Author: Tim van Walsem Master Thesis: Business Analytics

Demand Forecasting in a Disrupted Bicycle Market

Demand Forecasting in a Disrupted Bicycle Market

"What methods can best be used to improve demand forecasting in the disrupted bicycle market and in what capacity does the Bullwhip Effect influence these models?"

Tim van Walsem

Vrije Universiteit Amsterdam
Faculty of Science
Business Analytics
De Boelelaan 1081a
1081 HV Amsterdam

Host Organisation:
Cannondale CSG
1 Cannondale Way
CT 06897, Wilton
United States

February 7, 2024

1 Preface

This research is conducted for the Master Project in Business Analytics at the Vrije Universiteit Amsterdam. It focuses on demand forecasting in an industry that has been heavily disrupted: the bicycle industry. Due to the recent COVID-19 pandemic and the subsequent supply chain crisis, historical sales exhibit divergent demand trends, which can consequently lead to misleading demand forecasts. The research delves into the challenges posed by these disruptions and investigates methods to enhance the forecasting performance of time series forecasting models.

The research problem at the core of this investigation revolves around developing robust forecasting models capable of accommodating the irregularities induced by COVID-19 and the following supply chain crisis. By addressing this challenge, this research strives to provide actionable recommendations for Cannondale to improve their demand forecasting. It has a focus on improving the current way of working which does not take into account the influences of exceptional events on sales.

The research is conducted at Cannondale, a renowned player in the bicycle industry, which is a part of the multinational Pon. This research is a collaborative effort between the Vrije Universiteit Amsterdam, the Pon Datalab and Cannondale. Hoping to bring together the business knowledge at Cannondale, the academic knowledge at the Vrije Universiteit Amsterdam and the technical expertise at the Pon Datalab. The engagement with the Cannondale team has provided a real-world context for the academic endeavors, allowing to bridge theory and practice.

I would like to express my gratitude to my supervisors for their guidance throughout this research journey. Special thanks to Mathisca de Gunst, representing the Vrije Universiteit Amsterdam, for her academic mentorship. Also I want to thank the second reader, René Becker, for his helpfull insights in structuring the research. And finally I want to thank Ellen Mik from the Pon Datalab for her practical insights and support.

I hope this report proves to be a meaningful contribution to the field of Business Analytics and provides actionable recommendations for enhancing demand forecasting strategies in the bicycle industry.

2 Summary

During the last couple of years, the bicycle industry has experienced some extraordinary times. Due to COVID-19, the demand for bicycles surged, and manufacturers were doing all they could to provide enough inventory. However, this was a daunting task due to a supply chain crisis that followed in 2021. Because of, among other reasons, COVID-19 and the Russian-Ukrainian war, along with labor shortages, bicycle manufacturers were experiencing challenging times in fulfilling all orders. This resulted in a period with low inventories, where bicycles were sold as soon as they became available. By now, the production of bikes has caught up, but the expected demand is significantly lower than anticipated. This has resulted in a surplus of inventory and exorbitant net working capital costs, forcing multiple bicycle manufacturers to sell their bikes at reduced prices. On the other hand, some new types of bicycles have rapidly gained interest over the past few years. E-mountain bikes, e-city bikes, and gravel bikes are quickly gaining popularity, creating opportunities for Cannondale to enter new markets and foster innovation, but also resulting in a highly competitive market.

Due to these unique influences on the bicycle market in the last couple of years, demand forecasting has been very challenging. Historical data has been greatly impacted by demand increases and stock shortages. Additionally, owing to the sudden spikes in demand, the market is likely to suffer from the Bullwhip effect. This effect involves increasing consumer orders causing amplified disturbances further up the supply chain. For Cannondale, it is essential to identify underlying trends and formulate realistic forecasts that are not unduly affected by exceptional events.

This research employs multiple methods to forecast demand in a disrupted market. Through demand unconstraining, anomaly detection, and compensating the bullwhip effect, an effort is undertaken to predict realistic demand using historical data. Subsequently, employing various time series forecasting methods, a projection is made for the coming six months. The focus for this report has been on forecasting the demand on Trail bikes.

The research uses multiple methods to account for the exceptional supply and demand in recent years. Using the order book to accommodate for stock-outs appeared to be difficult, since the order book has not always been representative for the actual demand. The research also dives into using Google Trends to adjust anomalies in the historical sell-in data. Which resulted in a substantial improvement of the forecasting accuracy.

To derive a reliable forecast, the research compares a Weighted Autoregressive Model, a Prophet Forecasting Model and a Random Forest Model. The Prophet Model showed the most reliable results, particularly when adjusting for outliers using Google Trends. Treating the COVID-19-affected period as a 'holiday' in the model, assigning it a lower weight in predictions, proved effective. This approach yielded a Weighted Average Percentage Error of 32% on a monthly basis and a mere 0.16% error over a six-month period. While the overall trend can be predicted with reasonable accuracy, forecasting monthly sales peaks remains challenging.

While the projected demand for Trail bikes appears promising, it should be taken into account that the models should be tested over a longer period of time and across a broader spectrum of models to get a better knowledge of the reliability of the method. Some bicycle models might be affected differently by COVID-19, the supply chain crisis and the current discounting in the market. Further research is therefore advised.

Contents

3 Introduction

The bicycle industry has been facing rapid changes over the last decade. Cities are improving their infrastructures, the electric bicycle market is rapidly growing, gravel and adventure bikes are gaining popularity, and more people turn to bicycles to reduce their carbon footprint. On the other hand, many new competitor brands have been introduced, the supply chain has been heavily disrupted, and the pricing has been increasingly competitive. These trends make it increasingly important for the supply chain to enhance its efficiency. To achieve this, reliable demand forecasting is crucial. And since demand forecasting is the first step in Cannondale's production planning, it is a fundamental step in optimizing their supply chain. [27][23]

In this research, an attempt has been made to forecast demand two year into the future. The models are built to predict the demand on Cannondale's different bicycle models for different countries. The forecasting model is mainly trained upon historical data. Since there were some turbulent circumstances in the last couple of years, the historical sales data has rarely been an unconstrained representation of the demand. Due to an exorbitantly high demand during COVID-19, the market experienced a Bullwhip effect. Due to the Supply Chain Crisis the sales has often been constrained. To obtain a reliable forecast, an attempt is made to uncover underlying trends using demand unconstraining and outlier handling. The resulting forecasting model is evaluated using both the Weighted Absolute Percentage Error and the bullwhip effect to which it is subjected.

Firstly in Section 4 an overview will be given on the business context. Here, the relevance to Cannondale and Pon Bike is described, and some of the used business terminology is explained. Then, in Section 5 literature is reviewed that forms the basis of this research. Demand unconstraining, outlier detection, hierarchical forecasting, the bullwhip effect, the used forecasting methods and the evaluation methods are shortly explained. In Section 6 the data preparation steps are described. Followed by an analysis of the datasets in Section 7. This provides insights in the effects of the COVID-19 sales boost, inventory shortages, open orders, unconstrained demand, product launches and the bullwhip effect.

Section 9 presents the results. The models were evaluated using the Weighted Average Percentage Error (WAPE). The best-performing model was the Prophet Model, with a WAPE of 29% and a absolute error of 0.16% over six months. The discussion and conclusion will be provided in Section 10, and recommendations for further research are outlined in Section 11.

