Business Economic
and Financial Data

2024-2025
Sales Forecasting

Prepared by

Daniel Gutierrez, 2108173
daniel.gutierrezvelez@studenti.unipd.it

Fabio Pimentel, 2110709
fabiomanuellenin.pimentelcaminero@studenti.unipd.it

Table Of Content

About the restaurant

DimSum Records is an Asian restaurant located in a busy area of Medellín, Colombia. The concept of the restaurant is to offer fast food inspired by Asian-style dishes such as dim sum, dumplings, baos, noodles, and more. However, the dishes incorporate local influences due to the availability of various ingredients and flavors. In addition to the food, there is a strong focus on the bar. The restaurant provides a relaxed ambiance where customers can hang out after their meal and stay late into the night. This setting allows for the sale of cocktails and beverages that complement the Asian food offerings.

The restaurant was founded in November 2021 and has a capacity for 52 customers at a time. It also offers home delivery through its own delivery services and popular food delivery apps. The restaurant is open throughout the week, with extended hours on weekends.

The purpose of this business is to generate revenue with a target net profit margin of 30%. If earnings meet expectations, the restaurant may expand and open new locations in the future. This document aims to analyze the restaurant's sales performance and explore the potential for developing a model to forecast sales in the coming years.

Executive Summary

In this work, linear and nonlinear models were applied to daily, weekly, and monthly data in the form of time series. The accuracy and performance of these models were evaluated to identify the most effective ones for explaining and predicting the data. To keep this document concise, we present only the most relevant results from the models deemed suitable for the available data.

The discussion is limited to describing the data source, the properties of the collected data, the forecasting goals, and the predictions of the best-performing models. Since numerical information about the past is available and it is reasonable to assume that some patterns will persist into the future, quantitative forecasting methods were employed to make predictions for the short, medium, and long term.

The performance of the models could only be properly evaluated after the forecast period's data became available. Therefore, these predictions were compared with the actual sales data for November and December 2024. The initial data for fitting the models was obtained in November 2024, and as a result, conclusive sales figures for November and December were not yet available during the modeling phase.

For readers interested in more detailed information, please refer to the GitHub repository, which contains the data and R code with an exhaustive implementation of the models.

GitHub repository: https://github.com/danielgvb/time_series_padova

About the data

For this project, data on daily restaurant sales was gathered from November 2021 to October 2024. Some days had null values, which were accounted for during the data preprocessing phase before implementing the models.
Additionally, external data was collected to assess whether external factors influence the behavior of restaurant sales. Specifically, the following data sources were used:

Economic Data:

- Economic Activity Index (ISE): A monthly index describing the overall behavior of the Colombian economy.
- Foreign Exchange Rate (USD/COP): A daily average exchange rate in the foreign exchange market.
- Inflation (IPC): A monthly indicator of overall consumer price inflation in Colombia.
- Unemployment: A monthly indicator of the unemployment rate in the Antioquia region, where the restaurant is located.

Additional Data:

- Google Trends: A weekly indicator of Google searches for “restaurants” in Antioquia.
- Rainfall: Daily rainfall data in Medellín and Antioquia.
- Temperature: Daily temperature data (mean and median) in Medellín.

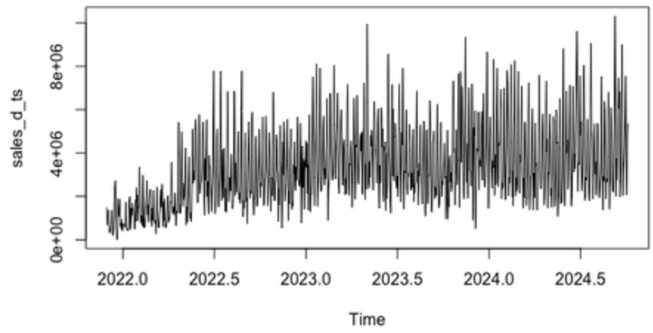
EDA

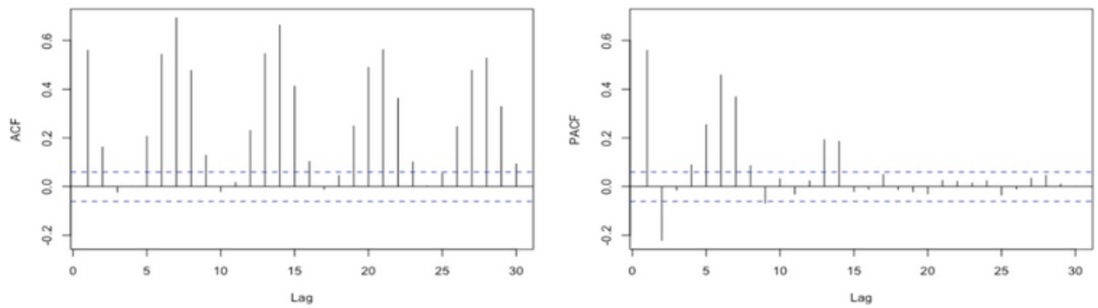
First we performed an exploratory data analysis to check how the target variable behaves, and also to check if it there is any kind of relationship with the external covariates.

Daily, weekly and monthly sales

Daily sales

There is a noticeable increase in daily sales from early 2022 to late 2022. This growth in the restaurant's sales can be attributed to the natural progression of a new business, where initial brand recognition is low, but marketing efforts and word-of-mouth gradually increase the business's popularity. After late 2022, sales show greater variability without a consistent upward or downward trend. This suggests a stabilization of growth, with large fluctuations in sales potentially tied to seasonal or event-driven factors.
The spikes, which indicate the presence of a seasonal pattern, may be associated with specific days of the week, such as Friday, Saturday, and Sunday, or particular periods like holidays. The frequency of these spikes suggests possible weekly seasonality, a common phenomenon in restaurants where weekend sales are typically higher than weekday sales.



