10-3-2024

# Competing Forecasting Models to Study Crisis Periods: The Case of Sweet Snack Sales

Sara Barradas
*ISEG Lisbon School of Economics & Management, Universidade de Lisboa*, l58277@aln.iseg.ulisboa.pt

Patrícia Martins
*Advance/CSG, ISEG Lisbon School of Economics & Management, Universidade de Lisboa*, patriciamartins@iseg.ulisboa.pt

# COMPETING FORECASTING MODELS TO STUDY CRISIS PERIODS: THE CASE OF SWEET SNACKS SALES

*Research full-length paper*

Barradas, Sara, ISEG Lisbon School of Economics & Management, Universidade de Lisboa, Lisboa, Portugal, l58277@aln.iseg.ulisboa.pt

Martins, Patrícia, Advance/CSG, ISEG Lisbon School of Economics & Management, Universidade de Lisboa, Lisboa, Portugal, patriciamartins@iseg.ulisboa.pt

## Abstract

*The COVID-19 pandemic had a significant impact on consumer purchasing behavior, particularly in the food retail sector. In response to this socioeconomic crisis, food retailers adjusted their strategies to align with new consumer preferences. During the pandemic, sweet snacks sales grew globally as it satisfied the desire for snacking for many families confined at home. Detecting the impact of COVID-19 on sales involves comparing pre-pandemic and post-pandemic sales trends to identify deviations attributed to the crisis. This study examines the sales evolution of sweet snacks at two major Portuguese retail chains from January 2018 to June 2023, split into pre-crisis, crisis, and post-crisis periods. The primary goal is to infer and compare forecasting models using ARIMA and Prophet time series models to assess consumer preference changes and, additionally, predict post-crisis future sales. The findings reveal a break in consumption patterns between periods. In the pre-crisis period, sales progressively increased until the lockdown, declining in the crisis period. At the end of the crisis, consumption patterns normalized, but post-crisis, retailers diverged due to their adaptability to new trends. ARIMA models provided better overall accuracy in predicting stable future sales, while Prophet models delivered more precise forecasts in the post-crisis sales patterns.*

*Keywords: Time Series Forecasting, Crisis Periods, Retailer Sales, Sweet Snacks*

## 1 Introduction

The outbreak of the COVID-19 pandemic in early 2020 led to profound changes in daily life worldwide, with a slowdown in the global economy, especially in private consumption. Like many other countries, food retailing was significantly impacted by lockdowns, social distancing measures and public health concerns. Consequently, this period of socioeconomic crisis reshaped the behavior of consumers regarding food consumption (Janssen et al., 2021), namely Portuguese consumers, in response to the unprecedented challenges posed by the pandemic (da Costa et al., 2023; Faghih & Forouharfar, 2022). Therefore, it is essential to understand if these effects indicate a permanent structural shift in consumers' purchasing patterns or a temporary change (Mehta et al., 2020). Additionally, it is important to understand the market dynamics, including how various retailers reacted to the crisis, as this can offer insights into competitive dynamics and market shifts.

Research indicates that the global pandemic had varying impacts on different food sectors. The crisis did not affect all food categories equally, with some experiencing more significant effects than others. (Dou et al., 2021; Janssen et al., 2021). Moreover, how families handled the disruptions in their daily routines affected the eating behaviors of both adults and children (Smith et al., 2021). Specifically, this influence was observed in the increased consumption of sugary sweet snacks (Jansen et al., 2021).

This research aims to study differences in food retail purchasing patterns of Portuguese consumers of sweet snacks resulting from the COVID-19 crisis. Therefore, the following research question guides our study "*What are the effects of the COVID-19 pandemic on sweet snacks sales in Portuguese retailers?*". Additionally, we delve into the effects of the pandemic on the food retailer market by producing future predictions of Sweet Snacks category sales for the period after June 2023. This analysis extends the assessment of sales performance in the post-crisis period and detects potential anomalies in the sector.

The study develops forecasting models to identify the time series forecasting models that most accurately describe the evolution of sales in the Sweet Snacks category. Therefore, this analysis compares competing time series forecasting models across three forecasting periods: the pre-crisis period from January 2018 to February 2020; the crisis period from March 2020 to December 2021; and the post-crisis period from January 2022 to June 2023. Comparing forecasting models between time periods is important for model performance evaluation, detection of structural breaks, assessment of model robustness, improved forecast accuracy, and for the identification of seasonal effects. Different forecasting models may perform better in different time periods due to changing patterns, trends, and external factors. This comparison helps to identify more reliable models and which ones are effective only in specific periods. Economic conditions, market dynamics, and consumer behavior evolve over time, demand adaptable forecasting methods for accurate forecasts. Structural breaks, i.e., abrupt changes in data patterns, can be detected through model comparisons, also providing further understanding of the impact of these breaks (Pardo-Jaramillo et al., 2023). Additionally, model comparison aids in selecting suitable models for future use. Finally, computing forecasting models across several time periods, particularly in different seasons, enable us to properly consider seasonal effects in forecasts (Box et al., 2015).

The analysis used sales data from two leading Portuguese food retailers provided through a partnership with a Category Management and Shopper Marketing consulting company. The goal was to identify unique sales trends and compare sales behavior between the two retailers during three different subperiods. After an exploratory data analysis, models were built to generate forecasts and extract insights using specialized R packages. The best predictive models provided numerical outputs and graphs for a robust analysis of performance metrics. This interactive approach highlighted the features of Autoregressive Integrated Moving Average (ARIMA) models, which have historically been popular, and Prophet, which has recently emerged as a potential alternative. The findings reveal differences between retailers regarding their sales strategy dynamics and recovery trajectories, analysis of model performance across retailers and forecasting periods, and in the evaluation of time series models, in which, ARIMA showed superior overall performance, but Prophet achieved more accurate predictions and provided narrower confidence intervals, enhancing the reliability of future forecasts.
The remainder of this paper includes background and context, followed by methods, results, and conclusions, limitations, and further research.

## 2 Background and Context

The COVID-19 pandemic revealed weaknesses in the global food supply chain. Disruptions in transportation, shortages in labor, and a surge in demand for specific products presented challenges to the food retail industry in Portugal. Two significant trends emerged from this situation: a change in consumer behavior and a growing dependence on e-commerce for grocery shopping (Faghih & Forouharfar, 2022; Roggeveen & Sethuraman, 2020). At the beginning of the pandemic, panic buying and stockpiling led to temporary shortages of essential food and household items. Retailers were forced to prioritize hygiene and safety measures such as contactless payment, curb-side pickup, and store sanitation. Consumer spending was significantly affected by the crisis, with rising unemployment rates and increased price sensitivity leading to an 8% reduction in European households' expenditure in 2020 (Zwanka, 2022). Portuguese consumers have reduced the frequency of dining out and increased eating at home, leading to a preference for private labels due to their lower prices compared to manufacturer brands (Pinto et al., 2022). Additionally, there was a growing demand for healthier and more sustainable food choices, such as organic and locally sourced products. Retailers responded by innovating their own-brand product range, including new 'free-from' products (no sugar, additives/preservatives, and allergens), meat substitutes, and options for specific diet regimes (da Costa et al., 2023). Also, consumers with higher disposable income occasionally indulge themselves by buying non-essential items such as gourmet goods. A McKinsey study revealed that 73% of Portuguese consumers have adopted new shopping behaviors, including trying own-brand products (McKinsey, 2020).