This research has the goal to answer the following question; "Which demand forecasting models perform best in the disrupted bicycle industry, and to what extent does the bullwhip effect influence these models?"

4 Business Context

To provide more clarity on the report, this section dives deeper into the business context. It will dive into how it contributes to the overall goal of Pon and why it is of importance to Cannondale. Also some of the business processes and important terminology will be explained.

4.1 Pon

Cannondale is part of Pon. With a total of 15 different bicycle brands, Pon Bike is the largest bicycle manufacturer in the world. Aiming to produce the best bike in any market segment. When producing bicycles on such a scale, the supply chain gets increasingly complicated. Leaving a lot of opportunities to increase efficiency over the whole chain. In Figure 1 a schematic overview is provided that is created by the Pon Datalab. It shows where the opportunities lay for Pon Bike. Demand forecasting is one of the crucial first steps on which the rest of the supply chain can be optimized.

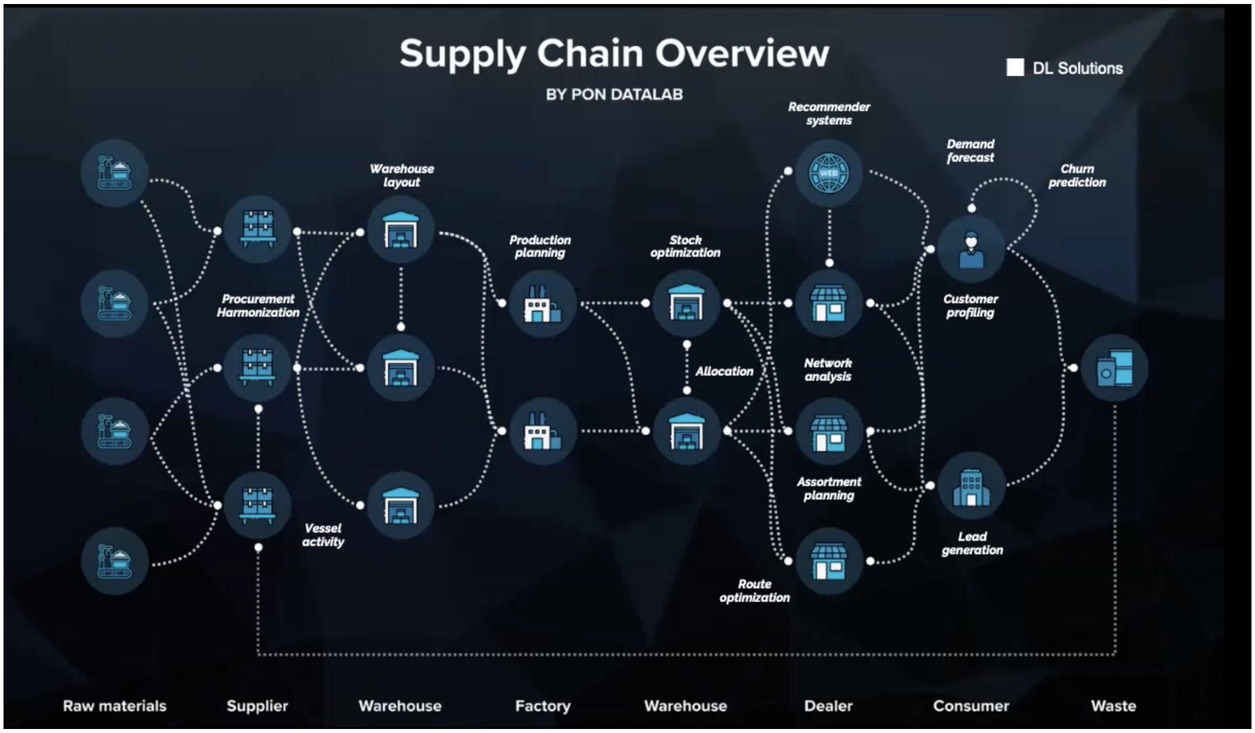


Figure 1: Strategy Pon Bike. *Source: Pon Datalab*

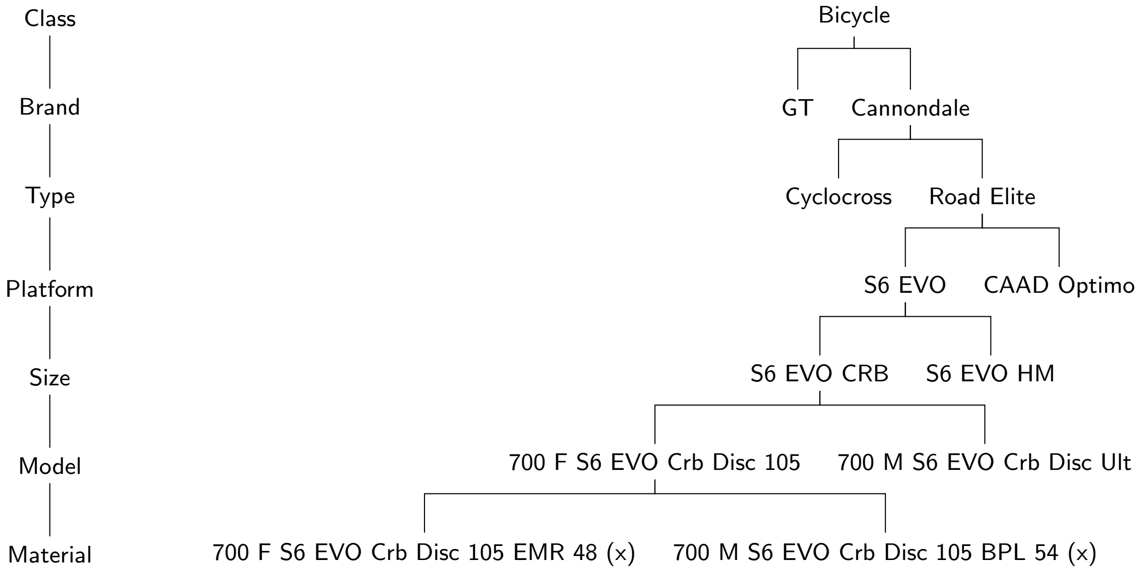
It is important to notice that the flow of goods goes from the manufacturer to the consumer. The flow of information however goes from consumer to manufacturer. The production must match the consumer's demand as close as possible. However, due to the delay of information about the demand, accurate forecasting is crucial in this process. If successful, demand forecasting would not only provide added value to Cannondale but also become a valuable tool for other companies within Pon Bike. A more accurate forecast could help reduce missed sales, increase production-to-dealer lead times and minimize the inventory holding period. Furthermore, it could be be a crucial improvement which helps optimizing the rest of the supply chain.

4.2 Sell-In, Sell-Out

In this report, there will often be mentioning of sell-in and sell-out data. It is very important to make a clear distinction between both terms. Sell-in refers specifically to the sales from the supplier (Cannondale) to the retailer (bicycle dealers/stores). Sell-out refers to the sales made from the retailer to its customers. Cannondale focusses on sell-in. However, some partners still provide information about there sell-out. It is important to note that a retailer can still sell their bikes to other retailers. During this research we do however make no distinction between those categories, and handle all sell-out data similarly.

