# Handling Covid-related Disruptions in Time-Series Forecasting

Master Thesis

presented by
Chien-Sheng Liu
Matriculation Number 1725115

submitted to the
Data and Web Science Group
Prof. Dr. Christian Bizer
University of Mannheim

January 2023

# Abstract

The time series forecasting of marketing has become considerably complicated after experiencing the ravages of Covid-19, as the regularity of the time series has been dramatically affected by various government epidemic prevention policies and people's changing consumption behaviours during the pandemic. For many businesses, it is imperative to urgently seek a new forecasting model in the post-epidemic era since forecasts often affect investor sentiment and the development of company strategies. Hence, a comprehensive comparison across Prophet, SARIMA, and Holt-Winters Seasonal Model was discussed in the paper. The robustness of each model to highly destructive time series is tested for different time periods, which are post-pandemic data, data from Jan 2018 to Aug 2022, and data from Jan 2018 to Aug 2022 with the removal of data within the pandemic. The results show that it is critical to identify the cut-off points in the series, as well as the timeframe of the model inputs.

**Keywords:** Time Series, Forecasting, Marketing, Prophet, SARIMA, Holt-Winters

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Time series analysis has been used in business to provide several statistically significant explanations to support effective and opportunity cost minimisation decisions. Time series forecasting models are also of immense influence in business, as it influences not only the planning of a company's strategies but also the evaluation of a company by investors. Marketing-related time series models are often driven by consumers, and the market's current economic situation makes marketing-related forecasting challenging since this type of forecasting often involves uncontrollable external factors, especially after the suffering from Covid-19. The stability of the time series is undoubtedly disrupted by the pandemic, as various uncontrollable factors emerged, such as the closure of cities and the presence of mutated viruses. The epidemic caused such a significant change in human consumption behaviour that the accuracy of the pre-epidemic time series model was reduced. Therefore, this paper uses the example of HelloFresh, a food technology company, to predict the number of new customer activations and utilises various time series data including the high-intensity disruptions during the epidemic as model inputs.

The seasonal forecasting techniques of Prophet, SARIMA and Holt-Winters are selected in this paper for experimental and inter-model comparisons, and their respective strengths and weaknesses are also illustrated. In particular, the impact of the input time-period of the models on their performance is discussed in depth in relation to changes in consumer behaviour patterns. In terms of forecasting, three different periods are addressed, the first of which is to take data from the spring of 2018 to the autumn of 2022 across different consumer patterns as model inputs, achieving by taking the entire period as usual. The second timeframe is the same as above but dragging data in epidemic from series to analyse if the model is robust to the different expenditure patterns. The third duration is when only the post-epidemic data is employed to examine the performance of the models to see whether the three models are robust to shorter time series. Looking at the results from the model inputs, the number of training points in the model had a strong influence. Although the data in the post-epidemic period did not show noticeable highs and lows during the epidemic, the models could only return general trends due to the scarcity of data points. Alternatively, models with the entire time period performed relatively well compared to the 2018-2022 data with the Covid-19 period removed, as the models had difficulty identifying cut-off points in the time series. From a modelling perspective, both Prophet and Holt-Winters seasonal models outperform SARIMA. Prophet's business-capable and customisable features allow the model to recognise changes in patterns in its time series. In contrast, Holt-Winters allows the model to learn recent trends due to the nature of its algorithm, which takes into account and gives weight to each data point.

## 1.1 Problem Statement

Processing time series accurately and effectively has always been challenging for users when forecasting. Particularly in the wake of the Covid-19 explosion, this issue has become one of the most pressing problems for analysts and data scientists to overcome. Undeniably, in the past, it was relatively manageable for researchers and companies to develop forecasting models if the time series of data are regular and without the involvement of numerous external factors. However, over past three years of Covid-19, the time series has been severely disrupted. Several external factors have emerged to disrupt the pattern that should be in place, such as the emergence of mutations that have significantly impacted the sales performance of major retailers.

The number of forecasting models with time series properties is innumerable and SARIMA, for example, is one of the ordinary but persuasive approaches. In recent years, Prophet has also become a contemporary model that is simple, fast and accessible in a wide range of domains, as its properties do not necessarily require advanced statistical skills for the person tuning the parameters. Holt-Winters Seasonal Method, in addition, also plays an essential role due to its exceptional capability of seasonality detection. In addition to the selection and application of time series forecasting models, searching for the best timeframe of input to models is another critical challenge. Over the past three years, the Covid-19 virus has mutated. The government has continuously implemented new and updated anti-epidemic measures, which have dramatically impacted people's lives and time series models. As can be imagined, the introduction of prevention policies has disrupted many time series patterns, resulting in substantial uncertainties when forecasting.

Marketing forecasting is particularly challenging, especially for the retail and food industries, which are particularly associated with people's livelihoods. They are undoubtedly the first to be impacted by the epidemic, as customers' consumption behaviours are no longer the same as before. In this paper, different inputs will be fed and compared with the example of HelloFresh's new customer acquisition in Germany, which surged significantly during Covid-19 but slightly dropped during the increase in vaccination rates and the loosening of preventive measures. Previously, HelloFresh's business performance had relatively consistent seasonal patterns and was comparatively accurate in business forecasts. However, within the Covid-19 and post-Covid periods, the company is now striving to determine the best strategy to resolve these issues under the corrupted time series and make predictions based on these data to optimise marketing strategies.

## 1.2 Literature Review

Time-series forecasting is vital in areas ranging from the marketing industry to the supply chain, from public health to disease control, as it is not only relevant to the cost of events but also decision-making. As a result, many scholars have developed different state-of-art methodologies of time series models, such as Taylor et al [42]., who designed Prophet, Box et al.'s [7] ARIMA prospective, and Holt and Winter's seasonal forecasting method [23] based on exponential smoothing to encode a large number of values from the past. However, each model has its applicable inputs and scenarios, so feature selection, data preprocessing methods, parameter selection, etc. must be considered before making predictions to minimise the possibility of model inaccuracy.

### 1.2.1 Marketing Forecasting

Advertising is inextricably associated with the direct-to-consumer FMCG industry, and the behaviour of consumers as a result of advertising is likewise critical in terms of company revenue. Regarding time series forecasting, it was not used as frequently in marketing in the past as nowadays. Initially, Dekimape and Hanssens [13] expounded on the difficulties encountered in the past application of time series analysis in the marketing field and demonstrated the relevance between time series and marketing. Understanding how customers respond to products is essential. Therefore, Urban et al. [44] presented a model to observe and predict transitions in consumer behaviour by analysing word of mouth, active consumer searches, dealer visits, and forum comments. This approach is similar to HelloFresh's marketing strategies, as HelloFresh relies heavily on online marketing platforms for new customer acquisition, therefore it is critical to track customer behaviour, including customer transitions, online word of mouth and customer funnels. Looking closer from the online marketing perspective, in Wang, Sun and others' [45] literature, a machine learning methodology was employed to predict CTR (click-through rate). The authors used xDeepFM model, a combination of compressed interaction network (CIN) and plain deep neural network DNN [10] proposed by Chen, Zhang and others in 2018 [9], to forecast CTR. The most crucial part of predicting CTR is to acquire the interaction characteristics behind the user's click behaviour [45], such as whether the webpage design corresponds to the user's preference, whether the webpage information attracts the user's interest to click, etc. However, the data contains many interaction messages, making it difficult to determine the order in which each feature was triggered. Therefore, they implemented a machine learning approach allowing the algorithm to capture the priority of triggered clicks automatically. Theoretically, Auc is often used to measure the accuracy of click-through rate prediction, and the results showed that the xDeepFM model performs well. However, this approach strongly relies on the test data sample, with slightly different test data sets that may lead to wildly varying results [30]. Therefore, the ROC curve is adopted here, showing that xDeepFM outperforms the SVM and NN models.

### 1.2.2 State-of-Art Time Series Forecasting Models Comparsion

Marketing forecasts are frequently non-linear, as consumer behaviour is often driven by unforeseen external factors, such as seasonality, marketing campaigns, and government policies, which may lead to substantial changes in purchasing. Hence, after reading the supermarket forecasting paper using Prophet [42], ARIMA [40] and Holt-Winters [46] methods, proposed by Kumar Jha and Pande [28], it contributes many inspirations to the research in this paper. In the experiment, they explored the scalability of each model and compared the accuracy using RMSE and MAPE. Kumar and Pande mention that Prophet's workflow [42] is designed to automate the surface problem [28]. However, analysts are responsible for constantly checking whether the predictions are reasonable. The experiment ended with Prophet achieving better results than the ARIMA and Holt-Winters seasonal models, with the most outstanding MAPE, RMSE and MSE. However, in this experiment, the authors argue that the scalability of Prophet could be improved, especially for large datasets. They proposed that if Prophet is combined with transfer learning methods, Prophet can be made more scalable for large datasets. Nevertheless, the proposal still needs to be verified. On the other hand, ARIMA [40], Prophet and LSTM [22] model comparisons were also adopted by Ning et al. [33] for oil production forecasting. From a resource science perspective, it is challenging to predict the production of unconventional resources due to the inhomogeneity of the sediments and the intricate flow pathways. This is analogous to HelloFresh's difficulties in accurately anticipating changes in the market, as the economic situation is constantly evolving. For example, the sudden appearance of Omicron in 2022 led to HelloFresh's results benefiting from various government policies to prevent epidemics. However, in the same year, the outbreak of war between Ukraine and Russia resulted in a global economic depression. Hence, a comparison is made between the proposed model by extracting significant features from existing time series and applying decreasing curve analysis and reservoir simulation modelling predictions. In this experiment, ARIMA and LSTM models outperform Prophet for oil rate forecasting, presumably as not all oil production data include seasonal effects. Also, Prophet model is capable of capturing seasonally induced production fluctuations, and ARIMA is very robust in predicting oil production rates from wells in unconventional reservoirs.

Arslan [3], on the other hand, proposed a hybrid predictive model based on LSTM [22] and Prophet [42] to forecast energy consumption, which is similar to marketing forecasting due to the non-linear characteristic of the predictor. While statistical methods effectively deal with linear time series, non-linear relationships have always been a complex problem [31]. The model proposed by Arslan [3] reduces the impact of irregular patterns on prediction by training a recurrent neural network with STL-decomposed [11] time series components based on the property that the neural network can simulate non-linear relationships and by stacked bidirectional LSTM [12] models. The process is to feed the LSTM with the trend and residual components derived from the STL decomposition, then re-seasonalise the seasonality obtained from the STL with the LSTM results generated in the preceding step, and incorporate the Prophet outcomes to produce the final prediction. In detail, the authors' model is built on two distinct sub-models, Prophet model, which is designed to preserve the seasonality of the time series, and a neural network model (stacked bidirectional LSTM), which is used for a de-seasonalised version of the data. As a result, seasonal patterns can be addressed better, and the model attempts to reduce the effects of seasonality in the prediction phase. In this concept, the seasonal patterns are derived from the raw data. The extracted seasonal data can be fused when the neural network model is in training, effectively reducing the training time. In addition, the authors also compared the hybrid model with advanced models such as ARIMA [40] and Holt-Winters [23]. The results of the RMSE of the hybrid model also showed a better performance than the Prophet and LSTM, both in terms of MAE and RMSE. The RMSE results are also the best in the hybrid model.

### 1.2.3 Distuptive Time Series Forecasting

On the other hand, uncertainty and interrupted time series are currently the most challenging part of forecasting. The series' stability has been extensively undermined, and the demand estimate has become different than before, especially during the pandemic outbreak. Similarly, forecasting the number of HelloFresh activations has been rendered problematic by the fact that the time series is not as steady as it once was and that the epidemic has brought many severe fluctuations and disruptions to the HelloFresh series. In order to have a more in-depth insight into how to deal with time series fluctuations and how to rectify destroyed time series, the following papers have been examined. In Yenidoğan's article [47], Prophet [42] and ARIMA [40] were used for the future value prediction of Bitcoin. In their experiments, the authors used a non-seasonal ARIMA model to predict Bitcoin. The paper suggests that the most challenging part of the model training is finding the most appropriate ARIMA parameters. Because Bitcoin's trading volume has changed significantly between 2016 and 2018, its time series has as much irregularity as the Covid-19 series. Therefore they chose to use auto.arima [25] to reduce the workload. On the other hand, the authors argue that Prophet's purely auto-forecasting approach is impractical in terms of incorporating meaningful assumptions, implying that the model contributor needs a solid data science background to be capable of making business assumptions. In contrast, the algorithm is robust at coping with missing data and capturing changes in trends and significant outliers. In addition, the authors specifically mention in the paper that the inclusion of additional factors is vital for the predictive model. Therefore, the experiment's Prophet and ARIMA models introduced additional regressors in the comparable bitcoin prices of the Japanese yen, British pound or Euro. The results indicate that Prophet has better results than ARIMA, with the former achieving 94% accuracy compared to 68% for the latter.

#### Time Series Disruption Caused by Covid-19

Furthermore, Covid-19 has significantly affected the forecasting of time series models, regardless of the industry. People's consumption habits have changed gradually with the emergence and fading of the epidemic. It is therefore a very practical issue to explore the extent to which Covid-19 affects time series and how accurate predictions should be performed. Silva et al. in 2022 [37], conduct a dynamic demand for hospital emergency departments throughout the year 2020 using Prophet [42], SARIMA [40] and random forest [8] methods. Due to the pandemic, the need for emergency departments has increased significantly. The strengths and weaknesses of the three models are described in the paper. The authors assume that government pandemic restrictions further reduce the number of emergency department visits, thus reducing the accuracy of the models as uncontrollable factors were added. They also suggest that if any of the components of the time series, i.e. trends, seasons and residuals, are disrupted, then the prediction may be subject to significant errors. In the experiment, the authors examined the stationarity of the data and recalibrated the parameters of SARIMA by means of the Augmented Dickey-Fuller test (ADF test) [14] in a stepwise manner. In other words, the data from 2016 were used initially, then the data from 2017 were included, followed by each year up to 2020, and the change in data stationarity was checked by this method. In contrast, the parameters of Prophet and Random Forest remain the same from year to year. SARIMA was chosen as the subject of this experiment because the authors needed a more robust differential model for non-linear forecasting, which is the same reason for using the SARIMA model in this paper [37]. In terms of results, Prophet was consistently the most potent model when the data on the epidemic were not combined. In contrast, when the model encountered structural changes in the data, Prophet's performance dropped from first to last place in an instant. However, SARIMA took first place at this point, although it had previously been in second place and was only about 1% behind Prophet's MAPE. On the other hand, Dr Gaur et al. [16] proposed a Covid-19 confirmed case trend prediction model based on ARIMA [40] and Prophet [42] during the epidemic. The paper compared the method's performance with a general prediction model based on linear regression and used the epidemic in India as the prediction target. This experiment was conducted at the beginning of the epidemic, in the first half of 2020. Although the author stated that ARIMA's expected results eventually returned a figure that did not differ much from the real world, he mentioned at the time that the epidemic would not have had much impact on the world if the epidemic prevention measures had remained unchanged [16]. However, looking at the results, we can see that the evolution of the epidemic was mighty from late 2020 to 2021. Therefore, after two years, the accuracy of the forecast using the ARIMA model alone has yet to be confirmed. This is also why the SARIMA model was selected for this study.

In addition to marketing forecasts being affected by the epidemic, transport patterns were also a casualty. The government's lockdown policy due to Covid-19 has curtailed the possibility of travel and commuting. While this measure was effective in preventing the spread of the infection, it also posed a challenge for transport authorities in planning construction. As a result, Tsai, Chen, and others [43] presented a multivariate long short-term memory model (LSTM) [22] to predict network-wide traffic situations under disturbance during the outbreak of Covid-19. At the same time, Smith et al. [38] also employed many different models, including linear regression and deep neural networks (DNN) [10], to predict public transportation demand. In addition to collecting many Covid-related data as potential predictors, such as the number of negative and positive Covid-19 tests, the number of confirmed cases and so on, they also used specific time units as possible predictors, like the day of the week, month, etc. From Smith's experiments, it is evident that incorporating other external epidemic-related factors into the model is desirable. Thus, the experiment to predict the number of activations of HelloFresh will also integrate data related to the epidemic, such as daily confirmations and deaths, into the experiment. Afterwards, they explored regularised linear regression, created a seven-layer neural network called Adam Optimiser [38], and constructed a model with a multi-layer stacking strategy, AutoGluon. When it comes to neural networks, CNN models [15] are integral to the prediction of time series. Convolutional neural networks also achieved a satisfactory outcome on the prediction of destructive time series in the power prediction model during the epidemic. Atik [4] fed hourly electricity consumption data as the model's input, decomposed these data into smaller units using the empirical mode decomposition method to extract deep features, and transformed these features into two-dimensional feature maps to build a CNN model. The proposed model has experimented with electricity consumption in Turkey, and a desirable upshot was acquired. Regarding results, the implementation of both LSTM and neural network models returned promising performance and impressive accuracy. Thus, processing highly destructive patterns using neural networks would help the algorithm learn each series's variation and capture the seasonal fluctuations more effectively.