The pandemic crisis led to a significant increase in e-commerce activity in the food retail sector, particularly in online grocery shopping and home delivery services (Gomes & Lopes, 2022). Moreover, fully online retailers have leveraged their share in the Portuguese industry (INE, 2021). When not purchasing online, consumers opted for more frequent shopping trips to smaller stores that are easily accessible from home, as opposed to larger hypermarkets and supermarkets. This growing demand for convenience and proximity has led to investments in new services, such as click-and-collect and pick-up points, which offer the benefits of online shopping with immediate access to products (Gomes et al., 2023). Food retail chains have been increasingly integrating omnichannel experiences into their strategies, allowing for data collection from online interactions and enhancing the relationship between retailers and shoppers in physical stores. Despite the rise of online sales, it is expected that physical stores remain the largest and most significant sales channel for retailers.

In Portugal, the retail food categories most affected by COVID-19 were tree nuts, due to increased snack consumption at home; alcoholic beverages, whose sales only increased in the first months due to government restrictions on purchases after 8 pm; and fresh produce, which initially saw greater demand for packaged goods due to uncertainty surrounding virus transmission. In contrast, ready-to-eat and ready-to-cook meals suffered a drop in sales as people had more time to prepare meals at home (USDA Foreign Agricultural Service, 2021). On the other hand, the sweet snacks category has seen a boost in consumption due to the increasing trend of enjoying snacks and pastries at home, with a focus on premium quality products. This includes the occasional purchase of premium products like chocolates and bonbons, as more people seek innovation, diversity, and indulgence in their at-home dining experiences. Five trends have been identified that reflect the post-COVID-19 behavior of Portuguese consumers and remain relevant in the current sweet snacks market. These trends include snacking at home, the "premiumization" of purchases, a shock to loyalty and shift toward private label brands, sustainability and ethical consumerism, and an alignment with health and wellness choices. Additionally, private labels have gained prominence in the Portuguese market by offering affordable alternatives targeted at budget-conscious consumers, which has led to a decline in market share for manufacturer's brands.

## 2.1 Forecasting models for sales

Forecasting retail sales is an important analysis that impacts decisions in retail businesses. Sales data are a type of time series encompassing identifiable trends and seasonal patterns, therefore requiring the development of forecasting models to effectively capture and predict future sales trends (Ramos et al., 2015). ARIMA and Prophet models have been widely used as effective time series models for forecasting retail sales (Ensafi et al., 2022; Fildes et al., 2022).

ARIMA models can handle non-stationary data, which is common in retail sales due to trends and seasonal variations. Based on historical data, ARIMA can provide accurate short-term forecasts, and include diagnostic tools to assess model fit and performance, ensuring reliable forecasts. As such, ARIMA are well-suited for forecasting retail sales dynamics. Prophet is a robust model for sales data forecasting because can handle various types of seasonality, including daily, weekly, and yearly patterns, accommodating holidays and special events that can affect sales which is an important feature in retail sales. This model is designed to manage missing data and outliers, making it resilient to anomalies. Additionally, it offers an intuitive and flexible framework, allowing users to incorporate other business-specific insights and trends into the model.

# 3 Methodology

This study's main goal is to understand the impact of COVID-19 on retailers' sales by comparing pre-pandemic and post-pandemic sales trends to identify deviations attributed to the crisis. We established a partnership with a Category Management and Shopper Marketing consulting company that provided us with sales data from a leading sweet snacks manufacturer of two market-leading Portuguese food retailers, over a five-and-a-half-year period, including the COVID-19 crisis period. For anonymization purposes, we refer to these retailers as Miam and Munch.

After cleaning and preparing the data and conducting an exploratory data analysis, we infer forecast models using various R packages, such as *forecast, tsibble, tidyverse, tseries, tidyquant, gridExtra, ggplot2, and prophet*. Finally, we present the results of the best predictive ARIMA and Prophet forecasting models using the numerical outputs and graphs generated by R Studio (Version: 2023.12.1+402).

## 3.1 Data

The source data was collected from the retailers' data management platforms with restricted access. From these platforms, the daily sales reports for Miam and Munch were extracted in Excel format for the period under analysis from January 2018 to June 2023. After data collection, the Excel files were aggregated into twelve folders organized by year and retailer. The final data set used in the analyses encompasses a total of 94625175 data points, 47312904 for Miam and 47312271 for Munch. The analysis of the collected data lead to the identification of ten different variables to use in the forecasting models. Table 1 presents the variables of our study.

| Variable Name | Description | Variable Type |
|---|---|---|
| SalesDate | Period under analysis. | Numeric Discrete |
| Retailer | Designation of the Portuguese retailer. | Categorical Nominal |
| Banner | Designation of the banner owned by a particular retailer. | Categorical Nominal |
| Region_Store | Store location in Portugal by geographical region (NUTS II). | Categorical Nominal |
| Store | Store identification code. | Categorical Nominal |
| Category | Designation of product category associated with a particular item. | Categorical Nominal |
| SubCategory | Product subcategory code associated with a particular item. | Numeric Discrete |
| Brand | Designation of the brand associated with a particular item. | Categorical Nominal |
| Item | Item identifier. | Numeric Discrete |
| SalesUnits | Number of units sold of a particular item. | Numeric Discrete |

*Table 1. Variables description of retailers' sales*

From the original dataset, we selected the most relevant variables in Table 1 to forecast the retailer's sales. The agreement with the Category Management and Shopper Marketing consulting company includes not disclosing information about data not contained in the study.

## 3.2 Modelling

### 3.2.1 ARIMA

A widely recognized contribution to the field of statistical methods in time series forecasting is the Box-Jenkins Method (Box et al., 2015). Box-Jenkins Method is versatile in handling different time series patterns across a wide range of models and provides high accuracy in short and medium-term forecasts (Makridakis et al., 2022). The Box-Jenkins class of models are composed of three key components, each of which helps to model a certain type of pattern: the Autoregressive component denoted by AR($p$) corresponding to the autocorrelation component of a time series using a linear combination of previous values to predict future values (Shumway et al., 2000); the Integration component denoted by I($d$) referring to the number of differences required to obtain a stationary time series (Box et al., 2015); and the Moving Average component denoted by MA($q$) connoting to the moving average component of a time series and uses a linear combination of past prediction errors, which constitute unknown factors that affect the time series but are not explained by their past values (Box et al., 2015).