4.3 Product Hierarchy

Forecasting can be done on different product levels. The sales of all bicycles can be predicted, but it also possible to create a forecast for a very specific model. The decisions made will be further clarified in Section 8. The hierarchy of the products is shown using the Master Data16.2, where all bicycle models are categorized. The product hierarchy is arranged as follows.


Class
Brand
Type
Platform
Size
Model
Material
Bicycle
GT
Cannondale
Cyclocross
Road Elite
S6 EVO
CAAD Optimo
S6 EVO CRB
S6 EVO HM
700 F S6 EVO Crb Disc 105
700 M S6 EVO Crb Disc Ult
700 F S6 EVO Crb Disc 105 EMR 48 (x)
700 M S6 EVO Crb Disc 105 BPL 54 (x)

4.4 Sales Versus Demand

Two often used terminologies that are used in this report are sales and demand. It is important to clearly define the differences between the two definitions. Sales concerns all bicycles that were actually sold in transaction. Demand on the other hand refers to the desire to buy a bike. When a bike is unavailable, there can be demand that will not be fulfilled as actual sales. Furthermore it is important to keep in mind that the demand of Cannondale's customers, the dealers, can deviate from the demand of the end consumer. It could occur that a dealer wants to buy bicycles in a certain period while its customers do not want to buy the bikes, leaving the dealer with surplus inventory. In this research the focus is mainly on the demand by dealers. Although the customer demand is also used to detect outliers and perform demand unconstraining.

5 Literature Review

In this section, the concepts that form the key components of this research will be reviewed. During the research multiple different models and techniques were used.

5.1 Demand Forecasting

As discussed in section 4, demand forecasting is a crucial step in optimizing the supply chain. The closer the production aligns to the actual demand, the fewer sales will be lost, the lower costs will be made on inventory and there will be less leftover inventory to be sold at an unfavorable rate. It also helps to better predict the future revenue and expenses. Therefore it can help provide some crucial information for the Sales & Operations department.

A supply chain is set into motion by a customer's request that has to be fulfilled. From there on the request goes upstream into the supply chain to put all parties involved into action. Therefore the information flows from customer to wholesaler to manufacturer and so on. In most cases forecasting gets more complicated higher up the supply chain due to informational delay.

J. Feizabadi[5] describes some key product characteristics that can influence the accuracy of a product's demand forecast. First of all the position in the supply chain is important. Closer to the consumer makes the forecasting easier since it generally is dependent on more and smaller orders. Furthermore, he states that products are easier to forecast when they have long product line life cycles, a constant demand, low margins, low inventory risk and less product variety. For Cannondale there is not one type of product. Some bicycle platforms have been around for decades, while other platforms such as gravel and electrical bicycles only recently entered the market and show more volatile demand patterns. Cannondale's models are reintroduced with minor changes on a yearly basis. Therefore the specific model has a short cycle, however the platform has a longer cycle.

Forecasting can be done in multiple ways. There are both quantitative and qualitative methods. Qualitative methods rely on expert opinion, market knowledge, surveys and the company strategy. Quantitative forecasting methods however, are based on historical data and are mathematically supported. This data-driven way of forecasting can be used to find more complex patterns and trends. Also it is easier to scale for more product groups and can be reused more often. [2]

The quantitative forecasting methods are often divided into two different categories. The traditional time series methods and the machine learning methods. Examples of the traditional methods are moving averages, time series decomposition and exponential smoothing. Examples of the machine learning models are Suport Vector Machines (SVM), Artificial Neural Networks (ANN) and Random Forests. Which models perform better can depend on the data, therefore is is beneficial to compare multiple methods.[21] [25] [19]

5.2 Demand Unconstraining

Demand unconstraining, or uncensoring as it also often is called in literature, is the activity of finding the actual demand based on constrained sales data. As discussed in 4 there is a difference between sales and demand. Jain et al. [12] describe that stock outs not only result in lost revenue, they also obscure the observations of the true demand. Additionally, products with a wide variety, short product lifetimes, and long lead times on their parts face increased availability vulnerabilities. To unconstrain the actual demand, they suggest using stock-out timing, defined as the moment a product reaches zero in the inventory.

Another industry with limited product availability is the aviation industry. Due to limited seats, they often have to determine demand based on constrained sales. I. Price et al.[18] use Gaussian process regression to forecast the demand on a given day. Gaussian process regression aims to find the underlying function that predicts the output function based on an input function.

An important note that has to be made in the case of bicycles, is what a dealer does when a material is out of stock. There are multiple options. Firstly, the dealer can switch to a similar bicycle model from the same

brand.Secondly, the dealer waits until the model is back in stock. Thirdly, the dealer does not buy any bikes at all. And finally the dealer goes to buy its bike at another brand. Although the last one seems to result in lost sales, it was just as often an advantage during the supply chain crisis, since other brands were suffering from similar problems. So by continuing all sales trend, there lays a risk of counting demand double.
[13] [14] [18] [12] [9]

5.3 Outlier Detection and Handling

Outliers can have a severe impact on the reliability of forecasting models. Unique events can result in extraordinary demand which, if not handled correctly, result in misleading forecasting results. N. Rennie et al.[20] describe how to detect outliers using both data analysis and time series decomposition.

Time series decomposition can be used to recognize the seasonal patterns and the long term trend in a data set. By compensating for the seasonal pattern and trend, the error can be acquired. If the error is outside of the set confidence interval, it is marked as an outlier.

5.4 Hierarchical Forecasting

Time series can often be structured in different hierarchies based on different dimensions in the data. For the creation of a forecast, it is important to choose a level of aggregation to perform the forecast on. Research by Syntetos et al.[24] provides a framework2 with four dimensions on which a forecast can be assessed. The echelon, the location, the product and the time. This framework can help identify in which dimension the forecasting is performed. For Cannondale, the echelon is as a manufacturer and a wholesaler. But the other dimensions can be aggregated as to what works best for the model.

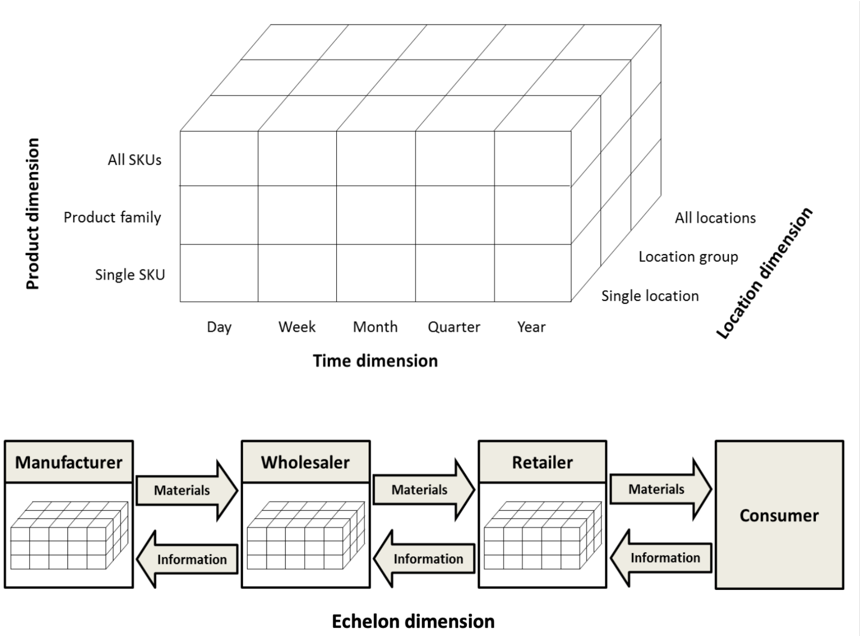


Figure 2: Supply Chain Structure: A FrameWork[24]

