekly correlation, although less strong than that of lag 1. The larger dispersion than in lag 1 indicates that the effect of the biweekly cycle is present, but weaker and less predictable.

Conclusion: These graphs confirm the patterns observed in the ACF, with a strong short-term linear correlation (lag 1) and clear weekly seasonality (lag 7), as well as a weaker correlation but present over a two-week cycle (lag 14). These results provide us with useful indications for including autoregressive (AR) components in future models, particularly for lag 1 and weekly seasonality.

SARIMAX model with parametrs optimization

We implemented a SARIMAX model to predict weekly restaurant revenues using a rolling window for model training. This approach allows the model to be continuously updated with the latest data and improve the accuracy of the predictions. Details of the code and results obtained are described below.

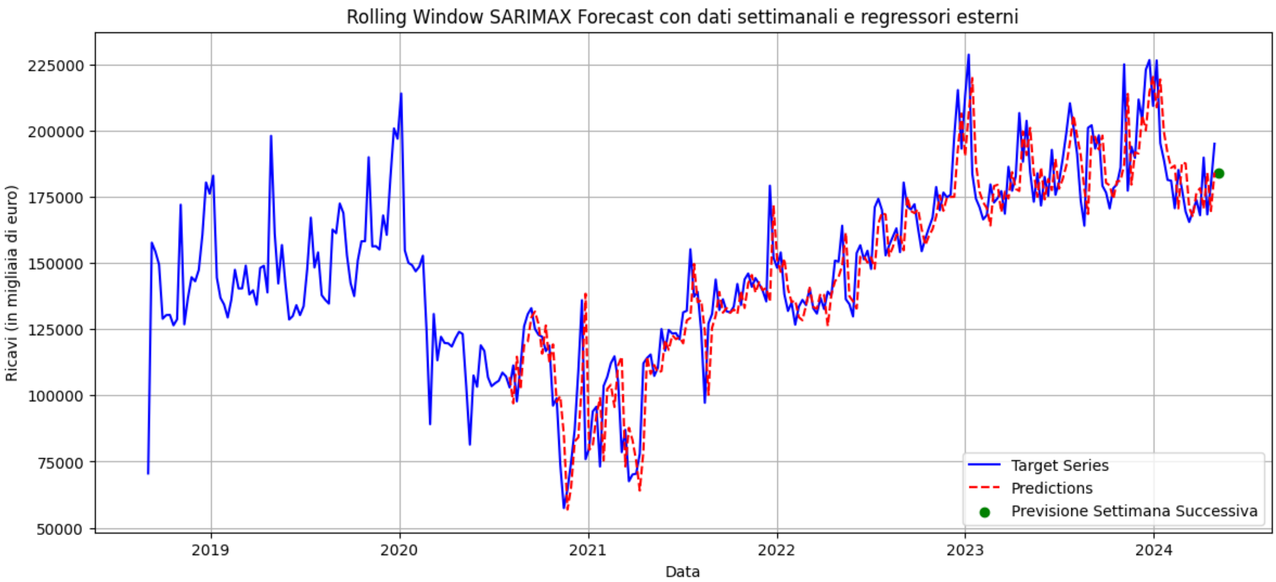


Figure 10: Rolling window SARIMAX with Weekly data and external regressors

Pipeline sarimax application to all restaurants

File	RMSE	MAPE (%)
Ristorante 1	14,443.09	7.92
Ristorante 2	26,786.89	6.62
Ristorante 3	11,509.51	7.98
Ristorante 4	14,197.53	6.18
Ristorante 5	15,095.26	6.45
Media	**16,406.46**	**7.03**

Figure 11: Results

TBATS

The TBATS (Trigonometric Box-Cox ARMA Trend Seasonal) model was used for forecasting the weekly time series. This model is particularly effective for handling complex seasonality and nonlinear variations in the data.