The ARIMA based models result from the combination of these Box-Jenkins components comprising three key components: *p* (autoregressive term); *d* (integration term); and q (moving average term) (Box et al., 2015; Shumway et al., 2000). Accordingly, the non-seasonal Autoregressive Integrated Moving Average model takes *p*, *d*, and *q* as parameters, represented by the formula in Eq. (1):

$$ARIMA(p,d,q): \phi p(L)\, \Delta^d y_t = c + \theta q(L)\, \varepsilon_t \qquad (1)$$

where $\phi p(L)$ corresponds to the non-seasonal autoregressive polynomial of order $p$ and $\theta q(L)$ corresponds to the non-seasonal moving average polynomial of order $q$; $\Delta^d\ y_t$ is the time series with $d$ differentiations; $c$ is a constant and $\varepsilon_t \sim N(0, \sigma^2)$ is a white noise series. Two alternative procedures can be employed in the process of building and fitting an ARIMA model for forecasting: Box-Jenkins method (Box et al., 2015) and Automatic ARIMA modelling (Hyndman & Khandakar, 2008).

### 3.2.2 Time Series Forecasting Process

To obtain a readable sales data structure for ARIMA models, the data set was converted into a *tsibble* object. Each row in the *tsibble* represents the unit sales of a specific product belonging to a brand, sold in a certain store owned by one of the retailers mentioned, located in a particular region and associated with a date of sale. The Sales *tsibble* was divided into three subsets corresponding to the three periods under analysis: pre-crisis period (2018/01/01 – 2019/02/29); crisis period (2020/03/01 – 2021/12/31) and post-crisis period (2022/01/01 – 2023/06/30). In turn, the Period Sales *tsibble* was filtered by 'BannerID' to obtain three *tsibbles* for each retailer, totalling six subsets. As such, Retailer Miam *tsibble* considers Banners 'B09', 'B10', 'B14' and 'B17' (corresponds to Miam Supers, Miam Hipers, Miam & GO and Miam Wellness), while Retailer Munch *tsibble* considers Banners 'B01', 'B02', 'B03', 'B05', 'B06' and 'B08' (Munch Hub, Munch Central, Munch Street, Munch HomeTech, Munch Office and Munch Online). Table 2 depicts the dimension of the ARIMA *tsibbles* and Prophet *data frames* for each period in R analysis.

| | # ARIMA observations | | | # Prophet observations | | |
|---|---|---|---|---|---|---|
| | Pre-Crisis | Crisis | Post-Crisis | Pre-Crisis | Crisis | Post-Crisis |
| Miam | 18067574 | 15095833 | 14149497 | 790 | 671 | 546 |
| Munch | 16165194 | 15653551 | 15493526 | 785 | 668 | 543 |

*Table 2. Dimension of the ARIMA tsibbles and Prophet data frames for each period in R analysis*

To visually identify patterns in the data, we produced a time series plot of total sales for each retailer over the different time periods. Furthermore, we decomposed the time series to analyze the presence of potential trends and seasonal patterns. It should be noted that the detection of seasonality was restricted in the crisis and post-crisis periods, given the data sets have less than two complete annual time periods, making the analysis only accurate for the pre-crisis period. The trend component varies according to the model and will be covered in the section of the Results Analysis. Afterwards, we applied a Yeo-Johnson transformation to stabilise the variance. This transformation was proposed by Yeo and Johnson (2000) as an extension of the Box-Cox transformation method, allowing for a wider range of input data including zero and negative values and making it useful for skewed. To evaluate the stationarity of time series, we computed unit root tests (ADF and KPSS tests), considering a significance level of 0.05. The results of these tests show that data is non-stationary. Additionally, the *unitroot_ndiffs()* function was applied to determine the recommended number of differences needed to achieve stationarity, which returned one order of differencing for yj-transformed time series data.

For model building, we plotted the ACF and PACF for the first differenced sales time series, with the aim of identifying the potential values for lags *p* and *q*. Then, we employed the *auto.arima()* function to facilitate the selection of the best ARIMA model based on the lowest AIC and BIC values for the yj sales time series. However, while this function automates the selection process, it might not always find the actual best model. Therefore, automatic selection was complemented by further manual evaluation of candidate models. The candidate models were chosen by adjusting the number of autoregressive and moving average terms by increasing or reducing the term by one or two, while maintaining the same differencing order ($d$=1). As such, the three data subsets were divided into a training set with the first 500000 observations and a test set with the last 200000 observations. During the model selection, the training and test sets were restricted to allow comparison of models with different parameter orders, given the computational limitations of R software.

### 3.2.3 Prophet

Prophet model is a non-parametric time series forecasting algorithm developed by Facebook in 2017 for business-related applications. It was created to optimise the wide range of business forecasting tasks with the following characteristics: regular nature; strong repeated seasonality; significant holidays; not too many missing observations or outliers; historical trend changes; and stochastic process (Taylor & Letham, 2018). The Prophet is represented by three main model components: trend, seasonality, and holidays, which are combined as in Eq. (2):

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \tag{2}$$

where *y(t)* corresponds to the observed value in the time series in period *t* ; *g(t)* corresponds to the general trend of the time series; *s(t)* corresponds to the seasonality component of the model; *h(t)* represents the influence of holidays and special events; and $\varepsilon_t$ represents idiosyncratic changes, assuming a normal distribution. Several studies conducted in the field of time series forecasting, demonstrate that Prophet provides significant improvements and greater accuracy for variables under study over ARIMA based models, particularly in the presence of seasonality (Duarte et al., 2021).

To implement the Prophet model, we converted the data set into a *data frame* structured with two columns: 'ds' assigned to the index 'SalesDate' and 'y' attributed to the measure 'SalesUnits'. Each row in the *data frame* represents the total sales for a retailer associated with a date of sale. We replicated the procedure applied to the ARIMA in Prophet to obtain six *data frames*. Moreover, to validate the consistency of seasonality and trend patterns observed in ARIMA, we produced time series graphs of total Sales, as well as the decomposition of time series. After that, we fitted Prophet Model attending to holidays and special events added as regressors, with an 80/20% split of the data set. The joint national holidays of both retailers are 'Valentines_Day', 'Carnival', and 'Easter', since Munch does not provide sales data for 25/12 ('Christmas') and 01/01 ('New_Year'). As such, the parameter 'extra_regressors' was set to 3. As a final note, Prophet has detected two types of 'seasonalities' (components of 'yearly.seasonality' and 'weekly.seasonality' are set to 1), which implies that models include seasonal patterns occurring on an annual and weekly frequency within the time series periods.

## 3.3 Evaluation and Forecasting

The Evaluation phase includes the model selection and evaluation and the final forecasting steps of the Time Series Forecasting Process.