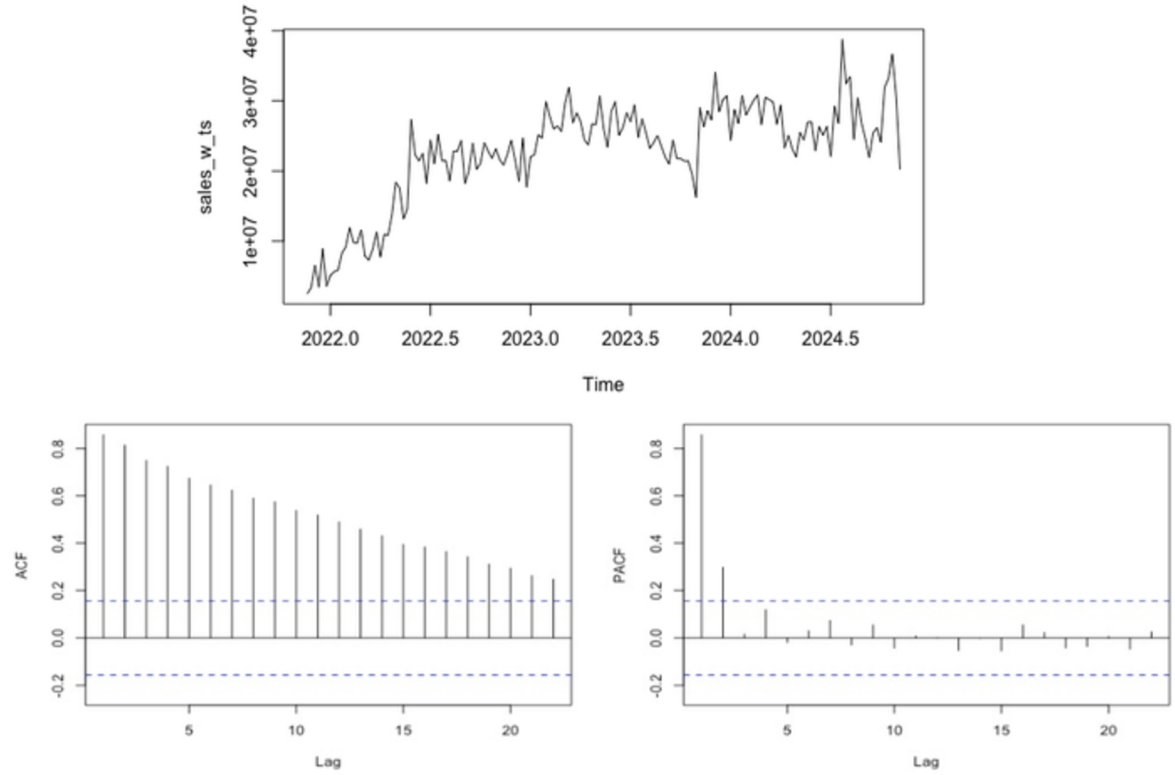


When data exhibit seasonality, the autocorrelation will be higher for seasonal lags (at multiples of the seasonal period) than for other lags. This is evident in the ACF plot, where significant and positive correlations are observed at weekly lags, such as 7, 14, and so on. Additionally, the autocorrelation for smaller lags is initially large and positive but gradually decreases as the lag increases, indicating the presence of a trend in the data.

Plotting the partial autocorrelation function (PACF) with confidence interval lines is a common method for analyzing the order of an AR model. To evaluate the order, the PACF plot is examined to identify the lag after which the partial autocorrelations fall within the confidence interval. This lag is likely to represent the order of the AR model.

Weekyl sales

The data shows a clear upward trend in sales from early 2022 to 2023, reflecting a consistent increase in weekly sales. This trend highlights the effectiveness of our marketing efforts, the growing popularity of our brand, and the rising demand for our products. By mid-2023 to 2024, the upward trend appears to stabilize, with sales fluctuating around a higher average. This may suggest that the growth rate is slowing or that the market has reached a saturation point.

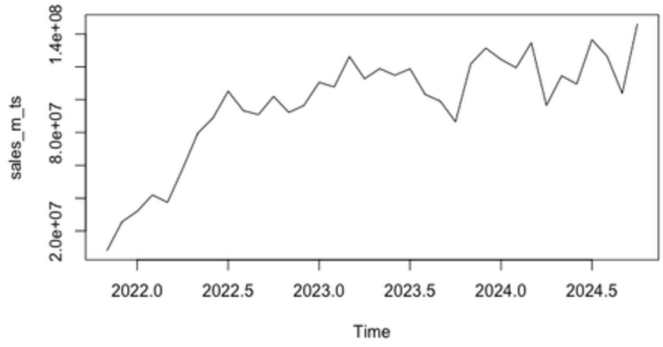


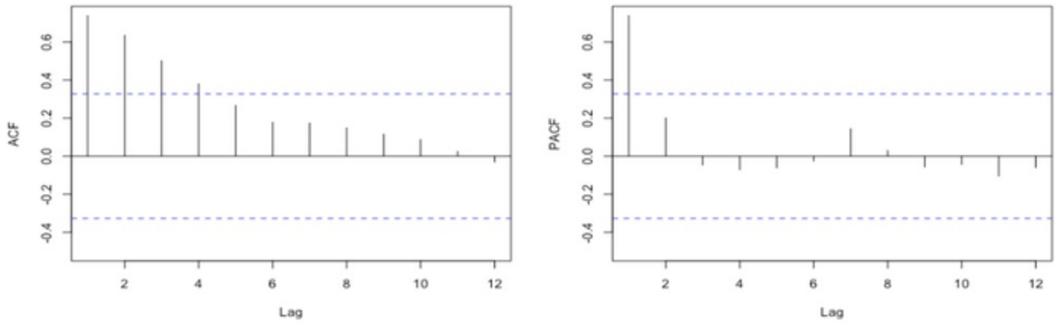
For weekly sales, the ACF plot does not reveal strong seasonal patterns, but it clearly shows that autocorrelation decreases slowly, indicating the presence of a trend in the data.

The PACF plot displays a significant spike at lag 1, followed by a quick drop to near-zero for subsequent lags. This suggests that the current value depends primarily on the immediately preceding value. Combined with the ACF, this implies that a simple autoregressive (AR) model might partially explain the data, although the trend would need to be addressed separately.

Monthly sales

The sales demonstrate a clear upward trend, particularly throughout 2022, indicating significant growth in the restaurant's monthly sales during that period. Around early 2023, the sales begin to level off, fluctuating within a higher range, which suggests that the initial rapid growth has slowed or stabilized. After 2023, the trend becomes more volatile, with noticeable peaks and troughs; however, the overall range remains consistently high.



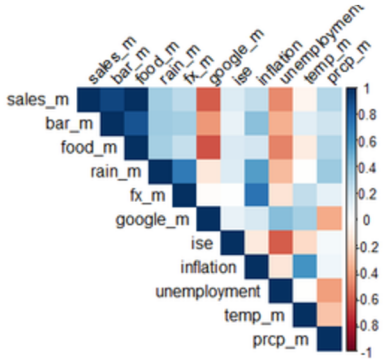


From the ACF wee can see how the trend effect rapidly decreases indicating that the data initially had some trend but then reached some sort of saturation in which the trend does not change much.

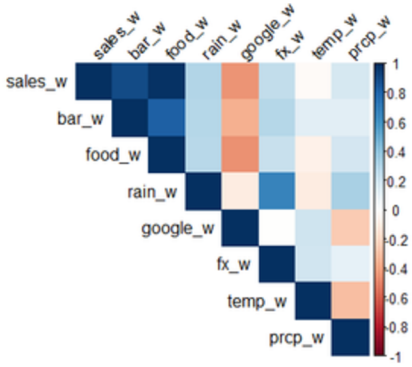
The PACF tell us that the first lag is strongly correlated with the series. The significant spike at lag 1 followed by a rapid decline to non-significance is characteristic of an AR(1) process, where only the first lag is important in modeling the time series.