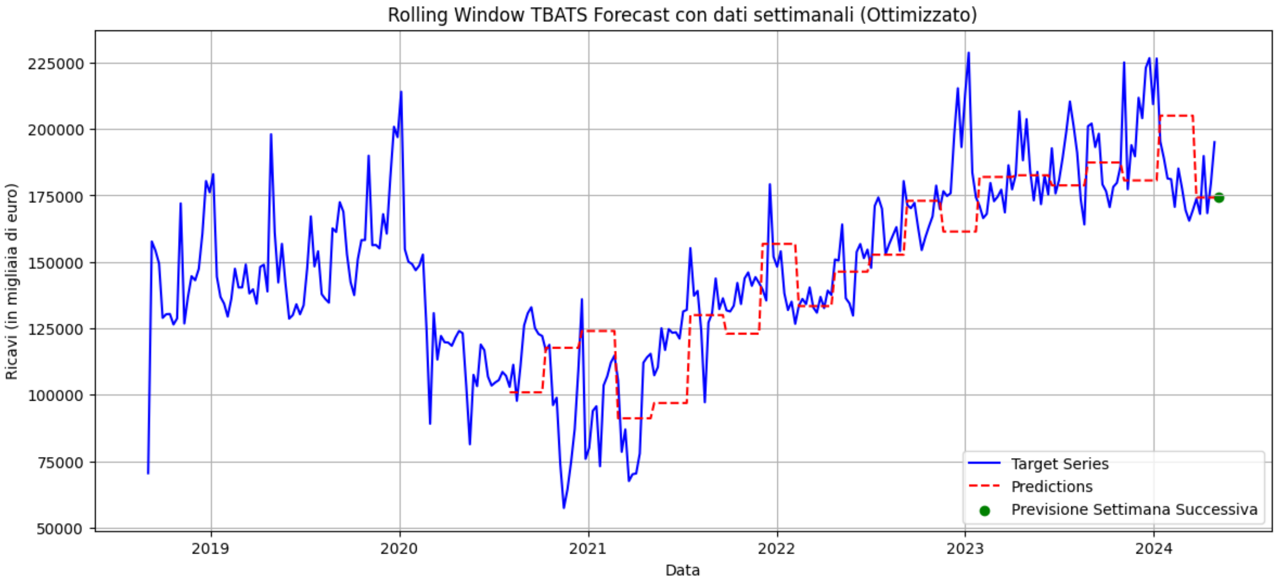


Figure 12: Rolling window forecats with weekly data

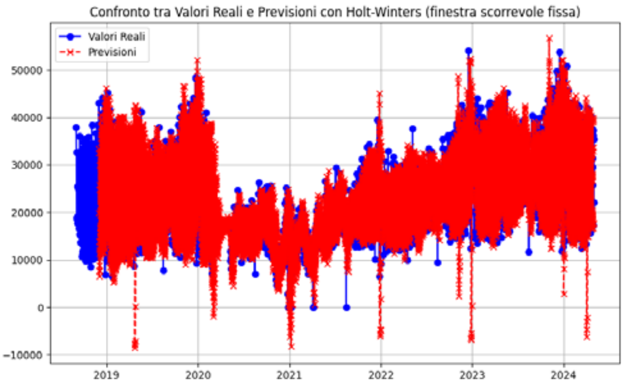
File	RMSE	MAPE (%)
Ristorante 1	21,186.91	12.55%
Ristorante 2	39,697.79	9.36%
Ristorante 3	15,137.35	11.72%
Ristorante 4	19,424.15	9.16%
Ristorante 5	19,399.86	9.08%
Media	**22,969.21**	**10.37%**

Figure 13: Results

The TBATS model provided reliable forecasts with a good ability to adjust for weekly seasonality. However, the SARIMAX model demonstrated a slight superiority in accuracy over TBATS

- Holt-Winters

The Holt-Winters model was used to make forecasts on time series with seasonality and trend. In this analysis, a fixed window (rolling window) was applied to train the model and generate forecasts over future periods.



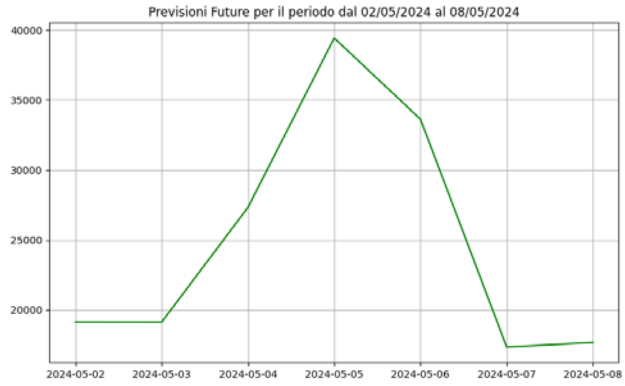


Figure 14: comparison between actual values and forecasts with Holt-Winters

Figure 15: future forecasts between 02/05/2024 and 08/05/2024

File	RMSE	MAPE (%)
Ristorante 1	4,897.41	17.61%
Ristorante 2	6,123.97	12.30%
Ristorante 3	3,321.11	23.15%
Ristorante 4	4,081.31	11.84%
Ristorante 5	3,899.50	12.53%
Media	**4,464.66**	**15.49%**

Figure 16: Results

the Holt-Winters model showed good adaptability to seasonal data and provided useful predictions for future periods. However, the SARIMAX model showed slightly better performance than Holt-Winters.