### 1.2.4 Model Selection for Paper

Machine learning methods such as LSTM, and various NN approaches are commonly used as models for marketing prediction. According to the above statements, these methods can be adapted to the complexity and variability of marketing time. As mentioned in the previous section, marketing forecasts are often non-linear, as the oscillations in the trend are often caused by exogenous. When it comes to the performance of LSTM, Lemus argue that LSTMs work well in volatile time series because of their inherent ability to quickly adapt to sharp changes in trends [29], but it requires a substantial amount of data to train it correctly as it is a neural network. Also, LSTM itself has a relatively complex model structure and is more time-consuming to train than other NNs [1], but it, in contrast, can handle volatile time series with uncertainty very well, compared to ARIMA [12]. On the other hand, with reference to SARIMA (or ARIMA), it appears from many arguments that the model is unable to accommodate abrupt changes in the time series [29]. However, it is also believed that the model can be used to make the series more stationary using differential methods, which can make the forecasts more robust to fluctuations in the trend [37]. In Silva's experiments, she found that SARIMA has better predictive capability than Prophet when it encounters abrupt cut-offs, as SARIMA can choose the best parameters by looking at the ACF and PACF. This argument is contrary to the one put forward by Lemua [29]. Therefore, this paper has chosen to adopt SARIMA as one of the experimental models in the hope of understanding the properties of SARIMA for marketing forecasting. Furthermore, Prophet's performance has been subject to very diverse conclusions in various experiments. In many works of literature, many scholars are sceptical about the performance of Prophet's fully automated prediction [28], but this comparatively also makes it faster and easier to predict and meet business needs [42]. Finally, when it comes to Holt-Winters, the algorithm is relatively simple to train because its moving average and exponential smoothing based properties prevent the model from being as time-consuming to train as SARIMA or Prophet. However, the limitation of the algorithm is that if the amount of data for the multiplicative model is too small, the accuracy distortion can be significant. For example, a timeframe with a data point of 10 or 1 might actually differ by 9, but the relative difference is about 1000%, so seasonality expressed as a relative term can vary dramatically [39].

**Why Prophet, SARIMA, and Holt-Winters?**

In the end, SARIMA, Prophet, and the Holt-Winters seasonal model were selected as the subject of the paper. The reason for choosing SARIMA is that, as mentioned above, the conclusions drawn from the experiments conducted by various scholars are pretty different. The ability of SARIMA to customise all parameters in a statistical sense may be practical for destructive time series, so incorporating SARIMA into the experiment will not only allow us to understand the fluctuations in the time series but will also enable us to find out the reasons for the success or failure of the prediction. Regarding Prophet, as the authors suggest, the model is suitable for business forecasting [42]. Prophet's strong time-series adaptability may be of immense help in marketing forecasting. Since the time series of HelloFresh has a great deal of seasonality and external effects, such as marketing campaigns, city closures, etc., the customisation provided by Prophet should hypothetically be able to cope with the destroyed time series, although Silva's experiments [37] disproved this conclusion. Another reason for using Prophet was to investigate the effectiveness of its automated predictions. The motivation behind this was to see if, as the authors claimed, the user could manipulate it without extensive statistical background or if, as Yenidoğan mentioned [47], overly complex time series would still require an extensive data science background. Finally, the Holt-Winters Seasonal Method was chosen because the model is relatively cheap to train and the learning time of the model is rapid, which fits well with business considerations.

In addition, the moving average feature also justifies itself in marketing terms, as past trends in activation numbers tend to influence the future, which should also be related to the lag effect of advertising.

### Methods of Dealing with Disruptions

The importance of adding external factors is directly mentioned in the articles of Yenidoğan [47] and Silva [37], which is also the most straightforward method besides applying the neural network model. This paper does not consider the inclusion of neural networks because the time cost of training neural network models is relatively high. HelloFresh requires finding a relatively simple and fast way to predict the number of activations. Arslan [3], Tsai [43], or LSTM, hybrid models, and other neural models proposed by others look reliable and worth to experiment, but these models need to consume more time. Overall, the paper will aim to get insights into comparing the performance between the three models and the behaviours of adding additional regressors when training the model if the input series has been disrupted.

# Chapter 2

# Theoretical Framework

This chapter presents the theoretical foundations applied to the HelloFresh activation forecasting model, beginning with business scenarios, followed by evaluation methods and experimental models. There are various model evaluation methods, but not all of them apply to the time series and HelloFresh scenarios. Hence, this chapter compares in detail some of the commonly applied model evaluation methods and explains why MAPE and RMSE were selected. On the other hand, the theoretical basis of the Prophet, Holt-Winters seasonal and SARIMA algorithms are also explained thoroughly in this chapter to facilitate understanding the reasons for selecting these three models.

## 2.1 Data Sources and Business Scenarios

Since this paper is in cooperation with HelloFresh, a leading company in the food technology industry, the experiment in this paper will focus on predicting the marketing activations the number of HelloFresh Germany and conduct an empirical study on the estimation of the number of marketing conversions based on the methodologies as mentioned earlier. As for the data on Covid-19, this paper used the data on the daily number of newly reported Covid-19 cases, deaths, tests, and quarantines by the EU/EEA country database from the European Center for Disease Prevention and Control, and vaccination rate data from Robert Koch Institute.

### 2.1.1 Business Scenarios

HelloFresh delivers consumers with cooking boxes containing prepared ingredients and recipes, which customers can acquire through a subscription model. HelloFresh's marketing penetrates the lives of all people, whether it is YouTube, Facebook, Instagram or Tiktok, and other major social platforms can find traces of HelloFresh advertisements. Since HelloFresh adopts a subscription-based business model, potential customers must become HelloFresh members to obtain HelloFresh products, and this series of processes is called *Conversions*. *Conversions* can be further divided into *Activations* and *Reactivations*. The first signifies that the customers are signing up for HelloFresh's service for the first time in their life, and the second implies that customers had subscribed before but cancelled it and now subscribe to the service again.Budgeting and new customer acquisition are usually highly correlated as the more money spent on campaigns, the more potential customers will be reached by advertising, and therefore the more new acquisition is likely to occur. During the epidemic, HelloFresh boosted its budget considerably compared to 2018 and 2019. However, during the epidemic, there were excessive external factors contributing to the activation pattern, and thus the impact of the marketing budget was less significant than other factors, such as the number of confirmed cases per day. Consequently, the budget is not directly taken into account in the experiment. Alternatively, the *Holidays* function is provided by Prophet to include several customised effects, such as Easter holidays, winter and summer holidays, and so on. In the experiment, the time period of HelloFresh's marketing campaigns was also accommodated in the model using this predefined method. As each campaign has a substantial budget, using Prophet's tailor-made feature enables the budget to be taken into account indirectly. *Activations* is the emphasis of this paper and the number that this paper intends to predict, as HelloFresh's business is constantly expanding, and there are a lot of innovative and new marketing campaigns spreading online every quarter. Hence, forecasting *Activations* is an urgent issue for the company to address in the current situation.

**Definition 1 (Activations)** *An activation means that a customer signed up for HelloFresh and bought its products for the first time.*

## 2.2 Evaluation Standards

After deriving the prediction results from the models, the next step is calculating the prediction error to evaluate the model performance and carry out model comparisons. In the article *'When is forecast accuracy important in the retail industry? Effect of key product parameters'*, Belt [6] argues that considering the bia-variance decomposition of the error measure is inevitable. Because inaccurate model assumptions cause bias in forecasting, this may prevent the model from capturing the actual pattern of the data, resulting in a reduction in accuracy, known as selection bias [21]. The primary reason for this is that the data samples selected for model training need to be sufficiently representative of the distribution of each series in real-world situations. Therefore, forecast bias is calculated using symbols rather than absolute error, which makes it indicative only of the direction of the forecast error, either positive or negative [21]. The mean error (ME) is defined in Equation 2.1 to assess the scale-dependent prediction error, where $y_i$ indicates the actual value, $\hat{y}_i$ represents the predicted outcome, and N is the number of evaluations.

$$ME = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i) \tag{2.1}$$

The MSE and MAE techniques of Equation 2.2 and 2.3 are evaluation methods extended from ME, but these two methods are more appropriate for assessing general regression models.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{2.2}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{2.3}$$

Over the years, many scholars have been researching and continuously proposing different error evaluation methods, but there still needs to be consolidated answers to address the issue of comparing across series. The main reason is that discovering a method suitable for cross-series is incredibly cumbersome. In order to achieve this purpose, scaling the series has become the most reasonable solution at present. However, it is still challenging to develop a scaling procedure for non-stationary or non-normality [21]. From Svetunkov's [41] book, although there are many different evaluation methods, the series in the real world is distinctive and may encounter specific conditions that make the proposed error method fail. For example, the time series data of a fast-growing food technology industry like HelloFresh often have vital trend components. Hence simply using MAE or MSE to calculate the error is not practical, as the methods mentioned-above need more time-series considerations. The data being statistically and practically robust is the key to choosing an excellent specific error measure, such as RMSE. However, after reviewing many relevant marketing forecast papers, a phenomenon raised is the same as that Hewamalage [21] explained in the article. Although many researchers are evaluating a specific series with some particular error measures and have given many suggestions and insights, the background and preconditions of these detailed series are often overlooked [21].

### 2.2.1 Category of Error Measurements

Error measures can be classified into Scale-dependent Measures and Percentage Errors Measures according to Hyndman's [26] paper. The former comprises popular methods like MAE, MSE mentioned-above, RMSE, RMdSE and MdAE. Such measures are more sensitive to outliers and therefore some scholars, such as Armstrong [2], have suggested that this type of measures should be discouraged from being used as an evaluation of forecast accuracy. However, historically, RMSE and MSE have been very well-received, mostly on account of their theoretical relevance in statistical modelling [26]. On the other hand, percentage error methods are frequently used for comparing the forecast performance of different data sets because of their scale-independent advantages. Common methods include MAPE, MdAPE, and RMSPE.

#### Error Measurements Based on Scale-dependent

1. Root Mean Squared Error

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \tag{2.4}$$

2. Root Median Squared Error

$$RMdSE = \sqrt{median(y_i - \hat{y}_i)^2} \tag{2.5}$$

3. Median Absolute Error

$$MdAE = median(|y_i - \hat{y}_i|) \tag{2.6}$$

#### Error Measurements Based on Percentage

1. Mean Absolute Percentage Error

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{Actual_i - Forecasted_i}{Actual_i}\right| \tag{2.7}$$

2. Median Absolute Percentage Error

$$MdAPE = median(\left|\frac{Actual_i - Forecasted_i}{Actual_i}\right|) \tag{2.8}$$

3. Root Mean Square Percentage Error

$$RMSPE = \frac{100\%}{n}\sum_{i=1}^{n}(\frac{Actual_i - Forecasted_i}{Actual_i})^2 \tag{2.9}$$

### 2.2.2 Choosing RMSE and MAPE for Evaluation Metrics

One way to assess how well a time series model fits a dataset is to calculate the root mean square error (RMSE), which represents the average distance between the model's predicted values and the actual values in the dataset. It means that RMSE is making a prediction of the standard deviation of the error or residual, which represents how the data are distributed around the line of best fit. The lower the RMSE, the better a given model fits the data set. The reason for applying RMSE in the experiments is that choosing the best-performing model among the different trained models on a given dataset is simply a matter of comparing the RMSE values of all models and selecting the model with the lowest RMSE value. At the same time, the RMSE also has the same proportion of valuable attributes as the target variable. It is, therefore, exceptionally straightforward to understand.

The Mean Absolute Percentage Error (MAPE) is the average of all absolute percentage errors between the predicted and actual values. This metric is practical for determining the performance of forecasting models as it returns errors as a percentage, making it understandable to users and allowing easy comparison of model accuracy across various scenarios and datasets. As the experiments in this paper were tested and compared over several different time periods and models, it is reasonable to consider the use of MAPE as a cross-model comparison of performance. To calculate this metric, the user first calculates the difference between the actual value and the predicted value, divides this by the actual value, and the result is averaged to give the MAPE; for example, if the resulting MAPE is 5%, then the average of the difference between the predicted and the actual target value is 5%.

In addition, from a statistical point of view, this paper aims to predict the number of new customer conversions using different series from time period-wise and model-wise. Therefore, using a measure with a root mean square for the error prediction is more appropriate for this scenario. RMSE method is an evolution of MSE, which, as mentioned above, is more suitable for general regression models, and RMSE is more reasonable for comparisons across models with same input. Since this experiment uses different models for testing, the use of RMSE facilitates the comparison of the experimental models. As for MAPE, this indicator can easily present the influence of the error value on the actual data as a percentage and compare the performance across series and models. Since this experiment tested the model inputs for three different time periods on three different models, it is easier to make fair comparisons using a percentage-normalized metric such as MAPE. This study's objective is to determine the extent to which the error affects the actual set. Calculating the proportion of the error at each prediction point to the actual data is effective in helping users to know whether the model captures different patterns of the time series well, which is particularly important for time series with high-intensity variations.

### 2.2.3 Prophet Cross Validation

Apart from the MAPE and RMSE, Prophet additionally provides a special cross-validation method for time series forecasting, which is performed by choosing cut-off points in history. Only data up to that cut-off point is utilised for each cut-off point to fit the model. This cross-validation procedure is automated using the *cross_validation* function against a series of historical cut-off dates. By specifying a forecast horizon, the size of the initial training period and the interval of the cut-off dates can then be determined. As a default, the initial training period is set to three times the horizon, with a truncation every half of the horizon.

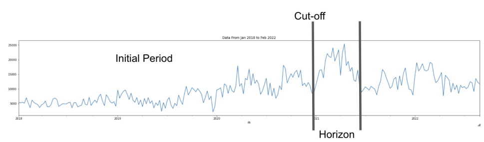


Figure 2.1: Prophet CrossValidation

## 2.3 Time Series Decomposition

From the perspective of time series properties, time series signals can be divided into three components: seasonality, trend and residual. The decomposition of STL is based on the seasonal trend decomposition process of LOESS, whereby the time series can be split into seasonal, trend and residual components [11]. Seasonal time series is the type of time series that most business fields want to explore in depth because it profoundly impacts the forecast accuracy of all professional fields, such as marketing.

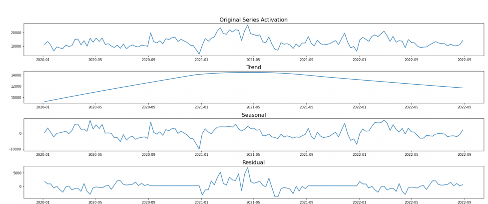


Figure 2.2: STL Decomposition

## 2.4 Forecasting Methods

### 2.4.1 Prophet

In the case of features that are not firmly time-dependent, disregarding the effect of time on the variables is a legitimate option to conserve time without compromising the accuracy of models [42], especially in a business environment where cost and efficiency are imperative. In numerous companies, the calibration and creation of forecasting models are commonly carried out by analysts who need sophisticated knowledge of time series statistics. Under general circumstances, the models can still be expected to have reasonable accuracy, as business frequently drives the numerical decisions in these conditions. However, after the ravages of Covid-19, time-based forecasting models are gradually surfacing in significant businesses, as Covid-19 has almost destroyed the high stability of time series that many companies had in the past.

As mentioned by Taylor and Sean et al. in Forecasting at scale [42], many firms were performing a large number of forecasts by people who might not have been well-trained in time series methods, and therefore a new model was proposed. The new model has sufficient flexibility to accommodate a wide range of business time series, and the individuals calibrating models do not necessarily require adequate knowledge of time series, nor do they need to be experts in the data field. As mentioned above, parameter tuning in Prophet is more straightforward than in other models. It supports the user in making adjustments without knowing the details of the underlying model, which is beneficial if the model needs to be adapted quickly and accurately in an efficient manner.

Based on Peters and Harvey's decomposable time series model [20], Prophet is built on this foundation. The components of the decomposable model include a trend function for non-periodic variability in time series values, a period function for regular changes (e.g. weekly and monthly seasonality), a holiday function for potentially irregular times of day or days, and finally, an error function representing any individual variability to which the model is not adapted. The formula of decomposable time series can be written as:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \tag{2.10}$$

As above formula, trend, period, holiday function, and errors are represented sequentially, respectively.

#### Saturating Forecasts

Typically, linear and logistic growth trend models are the two forecasting methodologies used by Prophet. In forecast growth, the total carrying capacity is paramount as the forecast must fall within this saturation range. For example, the size of the national population and the markets, etc., are the maximum potential points for forecasting results, and the model's outcome should be, at most, the maximum capacity point.

### Trend Changepoints

Significant abrupt variations in the trajectory of a time series are inevitable. Prophet's algorithm is designed to detect changepoints by identifying potential changepoints that allow rate changes and subsequently performing an L1 regularization, or sparse prior, on the magnitude of the rate change. As a default, Prophet identifies these changepoints automatically using known dates of business-related events [42] and enables the trend to be adjusted as appropriate. Prophet also supports the parameter $n\_changepoints$ to specify the number of potential changepoints, and by adjusting the regularisation the model can be better calibrated. It is worth noting that only the first 80% of the data will be incorporated into the potential change points, as the model must retain the remainder to preserve flexibility to future trends and to avoid overfitting at the end of the time series. Finally, the strength of the sparse prior can also be changed via $changepoint\_prior\_scale$ to account for over- and under-fitting of the model.