Correlation Analysis

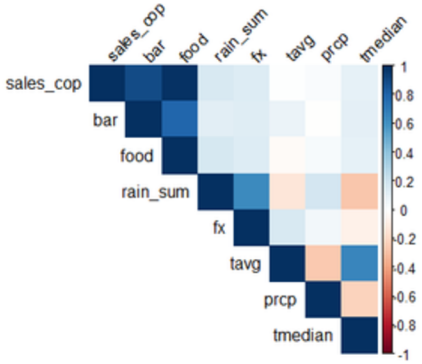
Montly Data



Weekly Data



Dailyly Data



At the monthly level, sales correlate negatively with Google searches and unemployment. At the weekly level, sales show correlations with rain, Google searches, foreign exchange rates (FX), and temperature. However, at the daily level, no clear correlation is evident in the plots. We plan to investigate this further by implementing linear models and analyzing their coefficients.
From the correlation plots, we observe that rain has a stronger correlation with sales than "prcp." Therefore, we drop "prcp" to avoid redundancy in using the same variable from two sources. Additionally, we exclude average temperature since median temperature appears to be more reliable.

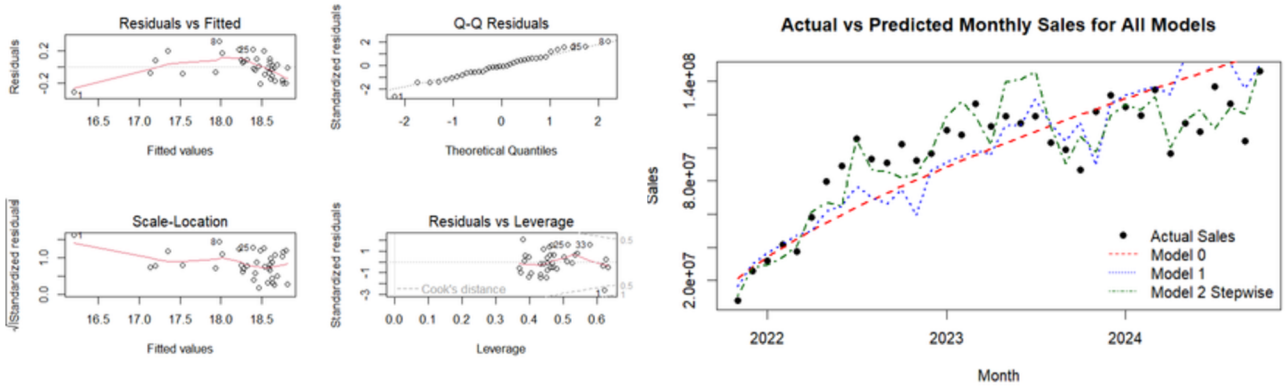
Linear Models (OLS)

The first set of models we implemented were the simplest: linear regression models. This modeling was performed iteratively. For each periodicity, we followed these steps:

1. Log-transform variables: The variables were log-transformed to produce more interpretable regression coefficients, particularly for covariates not directly related to time.
2. Fit a trend model: A model of sales was fitted against the trend, represented as linear time.
3. Fit a trend and seasonality model: A model of sales was fitted against both the trend and seasonality. For monthly and weekly periodicity, seasonality was represented by dummy variables for months (1 to 12) or days (Monday to Sunday). The first month or day was excluded to avoid perfect multicollinearity.
4. Fit a full model: A model of sales was fitted against trend, seasonality, and additional covariates. This step included a stepwise backward regression, selecting the best model based on AIC scores.

At the end of this modeling process, we obtained three models for each periodicity. These models were evaluated using R2 and RMSE to determine the best fit. Additionally, fitted values were plotted against the original values to visually assess model performance.
The results for the monthly periodicity are summarized in the table and charts below.

Model <chr>	R2 <dbl>	AIC <dbl>	RMSE <dbl>
Model all covariates step	0.9397340	-2.961086	13229295
Model trend	0.7750083	14.461543	21374338
Model trend + season	0.8123201	29.933834	21523336



The models effectively capture the trend in the monthly data but fail to account for some information, as indicated by a curved pattern in the residuals vs. fitted plot and a leverage point in November 2021, when sales were unusually low due to the restaurant's limited operation during its first week. To address these issues, we plan to fit time series models to better capture the data structure and improve residual behavior.

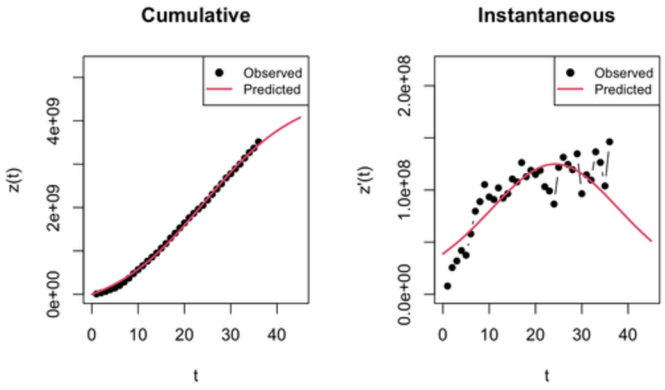
For the best-fit model on monthly data, stepwise regression revealed two significant predictors (p-value < 0.05):

- A 1% increase in Google Trends is associated with a 3.17% increase in sales.
- A 1% increase in the FX rate corresponds to a 2.72% increase in sales.

These effects are logical: Google Trends reflects consumer interest, while a higher FX rate benefits foreign customers, making prices more favorable and boosting sales.

Bass and Generalized Bass models

The cumulative curve (observed and predicted) follows a typical S-shaped pattern, characteristic of the Bass model. Early growth is slow due to the low p (innovation-driven phase). Growth accelerates in the middle phase as q (imitation) dominates. The curve eventually levels off as it approaches the market potential (m). We can see that the model fits the observed data well for cumulative adoption and adoption is likely nearing saturation, as the curve flattens toward m.



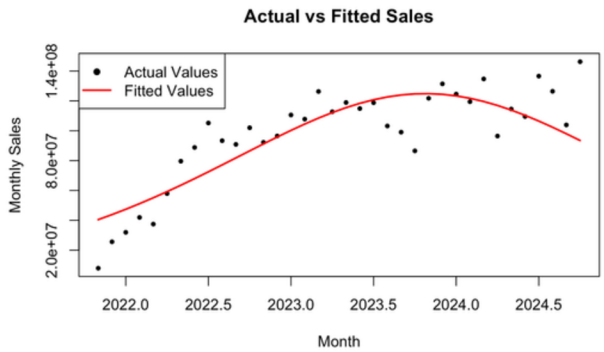
The instantaneous adoption rate shows the number of new adopters over time. The observed data (black points) roughly align with the predicted curve (red line), though there are some deviations. The peak of the curve represents the time of maximum adoption rate, after which the rate declines. The peak adoption occurs relatively early due to the relatively high q value (imitation-driven diffusion). After the peak, the rate of adoption slows as the market approaches saturation.