Prophet

The Prophet model was used to make forecasts on time series, applying a rolling window to improve the accuracy of forecasts on temporal data. This approach allows the model to be updated dynamically with the latest data, optimizing forecasts.

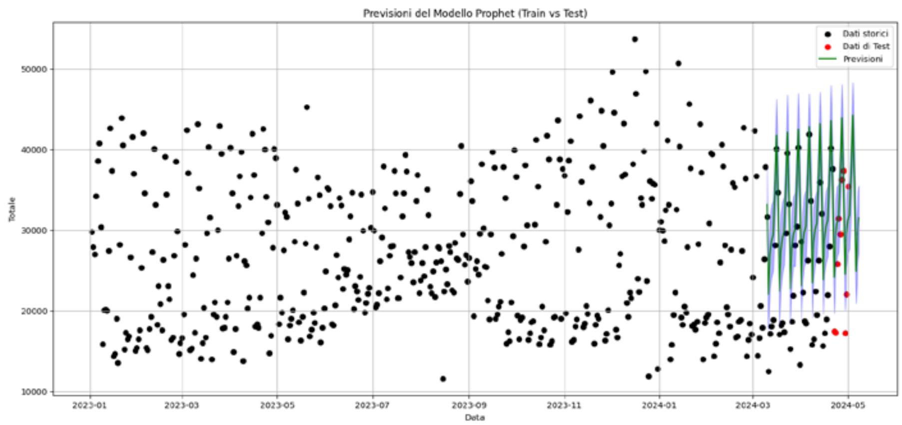


Figure 17: Forecasts of Prophet Model (Train vs Test)

The results were visualized through the graph showing the comparison between model predictions and actual values. This graph highlights the performance of the model over time and the accuracy of the predictions made.

File	RMSE	MAPE
Ristorante 1	6596.95	18.23%
Ristorante 2	6882.61	10.77%
Ristorante 3	3407.45	11.44%
Ristorante 4	6190.55	15.89%
Ristorante 5	5192.42	12.87%
Media	**5654.0**	**13.84%**

Figure 18: Results

Predictions of the Prophet model were compared with actual values to determine the accuracy of the model. The metrics shown show the model's ability to fit and predict the data, with an average RMSE of 4900 and an average MAPE of 6.48%.

Analysis consideration

The following analyses were done:

Creation of 2 models with weekly pooled data (for computational reasons) the SARIMAX and TBATS where the former uses external regressors and the latter does not and both use the expandable rolling window method as a cross validation model

Creation of 2 other models with daily data, the PROPHET and HOLT_WINTERS where again the former uses external regressors and the latter does not. Fixed rolling window is used as the cross validation method.

The purpose is meanwhile to see which model performs better by seeing RMSE and MAPE, which is the best computationally, and to see if the impact of the regressors is significant.

Cross-validation is an essential technique for evaluating the performance of a predictive model and ensuring that the model generalizes well to new data. In particular, for temporal data, specialized cross-validation methods that respect the sequentiality of the data are used. Two common approaches are the Fixed Window and the Expandable Window.