Hierarchical forecasting focuses on creating accurate forecasts, where the sum of forecasts at lower levels (like individual countries) must align with the forecasts at higher levels (such as continents). Hierarchical forecasting has been done in multiple ways. For example with bottom-up and top-down approaches, or combinations of both. With top-down approaches, the forecast is firstly done on the highest dimension. For example all sales in the US. Higher level forecasts are often less sensitive to incidental outliers and provides a more reliable broader picture. Bottom-up forecasting starts forecasting on the lowest level. For example on the level of one bicycle dealer. The forecasts are then aggregated upwards to build up to a broader level prediction. This can help to catch the more nuanced trend on a lower level. The top-down approach risks information loss, while the bottom-up approach is more error-prone.[4] [6] [8] [10] [11]

5.5 Bullwhip Effect

When a market experiences fluctuations, the Bullwhip Effect can occur. When the bullwhip effect occurs, minor fluctuations in consumer demand can result in amplified variation higher up in the supply chain. This effect mainly arises due to delivery delays, communication gaps and the way decision making processes work in a supply chain. For example, if a retailer suddenly experiences high demand on gravel bikes, he tends to immediately place an even larger order. Then the wholesaler will do the same, resulting in a ripple effect up the supply chain. Each stage in the supply chain overreact, resulting in excessive inventory. Something that can currently be observed in the bicycle industry.[22]

There are multiple ways to quantify the Bullwhip effect. For example the variance amplification factor, which calculates the increase of variation higher up in the supply chain. However, to calculate this, one needs insights in multiple levels of the supply chain, which is for this research very limited[3]. A more suitable method for this research is to calculate the Bullwhip effect based on the coefficient of variation.[16]

The coefficient of variation (CV) is given by:

$$CV = \frac{\sigma}{\mu}$$

The bullwhip effect can be calculated using the variance of orders:

$$\text{Bullwhip Effect} = \sum_{i=1}^{n} \text{Variance}_{\text{stage } i} \tag{1}$$

$$CV = \frac{S(O_{ikt})/[\sum_{t=1}^{T}(O_{ikt}) * \frac{1}{T}]}{S(D_{ikt})/[\sum_{t=1}^{T}(D_{ikt}) * \frac{1}{T}]},$$

Figure 3: Bullwhip Effect Formula

The bullwhip effect can be expressed mathematically using the following formula:

$$BE = \frac{1}{1-C}\left[\frac{\sigma_D^2}{\mu^2} + \frac{\sigma_O^2}{\mu^2} + 2 \cdot C \cdot \frac{\sigma_{DO}}{\mu^2}\right]$$

Where:

- BE is the Bullwhip Effect.
- C is the coefficient of correlation between demand and orders.
- σ_D^2 is the variance of demand.
- σ_O^2 is the variance of orders.
- σ_{DO} is the covariance between demand and orders.
- μ is the average demand.

5.6 Model Selection

For the selection of a suitable forecasting model, multiple factors are of importance. This research compares multiple models that all have different advantages and disadvantages. The models chosen are an autoregressive model, a random forest and the Facebook Prophet forecasting model.

- Autoregressive Model:
  - Pros
    - Simplicity of the model makes it easy to implement.
    - The model is based on seasonality and trend which are easily explainable.
    - Well-suited for Short-Term Forecasting.
    - Low Computational Requirements.
  - Cons
    - The model assumes stationarity.
    - The model is limited in Long-Term Predictions.
    - The model is sensitive to Outliers.
    - Difficult to add external factors.
- Random Forest:
  - Pros
    - Possibility to add many regressors.
    - The model gives insights in variable importance.
    - The model handles non-linearity.
    - The model is robust to outliers.
    - The model has a reduced Risk of Overfitting.
  - Cons
    - The model is difficult to explain (Black Box).
    - The model has intense computation power.
    - The model is not ideal for timeseries.
- Facebook Prophet:
  - Pros
    - The model is widely used and therefore trusted.
    - The model is easy to implement.
    - The model handles missing data well.
    - The model can handle holiday and special events.
    - The model has automatic seasonality detection.
  - Cons
    - The model is difficult to explain (Black Box).
    - There is limited control over model hyperparameters.
    - computationally Intensive for large datasets

5.7 Autoregressive Model

The autoregressive model (AR) is a time series forecasting model that predicts future values based on a linear combination of past observations. Mathematically, an autoregressive model of order p, denoted as AR(p), can be expressed as:

$$X_t = c + \phi_1 \cdot X_{t-1} + \phi_2 \cdot X_{t-2} + \ldots + \phi_p \cdot X_{t-p} + \epsilon_t$$

Here:

- X_t is the current value in the time series.
- c is a constant term.
- ϕ_i represents the autoregressive coefficients for each lag i.
- X_{t-i} are the past values up to lag p.
- ϵ_t is white noise, representing random error.

The autoregressive model assumes that the current value depends on its own past values, and the coefficients ϕ_i are estimated during the model training process.

The order p determines the number of past observations considered for prediction, influencing the model's complexity. The model is particularly effective when the time series exhibits a clear trend or pattern.

Stationarity is a common assumption for autoregressive models, ensuring that statistical properties remain constant over time. Model performance is often evaluated using metrics like mean squared error (MSE) to measure the accuracy of predictions against actual values.

Autoregressive models are widely used in various fields, including finance, economics, and signal processing. The model's simplicity and interpretability make it a valuable tool for time series analysis and forecasting. [7]

5.8 Prophet Model

Prophet is a forecasting model developed by Facebook that performs time series predictions. It's designed to handle datasets with daily observations and can incorporate seasonality, holidays, and special events.

Prophet decomposes time series data into three main components: trend, seasonality, and holidays. The trend captures the underlying growth or decline, seasonality accounts for periodic patterns, and holidays consider special events that might influence the data. One key strength of Prophet is its ability to handle missing data and outliers gracefully, providing more robust predictions. The model employs a decomposable time series model with components for daily effects, yearly seasonality, and holidays.

Mathematically, it can be expressed as:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t$$

Where:

- $y(t)$ is the observed value at time t.
- $g(t)$ represents the trend component capturing the overall growth or decline.
- $s(t)$ is the seasonality component, accounting for periodic patterns.
- $h(t)$ denotes the holiday effect, incorporating the impact of special events.
- ε_t is the error term, representing the unexplained variability.

Prophet utilizes an additive model, treating components as independent, simplifying the forecasting process. Additionally, it introduces a scalable method for Bayesian inference, improving computational efficiency while maintaining accuracy. The model's flexibility allows users to include custom seasonalities and adjust uncertainty intervals for a more comprehensive analysis.

Prophet's automatic detection of changepoints helps identify significant shifts in the time series, aiding in capturing sudden changes in the data. The model also incorporates holidays as potential drivers of changes, offering adaptability to various datasets.

5.9 Random Forest Model

Random Forest is an ensemble learning method widely used for predictive modeling. It operates by constructing a multitude of decision trees during training and outputs the average prediction of the individual trees for regression tasks or the mode for classification tasks.