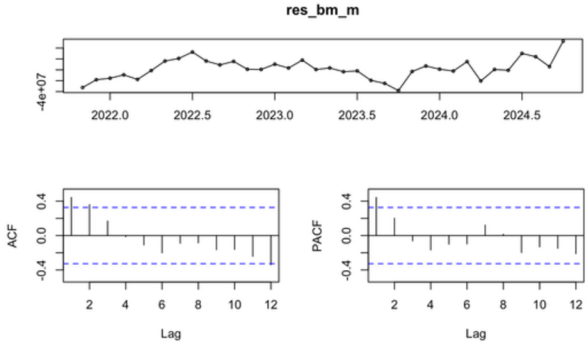
According to this, there are only 1m cop left to sell, this is less than a year / seems wrong. Fits well but the 30- onward is weird + sales might not be declining yet. Still reflects the innovation and copying in some sense.

Also the restaurants rely in word of mouth to reach full stage m = 4.664.000.000 COP, i.e 1 mm EUR approx. The restaurant has sold 3.515.788.885, according to this only in 1 year it should extinguish sells.

Then the p, the coefficient of innovation is 0.832%, this indicates that the adoption rate due to external influence is relatively low, but not uncommon for many markets. - it is actually relatively innovative.

Finally the q, the coefficient of imitation is 8.96%, suggests that imitation plays a larger role than innovation in driving adoption in this market.

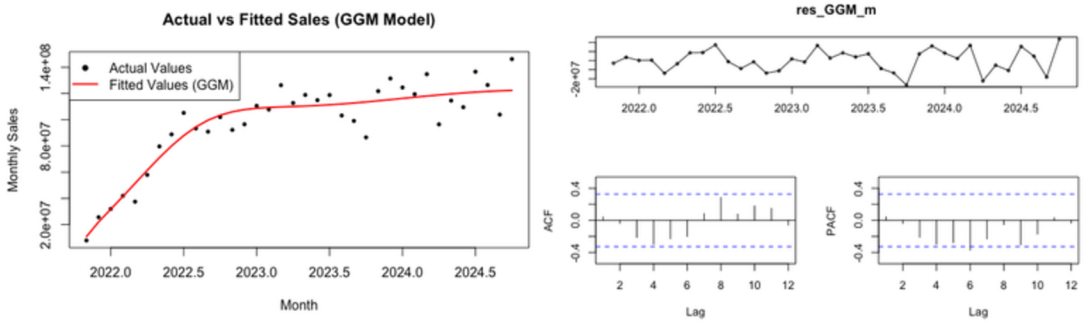




RMSE for Bass Model Predictions: 18.498.870.

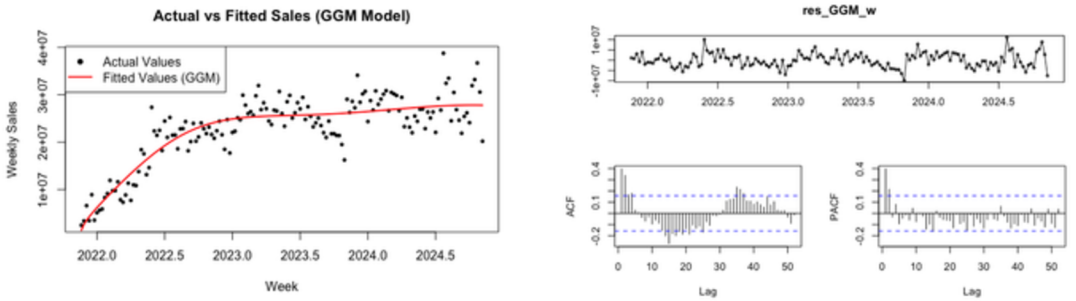
Guseo-Guidolin model

Guseo and Guidolin (2009) proposed a particular specification for m(t), by making the hypothesis that the development of the market potential depends on a communication process about the new product, which typically precedes the adoption phase and serves the purpose of “building” the market [Guidolin-2023].

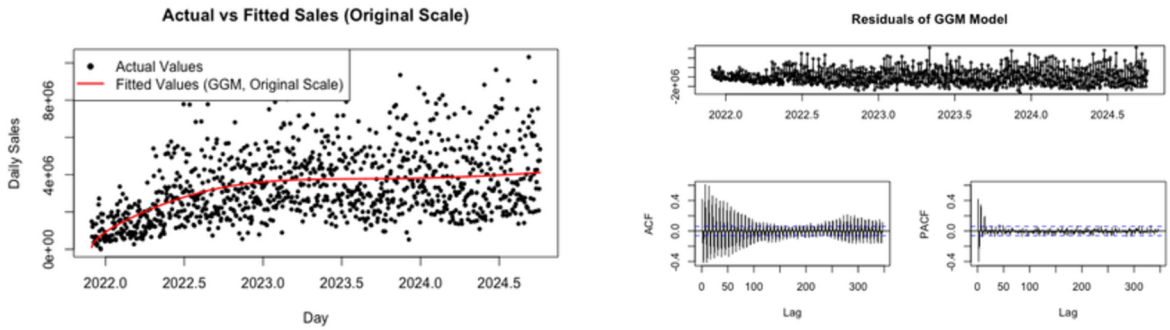


On the right, we see a GGM model fitted to the monthly sales data. On the left, the residuals of the model fluctuate around zero with no clear trend or seasonality, indicating that the model effectively captures the overall trend and seasonality in the data. However, the spikes in residuals suggest the presence of outliers or unexplained variability.
The ACF plot shows that most autocorrelations fall within the blue confidence bounds, indicating that the residuals are not significantly autocorrelated at most lags. This lack of significant autocorrelations suggests that the residuals behave like white noise, implying that the model successfully captures the serial dependencies in the data.

RMSE for GGM Model 1 (Base): 11.759.505



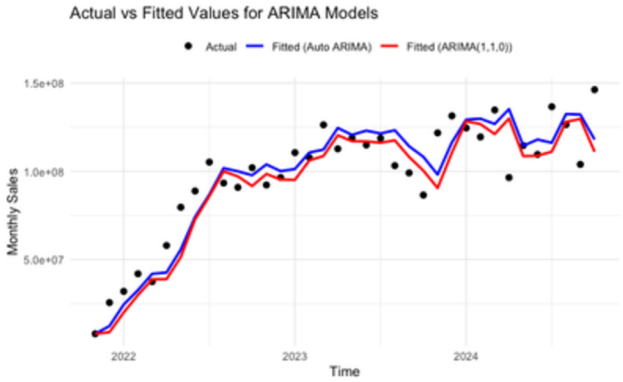
On the right the GGM model fitted to weekly data. The ACF plot shows a sinusoidal pattern with significant autocorrelation at the peaks of the "waves," it typically indicates the presence of seasonality or cyclic behavior in the time series data.
RMSE for Weekly GGM: 3.453.199