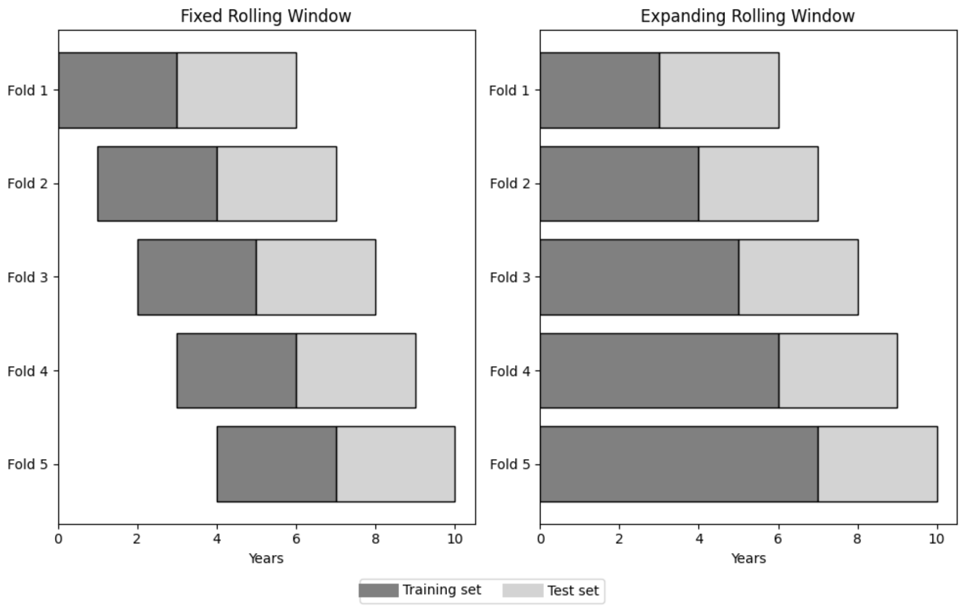


Figure 19: Fixed rolling window vs expanding rolling window

Fixed Rolling Window (Fixed Rolling Window)

Description In this method, the time window on which you train the model has a fixed size. As you move through the data, this window is updated by removing older observations and adding newer ones, thus keeping the number of observations used for training constant.

Starting Pattern:

Use the first n data points to train the model. Predict the value for the n+1 data. Window Moves:

Discard the oldest datum (1st datum) and add the next datum, forming a new window. Predict the value for datum n+2. This process continues until you have covered all the data.

Expanding Rolling Window (Expanding Rolling Window).

Description Unlike the fixed window method, here the window gradually expands by adding new observations without deleting old ones. Thus, the size of the window grows over time, including more and more data as you move forward.

Starting Pattern:

Use the first n data points to train the model. Predict the value for the n+1 data. Window expands:

Keep existing data and add new data. Predict the value for data n+2. This process continues, gradually increasing the window.

RESULTS

Prophet vs Holt-Winter

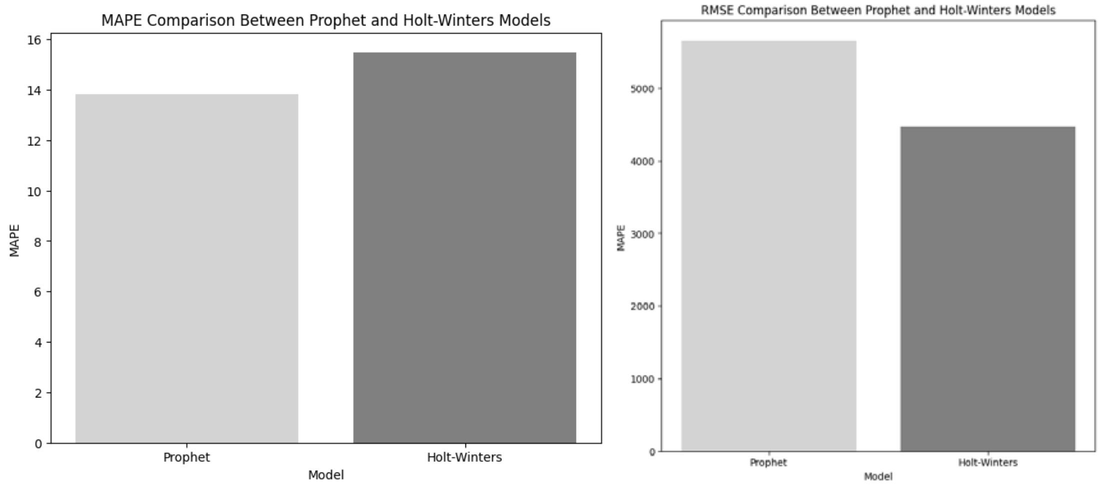


Figure 20: MAPE & RMSE comparison between Prophet & Holt-Winters

The MAPE of the Prophet model is lower than that of Holt-Winters. This means that, in percentage terms, the Prophet model makes fewer relative errors than Holt-Winters. In other words, Prophet tends to better predict the behavior of the time series in relation to actual values.