Each tree in a Random Forest is built using a random subset of the training data and a random subset of features at each split. This randomness enhances model diversity, making the ensemble robust and less prone to overfitting.

Mathematically, let's denote the training data as (X_i, y_i), where X_i is the input feature vector and y_i is the corresponding target variable. The algorithm constructs N decision trees, denoted as $T_1, T_2, ..., T_N$. For each tree, a random subset D_j of the training data and a random subset F_j of features are selected.

The prediction of the ensemble for a new input X is given by the average (for regression) or the mode (for classification) of individual tree predictions:

$$\hat{y}_{\text{ensemble}}(X) = \frac{1}{N} \sum_{j=1}^{N} \hat{y}_j(X)$$

Here, $\hat{y}_j(X)$ represents the prediction of the j-th tree. This ensemble approach leverages the wisdom of crowds, combining the strength of multiple trees to yield a more accurate and stable prediction.

Random Forests are known for their versatility, capable of handling various data types and complex relationships. They also provide feature importance scores, helping identify the most influential features in the prediction process. Overall, Random Forests are a powerful and user-friendly tool for predictive modeling in diverse scientific applications.[26] [19]

5.10 WAPE

To assess the performance of a forecasting model, various methods are available. Commonly used techniques include the Mean Absolute Error, the Mean Square Error, and the Mean Absolute Percentage Error, each with its own advantages and disadvantages. In this research, the Weighted Average Percentage Error (WAPE) is employed.

The WAPE is a metric used to assess the accuracy of forecasting models in scientific and business contexts, particularly suitable for situations where the magnitudes of the predicted and actual values vary significantly. WAPE is expressed as a percentage and is calculated as the weighted sum of the absolute percentage errors.

The formula for WAPE is given by:

$$WAPE = \frac{\sum_{i=1}^{n} w_i \cdot \left| \frac{y_i - \hat{y}_i}{y_i} \right|}{\sum_{i=1}^{n} w_i} \times 100\%$$

Where:

- n is the number of observations.
- y_i is the actual value at time i.
- $\hat{y}_i$ is the predicted value at time i.
- w_i is the weight assigned to the i-th observation.

WAPE measures the average percentage deviation of the predicted values from the actual values, with each observation's contribution weighted by w_i. The weights can be assigned based on the significance or importance of individual observations.

One advantage of WAPE is its sensitivity to errors, capturing both the direction and magnitude of forecasting inaccuracies. However, it is important to interpret WAPE cautiously, especially when dealing with situations where the denominator (y_i) is close to zero, as it can result in large percentage errors.

In summary, WAPE provides a comprehensive evaluation of forecast accuracy, considering both relative and absolute errors while allowing for the incorporation of weights based on the significance of individual observations. It is a valuable metric for researchers and practitioners aiming to assess the performance of forecasting models in diverse applications. [28] [1] [17] [15]

6 Data Preparation

For the data preparation, multiple steps were taken. This section shows the most significant steps, how they were performed, which decisions were made and what the connections between the different data sources are. A more extensive overview on the input data sets itself can be found in Section 16. The further pre-prossessing of the data will be discussed in the methodology8.

6.1 Data Cleaning and Validation

Most of the data used in this research was exported from SAP. With the exception of the Google Trends data and the sell-out data.

For the **sell-in data**, only the data on bicycles was provided. Leaving out accessories, components and helmets. The data provided was for the multiple brands that fall under the same bicycle group; GT, Schwinn, Charge and Cannondale. Only the Cannondale bicycles were used in the research. The data sets had no empty cells. To check for reliability, the monthly sales figures were compared to the internal reporting of the company. The numbers in those reports matched and were therefore deemed credible. For the sell-in data negative sales (which are returns), were excluded from the research and therefore filtered out. This was done since Cannondale wants to meet the original demand. The handling of returned bicycles is therefore out of scope for this research.

The **order data** was somewhat harder to use. Since there is a different way of using the order system in the USA compared to Europe. During inventory shortages, the USA would only accept orders when there were bicycles available, making it unclear which orders were missed. In Europe however, all orders were accepted, resulting in longer waiting times for already accepted offers. In Section 8 will be explained how the two different ways of working were handled.

The **stock data** provided insights in the availability of all Cannondale bicycle materials at all different storage locations at the last day of a given month. The data could not be acquired on a daily basis due to the size of the data set. Therefore, when aggregating on a daily level, the stock level would be based on the last available value that month. Cannondale only sends their bikes from the continent the bike is stored. Therefore, when looking at the stock, storages were grouped in an European, American and remaining storage locations.

The **customer data** provided insights into the dealers' locations. When training the models, the lowest level of detail considered was the customer's country. Forecasting on a more detailed location, for example the state, posed some difficulties. The sell-in data was linked to the customer location. However, for larger dealer groups only their central location was being used as the location of sales. Making it seem that all sales was in one state while it should actually be divided over multiple states where the dealer was active. Since this could potentially result in biases in the seasonal trends, the province/state was left out.

The **master data** was used to match sell-in and order data to the correct class, brand, type, platform, size, model and material. This enabled to choose different hierarchies in the product dimension.

The **sell-out data** was received from one of our sales partners, REI. A weekly overview was delivered which contained the sales and the remaining stock at that given moment in time. REI performs around 20% of the American sales.

The **Google Trends data** was externally acquired based on people searching for a specific bicycle platform. The provided data set gave a monthly overview which gave search activity on a scale from 0 to 100. For this research only individual search activity was acquired, so there was no comparison between different platforms. To ensure that the search term was representative with the demand, Figure 4 also compares the search term 'buy bicycle' and 'Cannondale Supersix Evo'. Here can be seen that the seasonal pattern is similar however the market trend may change.

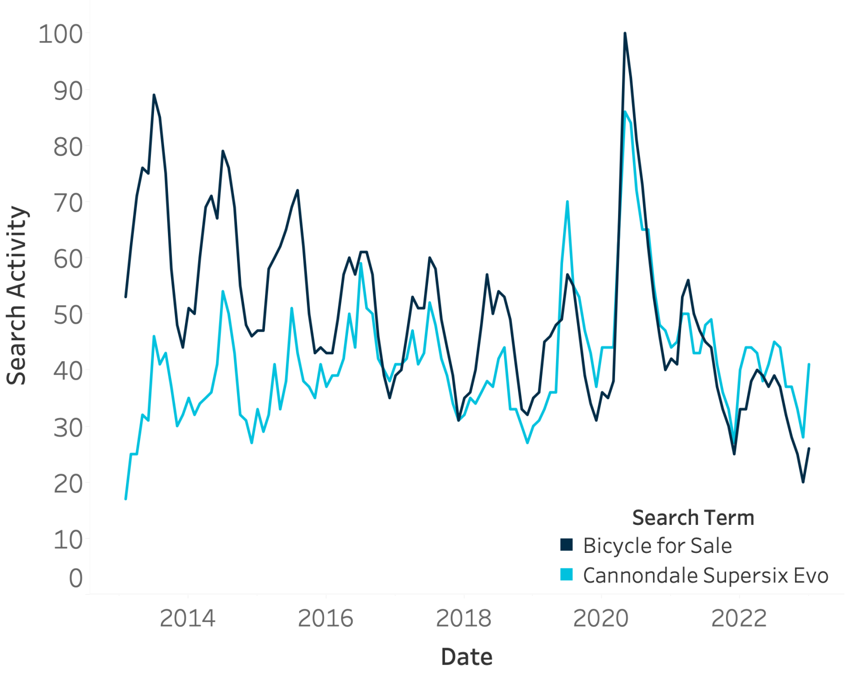


Figure 4: Search Trend

6.2 Data Integration

The data was integrated based on the links that could be made between the different data sources. A schematic overview of the data sets and the interconnections can be found in Section appendix15.