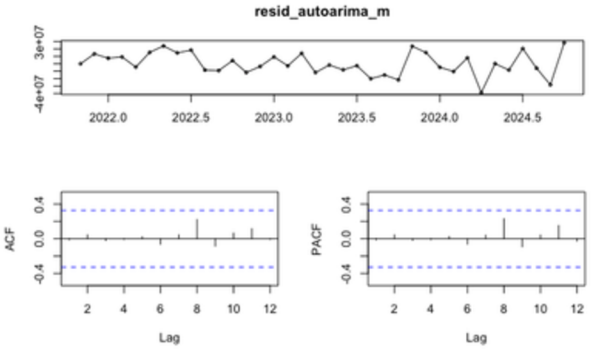


In the ACF we can see some sort of periodic behavior that the model has not fully captured. To address this, the model could be improved by incorporating a seasonal component or additional explanatory variables that account for recurring patterns in daily sales
RMSE for Daily GGM: 1.600.510

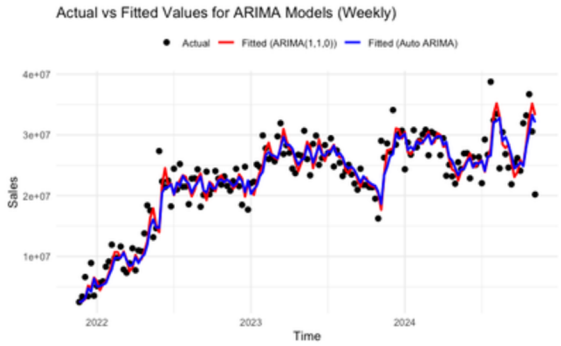
ARIMA models

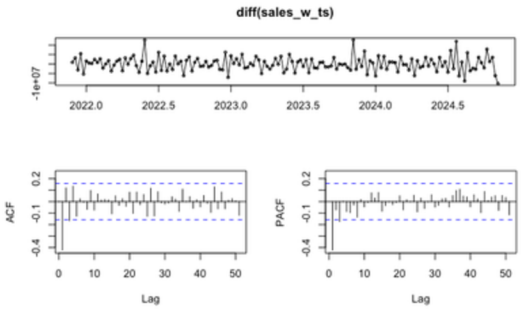
In this section we compare the performance of an ARIMA(1,1,0) which is a differenced first-order autoregressive model, and a fitted auto ARIMA model which is an ARIMA(0,1,1) with drift . The AIC of the the manual ARIMA model is 1265, while the one of the auto ARIMA is 1263, according to this the auto ARIMA model is most likely to be the best model for the monthly sales data set.



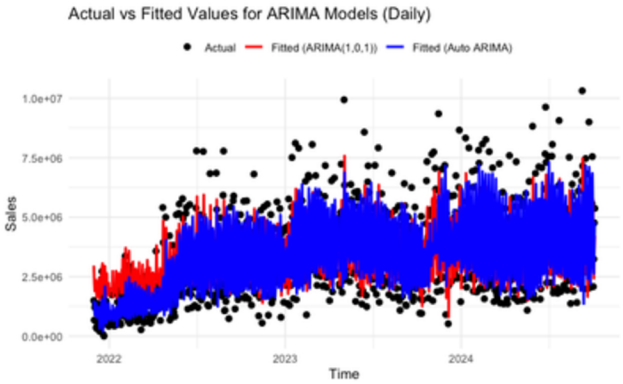


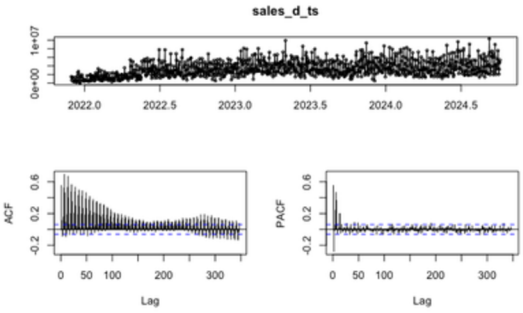
On the left we can see that the auto ARIMA model is pretty similar to the manual one, they both follow the data points in an acceptable way. On the right the plot of residuals, the ACF and PACF indicate that the model captures very well the trend in the data. We have also computed RMSE for Auto ARIMA Model which is 15.118.942 and RMSE for ARIMA(1,1,0) Model: which is 15.867.282.





We can not say the same for the autocorrelation of the ARIMA model applied to weekly and daily sales. As we can see in the ACF plots above and below. This is worse for daily data in which we can see that the autocorrelation plot does not look like white noise anymore, indicating that he model is not properly capturing the behavior in the data. Also the fitted plot is not readable. RMSE for ARIMA(1,1,0) Model (Weekly)is 3.385.012, while RMSE for Auto ARIMA Model (Weekly) is 3.328.293.



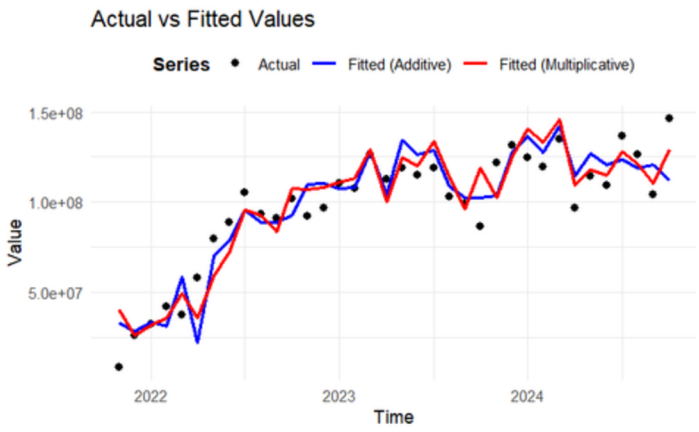


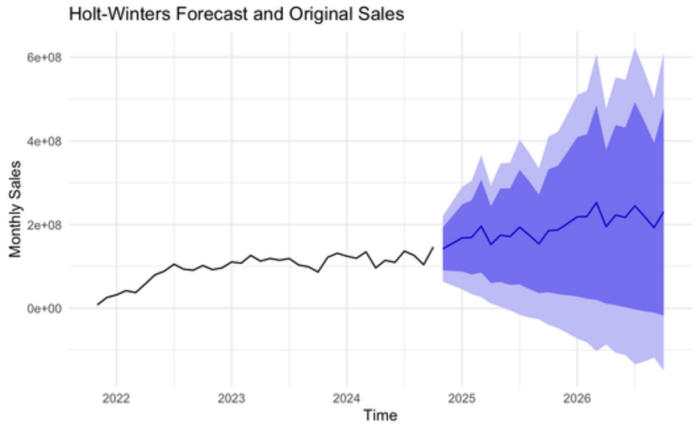
RMSE for ARIMA(1,0,1) Model (Daily): 1.475.571. RMSE for Auto ARIMA Model (Daily): 1.094.980.