The RMSE of the Prophet model is higher than that of Holt-Winters. The RMSE measures errors in terms of absolute differences, so this suggests that Prophet, while performing well in relative terms (MAPE), has larger absolute errors, which could be due to outliers or less accurate predictions over large values.

In general, it can be said that by managing outliers better Prophet definitely gives better results, also in computational terms since it is much faster. It seems, therefore, that the effect of outlier regressors is positive

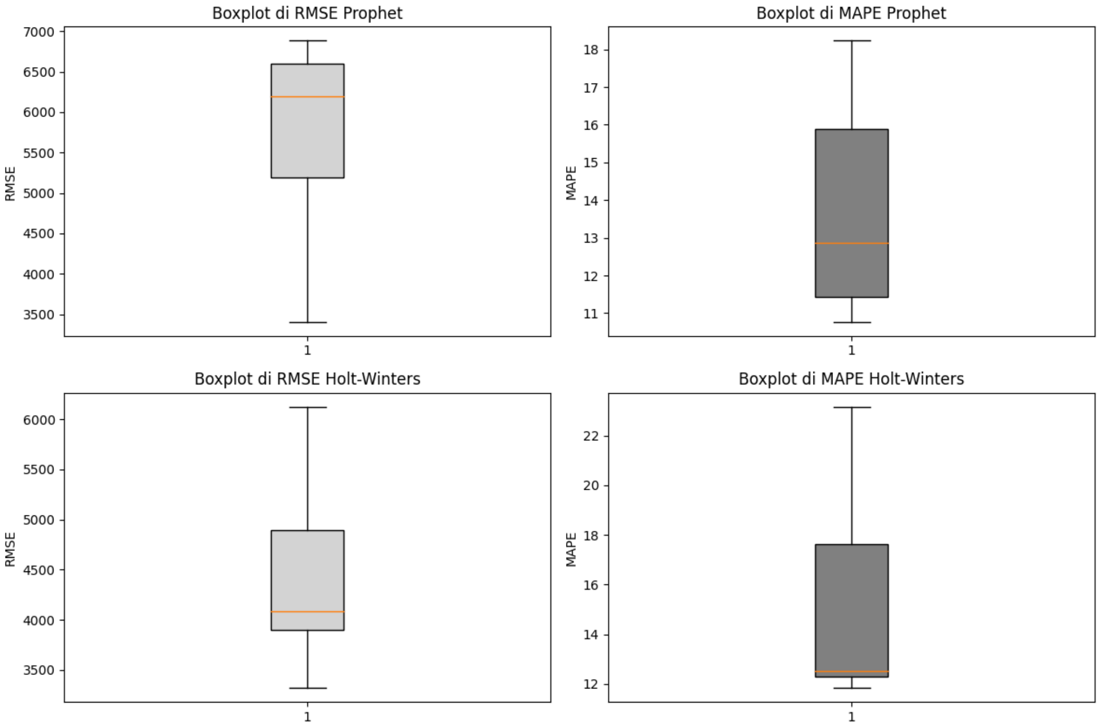


Figure 21: Boxplot MAPE & RMSE of Prophet & Holt-Winters

Prophet RMSE: The variance of the RMSE for Prophet is relatively high (198.8641), indicating that there is some variability in the results between restaurants. However, the range is smaller than for Holt-Winters, with values ranging from about 3,500 to nearly 7,000.

MAPE: The variance of MAPE is also quite low (9.90), with values ranging from about 11 percent to 18 percent. Prophet shows greater uniformity in MAPE values, although some restaurants have higher percentage errors than others.

Holt-Winters RMSE: The variance of the RMSE for Holt-Winters is slightly lower than for Prophet (117.8662), ranging from about 3,500 to 6,000. This indicates that despite the lower variance, Holt-Winters has a more even distribution of results among restaurants.

MAPE: The variance of MAPE is significantly higher (23.85) than Prophet, and the range of values is wider, varying from about 12% to over 23%. This suggests that Holt-Winters has more variable performance across restaurants in terms of percentage accuracy. Final Considerations.

RMSE: Prophet shows more variance in results than Holt-Winters, but has a narrower range. Holt-Winters, on the other hand, has a more consistent distribution but somewhat higher error values.