### 3.3.1 ARIMA

The *glance()* function was used to compute the information criterion of fitted alternative models. This function provides a concise summary of various model fit statistics including R-squared, AIC, BIC, log-likelihood and estimated residual standard deviation (sigma). Both AIC and BIC statistics perform well on large data sets, and therefore the best ARIMA model was chosen given the lowest values returned by two. Additionally, to compare simulated and observed values among the ARIMA, we computed the accuracy metrics of MSE, RMSE, MAE and MAPE.

Unlike the model selection process in which a limited training and test set was applied, the model fitting considered a training set of 80% and a test set of 20% of the total observations, to allow a final comparison between ARIMA and Prophet models. Once the best model was obtained, the pattern of the residuals was analysed using ACF plot and Portmanteau tests, specifically the *Box.test()* function with ‘Ljung-Box’ and ‘Box-Pierce’ type specified. The function determined the presence of autocorrelation between residuals at lags 10, 30, 50 and 70 for all models, since these are random character data. Nevertheless, it is important to note this does not invalidate the ARIMA for forecasting purposes.

For the forecasting phase, 366 periods-ahead predictions were generated to cover the period from July 2023 to June 2024. A 95% confidence interval was used, in which the actual value of total sales is expected to be within the range estimated by the model. For a complete analysis, a plot of the forecasted values was executed to monitor the ARIMA projections of sales over the post-crisis period.

### 3.3.2 Prophet

For Prophet models, we generated the predictions using the trained model on a single test set for which the performance was directly assessed. Then, manually calculated accuracy metrics for the entire test set. In the forecasting step, the *make_future_dataframe()* function was used to create a *data frame* with a column ‘ds’ containing the next 366 future timestamps for which forecasts will be generated. After this, the *predict()* function returned a *data frame* containing the forecasted values and bounds. As a final step, we visualized the Prophet projections with its components (date, day of week and day of year) in a plot.

# 4 Results

## 4.1 Evolution of Sales in the Sweet Snacks Category

### 4.1.1 Pre-crisis period: January 2018 to February 2020

The total sales of Miam and Munch behave similarly throughout the pre-crisis period (Figure 1). Sweet snacks sales showed fluctuations over time, following a clear upward trend. Between January 2018 and June 2018, there was an increase in sales until March 2018 (ranging from 43750 to 87500 for Miam and from 45000 to 90000 for Munch), followed by a decline until the end of the semester (from 37500 to 75000 and from 36250 to 75000, respectively). During this first subperiod, Miam had its pre-crisis sales peak, with sales reaching 97938 on 2018/01/27. The following semester showed another change in trend with an increase until the end of 2018. The demand for products in this category continued to behave similarly in 2019, with sales rising in the first two months of 2020. Munch reached its maximum value of 118059 units on 2020/01/18. The pre-crisis period ended with a range of values between 50000 and 87500 for the first retailer and between 60000 and 120000 for the second. Easter 2018 and 2019 (2018/04/01 and 2019/04/21) marked the lowest point in sales for both retailers during the pre-crisis period. Miam outliers are of 19201 and 14093 units, and Munch, outliers of 18448 and 14110. As a final remark, analyzing two consecutive periods of 2018 and 2019 revealed the presence of yearly seasonality in the pre-crisis period.

### 4.1.2 Crisis period: March 2020 to December 2021

During the crisis period, Munch assumed a similar pattern as Miam with emphasis from July 2020 (Figure 2). Overall, Miam followed an upward trend characterized by the projection of sequential growing cycles where the market tried to counter the declines throughout the entire period. Both retailers suffered a sharp drop in March 2020, after the high sales figures recorded at the end of the pre-crisis period, which were only maintained in the first half of this month at around 75000 and 105000 units. Between second half of March and June 2020, Miam’s sales dropped to 25000 but progressively increased to 65500. In contrast, Munch continued to show a downward trend in sales, with a range of 37500 and 75000. From the second half of 2020, retailer’s sales balance gradually improved until the end of the crisis period. The sales ranges for Miam and Munch are presented below, in accordance: from July 2020 to December 2020 (37500 to 75000 and 45000 to 90000) and from January 2021 to June 2021 (37500 to 62500 and 45000 to 75000). In the first semester of 2021, Miam signaled its high sales point of 89627 units on 2021/05/01. The last month of the period revealed markedly higher sales, between 50000 and 87500 for Miam and between 60,000 and 105,000 for Munch. Moreover, there are two common peaks observed in the last half of the crisis period, registered in 2021/10/30 (83943 units for Miam and 16045 units for Munch) and in 2021/12/23 (88745 and 113550 units). As in the previous two years, it was at Easter 2020 (2020/04/12) that sales reached the worst level recorded during the crisis period. This represents unit sales of 183 for Miam and 10703 for Munch.

### 4.1.3 Post-crisis period: January 2022 to June 2023

During the post-crisis period, there were contrasting trends in the sales of the two retailers, Miam and Munch. Miam experienced a decrease in sales, while Munch saw an increase (refer to Figure 3). For the first half of 2022, Miam's unit sales were between 50000 and 87500, and for the middle of the same year, the range was between 37500 and 81250. In 2023, sales were reported between 50000 and 75000 units, and in June 2023, sales slightly fell between 50000 and 62500. As for Munch, the sales volume started at 50000 and 112500 in the first half of 2022, and then 62500 and 125000 in the second half of 2022. In the last five months of the analysis, the retailer recorded sales between 62500 and 125000, and in June 2023, it showed a lower upper limit between 62500 and 100000, although still revealing an increasing pattern in sales. It's worth noting that both retailers experienced a peak in sales on 2022/12/23, close to Christmas Day. This increase in demand may be attributed to the attractiveness of sweet snacks as gift options. On this day, 97534 units sales were registered for Miam and 126280 for Munch. Similar to the pre-crisis and crisis periods, the post-crisis period also saw a decline in sales during the Easter holiday, except for 2021. Sales figures were 24215 and 23430 for Easter 2022 (2022/04/17), and 25240 and 24075 for Easter 2023 (2023/04/09).