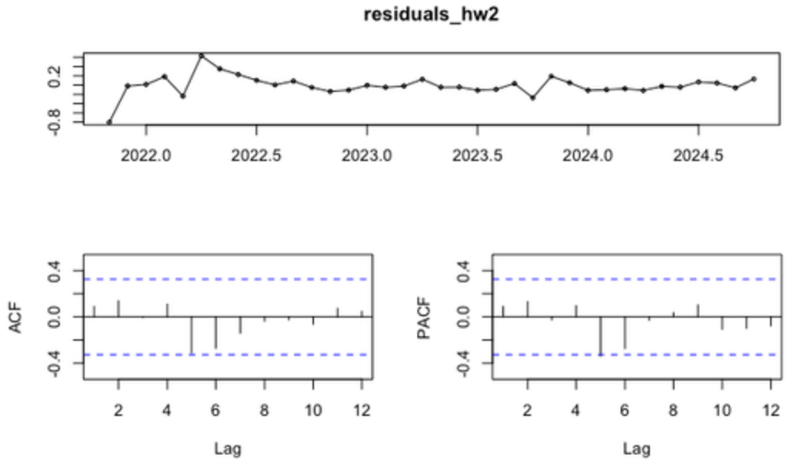
Holt-Winters

Triple Exponential Smoothing

In this section, we apply the Holt-Winters model to the monthly sales data. The actual vs. fitted values show that the multiplicative model generally follows the data more closely. Additionally, the forecast aligns well with the observed data, indicating a slight upward trend for 2025 and 2026.







In the correlation plot, ACF and PACF we can see that the models captures very well the seasonality and trend of the data since there are not significant autocorrelation and the residuals fluctuate around zero. The Augmented Dickey-Fuller Test confirms the residuals are stationary.

After evaluating the RMSE and the residuals of both models we conclude that the multiplicative version of the Holt-Winters model is better than the additive one. This model arises as a great solution to fit in a “smooth” way some series that might contain trend and seasonality.

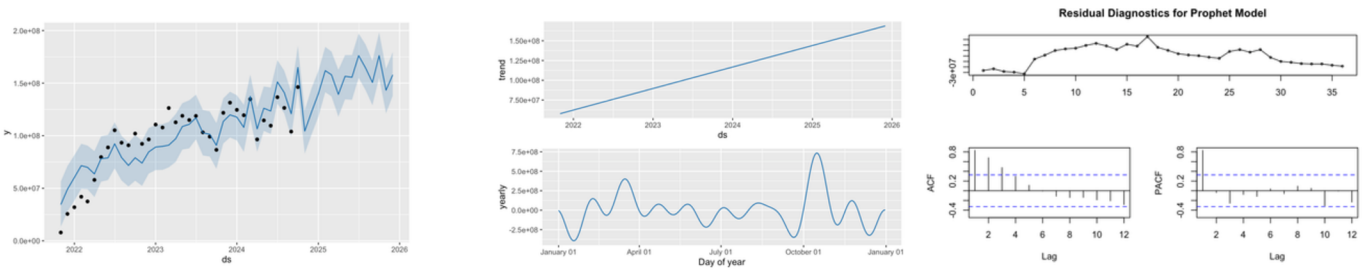
Unfortunately, the current implementation of R does not allow to fit a model with a periodicity larger than 24. In our case we have periodicity, or frequency of 12 for the monthly data, 52 for the weekly and 365 for the daily.

We could transform the weekly to 24 and interpret periods of 15 days at a time, but is not so coherent with the rest of our modelling efforts, and we also have some other tools to fit both the weekly and the daily frequencies.

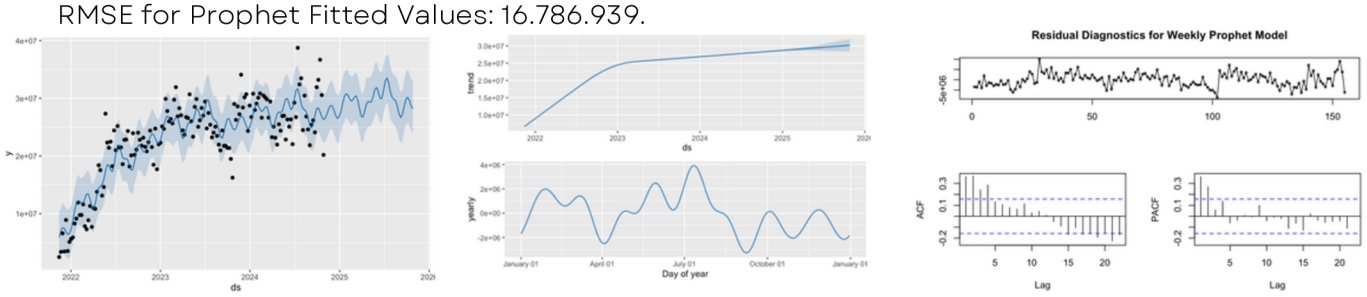
Prophet models

This model was introduced by Facebook (S. J. Taylor & Letham, 2018), originally for forecasting daily data with weekly and yearly seasonality, plus holiday effects. It was later extended to cover more types of seasonal data. It works best with time series that have strong seasonality and several seasons of historical data.

The next is the result of applying Prophet to monthly sales, we can see in the graph in the center that it projects a clear linear trend in sales. In the same graph below we can see how the Prophet finds yearly seasonality, showing some negative sales at the beginning of the year and a spike in sales in the third quarter of the year.



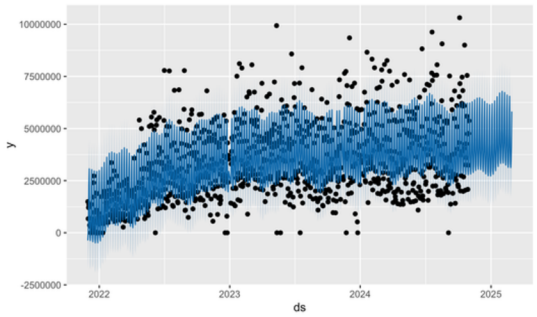
The residuals fluctuate around zero, which is a positive sign as it indicates no systematic bias in the forecast. However, periods of increasing variance, particularly toward the end, suggest the presence of heteroscedasticity (non-constant variance). This highlights areas where the model may underperform.
Additionally, significant spikes at lags 1, 2, and beyond the confidence intervals indicate that the residuals are not entirely random. This suggests that the Prophet model has not fully captured the temporal structure in the data. Adjustments to the model, such as incorporating external regressors or modifying seasonal components, may improve performance.
RMSE for Prophet Fitted Values: 16.786.939.