MAPE: Prophet has a more consistent distribution with lower MAPE values than Holt-Winters, which shows more variability and higher values, especially on some restaurants.

In conclusion, Prophet seems to be more consistent and produces more uniform results than Holt-Winters, especially in terms of percentage accuracy (MAPE). However, Holt-Winters may give similar or better results in some cases in terms of absolute error (RMSE).

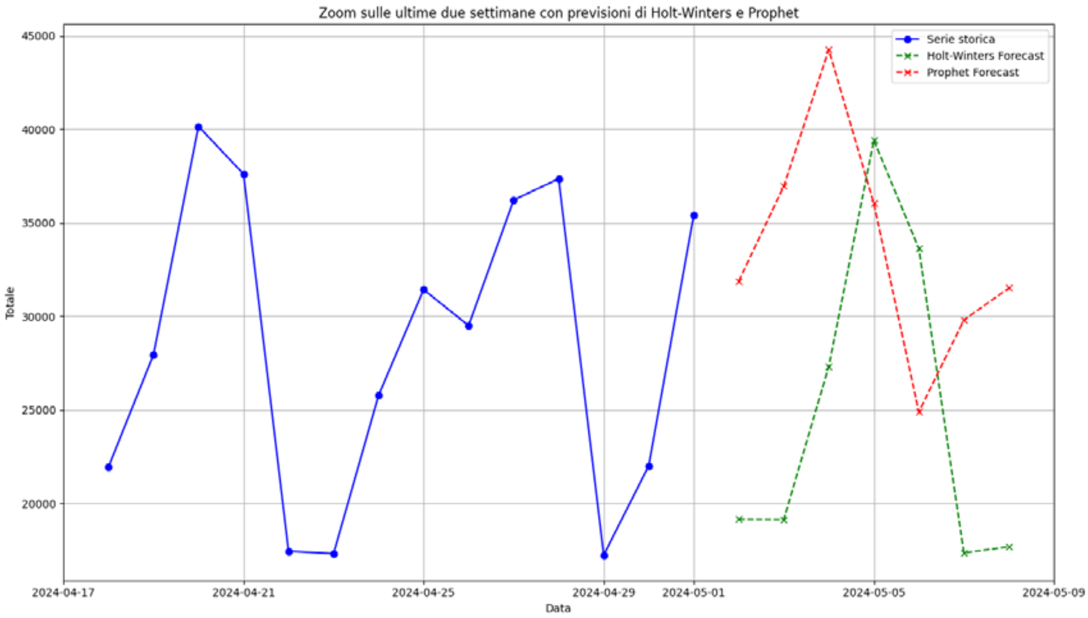


Figure 22: Last two weeks with forecasts of Prophet & Holt-Winters

SARIMAX vs TBATS

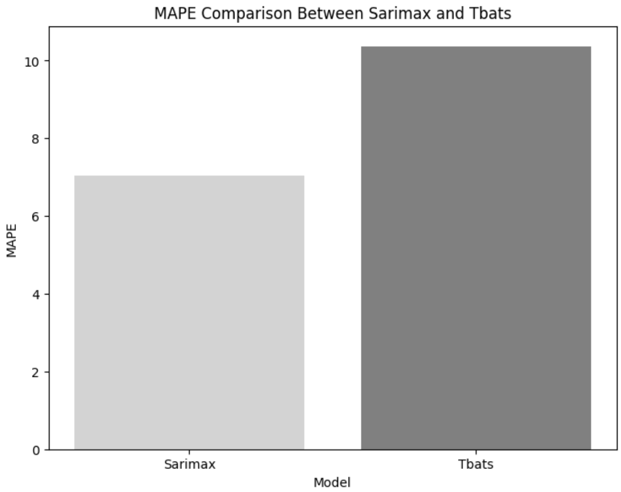


Figure 23: MAPE comparison between SARIMAX & Tbats

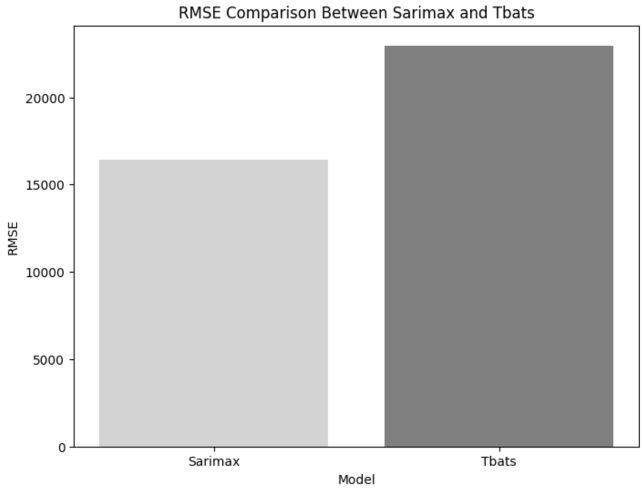


Figure 24: RMSE comparison between SARIMAX & Tbats

The MAPE (Mean Absolute Percentage Error) for SARIMAX is lower than for TBATS. This indicates that, in terms of mean percentage error, SARIMAX provides more accurate forecasts than TBATS. A lower MAPE for SARIMAX suggests that its forecasts are closer to actual values, in relative terms, than the TBATS model.

The RMSE (Root Mean Squared Error) for TBATS is higher than for SARIMAX. This implies that the TBATS model has larger absolute errors than SARIMAX. SARIMAX has a lower RMSE, which means it makes smaller absolute errors than TBATS. Comment on the results: SARIMAX seems