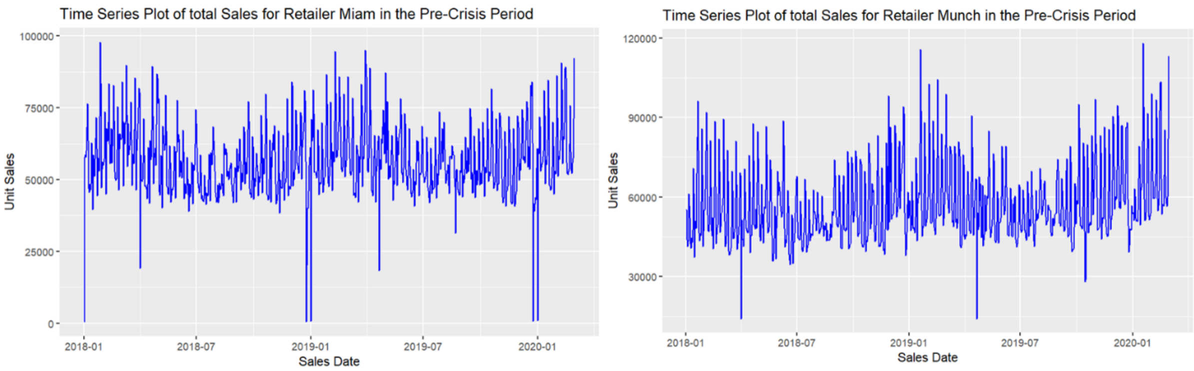


*Figure 1. Time Series Plot of total Sales in the Pre-Crisis Period (ARIMA Output)*

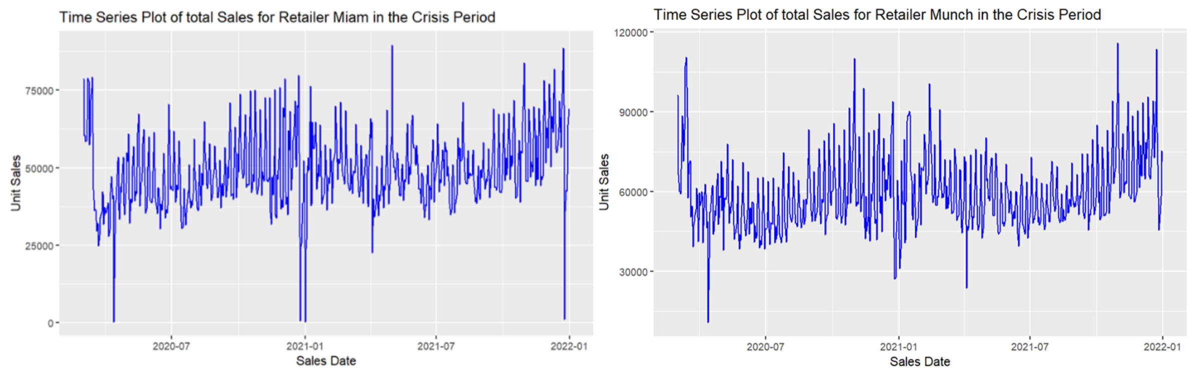


*Figure 2. Time Series Plot of total Sales in the Crisis Period (ARIMA Output)*

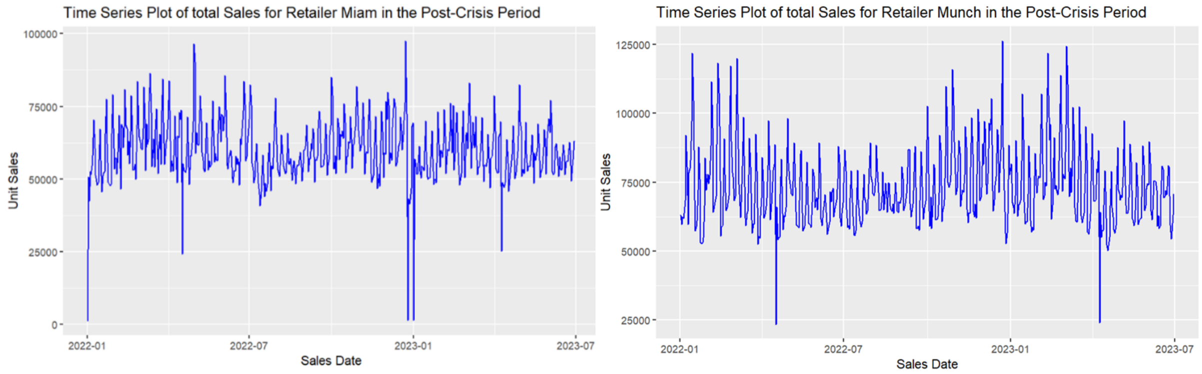


*Figure 3. Time Series Plot of total Sales in the Post-Crisis Period (ARIMA Output)*

In summary, during the pre-crisis period, both retailers depict similar sales patterns, steadily increasing until February 2020, with some fluctuations. At the very beginning of the crisis period (in March 2020, month in which national lockdown was declared), there was a sudden drop in sales due to the national lockdown. However, the market quickly recovered slightly. During the first three semesters of the crisis, sales were lower compared to the pre-crisis period, with Miam experiencing a more significant decline. Despite this, both retailers managed to recover in the latter half of the crisis, with sales volumes reaching similar levels to those before the crisis by December 2021, demonstrating their ability to adapt to challenges. In the post-crisis period, Munch's sales continued to improve and even exceeded pre-crisis numbers, indicating an ongoing recovery. On the other hand, Miam experienced a significant decline, suggesting a setback in its recovery process.

The number of outliers identified for Retailers Miam and Munch across the three forecasting periods is as follows: pre-crisis period (10 for Miam and 9 for Munch), crisis period (7 for Miam and 8 for Munch) and post-crisis period (7 for Miam and 20 for Munch). Outliers were detected by identifying unit sales values within each data subset that fall outside three standard deviations from the average for that subset. The computation of outliers in the Prophet model allowed the detection of peaks and drops in sales (Table 3). In terms of sales peaks, there isn't a particular season that stands out for sales in this category. However, the months with the highest sales tend to be January, October, and December. Conversely, both retailers consistently experience their lowest sales during the Easter holidays across all three periods, except for 2021. This could be due to concerns about a national public holiday potentially leading to limitations or gaps in sales recording. Additionally, analyzing two consecutive periods in 2018 and 2019 revealed the presence of yearly seasonality during the pre-crisis period.

| | Trend Component | Seasonality Component | Peaks | Drops |
|---|---|---|---|---|
| Pre-Crisis | | | | |
| Miam | Upward trend | Detected | 2018/01/27 | 2018/04/01 and 2019/04/21 (Easter 2018 and 2019) |
| Munch | Upward trend | Detected | 2020/01/18 | |
| Crisis | | | | |
| Miam | Downward trend until April 2020; Upward trend from April 2020. | Detected (*) | 2021/05/01 | 2020/04/12 (Easter 2020) |
| Munch | Downward trend until July 2020; Upward trend from July 2020. | Detected (*) | 2021/10/30 | |
| Post-Crisis | | | | |
| Miam | Downward trend | Detected (*) | 2022/12/23 | 2022/04/17 and 2023/04/09 (Easter 2022 and 2023) |
| Munch | Upward trend | Detected (*) | | |

*Table 3. Trend and seasonality components, peaks and drops of sales across pre-crisis, crisis, and post-crisis periods. (*) Seasonality detected only in the Prophet model.*