### Seasonality and Holiday Effects

Seasonality is an essential consideration when addressing time series in practice. To resolve business-related forecasting, the food technology industry must frequently consider the holiday effect to differentiate between low and high seasons for implementing different business strategies. Typically, the seasonal effect is recurring, such as during the winter and summer holidays, which are generally low seasons, and this downturn is a cyclical annual occurrence. The Fourier series is used within the Prophet model to moderate the recurrent seasonal influence. In adjusting the time series parameters in Prophet, Taylor et al. [42] found that setting N to 3 is the most appropriate value for weekly data and 10 for annual data. Furthermore, P represents the expected regular period of the time series. For example, if P is 365.25, the series has a period every 365.25 days, which is annual data. Similarly, it can be inferred that if P=7, it is the weekly data.[42].

$$s(t) = \sum_{n=1}^{N} (a_n \cos(\frac{2\pi Nt}{P}) + b_n \sin(\frac{2\pi Nt}{P})) \tag{2.11}$$

Taking an example from the above illustration, considering the series has the weekly seasonal effect, the formula, therefore, turns into below with an N of 3 and P of 7.

$$X(t) = [\cos(\frac{2\pi(1)t}{7}), ...., \sin(\frac{2\pi(3)t}{7})] \tag{2.12}$$

On top of this, the national holidays of the year are typically considered part of the off-season. Therefore it is unavoidable that holidays are accounted for in the model. Beyond regular seasonal effects, Prophet provides national holidays for various countries. Thus, the model can automatically accommodate the effect of these holidays on the time series simply by importing the holiday function. Furthermore, to cater for lagging effects in business models, specific parameters can be utilised to extend the time frame of holidays to match real-world business models. Naturally, multiple different seasonal effects could be presented in the time series. In that case, additional seasonalities are permitted to be introduced into the model, and additional seasonal functions are also possible to be customised for their seasonality to suit the different seasonal impacts on the data in each scenario.

### Additive and Multiplicative Timer Series

As known, time series can be divided into seasonality, trend and error. If a time series results from the summation of the above three attributes, then it is an additive time series, which is the default shape in the Prophet model. In a narrow sense, such a sequence represents a linear time series with a fixed ebb and flow and periodic pattern, such as a weather forecast.

$$y(t) = Trend(t) + Seasonality(t) + Error(t) \tag{2.13}$$

In contrast, if the series is generated by multiplication, it is a multiplicative series, which may be quadratic, higher-order or exponential, which in turn indicates that the ups and downs and periodicity of the time series will fluctuate over time [17], and is a non-linear time series, such as the sales index or stock price of most companies.

$$y(t) = Trend(t) * Seasonality(t) * Error(t) \tag{2.14}$$

The marketing forecasts for the food technology industry are non-linear, as the industry is strongly subject to time effects, such as poorer results in the winter and summer months compared to better results at the start of the school year. More importantly, the original time series of almost all industries has been compromised by Covid-19 over the last three years. Consequently, accurately breaking down the elements of time series is a critical issue nowadays.

### 2.4.2 SARIMA

SARIMA [40] makes predictions based on the past values of the data itself, that is, its lag and lag prediction error. Any time series with patterns and seasonality without random white noise can be modelled using the SARIMA model. Overall, SARIMA has seven parameters that need to be optimised, and the formula could be written as

$$SARIMA(p, d, q) * (P, D, Q) \tag{2.15}$$

and the *m-parameter* for adjusting the seasonally recurrent frequency. Before nailing on these seven parameters, smoothing the series is the crucial first stage, as the term autoregressive in SARIMA denotes that SARIMA is a linear regression model. The purpose of flattening the data is that the linear regression model is most effective when the predictive variables are uncorrelated and independent, as the model uses its lag as a predictor. The most commonly accustomed approach to smoothing the data is to differentiate it. That is, the previous value is subtracted from the current value. SARIMA (seasonal autoregressive moving average model), as its name implies, is a seasonal model composed of an autoregressive model and a moving average model. The p-parameter in SARIMA is derived from the following formula for the autoregressive model:

$$Y_t = a + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + ... + \beta_p Y_{t-p} + \epsilon_1 \tag{2.16}$$

and the q-parameter is inferred from the moving average formula below:

$$Y_t = a + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + ... + \phi_q \epsilon_{t-q} + \epsilon_t \tag{2.17}$$

Whether it is an AR model or a MA model, $Yt$ depends only on the value of its own lag, that is, $Yt$ is a function of $Yt$ lag. Afterwards, ARIMA in SARIMA model is as follows, which is a combination of AR and MA. where the time series is differentiated at least once, if the time series is not stationary. The equation thus becomes:

$$Y_t = a + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + ... + \beta_p Y_{t-p} + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + ... + \phi_q \epsilon_{t-q} \tag{2.18}$$

#### Choosing Parameters

After understanding the formula, identifying appropriate parameters is essential in implementing SARIMA. Adopting inadequate parameters may lead to distortions in the model's forecast, such as over- or under-fitting. Firstly, it is a prerequisite to examining the series for stationarity using the Augmented Dickey-Fuller test. Suppose the p-value of the test is less than the significant level, i.e. P-value less than 0.5. In that case, the series can reject the null hypothesis and infer that the series is stationary since the null hypothesis of the ADF test is that the time series is non-stationary. As mentioned earlier, if the P-value is more significant than 0.5, differencing the series would be a necessity. Nevertheless, if the series is over-differentiated, it may, in turn, harm the model, as the over-differentiated series may still be smooth.

The next step is to determine whether the model requires an AR. The appropriate p-value in SARIMA can be known by inspecting the partial autocorrelation (PACF) plot to find the required AR order. After excluding the effect of some lags, the partial autocorrelation can be imagined as a correlation between the series and its lags. Thus, the PACF transfer directly correlates with the lags and the series. Any autocorrelation in a smooth series can be corrected by adding sufficient AR terms. Consequently, placing the order of the AR term at or above the lag order of the significance interval in the PACF plot is the best strategy. Similar to checking the order of AR on the PACF, the order of the most appropriate MA can be found on the ACF. In other words, MA is technically an error in lagged forecasting.

### 2.4.3 Exponential Smoothing

Exponential smoothing was introduced by Robert G. Brown, who argued that the posture of a time series possesses stability or regularity [18] so that the time series can be reasonably extrapolated downstream. This methodology emphasises that recent past dynamics will continue to some extent into the recent future. Therefore weights can be assigned to highlight the significance of recent data.

Exponential smoothing is a standard method used in short and medium-term forecasting. The simple full-period average method utilises all past data in the time series equally, without omission. The moving average method leaves out more distant data and gives more weight to more recent ones in a weighted moving average, while exponential smoothing combines the strengths of both full-period and moving averages without discarding past data but gives only a diminishing degree of influence, which is, assigning weights that gradually converge to zero as the data move further away [19].

The basic formula for the exponential smoothing method is shown below, where $S_t$ represents the smoothed value at time *t*, $y_t$ is the actual value at time *t*, $S_{t-1}$ is the smoothed value at time *t-1*, and finally *a* indicates the smoothing constant, which ranges from [0,1].

$$S_t = a \cdot y_t + (1-a)S_{t-1} \tag{2.19}$$

It can be seen from this formula that $S_t$ is the weighted arithmetic mean of $y_t$

and $S_{t-1}$. With the change of the value of $\alpha$, it determines the degree of influence of $y_t$ and $S_{t-1}$ on $S_t$. When $\alpha$ is 1, $S_t = y_t$, whereas 0 is taken, $S_t = S_{t-1}$.

$S_t$ has a period-by-period retrospective nature and can be probed up to $St_{t+1}$. During the process, the smoothing constant decreases exponentially, hence the term exponential smoothing. The smoothing constant determines the level of smoothing and how quickly it responds to the difference between the predicted value and the actual result, so the smoothing constant is extremely critical. If the smoothing constant $\alpha$ is closer to 1, the more rapid the influence of the forward actual value on the current smoothed value will be. In contrast, if the smoothing constant $\alpha$ is closer to 0, the influence of the forward actual value on the current smoothed value decreases more slowly. Thus, when the time series is relatively smooth, a larger $\alpha$ is desirable. Conversely, when the time series has large ups and downs, a small $\alpha$ should be extracted so as not to ignore the effect of the forward actual value [18].

### Holt's Winters Seasonal Method

Holt [23], and Winters [46] extended the Holt method mentioned in [23] to facilitate the incorporation of seasonal factors into forecasting models. In addition to a forecast equation, two smoothing equations, a degree equation and a trend equation included in Holt's method, the methods of Holt-Winters' also naturally contain seasonal elements. The smoothing parameter is an additional $\gamma$ parameter to account for seasonal changes, in addition to the $\alpha$ and $\beta$ parameters already available. As shown in the formula below, this approach indicates the frequency of the seasons with an additional parameter of *m*, which means the number of seasons in a year. For example, if m=12, it represents monthly data, and if m=52 indicates weekly data.

From time series, it is known that seasonal changes can usually be classified as additive or multiplicative. When the seasonality does not oscillate discernibly in the time series, the additive model can be applied in this scenario. Conversely, the multiplicative model will be adopted if the series and seasonal variation have a level of upswing or downswing. When using the additive model, the seasonal component is represented as an absolute value at the scale of the observed series. Furthermore, in the level equation, the seasonal adjustment will be performed by deducting the seasonal component from the time series. The seasonality will finally accumulate to approximately zero each year. In contrast to the additive model, the seasonal component of the multiplicative model is specified as a percentage, meaning that the seasonal adjustment is determined by dividing the time series by the seasonal component, and the annual seasonal components finally add up to approximately *m* which is the seasonal frequency.

The equation below describes each component in the additive model's form.

To ensure that the estimates of the seasonal indices used for forecasting are derived from the last year of the sample, the forecasting equation achieves this by taking the integer part of $(h-1)/m$, which is also substituted by *k* in the equation. Subsequently, the level equation indicates the weighted average at time point t between the seasonally adjusted observations $y_t - s_{t-m}$ and the non-seasonal forecast $l_{t-1} + b_{t-1}$. The seasonality equation takes a weighted average of the current seasonal index $(y_t - l_{t-1} - b_{t-1})$. The seasonal index for the same season from last year accommodates seasonal changes, where *m* is the value for the *m* periods before. For example, if the input data is the 45th week of 2022, then an input of *m* as 52 will give us the 45th week of 2021, which can then be used to facilitate subsequent weighted averaging [24].

Forecast Equation:

$$y_{t+h|t} = l_t + hb_t + s_{t+h-m(k+1)} \tag{2.20}$$

Level Equation:

$$l_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1}) \tag{2.21}$$

Trend Equation:

$$b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)b_{t-1} \tag{2.22}$$

Seasaonality Equation:

$$s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m} \tag{2.23}$$

On the other hand, the individual components of the multiplicative model can be represented by the following equations. Essentially, the principles of multiplicative and additive properties are generally identical, the primary distinction depending on whether the series will rise or fall with time.

Forecast Equation:

$$y_{t+h|t} = (l_t + hb_t)s_{t+h-m(k+1)} \tag{2.24}$$

Level Equation:

$$l_t = \alpha\frac{y_t}{s_{t-m}} + (1-\alpha)(l_{t-1} + b_{t-1}) \tag{2.25}$$

Trend Equation:

$$b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)b_{t-1} \tag{2.26}$$

Seasaonality Equation:

$$s_t = \gamma \frac{y_t}{l_{t-1} + b_{t-1}} + (1-\gamma) s_{t-m} \tag{2.27}$$

Likewise, the Holt-Winters' approach allows for trend smoothing using damping in the same manner as the linear approach above, whether it is additive or multiplicative. By introducing damping parameter, the Holt-Winters' technique with decaying trends and multiplicative properties is able to perform more robustly on seasonal data [24].

$$y_{t+h|t} = [l_t + (\phi + \phi^2 + \phi^3 + ... + \phi^h b_t] s_{t+h-m(k+1)} \tag{2.28}$$

$$l_t = \alpha \frac{y_t}{s_{t-m}} + (1-\alpha)(l_{t-1} + \phi b_{t-1}) \tag{2.29}$$

$$b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*) \phi b_{t-1} \tag{2.30}$$

$$s_t = \gamma \frac{y_t}{l_{t-1} + \phi b_{t-1}} + (1-\gamma) s_{t-m} \tag{2.31}$$

# Chapter 3

# Experimental Research

This chapter aims to perform experiments on destructive time series with Prophet, SARIMA and Holt-Winters Seasonal methods. Commencing with an exploration of the data, the chapter explains the properties of the activation numbers, followed by a description of the three different timeframes to be fed into the models. The second part of the chapter proceeds with the experiments for Prophet, Holt-Winters seasonal method and SARIMA, with weekly data for the entire period from January 2018 to August 2022, weekly data for the same duration as above but with the elimination of pandemic data from March 2020 to February 2022, and finally, daily data for the post-epidemic period starting in March 2022 and ending in September of the same year as the training data for the model. It is worth mentioning that Holt-Winters was not carried out the test of the entire period with the removal of Covid-19 data, as the properties of the algorithm do not accommodate this input, which will be explained in detail in chapter 3.3. In addition, Prophet's algorithm provides several objects that can be calibrated to account for business scenarios. Therefore, all Prophet experiments started with the baseline model and superimposed additional functionalities such as seasonality, holidays, change points, and so on, depending on the situation of the model fit.

## 3.1 Data Exploration

### 3.1.1 Activations

As mentioned above, the number of activations is the predicted target of this paper. This subsection takes the input timeframe of the model as the basis and conducts data exploration in two parts, including the complete data from 2018 to Aug 2022 and the data in the post-epidemic period. The activation numbers discussed here are artificial, whereas the components of the time series, such as trend, residuals, and seasonality, are similar to the original data.

### Weekly Data From Jan 2018 to Mar 2022

As can be seen from table below, the average of series is 10080 with a standard deviation of 4762. Overall, the pattern is a low season during Christmas and sum-

| | Jan 2018 to Mar 2022 |
|---|---|
| Count | 243 |
| Mean | 10080 |
| Std. | 4762 |
| Min | 2147 |
| 25% | 6081 |
| 50% | 9673 |
| 75% | 13000 |
| Max | 25378 |

Table 3.1: Description Data From Jan 2018 to Feb 2022

mer vacations every year, especially the Christmas holiday, which is the lowest point of the year. This is in accordance with HelloFresh's business model, as the company's meal box subscriptions are subject to solid seasonal influences yearly. Theoretically, Christmas is the worst time of year for sales as families travel or prepare for Christmas dinner. From another perspective, it is apparent from figure 3.1 that the number of *Activations* has increased dramatically since 2020 with the outbreak of Covid-19. While the amount of *Activations* has been seasonal in the past, it previously hovered around 5000 per week, with the difference between the highest and lowest peaks being around 5000 generally. By spring 2021, however, the peak was around 25000, fluctuating over 15000, indicating a more substantial seasonal effect during Covid-19. This illustrates the challenges that HelloFresh has encountered in forecasting *Activations* with data points from the Covid-19 period, as the entire business pattern differs before and after Covid-19. Furthermore, upon the coming of post-pandemic period , the pattern evolved again. From figure 3.2, the data deviate from the normal distribution with a Skewness of 0.707 and Kurtosis of 0.034. Meanwhile, data from 2020 onwards significantly increases, especially in the spring and summer of 2021. Afterwards, the trend had a downward trend till the end of 2022.

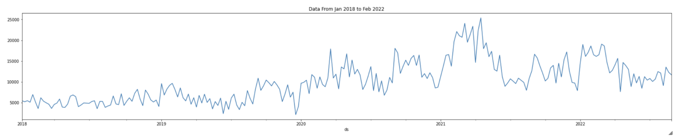


Figure 3.1: Trend of Data From Jan 2018 to Mar 2022

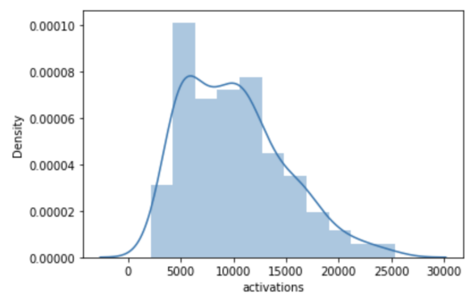


Figure 3.2: Density of Data From Jan 2018 to Mar 2022

### Daily Data From Mar 2022 to Sept 2022

As described in 3.2, the daily data from Mar 2022 to Sept 2022 has an average of 2396 with a standard deviation of 344. On the other hand, the max value of the series is 3313, and the min number is 1560. The reason for choosing daily data here is that there are fewer weekly data points after March 2022, so if the data of this period is used, daily data will be used to provide the algorithm with more basis for prediction. It can be seen from the figure that the number of Activations decreased after March 2022 and continued to fluctuate until August before a slight increase. The seasonality of the time series during this period could be more apparent, only the low point of summer vacation is easily observed.

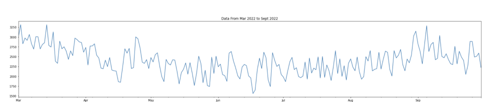

Figure 3.3: Trend of Data From Mar 2022 to Sept 2022

| | Mar 2022 to Sept 2022 |
|---|---|
| Count | 214 |
| Mean | 2396 |
| Std. | 344 |
| Min | 1560 |
| 25% | 2161 |
| 50% | 2386 |
| 75% | 2629 |
| Max | 3313 |

Table 3.2: Description Data From Mar 2022 to Sept 2022

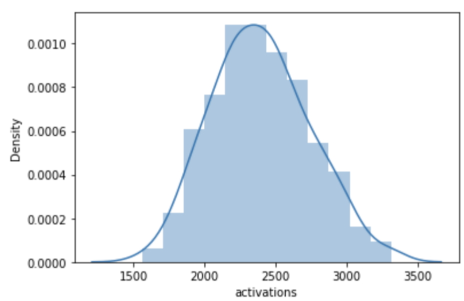


Figure 3.4: Density of Data From Mar 2022 to Sept 2022

### 3.1.2 Covid-19 Data

The Covid-19 outbreak occurred around January 2020, and on the 30th of the same month, the World Health Organisation officially declared the virus a public health emergency of international concern. As for Germany, the outbreak started to be significantly severe in March 2020 and continued until February 2022. Therefore, the time frame for the Covid-19 data presented in this paper starts on 1 March 2020 and ends on 28 February 2022.