6.3 Data Aggregation

As a preparation for the data models, new data sets were created based on daily, weekly and monthly aggregations. For every time period in each aggregation, different rows would be created for each country and each product size. Making it possible to forecast based on a higher or lower granularity.

6.4 Data Splitting

The splitting of the dataset was done using in Python after the data aggregation. The split was done so that the test period was exactly 1 year. So either 12 months, 52 weeks or 365 days, based on the aggregation. One of the issues faced was that during the beginning phase of the research, the most of the test data was in a period of inventory shortages. However, when newer data was added this problem was resolved. A broader analysis on the data split will be shown in section 7. Splitting the dataset randomly was not done, since that made it possible to learn from future trends and thus not reliable and repeatable for the future.

7 Data Analysis

Confidential

8 Methodology

This section will go into the steps that were taken to get to the final model. Often multiple possible steps were tried and either kept or discarded based on the effect on the forecasting results. Section 8.2.1, 8.2.2 and 8.2.3 go into the demand unconstraining and outlier detection. Sections 8.2.4, 8.4 and 8.3 go into the forecasting models and section 8.5 addresses the performance measurement. Any preceding steps are described in sections 6 and 7.

8.1 Programming Tools

The tools used in this research are Tableau and Python. Tableau was used for the data analysis and the visualisation of the results. Tableau was chosen since it is a useful tool to create a interactive visualisations that are easy to use by the business. Python was used for all steps ranging from the data preparation until the model evaluation. Python was chosen for the ability to handle statistical modeling and the broad availability of external packages. The external python packages that were used can be found in the Appendix 15.

8.2 Data Preprosessing

After the data preparation as described in Section 6, multiple preprosessing steps were performed to uncensor demand, to compensate for incidental events and delivery delays and to minimize the bullwhip effect. The European market and the American market used different ways of working for handling orders during inventory scarcity. On the American market, orders were not yet accepted until they were able to deliver bicycles. On the European market they accepted the orders and allocated the available bicycles whenever they were available. Because of this reason, two ways of demand uncensoring were attempted. For the European market based on late deliveries, for the American market based on stock-outs

8.2.1 Late deliveries

During COVID-19, due to the increased demand, Cannondale was facing difficulties to fulfill all incoming orders. Inventory was low and all incoming bikes were immediately being sold. This often resulted in dealers that had to wait months before they could receive their order. In those cases, the sell-in data does not fully represent the demand of the bicycles. To make up for that, sales that were delivered late were shifted from the delivery date to the original request date. This provided insights in the delays and at what time there actually was demand.

8.2.2 Stock-outs

On the American market however, the request date was not always representative for when there was actual demand. Therefore, the sell-in data combined with the order data still represents an underestimate of the actual demand. During stock-outs, the actual demand was likely to be censored. Therefore, for this research, the lost sales had to be compensated.

Two methods were tried and compared. The first method was to give the expected months of inventory as a feature for the machine learning model. However, this could only be done by predicting the future months of inventory. Therefore this method posed difficulties in predicting future demand.

The second method tried was to use Google Trends data to give an uncensored demand. As soon as the demand was below the monthly expected sales, the sales would be based on the google trends data compared to the months that year that the data was not constrained. This resulted in a data set which better represented the unconstrained demand.

8.2.3 Outliers

Due to COVID-19, the demand was far higher than it would have been under normal circumstances. Therefore the outliers have to be compensated to better represent the overall trend in bicycle demand. Since the sell-in data only incorporated some years of history and has a very incosistent monthly pattern, it was difficult to find the outliers. Therefore the Google Trends data was used. This gave ten year of historical interest in bicycle platforms and gave clear insights in when the demand got to exceptional levels.

To identify the outliers, the ThymeBoost package was used. By combining time series decomposition with gradient boosting, a seasonal pattern and a trend is identified. Using a 90 % confidence interval, the trend and pattern can be used to see if historical data points are considered outliers. If this is the case, the deviation is calculated by taking;

$$deviation = actualvalue/predictedvalue$$

Where the actual value is the google search trend and the predicted value is the value predicted by the ThymeBoost. Then, to compensate for the exceptional demand, the sales is divided by the deviation. This is done only in the months where the google search trend is considered an outlier.

8.2.4 Random Forest

To implement the random forest, the Scikit Learn package RandomForestRegressor is used. The model used a input table with multiple different columns. For the random forest, The month, quarter and year were given as an input. The month and quarter were given in different variables for each month and quarter, which was either 1 if it was that month or quarter, 0 if it was not.

The random forest model was tried using multiple methods. First of all the model was used without any data manipulation. The forecast was based purely on the historical sell-in on a monthly level. Secondly the model was used after compensating for the stock-outs. Thirdly the forecast was performed compensation for the trend outliers and lastly it was done by compensating for the trend outliers and stock-outs combined. For all methods, the predictions were made on a test period of 6 months.

For the random forest model it was also tried to give the inventory and stockout as a input variable. This way the model could help predict future values based on the expected inventory and demand. However, these values were quite difficult to decide on and this method was therefore discarded.

8.3 Autoregressive Model

For the autoregressive model, the ForecasterAutoreg package was used. The model takes the time as input. And forecasts the coming 6 months. The forecast was performed using the same preparation steps ad described for the Random Forest. For all four methods, also an extra method was tried. By providing a weight to a historical interval, the model can learn more on the period where the sales was not influenced by exceptional events. Therefore, the period highly effected by COVID-19 was given a lower weight of 0.25. This was done from March 2020 to June 2022. This resulted in eight different models. For all methods, the predictions were made on a test period of 6 months.

8.4 Prophet Model

For the Prophet model, Prophet was used. The model takes a timestamp and the sales as input for the forecast. Again, the same four preparation methods were compared. One of the features of Prophet is to give a holiday period as an input. This period is then given a lower weight in prediction future demand. For the same four preparation methods as described above, there was also tried to give a holiday input. The holiday was the COVID-19 influenced period, ranging from March 2020 to June 2022. This resulted in eight different models. For all methods, the predictions were made on a test period of 6 months.

8.5 Model Evaluation

8.5.1 WAPE

To analyze the performance of the models, two different metrics are used. Firstly, the WAPE was calculated on a monthly level. All six predicted months were compared to the actual sales. However, since the month when sell-in peaks can vary, the research also looks at the percentage error. By comparing the sum of the total predicted sell-in and the total sell-in of the test set, the percentage error was calculated. Since the bicycles are ordered on a yearly basis, it is more important that the total predictions are close to the actual sales than that each month is predicted precisely. The percentage error is calculated using the following formula:

$$PercentageError = \frac{\text{Total Predicted Sell-in} - \text{Total Sell-in of the Test Set}}{\text{Total Sell-in of the Test Set}} \times 100$$

9 Results

Confidential

10 Conclusion and recommendations

10.1 Summary of Main Conclusions

This research focuses on forecasting future sell-in in a recently highly disrupted bicycle market. By proper preprocessing of historical data and comparing multiple different forecasting models, it aims to find a robust model that is less effected by the exceptional sales due to COVID-19 or missed sales due to a disrupted supply chain. The key conclusions are that the methods do help to improve the forecasting methods, however there is still need to further investigate if the models perform well on different bicycle models and different markets.