## 4.2 Time Series Analysis: Pre-Crisis, Crisis, and Post-Crisis

The next step was assessing the performance of the two time series forecasting models, ARIMA and Prophet, across the three periods. Table 4 shows the outputs for two ARIMA models, with the first returned by the *auto.arima()* function and the second representing the best model selected with the lowest information criterion. For the pre-crisis period, ARIMA(2,1,1) was allocated to Miam, and ARIMA(3,1,5) to Munch. In addition, the accuracy measures presented in Table 5 were computed to compare simulated and observed ARIMA values, with Miam having a lower error margin. The Prophet model was considered more accurate for Miam, as shown by lower precision errors for the MSE, RMSE, and MAE metrics, except for MAPE. ARIMA achieved competitive results when compared to Prophet for both retailers. For the low error measures, both ARIMA models from the retailer's data set performed best for the pre-crisis period. However, it is important to note that the 'Inf' output for MAPE in the ARIMA suggests that in some cases, actual values might be zero, resulting in division by zero and an infinite metric result. This can lead to potentially unrealistic metrics in ARIMA, while the Prophet model has higher error values, but it provides finite and more interpretable error metrics.

For the crisis period, results indicate assigning ARIMA(4,1,1) to Miam and ARIMA(2,1,3) to Munch. In addition, the accuracy measures presented in Table 5 indicate that Munch has a lower error margin. In contrast, Prophet Model was perceived as more accurate model for Miam, as evidenced by lower precision errors for the MSE, RMSE and MAE. Once again, the ARIMA models proved to be the best predictive models of the crisis period.

For the post-crisis period, the results led to assign ARIMA(3,1,1) to Miam and ARIMA(1,1,2) to Munch. In addition, the accuracy measures presented in Table 5 indicate that Munch has a lower error margin. The Prophet model was considered to be more accurate for Miam, with lower precision errors for the MSE and RMSE metrics, except for MAE and MAPE. However, the difference in performance between the Prophet models was narrower. Table 5 summarizes the results of the models showing that, similarly to the other periods, ARIMA models are the most competitive choice for predicting the post-crisis period.

The data set used to select the ARIMA models was a subset of the period's data set. For robustness check, we tested a different data set split for the pre-crisis period. We used the same training set with the first 500000 observations from January 2018 and a test set with the first 200000 observations from August 2019 (2019/08/01 – 2019/08/09 for Miam and 2019/08/01 – 2019/08/10 for Munch). Despite this adjustment, we obtained a similar information criterion, although the models showed slightly lower error metrics. These results indicate that the data subset used for model computation does not impact the ARIMA model selection process.

| Pre-Crisis | *auto.arima*(1,1,1)/ *auto.arima*(2,1,5) | | ARIMA(2,1,1)/ ARIMA(3,1,5) | |
|---|---|---|---|---|
| | AIC | BIC | AIC | BIC |
| Miam | 1023795 | 1023828 | 1023398 | 1023443 |
| Munch | 1132962 | 1133051 | 1132838 | 1132938 |
| Crisis | *auto.arima*(3,1,1)/ *auto.arima*(2,1,2) | | ARIMA(4,1,1)/ ARIMA(2,1,3) | |
| | AIC | BIC | AIC | BIC |
| Miam | 1193570 | 1193626 | 1193528 | 1193595 |
| Munch | 1259780 | 1259835 | 1259617 | 1259684 |
| Post-Crisis | *auto.arima*(1,1,3)/ *auto.arima*(2,1,2) | | ARIMA(3,1,1)/ ARIMA(1,1,2) | |
| | AIC | BIC | AIC | BIC |
| Miam | 915526.9 | 915582.5 | 915475.4 | 915531 |
| Munch | 800904.1 | 800959.7 | 800618.2 | 800662.7 |

*Table 4. Information criterion for the ARIMA models in the pre-crisis, crisis, and post-crisis periods*

| Pre-Crisis | ARIMA(2,1,1)/ARIMA(3,1,5) | | | | Prophet | | | |
|---|---|---|---|---|---|---|---|---|
| | MSE | RMSE | MAE | MAPE | MSE | RMSE | MAE | MAPE |
| Miam | -0.0193 | 0.6597 | 0.4345 | Inf | 88480474 | 9406.406 | 6312.42 | 86.0198 |
| Munch | 0.0149 | 0.7819 | 0.4732 | Inf | 239240104 | 15467.39 | 10851.28 | 16.6182 |
| Crisis | ARIMA(4,1,1)/ARIMA(2,1,3) | | | | Prophet | | | |
| | MSE | RMSE | MAE | MAPE | MSE | RMSE | MAE | MAPE |
| Miam | 0.0735 | 0.9881 | 0.5384 | Inf | 100675107 | 10,033.7 | 7,043.55 | 53.76767 |
| Munch | 0.0422 | 0.8021 | 0.5355 | Inf | 131775185 | 11479.34 | 8248.461 | 11.64143 |
| Post-Crisis | ARIMA(3,1,1)/ARIMA(1,1,2) | | | | Prophet | | | |
| | MSE | RMSE | MAE | MAPE | MSE | RMSE | MAE | MAPE |
| Miam | 0.0161 | 0.5975 | 0.4182 | Inf | 52109041 | 7218.659 | 5825.868 | 10.88 |
| Munch | 0.0413 | 0.5233 | 0.3753 | Inf | 65573201 | 8097.728 | 5563.366 | 9.618567 |

*Table 5. Statistical comparison between ARIMA and Prophet models for predicting sales across pre-crisis, crisis, and post-crisis periods*

In summary, ARIMA models represented the best selection based on the lowest information criterion. Consequently, ARIMA(2,1,1) was allocated to Miam, and ARIMA(3,1,5) to Munch for the pre-crisis period. In turn, Prophet model exhibited higher accuracy for Miam, as evidenced by lower precision errors for the MSE, RMSE and MAE metrics, except for MAPE. During the crisis period, ARIMA(4,1,1) was assigned to Miam and ARIMA(2,1,3) to Munch. Similarly, the Prophet demonstrated superior accuracy for Miam. In the post-crisis, ARIMA(3,1,1) was assigned to Miam and ARIMA(1,1,2) to Munch, with Prophet once again showing higher accuracy for Miam. In this case, the performance discrepancy between Prophet models was less pronounced. In summary, ARIMA models consistently emerged as the best predictive models, outperforming Prophet across the three forecasting periods.

## 4.3 366-period Forecasting: July 2023 to June 2024

The previous accuracy metrics results suggest a solid foundation for future projections considering overall model quality. To conclude the analysis, the 366-period projections for ARIMA and Prophet models were performed, covering the period from July 2023 to June 2024 with a 95% confidence level, as shown in Figures 4 and 5 respectively.