As can be seen from the graph 3.5 of the number of confirmed Covid-19 cases, the number of new infections in a single day in 2021 is remarkably lower than in 2022. The end of 2020 to the beginning of 2021 was the first wave of peaks, and the German government implemented a strict city closure policy that prevented people from eating in restaurants. Strict border controls worldwide have also prohibited international movements from taking adequate precautions against the disease. As a result, the number of new HelloFresh users has grown immensely during this period compared to previous years. Immediately after the peak at the beginning of 2021, HelloFresh experienced a sharp drop in the number of new subscribers, but the number of new sign-ups was still higher than ever due to Covid-19. At the end of the same year, HelloFresh also acquired a large volume of new customers due to the Omicron mutation, as the number of new confirmed cases per day kept breaking new records. It was only around the Easter holidays that vaccine policies and border controls progressively eased worldwide, and international mobility gradually resumed. Covid-19 was slowly disconnected from HelloFresh's pattern of customer acquisition as people's lives entered the post-epidemic phase.

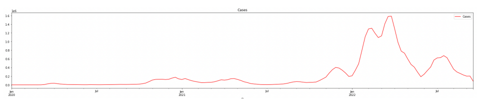

Figure 3.5: Covid Cases

From the perspective of the number of deaths, as shown in 3.6, there was a large number of deaths from the end of 2020 to the beginning of 2021, after which the number continued to decline and remained stable. There was a slight increase until the second half of the same year, then it fell again by the end of the year and fluctuated around. This trend differs considerably from HelloFresh's new customer acquisition pattern, as the decline in deaths is associated with higher vaccination rates, and those who die from Covid-19 are typically severely ill. From a business logic point of view, people diagnosed with Covid-19 have to be self-quarantined for 7-14 days according to current policy, and HelloFresh's home delivery service meets the needs of quarantined patients. Therefore, the number of cases is relatively correlated with HelloFresh's new customer acquisition pattern as opposed to the number of deaths.

In addition to the number of confirmed cases and deaths, vaccination rates per day, daily quatantines, daily recoveries, and daily Covid-19 tests were also factored into the feature engineering in the experiment. However, only the number of confirmed cases per day was comparable to the trend of HelloFresh's Activations. Therefore, in the following experiments, only the external variable of the number of confirmed patients per day was considered. All these data were downloaded from European Centre for Disease Prevention and Control(EU Disease Agency), the German Federal Ministry of Health(Bundesministerium für Gesundheit), and the Robert Koch Institut(RKI).

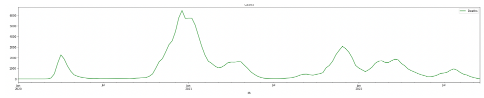


Figure 3.6: Covid Deaths

## 3.2 Prophet

The forecasting methodology with Prophet is performed and presented in this section. Since Covid-19 significantly impacts the time series and the pattern of *Activations*, it is inevitable to experiment with the input of the model with data from different time horizons. Although the outbreak of Covid-19 began in Germany around the end of January 2020, it was not until March of the year that the Covid-19 episode began to influence the time series dramatically. Six months after Omicron, the Covid-19 effect faded in March 2022 as the world gradually regained order with the opening of national borders. Here, three different periods, 2018 to 2022 with Covid, 2018 to 2022 without Covid, and after Covid-19, are presented in the following pages.

### 3.2.1 Data From 2018 Onwards

The first experiment investigated whether removing the Covid-19 data from the time series and performing the forecast gave better prediction results, meaning that the time range of the input data is January 2018 to February 2020 plus March 2022 to August 2022 (because data from Mar 2020 to Feb 2022 is in-pandemic data). In the other case, the complete series from 2018 to August 2022 was considered the input to the model, and the daily cases of Covid-19 in Germany were incorporated as an independent regressor.

#### Without Covid-19 Data

First, the model's input is weekly data of *Activations*. It, therefore, has 139 data points for model training. Prophet does not require users to split the dataset as a train, test, and validation set because Prophet has the functionality for time series cross-validation, which has mentioned above, to assess the quality and accuracy of forecasting errors using historical data. In line with best practise in machine learning, the model was initially fitted without adjusting any manual parameters or adding different seasonality. The results showed an increasing trend in activations year on year, which did not fit the company's business pattern of *Activations*. The trend can see in figure 3.7.

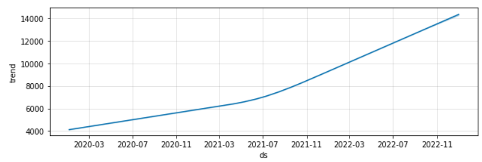


Figure 3.7: Trend: Prophet with No Covid Series - Baseline

Here, cross-validation was performed to evaluate the 180-day forecast performance, commencing with the first cut-off of 540 days of training data, followed by a forecast every 120 days. A total of 70 forecasts were obtained from the cross-validation over this two-year and eight-month time frame. The MAPE and RMSE of these 70 predictions were averaged and used as the evaluation criteria, obtaining 0.335 and 3809.09, respectively. Meanwhile, when comparing the overall historical forecast with the actual value, the MAPE result of 0.22 and RMSE of 1745 are gained.

| | Cross Validation | Historical Forecast |
|---|---|---|
| MAPE | 0.33 | 0.22 |
| RMSE | 3809 | 1745 |

Table 3.3: Evaluation: Prophet with No Covid Series - Baseline

From a business perspective, on- and off-season marketing campaigns and holidays strongly affect the number of *Activations*. For example, people are less inclined to buy HelloFresh during the holidays as HelloFresh offers a service that brings boxes of ingredients to their homes, whereas people tend to travel more at these moments. Additionally, the marketing campaigns of the company each year also play a substantial part in influencing the willingness of customers to purchase, as when the market is full of HelloFresh advertisements, there is a high likelihood that people will be aware of the company and have an impression of it, which in turn will lead to the purchase of the product.

In order to cope with the external factors mentioned above, it is feasible to set up a recurring holiday effect in Prophet manually. In this regard, historical data indicates that the summer and winter holidays and Easter and Christmas holidays constitute the off-season for the business. In contrast, the weeks following the holidays represent the high season. Regarding marketing campaigns, it is tempting to recognise that an extensive marketing campaign generally follows the summer and winter holidays. Furthermore, a series of marketing initiatives at the end of an off-season would typically result in a higher customer reach and *Activations* of new users. In addition, to take into account the lag and advance effects of national holidays, a buffer of 3 days was set before and after the holidays, while 15 days for winter and summer holidays catered to many people's travel plans. Meanwhile, according to research, the most effective time frame for a marketing campaign was typically within 15 days of launch. Therefore setting the lower window at 15 was the most appropriate number to fit the model.


Figure 3.8: Trend: Prophet with No Covid Series - On- Off Seasona and Holiday Effect

It can be recognised from figure 3.8 that the trend was stably increasing. According to these forecasting components, worse results in cross-validation, including holidays and marketing campaigns, were therefore acquired with the RMSE of 4123 and MAPE of 0.36 compared to the baseline model. However, if considering the historical forecasting purely, both evaluation standards all performed better than the baseline model.

Prophet's Cross-validation is not only possible for model soundness assessment but also for tuning the hyperparameters of the model, such as *changepoint_prior_scale*, *holidays_prior_scale* and *seasonality_mode*. During this step, a 5*4*2 grid with the above three parameters was constructed based on the external variables explored above, accompanied by parallelisation in the cut-off values. The parameters here

| | Cross Validation | Historical Forecast |
|---|---|---|
| MAPE | 0.36 | 0.21 |
| RMSE | 4123 | 1622 |

Table 3.4: Trend: Prophet with No Covid Series - Complete Model

were measured based on the average MAPE over 180 days, and the model was fitted with the best set of parameters in this section.

The time series' seasonal component in business data often conveys consumer behaviour over the seasons. In addition to the seasonal holiday effects mentioned above, the pattern of *Activations* also has a quarterly seasonality, as it tends to start with a low point in the number of *Activations* at the beginning of each quarter, gradually ascending, and then declining sharply to a low peak again. Therefore, it was to be expected that setting a different seasonality would be beneficial to the model fit. It is worth mentioning that the seasonality effect of Prophet is an estimation made by using partial Fourier sums. Per Fourier principles, it was assumed that the number of terms in the order (partial sum) of the Fourier was the parameter that determined the rate of change of seasonality. In the observed seasonality, the seasonality of seasons was included in the model fitting process, with a Fourier order of 5. The procedure here was analogous to the abovementioned paragraph in that the grid search was performed to identify the best parameters. However, in this case, a 5*5*4*2 grid was applied as *seasonality_prior_scale* was taken into account here.

The trend in figure 3.9 was diverging from the actual pattern of *Activations*, as the trend continued to climb over time. Meanwhile, concerning figure 3.10, the entire trend in *Activations* can be perceived to be reaching a historical peak in the data set around October 2022. Likewise, the interval in light blue in the chart was relatively broad, implying that there might be a significant deviation between the forecast and actual values. In the model evaluation, as seen from table 3.5, when the Cross-Validation was carried out, there was a slight degradation in all indicators when seasonality was introduced manually, compared to just the holiday and high- and low-season effects. Regarding evaluation data, RMSE went up slightly from 4123 to 4257, whereas MAPE remained constant. However, in terms of historical data, there was a noticeable improvement in measurement, with RMSE dropping from 1622 to 1248.

Using data from 2018 to 2022, combined with removing data points from the Covis-19 period, is not a sensible approach from the above experiments. From a business point of view, HelloFresh's business is growing year-on-year, and the pat-

| | Cross Validation | Historical Forecast |
|---|---|---|
| MAPE | 0.36 | 0.16 |
| RMSE | 4257 | 1248 |

Table 3.5: Evaluation: Prophet with No Covid Series - Complete Model

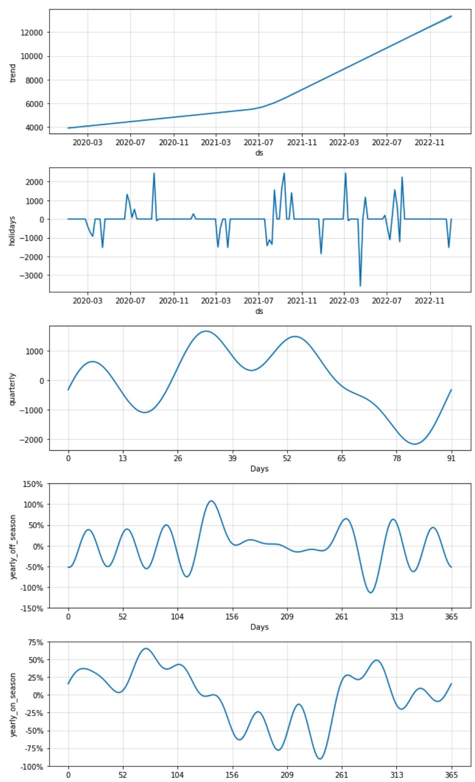


Figure 3.9: Prophet Without Covid-19 Data With All Effects

tern of *Activations* is still affected by the lag of the epidemic. Simply eliminating the data points between the epidemic and combining the data points before and after the epidemic can be expected that there will be a distinct disparity between the data before the epidemic in February 2020 and after the epidemic in March 2022. Additionally, there is no doubt that HelloFresh is one of the beneficiaries of the epidemic disaster, as customer consumption habits have been altered radically due to the strict German epidemic lockdown policy. Germans, who tend to travel on holi-

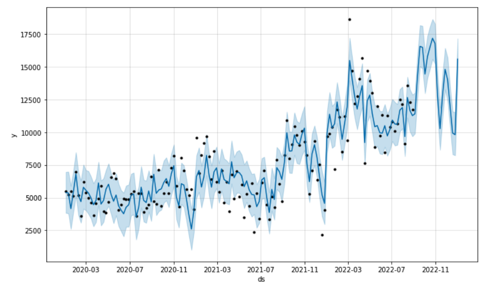


Figure 3.10: Prophet Forecasting without Covid-19 Data with All Effects

days, were unable to go out for meals and excursions under the stringent measures imposed by the government, thus causing a decisive alteration in the time series of *Activations* within the company. Hence, a traditional forecasting experiment with complete data set from Jan. 2018 to Sept. 2022 is conducted here.

### With Covid-19 Data

To improve the model's accuracy, it is worthwhile to incorporate data from the epidemic period into the dataset and to consider the COVID-19 confirmed cases to examine whether this additional variable would enable the model to learn the pattern more accurately over the epidemic period or not. Likely in the previous section, this one also included incremental performance monitoring in line with machine learning best practices to identify the internal and external factors that affect the model. In this context, the most fundamental baseline model was constructed first, and the seasonal model, the model with holiday effects, was constructed in this order. Crucially, as the model incorporated data from the period of the Corona, the number of confirmed cases per week was treated as an external variable for reference. At the same time, consumer behaviour was strongly driven by government control policies. Therefore manually setting the timing of each lockdown was inevitable. The epidemic was also subject to fluctuations depending on the presence of mutations. Hence the setting of changepoints was another feature of the model. By specifying the changepoints, the model can sensibly identify changes in the period, allowing it to forecast more accurately.

Through cross-referencing the actual *Activations* trend in figure 3.11 with the baseline model in figure 3.12, it is straightforward to notice that the two trends have a similar trajectory. Both graphs illustrate that the number of *Activations* started to increase year on year in 2018 and reached a peak in 2021 before gradually declining until the last data point in the dataset. Similar to the above, a cross-validation technique with an initial point of 730 days, a period of 365 days and a horizon of 120 days were implemented to verify the model's accuracy. With 47 results of validations, a MAPE of 0.31 and an RMSE of 3770 were obtained. Compared to the historical forecast, the MAPE was 0.19, and the RMSE was 1985. It is worth mentioning that without adding additional regressors and adjusting parameters, the most basic model incorporating Corona has performed better than the model that removed data from the pandemic period and added other effects.

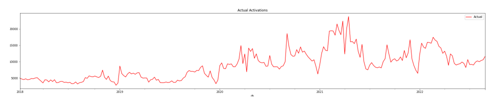

Figure 3.11: Actual Pattern of Activations

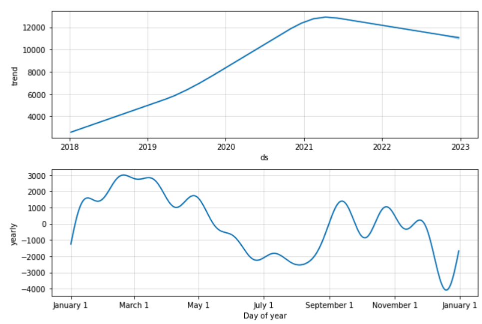


Figure 3.12: Trend: Prophet with Covid Series - Baseline

| | Cross Validation | Historical Forecast |
|---|---|---|
| MAPE | 0.31 | 0.19 |
| RMSE | 3770 | 1985 |

Table 3.6: Evaluation: Prophet with Covid Series - Baseline

Unlike the preceding exercise, seasonal effects were prioritised for exploration in this experiment. As seasonality matters more as a consequence of the pandemic, prioritising the effect of seasonality on the time series is more conducive to the fit of the model. Upon closer scrutiny of Figure 3.13, the first sub-plot reveals the trend in the time series, and it is unsurprising to discover the fact that the *Activations* pattern steadily declined from the beginning of 2021, with a relatively modest decline when the line peaked at the beginning of 2021. The hypothesis that seasonality has a potent effect on this time series also stands, given the cross-validation results, with RMSE falling by 27% to 3000 and MAPE improving by 4%. In addition to the cross-validation outcomes, the indicators in the historical forecasts have also markedly enhanced, thus reinforcing the fact that the inclusion of seasonal variables has resulted in the model being more effectively trained to deliver more accurate forecasts.

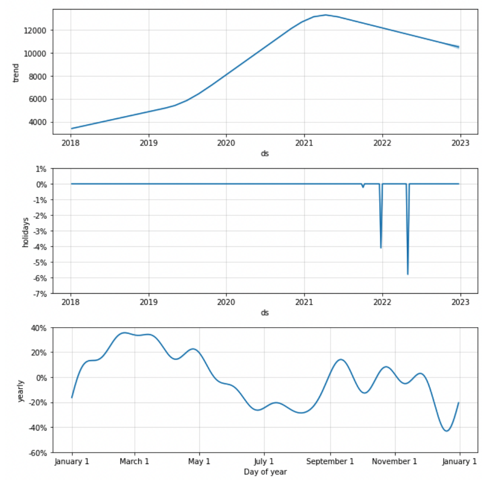


Figure 3.13: Trend: Prophet with Covid Series - Seasonality

| | Cross Validation | Historical Forecast |
|---|---|---|
| MAPE | 0.26 | 0.14 |
| RMSE | 3000 | 1685 |

Table 3.7: Evaluation: Prophet with Covid Series - Seasonality

It is worth mentioning that the holiday effect in figure3.13 has a pronounced reduction around 2022. Presumably, these would be Christmas and Easter, respectively, as the German border was opened in mid-2021, and the effect of retaliatory tourism should not be underestimated. Therefore, manual adjustment of holiday effects using the *holidays* parameter in Prophet is a mandatory requirement. Christmas and Easter are annual recurring holidays, which makes it implausible that the seasonal effects will only be varied around 2022. As such, it is a legitimate reason to introduce a data frame manually with Easter and Christmas dates for each year from 2018 to 2022 into the Prophet object for model fitting. As with the previous experiment, Easter and Christmas would have a three-day buffer day before and after the holiday to allow for the early and lagging effects. The summer and winter holidays are also the off-season for the company. Hence it would be reasonable to incorporate the summer and winter holidays as part of the holiday effect with a residual effect of 30 days. Finally, the model also considered the dates of the company's marketing campaigns to accommodate the corresponding buying sentiment. Furthermore, as the study said the company usually has a 15-day buying stimulus period, a 15-day lag effect was considered here. Moreover, in a nutshell, after September and before May was when the acquisition rate of new customers was relatively high, thereby integrating this phenomenon into the seasonal effect.