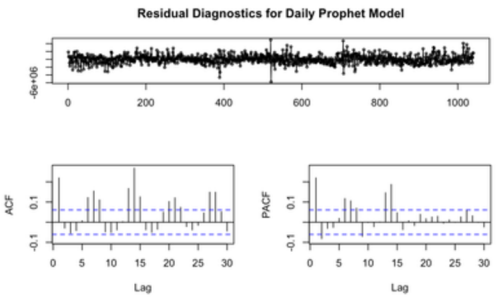


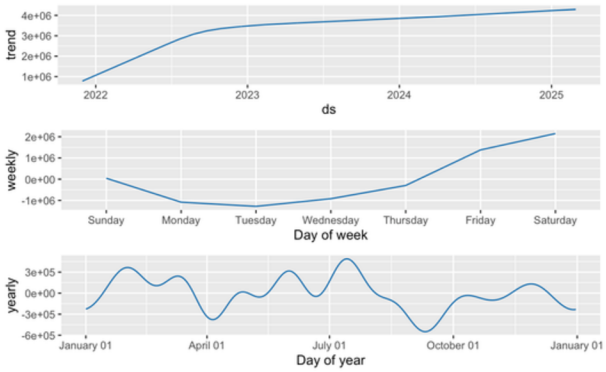
The graphs above show the result of the model on weekly sales. We can see that Prophet identifies some saturation in sales with slow increase after 2023. For 2025 there is some uncertainty on whether sales will increase or stay saturated. According to the model there is a spike in seasonality during summer while sales decreases for the last months of the year. In the ACF plot the first few lags have significant positive autocorrelation, indicating that the residuals are not entirely random. This suggests that the model has not captured all temporal dependencies, meaning there is still some predictable structure left in the residuals.

The PACF plot shows that first lag is strongly significant, suggesting a short-term relationship between past and current residuals. Some smaller significant lags indicate that there might be missing seasonality or trend components in the model.
RMSE for Prophet Fitted Values (Weekly): 3.266.752

Finally for daily sales, in the next plots we can see that the residuals oscillate around zero, which suggests that the model is not biased. The spread of residuals appears fairly consistent, although some spikes indicate occasional large errors. There is no strong visible pattern, which is a good sign that the model is capturing most of the trends. The ACF confirms the presence of weekly seasonality since we have periodic spikes at lags 7, 14, and 21. We can also see in the third plot how the model accurately identifies the trend through the week in which Monday, Tuesday and Wednesday are the worst day but then there is a increase in sales on weekends, which is a expected behavior. RMSE for Prophet Fitted Values (Daily): 1.071.298







Mixture Models:
GGM with SARIMA Refinement

In this section, we aim to maximize the predictive power of two approaches: adoption models, using the GGM as the base model, and SARIMA models, which capture intrinsic patterns in the seasonality and volatility of the stochastic processes underlying our time series.

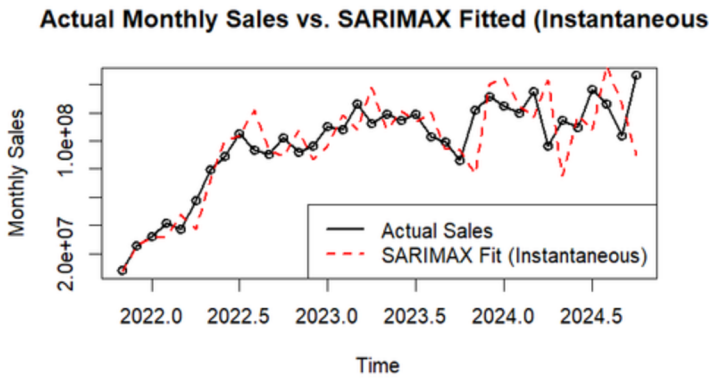
We first fit a GGM to the monthly, weekly, and daily sales data. The fitted values from the GGM are then used as an external regressor in a SARIMA model for cumulative sales, enabling us to uncover both the trend and seasonality of the series. The results are displayed in the plot below, focusing on the monthly and weekly series.

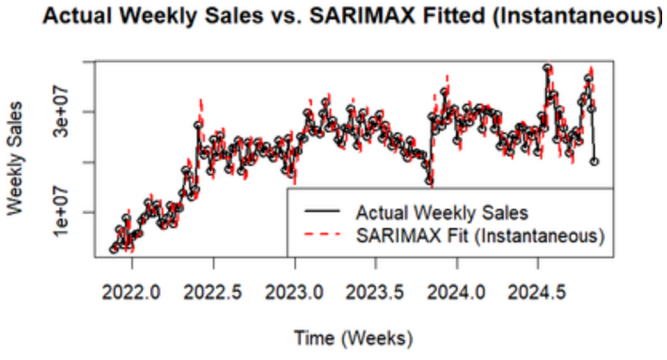
The fitted values of the mixed model closely follow the original data series, suggesting its potential usefulness for forecasting over longer timespans out of sample.

While the model demonstrates a good visual fit, an analysis of residuals reveals the following:

- For the monthly series, the residuals behave like a stationary process and show no significant serial correlation.
- For the weekly and daily series, the residuals also appear stationary but exhibit strong serial correlation in the correlograms.

Although serial correlation could not be fully eliminated with these models, stationarity was achieved according to the Augmented Dickey-Fuller (ADF) test.

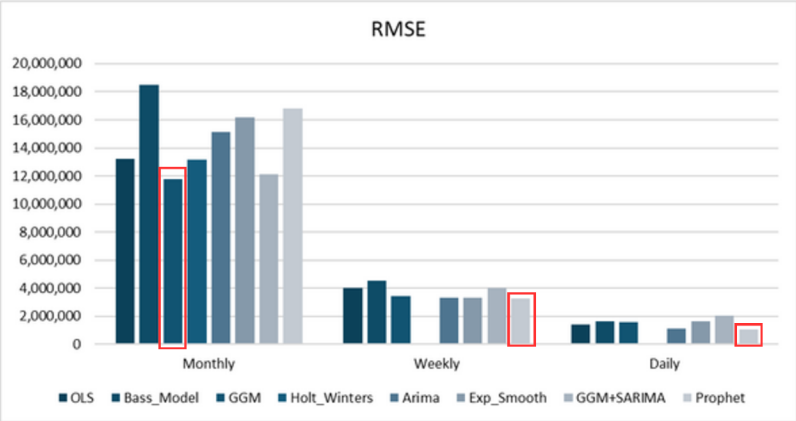




Model Selection

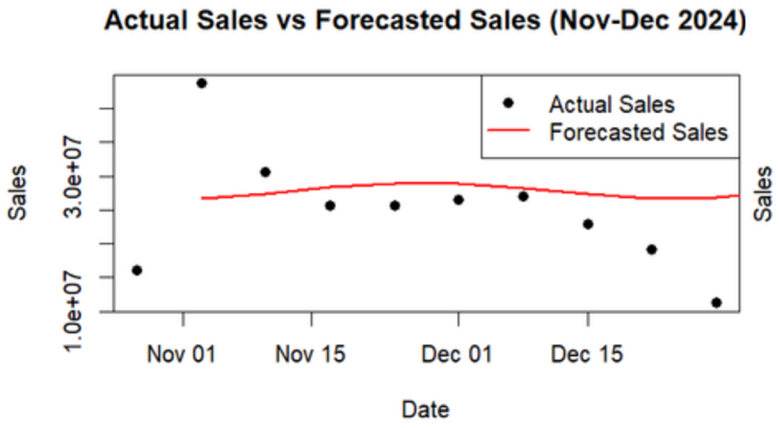
The following table summarizes the fitted models RMSE. We selected the best model in each periodicity to forecast the missing two months (November and December 2024). Finally, we evaluated how are they compared with the real sales.