Regarding ARIMA forecasts, the sales projections indicate trend stabilization for both retailers. For Miam, the ARIMA(3,1,1) model suggests a slowdown in the downward post-crisis trend, with expected values around 62050 units (95% confidence interval: 19942 to 104157). For Munch, the ARIMA(1,1,2) model points to steady sales after a period of continuous post-crisis growth, setting values around 80462 units (95% confidence interval: 45974 to 114950). In the transition to Prophet, the sales projections align with the post-crisis pattern. For Miam, the model continues to follow the downward trend in sales observed during the recent post-crisis period, with forecasted values around 59778 (95% confidence interval: 49719 to 69832). For Munch, the model suggests continued growth from the post-crisis period, projecting values around 76968 (95% confidence interval: 66098 to 87832). In terms of seasonality, sales exhibit volatile behavior throughout the months. Notably, July and August experience drops, while the highest peaks in category's sales for both retailers occur in November and December.

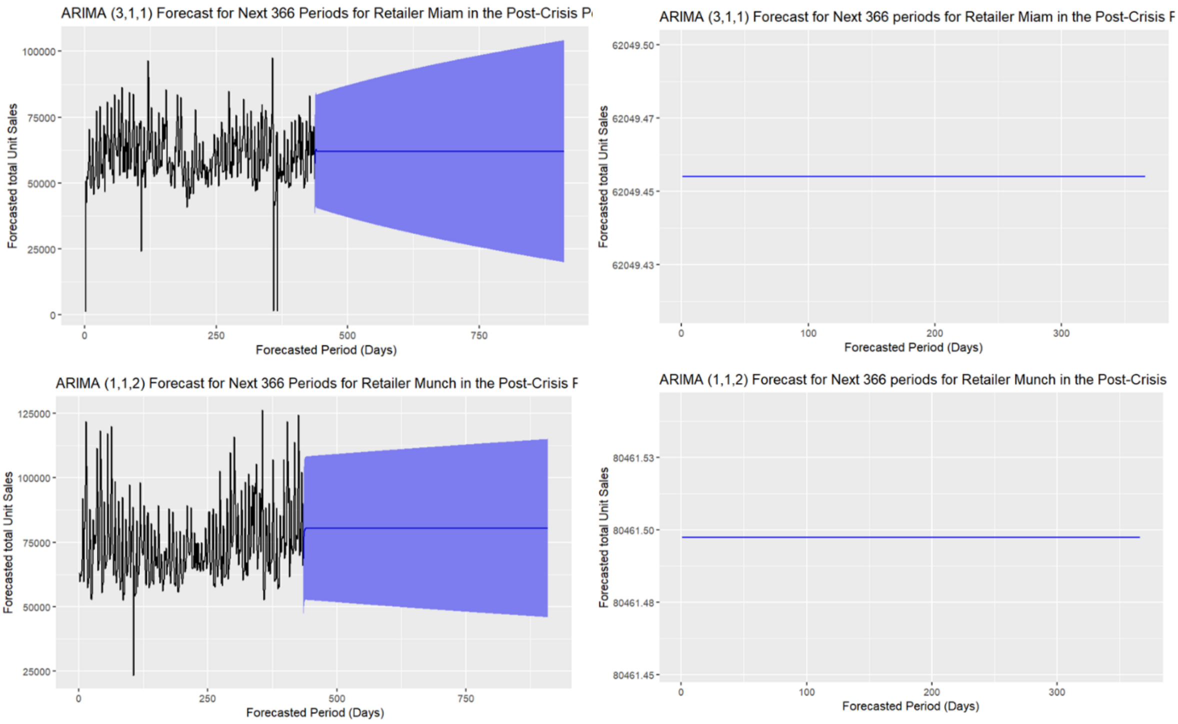


*Figure 4. ARIMA forecast of total sales between July 2023-June 2024 (ARIMA output)*

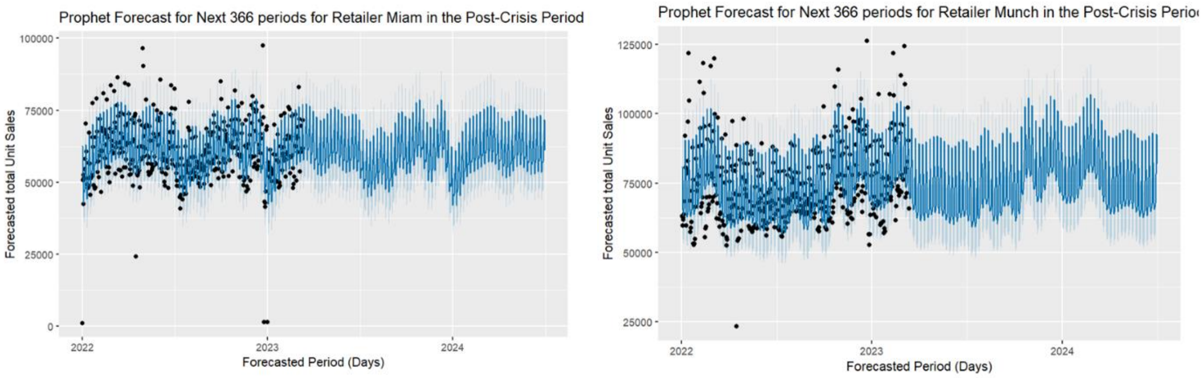


*Figure 5. Prophet forecast of total sales between July 2023-June 2024 (Prophet output)*

# 5 Conclusions, Limitations and Future Research

This study investigated the sales evolution of the sweet snacks category for two major Portuguese retailers over a five-and-a-half-year period, including the COVID-19 crisis period. This work aimed to identify the best forecasting model for three periods under analysis, pre-crisis, crisis, and post-crisis periods. The main findings of our study include the identification of different retailers' sales strategy dynamics and recovery trajectories. Both retailers experienced a recovery in sales at the end of the crisis period. However, while Munch sustained growth, Miam faced a setback in post-crisis period. This divergence in performance can be attributed to the different sales strategies employed by each retailer. Miam opted for a more complex ARIMA model during the crisis period and transitioned to a simpler model in the post-crisis period, changing only the AR component between periods. Conversely, Munch adopted a strategy of continuity since the pre-crisis period, reflected in progressively less complex models. This suggests that Miam's sales strategy was not aligned with the new post-crisis consumption patterns. Additionally, two overall findings were detected regarding the models' performance. In the first place, ARIMA revealed greater accuracy in capturing Munch's sales patterns, while Prophet performed better for Miam throughout the period under analysis. Second, it was found that the post-crisis period returned the highest predictive ability, as evidenced by the lowest error metrics obtained in both models. Lastly, while ARIMA showed superior overall performance based on statistical comparison, Prophet achieved more accurate predictions of the observed sales pattern. Unlike ARIMA, Prophet provided considerably narrower confidence intervals for both retailers, enhancing the reliability of future forecasts and reaffirming its superiority in forecasting Sweet Snacks sales.