Likewise, in figure 3.14, the trend towards 2022 was consistently degraded in this experiment compared to the previous seasonal-only model. However, it is intriguing to note that the holiday effect was ideally taken into account here. What can be observed was that the Easter and Christmas holidays represented a low point in the conversion rate each year, and the latter's impact was increasing every year, which can reasonably be assumed to be a consequence of revenge tourism. According to the table 3.8, a 1% increase of MAPE in over 47 results of Cross-

| | Cross Validation | Historical Forecast |
|---|---|---|
| MAPE | 0.25 | 0.12 |
| RMSE | 2920 | 1622 |

Table 3.8: Trend: Prophet with Covid Series - Seasonality and Holiday Effect

Validation was obtained with the accompaniment of the decrease of 80 in RMSE. In the historical forecast, either evaluation metrics won improvements, especially MAPE, achieving around 12%. This phenomenon illustrates that this model is robust to the historical forecast, but predictive one should be further adjusted by adding other regressors, such as Covid-19 data, and tuning other parameters.

Furthermore, looking into figure 3.15, the light blue interval is significant due to the substantial uncertainty of the forecast. It indicates that the black dots, forecasted points, may reasonably fall between the upper and lower bound, representing the considerable possible bias. Hence, The next model aimed to decrease the uncertainty interval lest the inaccuracy rate rises. Simultaneously, as seen from the same graph, pandemic periods possess many black dots that fall outside the interval. Enforcing them into the certainty area was another crucial part that needs to be tackled. If quarterly seasonality with a Fourier order of 5 was added to the model, MAPE increased by 2% to 24%, and RMSE dropped from 2920 to 2865. An assumption is that quarterly seasonality can occur every quarter and that the campaign will be most effective in reaching customers and stimulating buying within

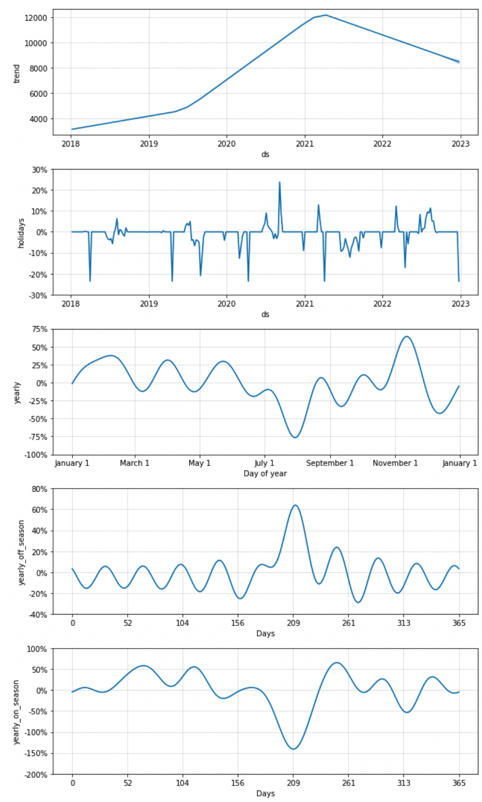


Figure 3.14: Evaluation: Prophet with Covid Series - Seasonality and Holidat Effect

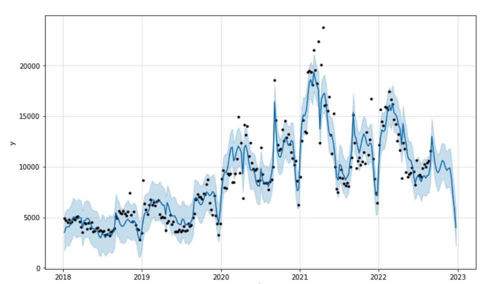


Figure 3.15: Prophet Forecasting with Covid-19 Data with Seasonality and Holidat Effect

two weeks and one month of the campaign launch.

It is well-known that Covid-19 severely impacted consumer *Activations* be-

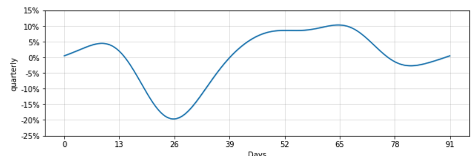


Figure 3.16: Trend: Prophet with Covid Series - Seasonality, Holiday, and Quarterly Effect

haviour. However, with seasonal and holiday adjustments to the parameters, the model still had potential for improvement based on the results returned from the cross-validation. Therefore, this experiment incorporated daily confirmed data on Covid-19 to measure whether it positively affected model fit.

The German authorities implemented various measures during the pandemic to prevent the epidemic. The lockdown policy had the most decisive impact on the number of new customers acquired by the company. It is reasonable to incorporate a customised holiday effect into the series of the lockdown period. According to Prophet's official documentation [42], to accommodate time series changes during the lockdown period, users are allowed to treat each closure as a one-off long holiday, specify a start and end point, and set up a buffer window to address lagging effects. Therefore, the experiment here inherited the basis of the preceding model and incorporated three periods of lockdown or strict measures in Germany as a one-off holiday. In brief, the three periods were, according to the chronology, the first lockdown in spring 2020, followed by a strict lockdown from winter 2020 to summer 2021, and a 2G+ *(3 Vaccinated, Recovered, 2 Vaccination with Rapid Test )* policy with the emergence of Omicron mutation in winter 2021. The reason for including only Omicron is that it has been observed that other mutations of the virus did not exhibit the dramatic spike in time sequences that Omicron did. This is further explained by the fact that Omicron was more infectious than other viruses, but the mortality rate decreased significantly as vaccination rates increased. As a result, the data on the number of confirmed cases suggests that many more people were quarantined at home during this period than in the past. This is a reasonable assumption for the most significant number of *Activations* recorded during this period. Moreover, due to the nature of this mutation mentioned above, the government did not impose a mandatory and strict moratorium on all commercial activities. It prohibited all non-essential social interaction during the period when the virus was spreading furiously. As such, whilst the city was not closed for the winter of 2021, the increase in quarantines meant that the epidemic was considered another one-off holiday. In another part, Prophet enables the detection of change points by assigning a significant number of possible change points for rate changes in advance. Then it applied a sparse prior (equivalent to L1 regularisation) to the magnitude of the rate change, which in essence, implies that Prophet had a wide range of possible places to alter the rate [42]. Hence, this experiment also took into account the parameter *change_point_scale* to adjust the strength of the sparse prior.

First of all, the figure 3.17 below is the result obtained without adding the number of confirmed cases of Covid-19 but adding the lockdown and Omicron period as a one-off holiday and the *change_point*. The *change_points* was first added to the model, and after a 5*5*4*2 grid search, the best model was identified when the *change_point* was 0.5. Meanwhile, the performance of the model assessment scores was noticeably raised. Furthermore, the holiday component could be recognised as reflecting the effects of the Covid-19 closure and Omicron, particularly in the immediate aftermath of the outbreak in 2020 and the emergence of the Omicron mutation in late 2021. In the experiment, figure 3.18 indicates that the black forecasted points almost fell into the light blue uncertainty interval. Surprisingly, the light blue interval became relatively small compared to the above approaches. Although some forecasted points dropped outside the certainty area, the evaluation metrics returned super well-performed numbers with the MAPE of 0.04 and the RMSE of 529 in the historical forecast. However, the results of cross-validation with the initial period of 730 days, a period length of 120 days, and a horizon of 90 days could be further improved with an RMSE of 3760 and a MAPE of 0.26.

| | Cross Validation | Historical Forecast |
|---|---|---|
| MAPE | 0.25 | 0.04 |
| RMSE | 3760 | 529 |

Table 3.9: Evaluation: Prophet with Covid Series - Seasonality, Holidat Effect, Quarterly Effect, and Change Points

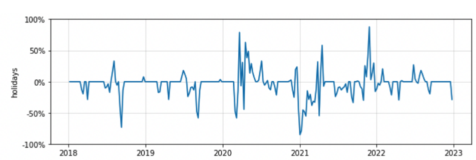


Figure 3.17: Prophet Trend with Lockdown and Omicron

Next, an additional regressor component was built into Prophet by including the number of Covid-19 weekly confirmed cases from March 2020 to February 2022 in the model. As can be noticed in figure 3.19, this regressor illustrates

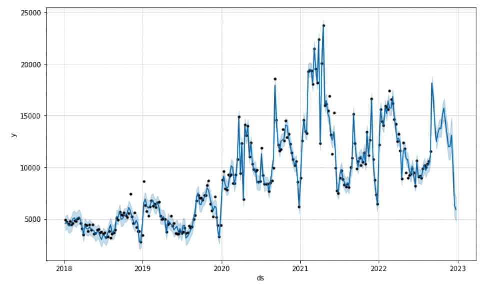

Figure 3.18: Prophet Forecasting with All Covid-19-related Effect, Seasonal and Holiday Effect

the tendency of fluctuations in the pandemic in its entirety, and it can be understood that the regressors were integrated into the model fitting process, allowing the model to learn the patterns of activations over this period more effectively. On

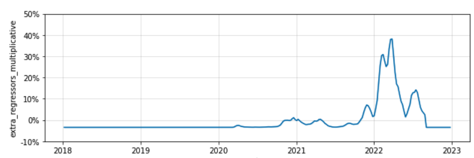

Figure 3.19: Covid-19 Regressor

the other hand, it can be observed from figure 3.20 that when the data on the number of confirmed cases was not added, the results fitted by the model had a higher trend. On the contrary, with the inclusion of data, the upward trend of the model trend is relatively conservative and moderate. This change can be interpreted as the model recognising that the additional increase during the pandemic is a one-off effect rather than a recurring one. Therefore, the model trend tended to flatten when confirmed cases are included. When referring to the model evaluation criteria, using the same settings as above for cross-validation, there was a marginal improvement in both MAPE and RMSE over the model without confirmatory data, from 3760 to 3513 and from 25% to 23%, respectively.

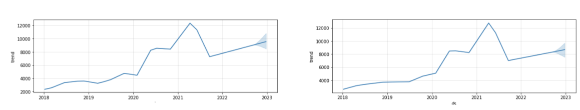


Figure 3.20: Trend Change With(out) Covid-19 Data

| | Cross Validation | Historical Forecast |
|---|---|---|
| MAPE | 0.23 | 0.04 |
| RMSE | 3513 | 525 |

Table 3.10: Evaluation: Prophet with Covid Series - Covid Effects and All Time-Series Effects

### 3.2.2 Post-Covid as Input

After two years of the Covid-19 epidemic, the world has largely subsided. The following experiment presumes that the *Activations* pattern beyond March 2022 was no longer affected by the pandemic. Thus the model inputs for the experiments in this section started in March 2022, and the cut-off point fell on September 30, 2022. Note that the cut-off point for the input data in this experiment is one month later than all the above experiments to provide the model with more clues to fit the real world better. Due to the relatively short duration of the input, the date unit of the experiment was based on daily rather than weekly data.

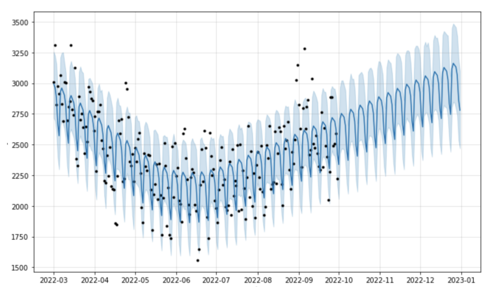

Figure 3.21: Trend: Prophet with Post-Covid -Baseline

As with other experiments, the baseline model was the first to be executed. Since the total data length was 214 days, the initial date point was set to 170 days for the cross-validation, along with a forecast horizon of 30 days and a period of 40 days. It can easily be seen from figure 3.21 that the model returned a multiplicative result with very regular fluctuations. Although the model was not entirely accurate at this point, it was possible to identify the tendency of the *Activations*. Twenty-seven validation results returned exceptionally impressive metrics, with MAPE averaging 0.07 and RMSE 230. Regarding historical forecasts, RMSE performed slightly weaker, reaching 214, while MAPE was flat at 0.07.

Next, due to the short time frame of the data entry, it was assumed that the sea-

| | Cross Validation | Historical Forecast |
|---|---|---|
| MAPE | 0.07 | 0.07 |
| RMSE | 230 | 214 |

Table 3.11: Evaluation: Prophet with Post-Covid -Baseline

sonality effect would be relatively more influential than the holiday effect, as the series has a quarterly seasonality. As mentioned above, the pattern of *Activations* was significantly seasonal. Therefore the seasonality factor was prioritised here to include in the model. Similarly, the period after September and before June each year was regarded as the peak season, so additional seasonality was incorporated manually into the model in order to facilitate learning of the curves by Prophet. Quarterly seasonality with a Fourier order of 10 was also considered. After a grid search with a seasonal list of 5 possibilities multiplied by a time series of 2 possibilities *(additive, multiplicative)*, the model yielded the results shown in Figure 3.22.

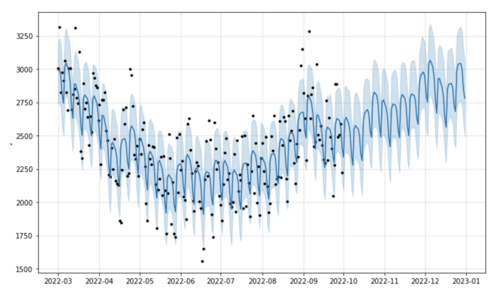


Figure 3.22: Trend: Prophet with Post-Covid -Seasonality

The inclusion of seasonality in the model did not result in significant progress, either in terms of RMSE or MAPE. However, it can be noticed that the light blue uncertainty zone was still relatively unstable and that many of the black forecast points still landed outside the light blue zone. There were two possibilities to consider, one being the introduction of change points and the other being the holiday pattern. As can be observed from the graph, the black points that fell outside the stable interval were not as temporally regular, so the inclusion of change points might be able to fine-tune the accuracy of the forecast.

Figure 3.23 illustrates the result of applying the *change_points* parameter, which was derived from the grid search in the same way as the model above. As can be seen from the graph, the recognition of change points did not work very effectively.

| | Cross Validation | Historical Forecast |
|---|---|---|
| MAPE | 0.06 | 0.06 |
| RMSE | 217 | 188 |

Table 3.12: Evaluation: Prophet with Post-Covid -Seasonality

Despite manually adding a change list to inform the Prophet of where the more prominent changepoints were, there was still no apparent improvement. This can be interpreted as a result of the fact that the algorithm could not accurately project each change point as the trend of *Activations* altered significantly over this period, which in turn resulted in the change points not being very useful in this model. In summary, each of these scores was unchanged, with no growth or decline, whether based on cross-validation or historical forecasts.

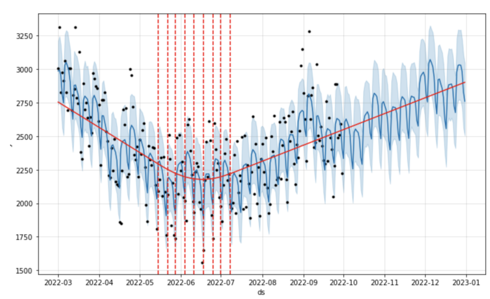


Figure 3.23: Trend: Prophet with Post-Covid -Seasonality and Change Points

Finally, after including the holiday effects of public holidays, marketing campaigns and summer and winter vacations in the model with the corresponding buffer times, it can be observed that most of the points fell within the light blue confidence interval. Although nearly half of the black dots still did not overlap with the solid blue line, an improvement in accuracy can be detected. At the same time, due to the addition of seasonality, the trend of the model was no longer regular, so the model could better learn the impact of the holiday effect on the time series. From figure 3.24, the trend of the series under this model was less aggressive than the model above because the Christmas effect was factored into the model, which was indeed more in line with the business view of the company. Although the cross-validation results regressed slightly, with RMSE and MAPE rising to 233 and 0.07, MAPE reduced the margin of error to 0.04 and even reached the lowest point of 132 only in the historic forecast. Looking closer from figure 3.26 at the predicted value and the actual value from a visual point of view, it can be found that

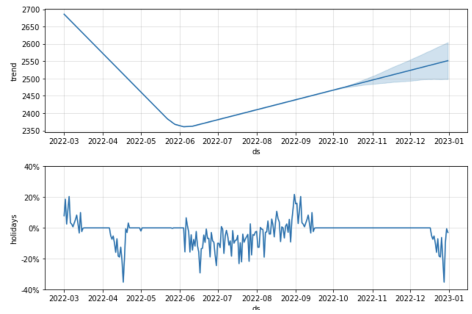


Figure 3.24: Trend: Prophet with Post-Covid -All Effects

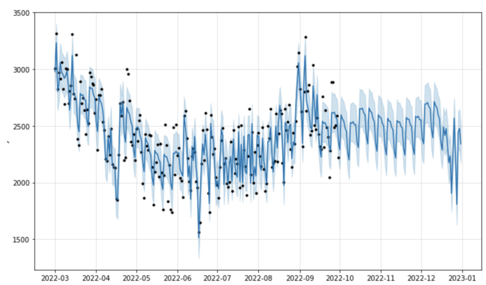


Figure 3.25: Prophet Forecasting with Post-Covid Data with All Effects

the rise and fall of the predicted results were generally consistent with the actual results. However, the predicted ones were not consistent at some time points with large fluctuations. This demonstrates that holiday and seasonal effects contribute favourably to the modelling of the post-epidemic period.