to perform better in terms of both relative error (MAPE) and absolute error (RMSE). This suggests that the SARIMAX model is more accurate in predicting both percentage-scale values and absolute differences between predicted and actual values. So here it is clear that the impact of the regressors always positive for the accuracy of the model.

It can be said that it tends to be the case that all models have good MAPE and RMSE, especially MAPE which is below the 20% threshold, while only sarimax has it excellent i.e., below the 10% threshold. It would seem that the models with the expandable rolling window perform better, although to be certain one would have to compare them with the same distribution, i.e., all weekly or all daily. (Unfortunately, for computational reasons this was not possible).

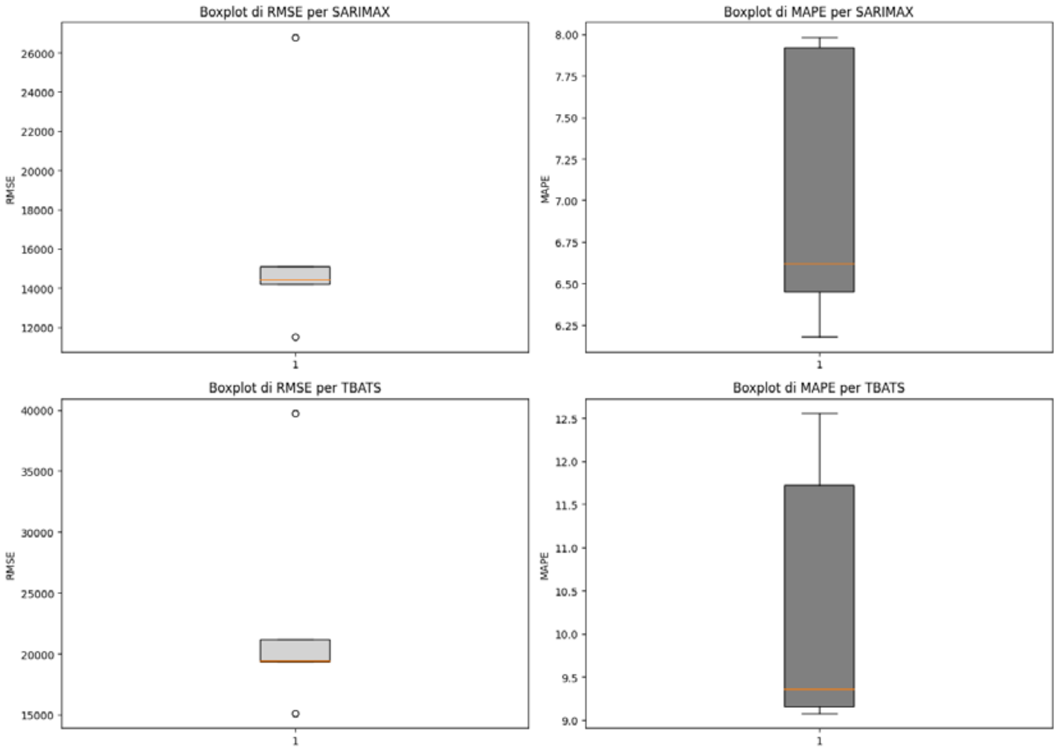


Figure 25: Boxplot of RMSE & MAPE for SARIMAX & Tbats

Box plot of RMSE for SARIMAX: The distribution of RMSE values for SARIMAX shows moderate variation, with most values concentrated between about 12,000 and 16,000. Two outliers are noted: one very high, near 26,000, and one very low, near 12,000. This indicates that although SARIMAX tends to provide fairly consistent results, there are some predictions that deviate significantly, with a much higher error (rist2_final.csv).