RMSE			
Model	**Monthly**	**Weekly**	**Daily**
OLS	13,229,295	4,000,370	1,421,301
Bass_Model	18,498,870	4,542,019	1,651,896
GGM	**11,759,505**	3,453,199	1,600,510
Holt_Winters	13,169,921	NaN	NaN
Arima	15,118,942	3,328,293	1,094,980
Exp_Smooth	16,189,623	3,331,847	1,642,520
GGM+SARIMA	12,123,823	4,027,032	2,053,161
Prophet	16,786,939	**3,246,088**	**1,071,375**

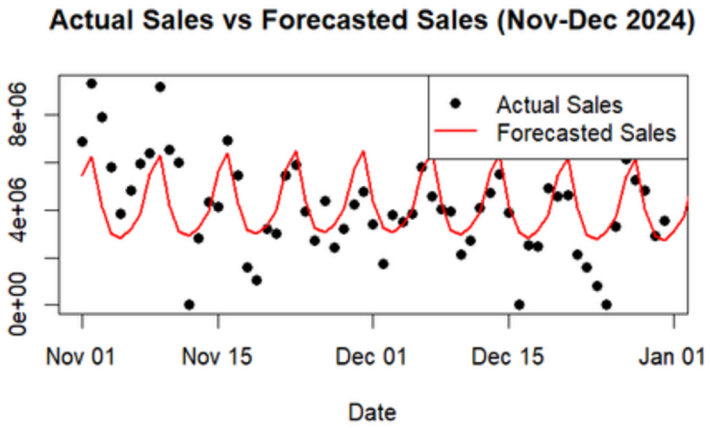


For the monthly periodicity the best model was the GGM, followed by the Holt-Winters.
For the weekly and daily periodicity the best models were both the Prophet.
The following plots compare the model forecast with the real life values:

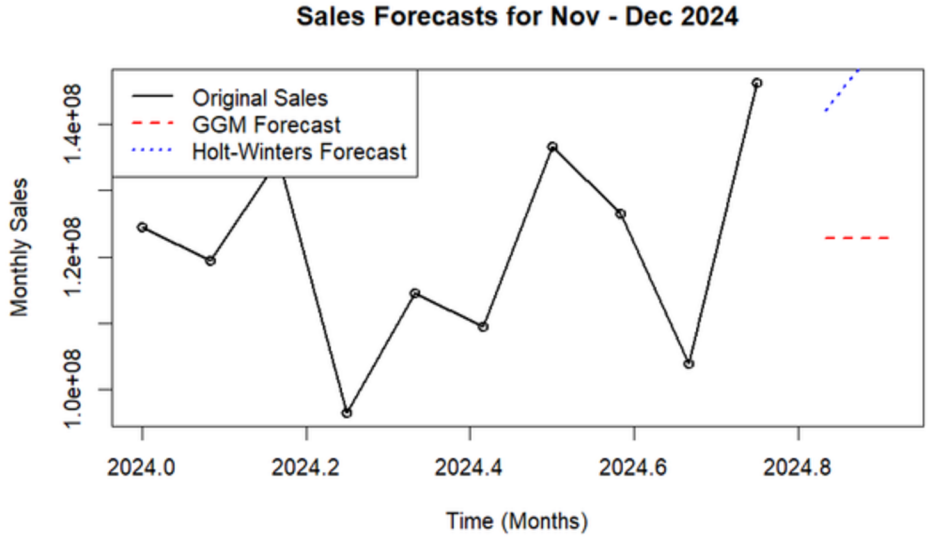
Weekly



Daily



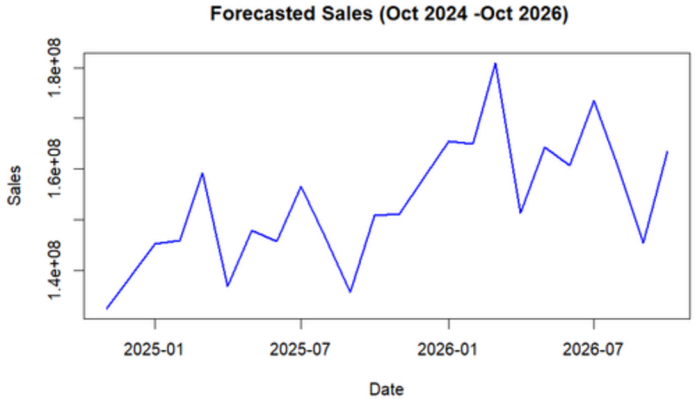
Monthly



Long Term Forecast

To evaluate the restaurant's performance, we conduct a long-term forecast using the monthly models. For this, we employ both the GGM and Holt-Winters models. The outcomes of both models are averaged because the GGM tends to over-smooth the series and naturally predicts a decline in sales, which is unlikely for a restaurant at this stage. Using the sales forecast, we develop a financial model (a three-statement model) to value the company. This represents a step beyond traditional time series analysis, as it uses the forecast to assess the investment potential of the restaurant.

For the valuation, we apply two methodologies: the Free Cash Flow (FCF) method and the EV/EBITDA method. Both approaches illustrate how the restaurant's value has changed over time. We perform a five-year forecast, leading to the following results:



Item	EUR
Free Cashflow Method	
EV FCF	**230,844**
Cash	0
Debt	0
Equity Value	**230,844**
EBITDA Method	
Exit Multiple	7.00x
FWD EBITDA	98,818
EV	**691,723**
IRR FCF	**26.6%**
IRR EBITDA Multiple	**33.6%**

The plot on the left shows a steady increase in sales based on the forecast generated through time series analysis. On the right, the table summarizes the valuation results, with yearly sales as the primary input. The restaurant's initial investment was just under €70,000. The financial model spans a five-year period and estimates a return on investment (ROI) between 26.6% and 33.6%, depending on the valuation methodology. Despite the variation, both metrics are of a similar order of magnitude.

Conclusions

Throughout this report, we have outlined several methods to better understand sales behavior and forecast sales across three periodicities:

- Monthly data: The analysis suggests that sales may be stagnating, likely due to the limited availability of tables in the restaurant and the natural constraints faced by all businesses.
- Weekly data: The model predicts similar values for the upcoming weeks, which is reasonable given the lack of strong seasonality identified in this periodicity.
- Daily data: A robust predictive model was developed, accounting for the strong seasonality observed on weekends. While this pattern was already known, the model provides valuable insights for optimizing staff scheduling and inventory management.

Key Model Performance:

- Short-Term Forecasting: Holt-Winters and ARIMA performed well but required careful tuning.
- Long-Term Forecasting: Diffusion models offered valuable insights into the product lifecycle and market saturation.
- Complex Patterns: Prophet emerged as the most robust model, effectively handling seasonality, holidays, and change points with minimal effort.
- External Factors: Regression models provided explainability but were more challenging to optimize.