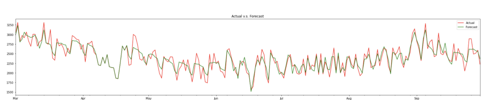

Figure 3.26: Prophet: Post-Covid - Comparison of Actual and Forecast

| | Cross Validation | Historical Forecast |
|---|---|---|
| MAPE | 0.07 | 0.04 |
| RMSE | 233 | 132 |

Table 3.13: Evaluation: Prophet with Post-Covid -All Effect

## 3.3 Holt-Winters' Seasonal Method

Holt-Winters technique employs numbers from the entire time series, and it is clear from the equations that historical data can have a powerful influence on future predictions, especially in the recent past. As with Prophet, three time periods are applied here for future prediction, and the feasibility of this algorithm is also explored in the following paragraph. The series was divided into two parts, 80% of the data was regarded as the training set for model training, and the other 20% was used as the test set.

#### Data from Jan. 2018 to Aug. 2022

First, taking weekly data from January 2018 to August 2022 as an example, using STL time series decomposition in figure 3.27 allowed us to discover that the trend curve rose noticeably with the outbreak and fell after the epidemic subsided. Seasonality fluctuated relatively steadily, with peaks and troughs occurring at roughly half-yearly intervals. Table 3.14 shows the $\alpha$, $\beta$ and $\gamma$ values derived from the

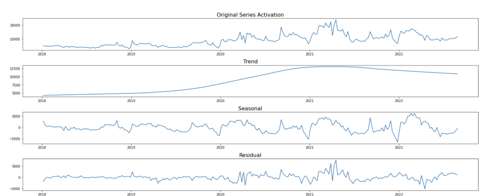


Figure 3.27: Time Series Decompositionn of data between 01.2018 and 08.2022

experiment, representing the level, slope and seasonal components, respectively. Three parameters hold values between 0 and 1, with values close to 0, signifying that relatively less weight is given to recent observations when forecasting forthcoming values. In the multiplicative model, a small $\gamma$ value denotes that the seasonal component hardly changes over time. In contrast, a small $\beta$ value in an additive model means that the slope component does not fluctuate over time, while a larger $\gamma$ value means that the series does not fit the multiplicative model. From the $\gamma$ values of the additive model in the experiment, it is clear that using the multiplicative model is a more satisfactory choice. With the decomposition of the STL above, it is also evident that the seasonality did oscillate significantly over time, while the minimal $\beta$ here suggests that the slope component was stable. On the other hand, an extensive $\gamma$ has a significant time-dependent seasonal component in a multiplicative model [35]. The first 80% of the entire dataset was allocated to the model training data and the remainder to the test data. The table 3.15 below

| Parameter Setting | Alpha | Beta | Gamma |
|---|---|---|---|
| mul-trend + add-seasonality | 0.464 | 0.0001 | 0.102 |
| mul-trend + mul-seasonality | 0.393 | 0.0001 | 0.137 |

Table 3.14: Parameters with Covid Data

shows that the model fits very agreeably in the training set. However, the average difference between the predicted and actual values in the test set is significantly more expansive, implying that the model is unsuitable for forecast based on this scenario.

| | Training Set | Testing Set |
|---|---|---|
| MAPE | 0.13 | 0.16 |
| RMSE | 1822 | 2947 |

Table 3.15: Evaluation: Weekly HW with Covid

The method of removing data points from the Covid-19 period based on the time range between January 2018 to March 2022 is not reasonable in the Holt-Winters method since the algorithm is established on exponential smoothing. Brown [34], who proposed the concept of exponential smoothing, argued that the dynamics of a time series are stable and regular, and therefore the time series can be reasonably deferred. In other words, he believed that the recent past would have a certain degree of influence on the future so that more recent data would be given more weight and more distant data would be given less weight, with a gradual convergence to zero [34]. Hence, removing the time series from the epidemic period from Mar 2020 to Feb 2022 naively would eradicate the time series curve. The first explanation for this is that the missing intervening period would affect the future according to Brown's concept. However, the removed time would be replaced by the pre-epidemic series so that a number that should have converged close to zero would have more weight and impact the forecast. Second, the effect of Covid-19 has a lagging effect. Simply removing one of the intervening periods from the series would result in a spike in the pre-and post-epidemic data. Since HelloFresh had a very significant growth during the epidemic, removing the middle would be seen as an instantaneous growth in the company's performance without any external factors. This effect in exponential smoothing also affects future forecasts.

**Post Covid-19 Experiment**

As it is conceivable that the data during the epidemic are full of uncertainties, hence this section attemptes to use daily data from March 2022 onwards as input to the model and to extend the data endpoint to 30 September 2022 in order to obtain more observations for the model to learn the patterns sufficiently. It can be caught

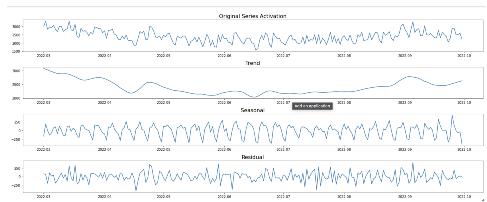


Figure 3.28: STL Decomposition - Post Covid-19

from the table 3.16 that, in this case, there is no meaningful distinction in whether the additive model or the multiplicative model is used. Further analysis, looking at the outcomes of the time series decomposition from graph 3.28, utilising an additive model is a more satisfactory option because the seasonality does not follow the changes in time with apparent highs and lows. The graph below displays that

| Parameter Setting | Alpha | Beta | Gamma |
|---|---|---|---|
| mul-trend + add-seasonality | 0.183 | 0.0001 | 0.06 |
| mul-trend + mul-seasonality | 0.182 | 0.0001 | 0.07 |

Table 3.16: Parameters for post Covid-19

the predicted and actual values curves are broadly in line with each other, although there is a slight discrepancy in the predictions. When looking at the diagnostic figures 3.29, although MAPE was functioning exceptionally well, there was still room for improvement in RMSE. In the longer term, this model is anticipated to perform better, as it has achieved better than the previous model with insufficient data points after an influential shock.

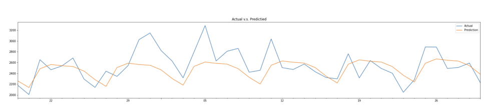

Figure 3.29: HW - Daily with Post-Covid Comparison between Reality and Prediction

| | Training Set | Testing Set |
|---|---|---|
| MAPE | 0.08 | 0.07 |
| RMSE | 213 | 247 |

Table 3.17: Evaluation: HW - Post Covid-19

## 3.4 SARIMA

The tuning of SARIMA parameters is especially problematic in destructive time series because of the number of cut-off points in the target variables. The two most crucial points in SARIMA forecasting are the time series' stationarity and the data's simplicity. The former is that if the time series of the target variable is severely time-influenced, meaning that it has significant ups and downs, the user must differentiate the series to render it relatively smooth. While differentiating is a simple task, models often return suboptimal results despite using the optimal differentiation parameters. The latter is that if the data are too sophisticated, the best P and Q are incredibly resistant to optimisation, despite using PACF and Partial ACF to check the pattern, respectively. Since SARIMA has seven parameters that must be determined, and it is not straightforward to optimise the parameters represented by ACF and PACF artificially, the data were divided into three groups of train, validation and test when building the model. Training set comprises 80% of the data, followed by 10% and 10%, respectively according to HelloFresh's activation pattern. Furthermore, grid search was also employed to find the best model.

#### Post Covid-19 Experiment

The forecast of SARIMA requires sufficient clues for the algorithm to learn past patterns; hence supplying adequate data and time series patterns is necessary. However, the data in the post-epidemic period only start from March 2022, so it is conceivable that although the historical data has a better MAPE or RMSE performance during the fitting process, there may still many uncertainties when making future predictions.

Confirming the stationarity of the time series is the first step in SARIMA forecasting. The definition of stationarity data possesses the following three characteristics: the series has a constant mean value, the auto covariance does not depend on time, and it has a constant variance, respectively. First, the trend of the moving variances or moving averages is visualised using the rolling statistics method. The series is defined as unstable if the average or moving variance fluctuates over time. As seen from the red moving average curve and the black moving variance curve in Figure 3.30, the series has a small range of highs and lows over time. In order to determine more accurately whether the series is stationary, the Augmented Dickey-Fuller test [36] is operated to provide detailed statistical information. This approach presupposes that the data are non-stationary and defines stability by the p-value. Suppose the p-value is more significant than 0.05. In that case, the null hypothesis is valid, meaning that the data is non-stationary, and the use of differencing is desirable for this series. Conversely, if the p-value is less than 0.05, then the data is a stationary series. After performing the ADF test, the ADF statistic is

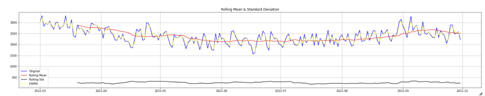

Figure 3.30: SARIMA - Post Covid Checking Stationary

-2.69, which is less than 10% of the critical value but higher than 5% and 1% of the critical value. Meanwhile, the p-value was 0.08, slightly above the threshold of 0.05. It is difficult to specify whether this series is perfectly stationary or oscillating, as the p-value is above the threshold, but the ADF statistic is below the 5% threshold. Therefore, when fitting the model, d (the differential parameter), can be set to 0 or 1 to check the individual results.

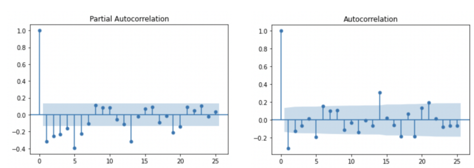


Figure 3.31: SARIMA - Post Covid PACF and ACF

In order to find out the best parameters of moving avergae's and autoregressive's component, plotting ACF and PACF is critical for modelling SARIMA. ACF implies the relationship between the lagged value and the present value of a given time series. In the ACF, the x-axis represents the number of lags, and the y-axis represents the correlation coefficient. PACF is a partial autocorrelation function that accounts for the partial correlation between the series and the lag. Explained in terms of linear regression, PACF predicts *y(t)* based on $y(t-1), y(t-2), y(t-3)$ . Both ACF and PACF plots should be taken into account. For AR, an ACF plot that decreases gradually and a PACF that decreases sharply after a significant lag in p is the ideal situation. To define the MA process, the desired ACF and PACF plots are the opposite of the AR; ACF should show a significant decrease after a lag of q, while PACF should show a geometric or gradual decrease [32]. According to the above explanation, setting ARMA to (5,1) or (1,1) was a parameter set that needs to be tried. On the other hand, it can also be noticed from the figure 3.31 that the time series was highly complex, so it was not straightforward to specify the best parameters. Consequently, in addition to the two sets of parameters mentioned above, multiple experiments will also be carried out. Since SARIMA also has to

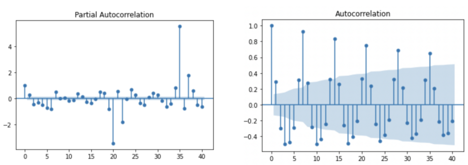


Figure 3.32: SARIMA - Post Covid Seasonal PACF and ACF

determine seasonal AR and MA parameters, the seasonal component has been obtained by decomposing the time series using STL. As can be seen from the figure 3.32, the seasonal PACF suddenly stood out in 20 and 35 without any regularity, while the ACF was repetitive, with a peak after every seven lags and a slow decline. Therefore, in SARMA, the groups (0,1), (0,0), (1,1), (1,7) etc., needed to be tried. It can also be seen from the graph that the given series has a seasonality of 7 lags. Thus the seasonal frequency m can be assigned as 7.

| | Training Set | Validation Set | Testing Set |
|---|---|---|---|
| MAPE | 0.08 | 0.11 | 0.05 |
| RMSE | 337 | 332 | 175 |

Table 3.18: Evaluation: SARIMA Post Covid-19

According to the table 3.18, the fitting process of this model has shown remarkable outcomes in the training, validation, and test sets. The RMSE in the test set even reached below 200. Comparing it with the graph 3.33 below of actual and predicted *Activations*, SARIMA roughly imitated the trend of the real world. However, SARIMA cannot perfectly explain the ups and downs of the details of each period. The above results validate the properties of SARIMA described above. The lack of sufficient informativeness may prevent accurate predictions from being achieved, while uncertainty in the parameters of complex datasets may result in a loss of accuracy. Therefore, it is imaginable that SARIMA is indeed a simple model that can provide a general trend, but if the goal is to provide detailed and high-precision forecasts, the practicality of SARIMA is still being determined.

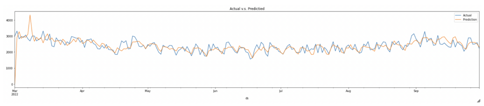

Figure 3.33: SARIMA - Post-Covid Comparison between Reality and Prediction

**Data from Jan. 2018 to Aug. 2022 with Covid-19**

As mentioned above, SARIMA places particular importance on the adequacy of the amount of information available. Therefore, four years and nine months of data were taken as input to the experiment, and data from Covid-19 were also incorporated. Unlike the Holt-Winters method, the exogamous variable of the number of confirmations of Covid-19 was also included in this experiment in the hope that this additional factor would allow the algorithm to better shape the predictive model.

From figure 3.34, the data before the pandemic appeared relatively flat, but after the outbreak, the moving averages and moving variances began to fluctuate dramatically. With the ADF test, a p-value of 0.15 was obtained, and the ADF statistic of -2.38 was larger than the 10% critical value of -2.57. Therefore, a differentiation of the series is mandatory. As can be seen from the figure 3.35,

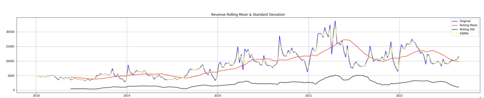

Figure 3.34: SARIMA - with Covid Checking Stationary

the results of both ACF and PACF showed a distinct moderation after the first lag. Therefore, ARMA(1,1) was a worthwhile set of parameters to test. However,

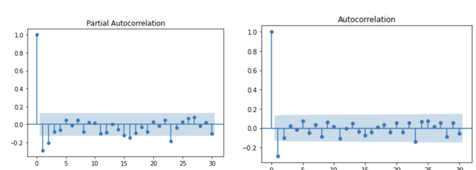


Figure 3.35: SARIMA - with Covid PACF and ACF

when determining the seasonal set of parameters, it can be concluded from the figure 3.36 that no consistent pattern can be identified for either ACF or PACF. In the PACF, AR(1) appeared to be the more promising parameter, but the lags 6, 17, 24, and 29 emerge after a period of remission. In the ACF, 1 and 15 were both appropriate points to attempt, as they were the two peak values of positive and negative correlation, respectively. Regarding seasonal frequency m, 12 was defined according to the STL decomposition due to the existence of quarterly cycles.

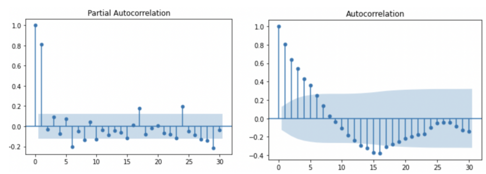


Figure 3.36: SARIMA - with Covid Seasonal PACF and ACF

As can be gleaned from the table 3.19, the model performs relatively well, with MAPE reaching below 10% in the test set and RMSE being lower than in the validation set. It suggests that the model is more robust to future predictions after turbulence due to SARIMA's superior capability to find cut-offs and learn critical factors when fitting the model. The results in the figure 3.37 below show

| | Training Set | Validation Set | Testing Set |
|---|---|---|---|
| MAPE | 0.14 | 0.19 | 0.09 |
| RMSE | 1910 | 2578 | 1040 |

Table 3.19: Evaluation: SARIMA with Covid-19

the need for SARIMA to possess a large number of data points. With more than four years of data, the algorithm was capable of learning the oscillations of the time series across the data satisfactorily. It is evident that the prediction curve is relatively fine-grained and that the performance varies with different points along the time scale. Moreover, despite occasional discrepancies, most predicted trends are consistent with the real world.

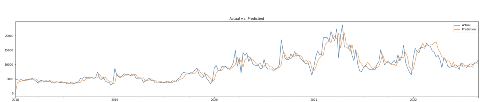

Figure 3.37: SARIMA - with Covid Comparison between Reality and Prediction

### Data from Jan. 2018 to Aug. 2022 without Covid-19

The complexity of the data, the exponential nature of the time, and the number of data points, as described in the above section, can seriously affect the results of

SARIMA predictions. As can be imagined that the fluctuations in the series could be pronounced because of the direct truncation of all data during Covid-19. Hence, there is bound to be a large gap between the two series, as in the case of Prophet. With this experiment, the influence of SARIMA on the above three factors will be verified.

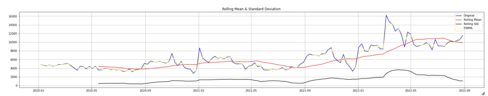

Figure 3.38: SARIMA - without Covid Checking Stationary

It is clear from the figure 3.38 that the moving average and the moving variance have a tendency to increase over time, so differentiation is necessary. On the other

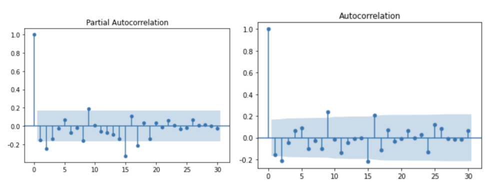


Figure 3.39: SARIMA - without Covid PACF and ACF

hand, as can be seen from the graph 3.39, although the ACF and PACF have been plotted, the best P and Q were challenging to determine. In general, the PACF diagram weakened subsequent lags when the lag was 2 or 15; the same applies to the ACF when the lag was equal to 2, 9, or 15. It was necessary to try these parameter sets in permutation. Regarding the seasonal PACF, the correlation coefficients were progressively converging at none of the points, regardless of the lag. This clearly demonstrated the weakness of SARIMA for exponential time complexity, in that P and Q are poorly determined, as no regularity existed in the time series, which implied that the residuals are too significant.