10.2 Relationship between Conclusions and Literature

An exploration of the existing literature provided various sources on time series forecasting, however sources that focused on the data preparation were more scarce. The research therefore hopes to have a solid foundation based on previous forecasting models, but to enrich the field with new methods that contribute to a realistic representation of historical sales trends.

10.3 Significance to Cannondale

For Cannondale to use the forecasting models, there should be a further investigation that tests the performances over a longer time and over a wider model scope. However, the learnings on the trend manipulation can already provide useful insights that can be taken into account in the current production forecasting. Using Google Trends to pinpoint outlying demand and using the ThymeBoost model to calculate the impact of the outliers showed promising results. These insights contribute to a nuanced comprehension of market trends and consumer behavior, offering strategic advantages for informed decision-making.

10.4 Discussion of Conclusions Contrary to Hypotheses

While some conclusions align with initial hypotheses, certain methods did not yield the expected results. Using the open orders to compensate for the stock-out did not seem to improve the forecast. A reason for this could be that bicycle dealers were just placing multiple orders to get their hand on whichever bicycles they could acquire. Therefore the order book could be exaggerating the actual demand.

10.5 Limitations and Their Impact on Reliability

The research focused on the Trail bicycles, a more mature bicycle model with a stable yearly sales. Other models will have less historical data, more exceptional innovation and be more sensitive to the competitor's behaviour. This can make the market even more volatile and harder to predict. Therefore, reproducing the method for other bicycle models requires further investigation on the reliability and robustness.

10.6 Recommendations for Business Processes

In light of the research outcomes, subtle adjustments to the business processes within the bicycle industry could enhance the accuracy of sell-in forecasting. By compensating the exceptional demand that COVID-19 caused, a more realistic market trend can be forecast.

10.7 Suggestions for Further Research

While this study has delved into multiple methods for sell-in forecasting, there are various subjects for further exploration. These methods will be further discussed in Section 11.

11 Further Research

This research aims to create a forecasting model that is able to deal with a volatile historical market. To acquire a reliable model it was mainly important to achieve a robust model that found the underlying trends. Therefore, it may have limitations in accurately predicting exceptions in demand. Forecasting will never be flawless, since the actual demand will always be depending on an endless amount of factors. Yet, some improvements could clearly help reveal underlying trends in the bicycle industry even better. In this section highlights multiple components of the research that could be improved.

11.1 Market data

Most data used in this research is focused on Cannondale's historical performances. However, as a source this can be prone to incidental influences such as a good year in the Tour de France, one new design or a groundbreaking innovation. These do however not always last and can be misleading for the overall interest in bicycles. Data on the overall bicycle market shows a more stable trend and is therefore less prone to exceptional occurrences that happen on a small scale. By combining the intelligence from other Pon companies, Pon's partners and available market sources, a more stable trend could be analyzed. This would create a model that is closer to the expected trend and leaves it up to the business to decide if future demand will again diverge due to exceptional conditions.

11.2 Sell-out

The sell-out data available was limited to only one store. other available sell-out data sources are limited or also biased due to the dealer's market focus. A better insight on the sell-out patterns could be very useful in understanding the client's demand. Which has a much more stable seasonality and can therefore be predicted easier. When Cannondale knows the client's demand, it can collaborate with the dealers to push their deliveries at the right moment to the right dealers. This would help Cannondale plan ahead further and more accurately.

11.3 Internet Traffic

The internet traffic that was analyzed in this research was based on Google Trends. However it can be difficult to pinpoint exactly the motivation of the individual that searches for a bike. Was it with the intention to buy a bike, or merely out of curiosity. To uncensor the actual demand on a bicycle there might be more reliable sources that are more likely to be used by prospective buyers. One of out partners, 99 Spokes, is a website that compares bicycles on their prices and spec levels, across the whole industry. The activity on that website might be a closer representation of the demand. However, this has to be investigated.

11.4 Forecasting Models

In this research, multiple different machine learning and time series models were investigated. This research field is however changing rapidly and new models are being introduced regularly. Methods to improve your models such as ensemble methods (Bagging, boosting, stacking) could help increase the accuracy further. These could be combined into a hybrid model with both the machine learning models and the time series models.

11.5 Long-term Forecasting

With more data input, a forecast could be made for a longer period of time. This could help make even larger decisions such as choosing where a factory should be opened or were product development should focus on. This could also help provide more insights in the years without COVID influences

11.6 New Products

With the methodology used in this research, innovative product launches or niche products were difficult to forecast. A future study might be focusing on better understanding the expectations for a newly launched product and predict sales even before creating a new type of bike.

12 References

References

[1] N. K. Ahmed, A. F. Atiya, N. E. Gayar, and H. El-Shishiny. An empirical comparison of machine learning models for time series forecasting. *Econometric reviews*, 29(5-6):594–621, 2010.

[2] J. S. Armstrong. *Principles of forecasting: a handbook for researchers and practitioners*, volume 30. Springer, 2001.

[3] A. Boudghene Stambouli, D. Zendagui, P.-Y. Bard, and B. Derras. Deriving amplification factors from simple site parameters using generalized regression neural networks: implications for relevant site proxies. *Earth, Planets and Space*, 69(1):1–26, 2017.

[4] J. B. Edwards and G. H. Orcutt. Should aggregation prior to estimation be the rule? *The Review of Economics and Statistics*, pages 409–420, 1969.

[5] J. Feizabadi. Machine learning demand forecasting and supply chain performance. *International Journal of Logistics Research and Applications*, 25(2):119–142, 2022.

[6] G. Fliedner. An investigation of aggregate variable time series forecast strategies with specific subaggregate time series statistical correlation. *Computers & operations research*, 26(10-11):1133–1149, 1999.

[7] P. H. Franses and R. Legerstee. A unifying view on multi-step forecasting using an autoregression. *Journal of Economic Surveys*, 24(3):389–401, 2010.

[8] Y. Grunfeld and Z. Griliches. Is aggregation necessarily bad? *The review of economics and statistics*, pages 1–13, 1960.

[9] H. Huang and Q. Liu. Intelligent retail forecasting system for new clothing products considering stock-out. *Fibres & Textiles in Eastern Europe*, (1 (121)):10–16, 2017.

[10] K. Hubrich. Forecasting euro area inflation: Does aggregating forecasts by hicp component improve forecast accuracy? *International Journal of Forecasting*, 21(1):119–136, 2005.

[11] R. J. Hyndman, R. A. Ahmed, G. Athanasopoulos, and H. L. Shang. Optimal combination forecasts for hierarchical time series. *Computational statistics & data analysis*, 55(9):2579–2589, 2011.

[12] A. Jain, N. Rudi, and T. Wang. Demand estimation and ordering under censoring: Stock-out timing is (almost) all you need. *Operations Research*, 63(1):134–150, 2015.