Box plot of MAPE for SARIMAX: The MAPE for SARIMAX has a wide distribution, varying between 6% and nearly 8%. The median is positioned near 6.5%, indicating that most SARIMAX predictions have relatively low percentage error. However, the variability is considerable, suggesting that the relative accuracy of the model can vary significantly from restaurant to restaurant.

Box plot of RMSE for TBATS: The distribution of RMSE values for TBATS is higher than for SARIMAX, with most values between 19,000 and 25,000. Two outliers are also noted here: one very low, near 15,000, and one very high, near 40,000, suggesting that TBATS has greater variability and some cases where the absolute error is very high (rist2_final.csv). TBATS, in general, seems to have higher absolute errors than SARIMAX.

Box plot of MAPE for TBATS: MAPE for TBATS also has a wider distribution than SARIMAX, varying between about 9% and 12.5%. The variability is high, suggesting that the performance of the model in terms of relative accuracy changes considerably from restaurant to restaurant. The median is higher than SARIMAX, indicating that TBATS tends to make higher average percentage errors.

In general, SARIMAX appears to be the more stable and accurate model, especially when mean percentage errors are considered, while TBATS may be more susceptible to larger errors under certain conditions or for some specific restaurants.

The resulting graph compares the SARIMAX and TBATS model forecasts with the weekly time series for the specified restaurant

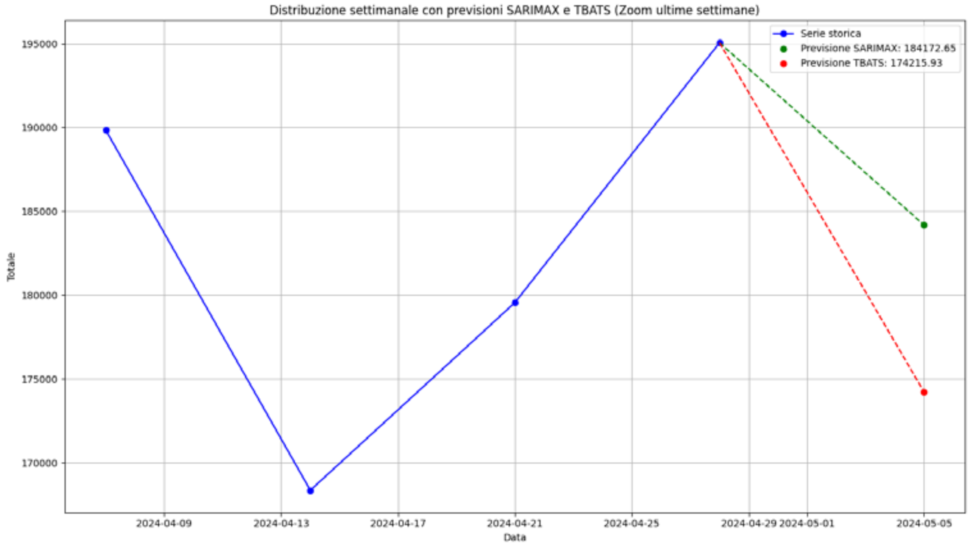


Figure 26: Last two weeks with forecasts of SARIMAX & Tbats

FUTURE DEVELOPMENTS

To improve forecast accuracy, new regressors such as promotions, local events, and price changes are planned to be included. This will enrich the models with additional factors affecting turnover, providing more detailed forecasts.

For fair model evaluation, uniform cross-validation will be implemented on all models considered. This approach will make it possible to directly compare the performance of different models and select the most accurate one.

Extending the models to daily data will provide more detailed and useful forecasts for day-to-day management, such as staffing and inventory planning. This will require modifications to the models to handle the granularity of daily data and optimize forecasts.

These developments aim to improve the accuracy and usefulness of forecasts, providing more effective tools for restaurant operations management and optimizing business strategies.