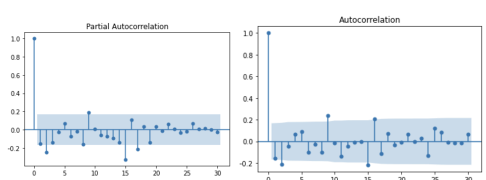


Figure 3.40: SARIMA - without Covid Seasonal PACF and ACF

Figure 3.41 reveals that the model fits well. However, when looking for the best parameters, it was found experimentally that this was challenging. From figure 3.39 can be seen that the lag situation is too complicated to determine the best one as it has no regularity. After testing more than 15 sets of parameters after applying the grid search, 70% of possible input-sets all returned similar results. Meanwhile,

| | Training Set | Validation Set | Testing Set |
|---|---|---|---|
| MAPE | 0.14 | 0.08 | 0.04 |
| RMSE | 1371 | 843 | 572 |

Table 3.20: Evaluation: SARIMA without Covid-19

the performance of RMSE and MAPE is optimistic, but it is doubtful whether the model is effective in withstanding rapid changes in serial patterns. This will be explained in more detail in the next chapter.

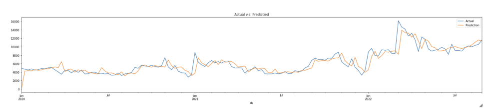

Figure 3.41: SARIMA - without Covid Comparison between Reality and Prediction

## 3.5 Empirical Conclusion

To conclude, this chapter examined the characteristics of the models using three different periods as inputs to three experimental models. It can be recognised from the empirical works that Prophet's model construction is more resilient and tailor-made than others, as its versatility enables users to determine what objects should be included in the model when fitting the model depending on the business requirements. In order to carefully and comprehensively investigate the impact of each factor on the input time series, it was essential to build the baseline model and add the potential factors one by one in turn. According to the result, it indicates that seasonality and holidays significantly impact the number of HelloFresh activations. The Holt-Winters seasonal model is the speediest model to fit since it is based on the moving average method, and therefore its simplicity makes it relatively easy to fit. In addition, the Holt-Winters model, which predicts the future by aggregating every point in the past, cannot be successfully carried out in the experiment by taking out the Covid-19 time series from January 2018 to August 2022, as there is a two-year disruption in the time series. The aforementioned approach severely compromises Holt-Winters' properties using moving averages [24], which would also make the HelloFresh activation model inaccurate as there is a significant change in the activation pattern before and after the epidemic. On the other hand, SARIMA encountered several challenges in the modelling process, such as the complexity of the input series, which made it problematic for the PACF and ACF to provide a reasonable lag for the user to formulate the model parameters for the algorithm. Furthermore, although SARIMA is technically capable of incorporating additional regressors to enable the algorithm to learn the model more effectively, in the case of the HelloFresh activation number, adding a series of daily confirmations could have been more useful compared to Prophet. Also, this experiment tried to apply *auto.arima* [25] to automatically select the parameters, but it took more than half a day to run the algorithm, so it finally failed. Moreover, although the test sets returned by SARIMA have excellent evaluation figures, the modelling process is time-consuming.

# Chapter 4

# Methodologies Evaluation

As three models with three different inputs of time scenarios were conducted, the performance between models is evaluated and described in this chapter. The chapter compares models with the same input training data set using statistical model evaluation metrics. Furthermore, chapter 4.2 provides insights from the input timeframe point of view and model-wise perspectives, allowing people to have a comprehensive understanding regarding the performance of each model apart from only seeing numbers as models' assessment not only relies on statistical approaches but also considers the congenital characteristics of the algorithms.

## 4.1 Empirical Results of Model Performance

The evaluation criteria here are based on MAPE and RMSE. The comparison points for the post-epidemic period are 57 data points, from 1 October 2022 to 26 November 2022, while the data length of the other two input sets is 12, from the first week of September 2022 to the third week of November 2022.

### 4.1.1 Using 2018 Onwards data

| | Prophet | SARIMA | Holt-Winters |
|---|---|---|---|
| MAPE | 0.12 | 0.15 | 0.06 |
| RMSE | 1868 | 1984 | 925 |

Table 4.1: Evaluation of 2018 Onwards with Covid Model Across Models

From the data points after 2018, it can be shown that Holt-Winters performed relatively favourably, and the algorithm yielded the optimal MAPE and RMSE. However, there were still significant discrepancies between the actual and predicted values, and the predicted trend was only partially in line with reality. The deviations were mainly within 10000. However, when examining the forecasts from the perspective of Prophet, it was readily apparent that the forecast and reality lines were more aligned, although some points inflate the MAPE and RMSE over time. In this case, the performance of SARIMA is doubtful due to its inconsistency of the trend line. As can be seen from the graph 4.1, the two lines partially have some similarities in the trend, whereas the gap between them cannot be disregarded.

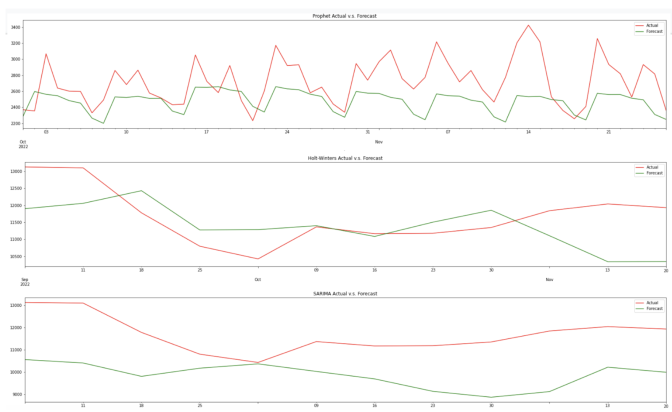


Figure 4.1: 2018 Onwards Performance across three models

In this experiment, 4.5 years of weekly data points was considered when building the model, which indicated that the model could have more information to learn the trend of the previous patterns. Especially Prophet, this methodology enables users to integrate as many external factors as they wish, which may lead the algorithm to consider many external interferences better to adjust the cut-off points. Prophet, from the experiment, provided a more promising trend for future prediction, which implied that this approach is robust to noises and the nonstationary time series. Compared to SARIMA, as this algorithm has fewer parameters that allow users to calibrate, it is, therefore, incapable of shaping the trend and adapting fluctuations ideally since it has low flexibility. On the other hand, insufficient data points of post-covid data may be problematic for SARIMA. The complexity of the data is another shortage that SARIMA has. As at least 10 cut-off points existed throughout the dataset, the model suffered from having stable patterns that permit the algorithm to learn the trend. The post-covid period was particularly critical when learning the model since the lack of data from the post-period led to difficulty in simulating the pattern. Moreover, Holt-Winter's seasonal method is exceedingly complicated to define performance as it possesses the best evaluation performance in MAPE and RMSE. In contrast, its alignment between reality and prediction is not completely optimal. This phenomenon can be illustrated by the need for more flexibility in calibrating parameters since adding external regressors is impossible in this algorithm. The algorithm only has some parameters that emphasise the general trend, such as trend damping and the mode of seasonalities. On the other hand, this algorithm considers every data point, and it causes the effect of some closer extreme points to be amplified, such as data points in omicron outbreaking due to its weighting characteristic.

### 4.1.2 Using 2018 Onwards data without Covid-19

| | Prophet | SARIMA |
|---|---|---|
| MAPE | 0.33 | 0.08 |
| RMSE | 4327 | 1030 |

Table 4.2: Evaluation of 2018 without Covid Onwards Model Across Models

As mentioned in the above chapter, conducting Holt-Winters using 2018 onwards data without Covid-19 is illegitimate due to the algorithm's characteristic of weighing every data point. Hence, only Prophet and SARIMA were experimented with and assessed in this paper.

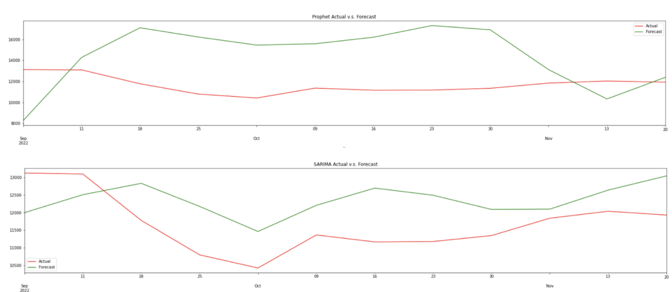


Figure 4.2: 2018 Onwards w/o Covid Performance across three models

It can be seen from the table 4.2 that both MAPE and RMSE are not desirable in Prophet compared to SARIMA. The result can be inferred as the Prophet only follows the changes when the trend changes, and the *seasonal_prior_scale* is ineffective even though manually defining the parameter is executed. A similar occurrence is described in Jung et al.'s paper [27] *"A probabilistic time series model* *for sales forecasting considering the title of Prophet's paper"*. Similar to this literature, Prophet did not consider recent data but consistently modeled predictions with peaks at the intersection of pre- and post-Covid. Before the epidemic, weekly *Activations* were typically less than 10000 and hovered around 7500, while the data after March 2022 often oscillated around 12500 due to the lag effect. Bringing together two radically contrasting data patterns inevitably resulted in a dramatic spike. The pre-epidemic Christmas holiday was particularly low, and the direct link to the post-epidemic March 2022 data produced a virtually non-existent increase. Prophet's inability to take into account the shortcomings of the recent data has affected the algorithm in this wave of growth and has seriously impacted future forecasts.

Furthermore, throughout the period in the graph 4.2, Prophet almost every moment has a significant deviation between reality and forecast, reaching 4000 or above. A different result, however, is obtained in SARIMA with a more promising alignment between two lines simultaneously. When implementing SARIMA, it is essential to specify the appropriate $p$ and $q$ parameters to understand the extent to which the future series is affected by the previous through the ACF and PACF plots. In addition, the ADF test also helps users comprehend whether the data is stationary. The predictions were made relatively stable by manually detecting several SARIMA-specific points of the series one by one. Also, due to the absence of pandemic data in the series, fewer cut-offs are produced, and the complexity of the data is also decreased. Consequently, SARIMA has a chance to return a more reasonable prediction value.

### 4.1.3 Using Post-Covid Data

Employing daily post-covid data as input is moderately identical to the usual forecasting since the time series does not suffer from the disturbance by Covid-19, except for the lag effect. Technically, the table 4.3 illustrates the performance across three models, and both Prophet and Holt-Winter's methodologies yielded promising numbers in MAPE, reaching 0.09 and 0.07, respectively. From figure 4.3, it is effortlessly discovered that all three forecasting approaches have a stable seasonality pattern, although variations between reality and forecast still materialise. Regardless, the trend line of the reality across models is aligned with the forecast, implying that the hypothesis of post-covid data being regular is valid. This led to the forecasting being more favourable and relatively uncomplicated due to the abatements of uncertainties and cut-offs.

In this experiment, the biggest shortcoming is insufficient training data. Since the training data does not reach one year, the seasonal parameters are difficult to specify ideally because the pattern of the whole year has yet to be revealed. If the

| | Prophet | SARIMA | Holt-Winters |
|---|---|---|---|
| MAPE | 0.09 | 0.12 | 0.07 |
| RMSE | 342 | 426 | 277 |

Table 4.3: Evaluation of Post-Covid Model Across Models

training data point has one year or even two years, then the seasonal trend is bound to be more consistent with the real-world rather than such a regular line. In other words, such low MAPE and RMSE may be due to the regular seasonality of the data itself, although the ups and downs still exist. The ideal situation should be that the two lines are comparable to the same state, but it is difficult to achieve this goal owing to the scarcity of a one-year training set. A training set of less than one year signifies that the current prediction cannot be extrapolated with reference to the data of the same period of the previous year or longer, and this will make the model unaware of what happened to the pattern of the same period in the past few years. Therefore, the algorithm has to utilise the data in the current training set to make deductions, so the uncertainty will likely increase accordingly.

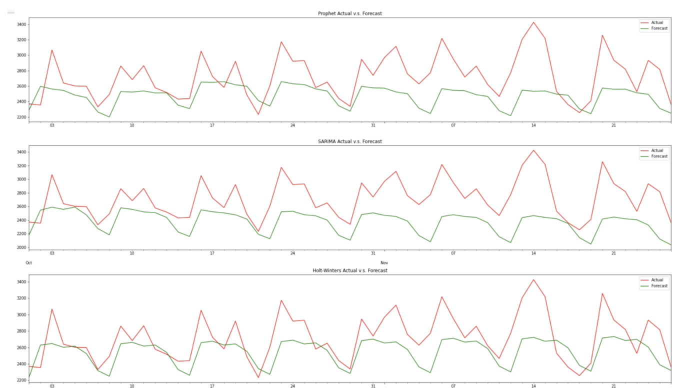

Figure 4.3: 2022 Onwards Performance across three models

## 4.2 Insights into Comparison

### 4.2.1 Time Period Prerspectives

One conclusion that can be drawn from various experiments is that having an appropriate period is an essential key when making a forecast. For example, as can be seen from figure 4.3, the prediction trend of the model using the post-covid period as input is roughly in line with the real world. However, inadequate points make the model unable to accurately predict its seasonality and ups and downs, which is particularly serious to the post-covid date frame as input. Since the pattern of HelloFresh's activations is affected by customer behaviours greatly, and its consumption patterns also suffer from the seasonality a lot, it is, therefore, critical to have more than a full yearly cycle of input data for the algorithms to learn the whole pattern otherwise the model has to make predictions based on unknown patterns.

When it comes to the disrupted time series as the input of models, the interference of the time series has seriously impacted the forecasting model because of the outbreak of Covid-19. From a business perspective, consumption behaviours dramatically change before and after the pandemic, referring the figure 3.1. It implies that removing the interval of the pandemic period and connecting two different patterns leads to a larger disruption in the time series as the time is cut off. In other words, all patterns during the period disappeared groundlessly. Nevertheless, customer patterns continuously change, and data points from the past and future are reciprocally influenced. However, SARIMA returned an outstanding outcome with this setting. In terms of results, SARIMA's ability to define the lag of AR and MA by means of the time series with breakpoints is of considerable benefit, however, due to the short time period evaluated, it is inconclusive at this stage whether this time period is applicable to HelloFresh's Activations model.

Compared to the truncated and post-epidemic time series, the time series from January 2018 to August 2022 is this experiment's most desirable input period, although the statistical evaluation numbers did not yield the best input outcome. This input collects the longer data length and the complete activation patterns, containing each transition without any break. It solves the shortages among other inputs but still needs to improve its issues, such as adding more customised regressors to allow the algorithm to learn the patterns from the disrupted period. Overall, if the series has more data points, feeding the forecasting model with the post-covid data could be the optimal solution as HelloFresh is now facing a different consumption pattern, and the effect of Covid-19 may be faded with time goes. Nevertheless, in the current situation, using complete series from 2018 to August 2022 as model input is the best solution according to the result of experiments.

### 4.2.2 Model Perspectives

#### Prophet

Like the purpose for which Prophet was created, this model is suitable for business use and users with less statistical background. However, this also implies that the model lacks the flexibility to allow users to define the most meaningful statistical parameters affecting the time series, such as lag parameters. Prophet is a thoroughly automatic model. However, when encountering more complex time series, such as the topic discussed in this paper, it will need more freedom to adjust the statistical significance behind the algorithm. On the other hand, since the seasonal parameters provided by the algorithm are relatively straightforward, it is problematic for the user to find out the real reason why the model is inaccurate, and the influence of *seasonal_prior_scale* is also minimal. The effectiveness of the automatically defining cut-off in Prophet is also doubtful because after trying multiple input periods, not all cut-offs were detected, especially ones within Covid-19, even though the cut-off has been manually defined. Nevertheless, Prophet is still an outstanding model time-wisely with the input of the complete duration of 2018 to 2022 and post-Covid set, reaching around 0.11 in MAPE, and it further supplies many functions that can meet business needs.

From the above experiments, it can be known that the accuracy of the model is significantly enhanced when seasonal and holiday objects are added. In addition, when the model was fitting, the learning capacity of the model also improved after incorporating customised events such as the annual fixed marketing campaign and the governmental city lockdown. The Fourier parameter in the custom seasonal object also allows the algorithm to learn the pattern in complex time series. These features supports the algorithm to solve the rapidly changing activation pattern of HelloFresh during the epidemic. Overall, Prophet provides the greatest support for marketing prediction among the three models, as marketing is indeed subject to numerous external factors, and HelloFresh's intricate time series is backed up by Prophet because of these abundant customisied functions.

#### SARIMA

SARIMA is a well-known model for time series. By manually calibrating its parameters, the algorithm has more flexibility to respond to changes in the series. Although the parameters can incorporate additional variables to facilitate the target variable to accommodate its prediction curve, the effect of this function did not play an influential role in this experiment. To elaborate further, the conclusion drawn by this algorithm in this experiment is that it cannot cope with too complex time series, especially the series full of many change points, which indirectly makes it very difficult to determine the optimal parameters. However, this possible reason also comes from the fact that although the algorithm has its capacity of freedom so that users can adjust the stationarity of the series and the parameters of AR and MA, too few parameters make prediction challenging [5]. For example, the closure of the city and the outbreak of mutant viruses are unconventional events. SARIMA is relatively ineffective in dealing with this aspect. Regarding advantages, SARIMA is a moderately easy-to-train model, and getting hands-on is relatively fast. If the model's input has a relatively simple and regular pattern, then SARIMA would be an ideal option. Taking an example of experiment of removing Covid-19 series from series, SARIMA performed better compared to Prophet, with a MAPE of 0.08. This result can be inferred from the fact that the entire series became more stationary than the original series, as many changepoints were eliminated with the removal of the epidemic period, so the prediction accuracy was significantly improved.