The findings of this research offer valuable contributions to both theoretical and practical aspects of business forecasting in the food retail industry. Our research emphasizes a time-series analysis and the investigation of evolution patterns in sales data, going beyond simple statistical analysis and offering a more nuanced understanding of sales dynamics over time. Additionally, it explores seasonality, identifies sales evolution patterns including peaks and drops, pinpoints the specific time periods when these occurred, and projects future sales patterns.

Our study comprehensively analyzes sales changes across various time periods between two main retailers using two time series forecasting models. This complementary approach using two different models allowed us to obtain more comprehensive results. Both models provide different aspects of forecasting, which is particularly relevant considering the impactful differences in the periods of analysis. This approach allowed us to have a broader perspective in our analysis. Together, these aspects form an innovative research framework that enhances our comprehension of the pandemic's impact on consumer purchasing habits and provide a tool for future forecasting in similar contexts. Our investigation makes practical contributions by providing retailers with models to predict future sales of any products within this category during periods of high market disturbance. The findings inform retailers about how sales fluctuations occur and how they recover from these disturbance periods. In turn, this has important consequences for sophisticated strategic decision-making and inventory management.

Despite the important findings, we identify some limitations in our study. Given the large data sets, we faced software performance restrictions, particularly in the R Studio environment, which led us to use limited training and test sets during the model selection process. Nevertheless, we executed several experiments that indicate good overall system performance, as evidenced by the robustness check analysis. Hence, further research could use more robust technological conditions to overcome computational restrictions and train and test larger data sets. Regarding the forecasting models, it is important to acknowledge that external factors impacting sales in the Sweet Snacks category were not incorporated in the forecasting period. Additionally, the study of the online channel sales data would provide valuable insights that complement our findings, given the significant increase in e-commerce during the crisis period. Furthermore, it would be of interest to investigate the Sweet Snacks category of private label brands and compare these results with those obtained for a manufacturer's brand in this research, to follow the growing trend towards private label loyalty witnessed since the pandemic.

References

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control*. John Wiley & Sons.

da Costa, R. L., Miguel, I., Pereira, L., Dias, A., & Gonçalves, R. (2023). Hábitos Alimentares Saudáveis para a Indústria Alimentar em Portugal. *RPER*(66), 133-152.

Dou, Z., Stefanovski, D., Galligan, D., Lindem, M., Rozin, P., Chen, T., & Chao, A. M. (2021). Household food dynamics and food system resilience amid the COVID-19 pandemic: A cross-national comparison of China and the United States. *Frontiers in Sustainable Food Systems*, *4*, 577153.

Duarte, D., Walshaw, C., & Ramesh, N. (2021). A comparison of time-series predictions for healthcare emergency department indicators and the impact of COVID-19. *Applied Sciences*, *11*(8), 3561.

Ensafi, Y., Amin, S. H., Zhang, G., & Shah, B. (2022). Time-series forecasting of seasonal items sales using machine learning–A comparative analysis. *International Journal of Information Management Data Insights*, *2*(1), 100058.

Faghih, N., & Forouharfar, A. (2022). *Socioeconomic Dynamics of the COVID-19 Crisis*. Springer.

Fildes, R., Ma, S., & Kolassa, S. (2022). Retail forecasting: Research and practice. *International Journal of Forecasting*, *38*(4), 1283-1318.

Gomes, S., & Lopes, J. M. (2022). Evolution of the online grocery shopping experience during the COVID-19 pandemic: Empiric study from Portugal. *Journal of Theoretical and Applied Electronic Commerce Research*, *17*(3), 909-923.

Gomes, S., Lopes, J. M., & Oliveira, J. (2023). Online Food Shopping: Determinants and Profile of Portuguese Buyers in the Pandemic Context. *Innovar*, *33*(87), 73-91.

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, *27*(3), 1-22.

INE. (2021). *Instituto Nacional de Estatística - Estatísticas do Comércio : 2020*. https://www.ine.pt/xurl/pub/280814598

Jansen, E., Thapaliya, G., Aghababian, A., Sadler, J., Smith, K., & Carnell, S. (2021). Parental stress, food parenting practices and child snack intake during the COVID-19 pandemic. *Appetite*, *161*, 105119.

Janssen, M., Chang, B. P., Hristov, H., Pravst, I., Profeta, A., & Millard, J. (2021). Changes in food consumption during the COVID-19 pandemic: analysis of consumer survey data from the first lockdown period in Denmark, Germany, and Slovenia. *Frontiers in nutrition*, *8*, 635859.

Makridakis, S., Fry, C., Petropoulos, F., & Spiliotis, E. (2022). The future of forecasting competitions: Design attributes and principles. *INFORMS Journal on Data Science*, *1*(1), 96-113.

McKinsey. (2020). *Consumer sentiment in Portugal during the coronavirus crisis* https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/survey-portuguese-consumer-sentiment-during-the-coronavirus-crisis

Mehta, S., Saxena, T., & Purohit, N. (2020). The new consumer behaviour paradigm amid COVID-19: permanent or transient? *Journal of Health Management*, *22*(2), 291-301.

Pardo-Jaramillo, S., Muñoz-Villamizar, A., & Gomez-Gonzalez, J. E. (2023). Unveiling the influence of COVID-19 on the online retail market: A comprehensive exploration. *Journal of Retailing and Consumer Services*, *75*, 103538.

Pinto, J. P., Veloso, C. M., Sousa, B. B., Valeri, M., Walter, C. E., & Lopes, E. (2022). Managerial practices and (post) pandemic consumption of private labels: Online and offline retail perspective in a Portuguese context. *Sustainability*, *14*(17), 10813.

Ramos, P., Santos, N., & Rebelo, R. (2015). Performance of state space and ARIMA models for consumer retail sales forecasting. *Robotics and Computer-Integrated Manufacturing*, *34*, 151-163.

Roggeveen, A. L., & Sethuraman, R. (2020). How the COVID-19 pandemic may change the world of retailing. *Journal of Retailing*, *96*(2), 169.

Shumway, R. H., Stoffer, D. S., & Stoffer, D. S. (2000). *Time series analysis and its applications* (Vol. 3). Springer.

Smith, K. R., Jansen, E., Thapaliya, G., Aghababian, A. H., Chen, L., Sadler, J. R., & Carnell, S. (2021). The influence of COVID-19-related stress on food motivation. *Appetite*, *163*, 105233.

Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, *72*(1), 37-45.

USDA Foreign Agricultural Service. (2021). *Portugal: The Portuguese food retail sector*. United States Department of Agriculture Foreign Agricultural Service. https://fas.usda.gov/data/portugal-portuguese-food-retail-sector

Yeo, I. K., & Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, *87*(4), 954-959.

Zwanka, R. J. (2022). COVID-19 Impacts to the Supermarket Industry: A framework of major long-term and short-term changes in consumer behavior. *International Journal of Business Marketing and Management*, *7*(1), 1-12.