From the perspective of marketing forecasting at HelloFresh, when dealing with such a complex time series of HelloFresh activations, SARIMA's inability to solve high-complexity time series is particularly obvious, especially with data during the epidemic. Finally, the algorithm returned the worst RMSE and MAPE among the three models for inputs between 2018 and 2022 and the post-pandemic period. This directly demonstrates the aforementioned weakness of SARIMA, which at this stage is not appropriate for the predictions required by HelloFresh's business situation. In the long run, if the time series becomes regular and stationary, then SARIMA would be an acceptable choice.

**Holt-Winters' Seasonal Method**

The Holt-Winters seasonality method is a remarkably effective algorithm if the time series pattern is similar to the previous one. It can be seen from the formula of the algorithm that both additive and multiplicative methods take into account each data point in the series and weigh each point to a different degree. However, the input time series of the model in this experiment has intricate and erratic patterns. Therefore, Holt-Winters is more demanding to deal with this scenario resulting from some past patterns that are no longer reasonable for future predictions.

As with the Omicron issue mentioned above, the mutation had a pronounced effect on the HelloFresh activation pattern in 2022, and this period is closer to the last point in the series. Thus Holt-Winter's seasonal algorithm would give higher weight to the data points during the Omicron period in terms of the algorithm. Although the method still has the most outstanding MAPE and RMSE in the model comparison for the complete time series from 2018 to August 2022, its performance is questionable as the data points for evaluating the model are too few due to the scarcity of post-epidemic data. In addition, the additive model defines a time series whose seasonality is constant over time, while the multiplicative suggests that its seasonality will increase proportionally consequently [24]. According to the above two rulings, the effectiveness of using this algorithm to address complex time series is doubtful. One reason is that the disrupted time series has no regularity. In other words, the series will have irregular changes over time so that accurate prediction would appear more complicated.

In evaluating the model, it would be arbitrary to use MAPE or RMSE alone, as there are only 12 points of weekly data to evaluate. However, increasing the number of evaluation points would conversely result in too little training data for the post-epidemic period, which would create an imbalance in the data before, during and after the epidemic, resulting in an uneven fit of the model.

In summary, Holt-Winters' seasonal model is unsuitable for HelloFresh's marketing model predictions if the input data points are 4.5 years long and the series contains many change points. However, it is suitable for short-term forecasting in the post-epidemic period, as the graph 4.3 shows that the time series in the post-epidemic period is relatively regular, and fewer factors interfere with the series as large as Covid-19. Therefore, if the training set is small, the advantages of Holt-Winters' seasonal approach are more prominent than for other models, as the algorithm is able to learn from recent data and give short-term feedback. In the case of long-term forecasting, the algorithm may need to obtain longer-term data for forecasting, which may result in the longer-term data no longer matching the current trend. Therefore, Holt-Winters' seasonal approach is appropriate for HelloFresh's short-term marketing forecasts based on its characteristics.

## 4.3 Experimental Summary

Taking HelloFresh's activation forecasts as an example, Prophet's and Holt-Winters' seasonal models yielded promising outcomes for both the complete time series and the post-epidemic period. Prophet's strength lies in the extensive functionality that allows users to customise it to suit their HelloFresh business needs. However, Prophet is weaker for time series with a two-year interruption length, presumably because of the different patterns before and after the epidemic, which resulted in learning difficulties for the algorithm. Overall, the advantages of Prophet are consistent with those required by the HelloFresh marketing prediction model due to its substantial customisation and freedom to adjust parameters. On the other hand, Holt-Winters seasonal model has a comparative advantage for the post-epidemic period as its extended algorithm based on moving averages permits short-term forecasting. The steady time series in the post-epidemic period also helps the algorithm to learn the trends of the components of the time series more efficiently, allowing the algorithm to return more reliable forecasts. Although Holt-Winters' model achieves excellent RMSE and MAPE for the inputs from 2018 to 2022, its trend deviates significantly from the actual figures, as can be seen in Figure 4.1. Therefore, the accuracy of the method is not conclusive at this stage, as the model only has 12 evaluation points. Therefore, if HelloFresh were to have more stable data to feed into the model in the future, Holt-Winters' model would be instrumental in predicting HelloFresh's marketing. Finally, training SARIMA models with many variations is a challenging task. Experimentally, SARIMA's RMSE and MAPE perform well, but they are time-consuming in determining the parameters and slow in fitting the model. Looking back at 4.1, it is also clear that there is still a significant discrepancy between the SARIMA predicted trends and the real world. In the post-epidemic period, SARIMA returns roughly the same trend as the other models, but its evaluation index is the weakest. The only model that withstands the two-year Covid-19 interruption series as input was SARIMA, supposedly because the most appropriate parameters of SARIMA can be artificially determined by the ACF and PACF plots, which helped the algorithm to recognise the patterns. In addition, since the inter-epidemic data is relatively unstable and has many cut-offs, removing the disruptions allows SARIMA to have more reliable data for prediction. In conclusion, Prophet and Holt-Winters seasonal models are slightly better than SARIMA for purely post-epidemic forecasting or forecasting using long-term series as a training set.

# Chapter 5

# Conclusion

The prediction of highly disruptive marketing time series has always been a delicate topic, especially in the aftermath of pandemic outbreaks worldwide. After conducting numerous literature reviews, it is easy to recognise that there is no definitive answer to this task. In terms of data science, the data during an epidemic may be treated as anomalous in the long run. The time series involves too much unpredictability and will lose its importance in the post-epidemic period. HelloFresh's number of Activation predictions is sensitive to this question, as the activation behaviour is essentially based on human behaviour, which comprises many uncertainties. Furthermore, epidemic preparedness measures have considerably influenced consumer behaviour during this period. In addition to the external factors brought about by the epidemic, such as the emergence of mutated viruses, changes in customer behaviour make time series prediction even more challenging.

The choice of models and timeframe of the training set is, therefore, particularly paramount. According to the experiment on HelloFresh's Activation prediction, Prophet's diverse functionalities are robust to severe changes in the Activation pattern. It applies to long-duration input with lots of cut-offs, such as data from 2018 to 2020 September and shorter period input, like post-Covid-19 data, reaching MAPE of 0.12 and 0.09, respectively. On the other hand, Holt-Winters' seasonal model works well with post-Covid input according to HelloFresh's scenario, reaching 0.07 in MAPE. Although a better MAPE of 0.06 was obtained when the input was using series from 2018 to September 2022, the performance is still doubtful. Holt-Winters considered every data point when fitting the model. At the same time, HelloFresh's pattern kept changing from time to time, which led the algorithm challenging to learn the pattern from historical points. This can be recognised from the disparate trends between forecast and actuality. Overall, Holt-Winters's algorithm is applicable for stable input, but the performance of input with lots of disruptions still needs time to affirm. Training SARIMA is the most challenging of HelloFresh's marketing activation forecasts, as the characteristics of its algorithm do not lend themselves to overly complex time series. However, this feature of SARIMA has conversely enabled the algorithm to achieve exceptional outcomes over Prophet and Holt-Winters for the input of removing epidemic data from the 2018 to 2022 September series, with a MAPE of 0.08 compared to Prophet's 0.33. Although SARIMA demonstrated impressive success in other periods, searching for appropriate parameters was time-consuming due to the series complexity. Furthermore, when the number of parameters gets larger, fitting the model will require more time than other algorithms. Hence, as a result, SARIMA did not dominate the HelloFresh scenario compared to the other two algorithms.

From the above experiments, as can be learned that incorporating a function that can detect cut-off points properly or manually configure them is expected to lead to more confident predictions. At the same time, adapting a model that can accommodate the diversity of patterns in the series is also to be anticipated as an advancement in prediction. In addition, since marketing forecasts tend to involve several exogenous, such as marketing campaigns and peak and low seasons, a model that considers other underlying influences on the target variables may provide the algorithm with more guidance in generating reliable forecasts. Although it is still very inconclusive to determine the three models' performance in the long run, it is known that over time, as more data points become available, models' predictions can be expected to become more accurate.

## 5.1 Future Work

There is still plenty of possibility for experimentation and exploration of models for solving this problem, along with a number of approaches for highly destructive time series. For example, incorporating machine learning algorithms such as LSTM and neural networks or finding a reasonable way to process data to detect anomalies may result in fewer cut-off points, which could also enhance the precision of predictions. The following are some possible subjects that could be further investigated in the future.

- Applying the machine learning approach to well-adapt the variations of patterns, such as LSTM and CNN.
- Incorporating statistical regressors for Covid period data to smooth the fluctuation, such as using Gaussian function to degrade the data.
- Waiting for more recent data and testing all models again to make comparisons.

# Bibliography

[1] aditianu1998. Understanding of LSTM Networks, May 2020. Section: Machine Learning.

[2] J Scott Armstrong. Principles of Forecasting: A Handbook for Researchers and Practitioners.

[3] Serdar Arslan. A hybrid forecasting model using LSTM and Prophet for energy consumption with decomposition of time series data. *PeerJ Computer Science*, 8:e1001, June 2022. Publisher: PeerJ Inc.

[4] Ipek Atik. A New CNN-Based Method for Short-Term Forecasting of Electrical Energy Consumption in the Covid-19 Period: The Case of Turkey. *IEEE Access*, 10:22586–22598, 2022. Conference Name: IEEE Access.

[5] Aayush Bajaj. ARIMA & SARIMA: Real-World Time Series Forecasting, July 2022.

[6] Teemu Belt. When is forecast accuracy important in the retail industry? Effect of key product parameters. April 2017. Accepted: 2017-04-13T10:02:28Z.

[7] George Box. Box and Jenkins: Time Series Analysis, Forecasting and Control. In *A Very British Affair: Six Britons and the Development of Time Series Analysis During the 20th Century*, pages 161–215. Palgrave Macmillan UK, London, 2013.

[8] Leo Breiman. Random Forests. *Machine Learning*, 45(1):5–32, October 2001.

[9] Jingyuan Chen, Hanwang Zhang, Xiangnan He, Liqiang Nie, Wei Liu, and Tat-Seng Chua. Attentive Collaborative Filtering: Multimedia Recommendation with Item- and Component-Level Attention. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in*

*Information Retrieval*, pages 335–344, Shinjuku Tokyo Japan, August 2017. ACM.

[10] Krzysztof J. Cios. Deep Neural Networks - A Brief History, January 2017. arXiv:1701.05549 [cs].

[11] Cleveland, Robert B, Cleveland, William S, and Terpenning, Irma. STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Statistics Sweden (SCB)*, 6(1):3–73, March 1990.

[12] Zhiyong Cui, Ruimin Ke, Ziyuan Pu, and Yinhai Wang. Deep Bidirectional and Unidirectional LSTM Recurrent Neural Network for Network-wide Traffic Speed Prediction, November 2019. arXiv:1801.02143 [cs].

[13] Marnik G Dekimpe and Dominique M Hanssens. Time-series models in marketing:: Past, present and future. *International Journal of Research in Marketing*, 17(2):183–193, September 2000.

[14] D. Dickey and Wayne Fuller. Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *JASA. Journal of the American Statistical Association*, 74, June 1979.

[15] Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4):193–202, April 1980.

[16] Dr. Shikha Gaur. GLOBAL FORECASTING OF COVID-19 USING ARIMA BASED FB-PROPHET. *International Journal of Engineering Applied Sciences and Technology*, 5(2):463–467, June 2020.

[17] J Greunen, Andre Heymans, Chris Van Heerden, and Gary Vuuren. The Prominence of Stationarity in Time Series Forecasting. *Journal for Studies in Economics and Econometrics*, 38:1–16, April 2014.

[18] R Gustriansyah, N Suhandi, F Antony, and A Sanmorino. Single exponential smoothing method to predict sales multiple products. *Journal of Physics: Conference Series*, 1175:012036, March 2019.

[19] Seng Hansun. A Novel Research of New Moving Average Method in Time Series Analysis. *International Journal of New Media Technology (IJNMT)*, 1:22, August 2014.

[20] A C Harvey and S Peters. Estimation Procedures for Structural Time Series Models. 9(2):21.

[21] Hansika Hewamalage, Klaus Ackermann, and Christoph Bergmeir. Forecast Evaluation for Data Scientists: Common Pitfalls and Best Practices, April 2022. arXiv:2203.10716 [cs, stat].

[22] Sepp Hochreiter and Jürgen Schmidhuber. Long Short-term Memory. *Neural computation*, 9:1735–80, December 1997.

[23] Charles C. Holt. Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20(1):5–10, January 2004.

[24] Rob J Hyndman and George Athanasopoulos. *Principles and Practice (2nd ed)*. April 2018.

[25] Rob J. Hyndman and Yeasmin Khandakar. Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software*, 27:1–22, July 2008.

[26] Rob J. Hyndman and Anne B. Koehler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688, October 2006.

[27] Seungjae Jung, Kyung-Min Kim, Hanock Kwak, and Young-Jin Park. A Worrying Analysis of Probabilistic Time-series Models for Sales Forecasting, November 2020. arXiv:2011.10715 [cs, stat].

[28] Bineet Kumar Jha and Shilpa Pande. Time Series Forecasting Model for Supermarket Sales using FB-Prophet. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, pages 547–554, April 2021.

[29] G. Lemus. Why Financial Time Series LSTM Prediction fails, March 2019.

[30] Jorge Lobo, A. Jiménez-Valverde, and Raimundo Real. AUC: A misleading measure of the performance of predictive distribution models. *Journal of Global Ecology and Biogeography*, 17:145–151, January 2008.

[31] Francisco Martínez, María Pilar Frías, María Dolores Pérez-Godoy, and Antonio Jesús Rivera. Dealing with seasonality by narrowing the training set in time series forecasting with kNN. *Expert Systems with Applications*, 103:38–48, August 2018.

[32] Mohammad Masum. Identifying AR and MA terms using ACF and PACF Plots in Time Series Forecasting, June 2022.

[33] Yanrui Ning, Hossein Kazemi, and Pejman Tahmasebi. A comparative machine learning study for time series oil production forecasting: ARIMA, LSTM, and Prophet. *Computers & Geosciences*, 164:105126, July 2022.

[34] Eva Ostertagová and Oskar Ostertag. Forecasting using simple exponential smoothing method. *Acta Electrotechnica et Informatica*, 12(3), January 2012.

[35] Rob J Hyndman and George Athanasopoulos. *Forecasting: Principles and Practice (3rd ed)*. Melbourne, Australia, 2021.

[36] Dipanjan Sarkar, Raghav Bali, and Tushar Sharma. Forecasting Stock and Commodity Prices. In *Practical Machine Learning with Python: A Problem-Solver's Guide to Building Real-World Intelligent Systems*, pages 467–497. Apress, Berkeley, CA, 2018.

[37] Eduardo Silva, João Ferreira-Coimbra, Eva Oliveira, Mariana Henriques, and Nuno F. Rodrigues. COVID-19 Impact on Forecasting Emergency Department Visits Performance. *SSRN Electronic Journal*, 2022.

[38] Joshua Smith, Yingying Zhu, and Zheng Li. Public Transit Prediction During COVID-19 Pandemic. In *2022 IEEE 8th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS)*, pages 92–94, May 2022.

[39] Snehal_bm. Holt Winter's Method for Time Series Analysis | Holt Winter's Method, August 2021.

[40] Eric Stellwagen and Len Tashman. ARIMA: The Models of Box and Jenkins. page 7.

[41] Ivan Svetunkov. *Time Series Analysis and Forecasting with ADAM*.

[42] Sean J. Taylor and Benjamin Letham. Forecasting at scale. Technical Report e3190v2, PeerJ Inc., September 2017. ISSN: 2167-9843.

[43] Meng-Ju Tsai, Hsiu-Yuan Chen, Zhiyong Cui, and Yinhai Wang. Multivariate Long And Short Term LSTM-Based Network for Traffic Forecasting Under Interference: Experiments During COVID-19. In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, pages 2169–2174, September 2021.

[44] Glen L. Urban, John R. Hauser, and John H. Roberts. Prelaunch Forecasting of New Automobiles. *Management Science*, 36(4):401–421, 1990. Publisher: INFORMS.

[45] Peisong Wang, Minbo Sun, Zizheng Wang, and Yihang Zhou. A Novel CTR Prediction Based Model Using xDeepFM Network. In *2021 IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI)*, pages 635–638, September 2021.

[46] Peter R. Winters. Forecasting Sales by Exponentially Weighted Moving Averages. *Management Science*, April 1960. Publisher: INFORMS.

[47] Işil Yenidoğan, Aykut Çayir, Ozan Kozan, Tuğçe Dağ, and Çiğdem Arslan. Bitcoin Forecasting Using ARIMA and PROPHET. In *2018 3rd International Conference on Computer Science and Engineering (UBMK)*, pages 621–624, September 2018.

# Appendix A

# Data Sources

- Covid-19 confirmed cases, death, quatantines, recoveries, Covid-19 tests per day data European Centre for Disease Prevention and Control
- Vaccination data Bundesministerium für Gesundheit

## Ehrenwörtliche Erklärung

Ich versichere, dass ich die beiliegende Master-/Bachelorarbeit ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen. Ich bin mir bewusst, dass eine falsche Er- klärung rechtliche Folgen haben wird.

Mannheim, den 17.Jan.2023 Unterschrift