[13] R. Kalla, S. Murikinjeri, and R. Abbaiah. An improved demand forecasting with limited historical sales data. In *2020 International Conference on Computer Communication and Informatics (ICCCI)*, pages 1–5. IEEE, 2020.

[14] N. Kourentzes, D. Li, and A. K. Strauss. Unconstraining methods for revenue management systems under small demand. *Journal of Revenue and Pricing Management*, 18:27–41, 2019.

[15] H. L. Lee, V. Padmanabhan, and S. Whang. Information distortion in a supply chain: The bullwhip effect. *Management science*, 43(4):546–558, 1997.

[16] R. Metters. Quantifying the bullwhip effect in supply chains. *Journal of operations management*, 15(2):89–100, 1997.

[17] Z. Michna and P. Nielsen. The impact of lead time forecasting on the bullwhip effect. *arXiv preprint arXiv:1309.7374*, 2013.

[18] I. Price, J. Fowkes, and D. Hopman. Gaussian processes for unconstraining demand. *European Journal of Operational Research*, 275(2):621–634, 2019.

[19] S. Punia, K. Nikolopoulos, S. P. Singh, J. K. Madaan, and K. Litsiou. Deep learning with long short-term memory networks and random forests for demand forecasting in multi-channel retail. *International journal of production research*, 58(16):4964–4979, 2020.

[20] N. Rennie, C. Cleophas, A. M. Sykulski, and F. Dost. Identifying and responding to outlier demand in revenue management. *European Journal of Operational Research*, 293(3):1015–1030, 2021.

[21] J. Shahrabi, S. S Mousavi, and M. Heydar. Supply chain demand forecasting; a comparison of machine learning techniques and traditional methods. *Journal of Applied Sciences*, 9(3):521–527, 2009.

[22] M. S. Sodhi and C. S. Tang. The incremental bullwhip effect of operational deviations in an arborescent supply chain with requirements planning. *European Journal of Operational Research*, 215(2):374–382, 2011.

[23] Q. Sun, T. Feng, A. Kemperman, and A. Spahn. Modal shift implications of e-bike use in the netherlands: Moving towards sustainability? *Transportation Research Part D: Transport and Environment*, 78:102202, 2020.

[24] A. A. Syntetos, Z. Babai, J. E. Boylan, S. Kolassa, and K. Nikolopoulos. Supply chain forecasting: Theory, practice, their gap and the future. *European Journal of Operational Research*, 252(1):1–26, 2016.

[25] L. F. Tratar, B. Mojškerc, and A. Toman. Demand forecasting with four-parameter exponential smoothing. *International Journal of Production Economics*, 181:162–173, 2016.

[26] N. Vairagade, D. Logofatu, F. Leon, and F. Muharemi. Demand forecasting using random forest and artificial neural network for supply chain management. In *Computational Collective Intelligence: 11th International Conference, ICCCI 2019, Hendaye, France, September 4–6, 2019, Proceedings, Part I 11*, pages 328–339. Springer, 2019.

[27] M. D. Xames, J. Shefa, and F. Sarwar. Bicycle industry as a post-pandemic green recovery driver in an emerging economy: a swot analysis. *Environmental Science and Pollution Research*, 30(22):61511–61522, 2023.

[28] X. Yuan, X. Zhang, M. Wang, D. Zhang, et al. Quantifying the bullwhip effect in a reverse supply chain: The impact of different forecasting methods. *Mathematical problems in engineering*, 2022, 2022.

13 List of Figures

List of Figures

14 List of Tables

List of Tables

15 Appendices

15.1 Python Packages Used

- Pandas
- Numpy
- Sklearn
- statsmodels
- datetime
- matplotlib
- prophet
- math
- skforecast

16 Datasets

For the research, multiple datasets were used. In this section all the input data will be described and explained. In order to keep the overview comprehensible, irrelevant variables are omitted from the overview. The data sets that were provided by Cannondale were mainly focused on the historical performances.

16.1 Customers

Description: Cannondale sells its bicycles not to the end client directly, but through bicycle dealers. Those dealers are located all around the world. This data set contains address information on all the dealers Cannondale sells its bikes to. The data is exported from SAP.

Size of data set: 29,164 rows
Time of export: 01-07-2023
Variables: Sales Org, SAP #, Customer Name, Customer Group, Street Address, City, Region-State-Province, Postal Code, Country, Sales District, id

Table 1: Customers Data variables

Sales Org	SAP #	Customer Name	Customer Group	Country
Street Address	City	Region	Postal Code	Sales District

16.2 Material Master

Description: The Material Master data set contains information on all different bicycle models and the product groups they belong to. All bicycles are divided into different groups. The hierarchy from large to smaller groups in BPSA Group, Class Name, Brand Name, Type Name, Size Name. Material Name is the unique identifier for a bicycle with a specific size and color. The data is exported from SAP.

Size of data set: 22,400 rows
Time of export: 01-07-2023
Variables: Material, Material Name, Class Name, Brand Name, Type Name, Size Name, Created on, Type of Bicycle, BPSA Group, Frame Size, Type of Frame, Gender, Material Type, Model Year, Platform, Platform Name, Model Name, Category (Dorel), Category Name, Active Model Year, id

Table 2: Material Master Data Variables

Material	Material Name	Class Name	Brand Name	Type Name
Created on	Type of Bicycle	BPSA Group	Frame Size	Type of Frame
Material Type	Model Year	Platform	Platform Name	Model Name
Category Name	Active Year	id		

16.3 Order book

Description: The order book contains all the historical and pending orders from customers. It shows the quantity a dealer want to order on a material level. There are three different dates available. The first one is the 'Item Created On', which shows when the order was put into SAP. The second date is the 'Request Date', which shows when the dealer would like to retrieve its order. The third one is the 'Material Availibilty Date', which indicates when the bicycles are available for shipping. The data is exported from SAP.

Size of dataset: 115,343 rows
Time range: 01-01-2019 to 30-06-2023
Variables: Material, Sales Doc, Sales Org, Country, Item Created On, Material Availability Date, Request

Date, Pon Category, BPSA Category, Sales Group, Sold To #, Allocated Qty, Brand, Class, Model Name, id

Table 3: Order Variables

Material	Sales Doc	Sales Org	Country	Item Created On
Request Date	Pon Category	BPSA Category	Sales Group	Sold To #
Brand	Class	Model Name	id	

16.4 Sell-in

Description: The sell-in data contains all sales from Cannondale to the dealers. The sales are recorded on a Material level and show the quantity of the specific material that is sold to a dealer at a specific time. The data is exported from SAP.

Size of dataset: 1,806,984 rows
Time range: 01-01-2019 to 30-06-2023
Variables: Sales Document, Sold to party, Brand, Material, Model Name, Ship to Country, Pon Category, BPSA Category, Posting date, Sales Group, Unique ID, Units Sold, Sales Org, id

Table 4: Sell-In Data Variables

Sales Document	Sold to party	Brand	Material	Model Name
Ship to Country	Pon Category	BPSA Category	Posting date	Sales Group
Unique ID	Units Sold	Sales Org	id	

16.5 Sell-out

Description: Sell-out data is harder to acquire, since Cannondale does not have direct access to dealers their sales. The sell-out data used is gathered by combining weekly reports from one of our largest dealer-groups, REI. Although it gives a very insightful view of the seasonality of the sales at REI, it is to be noted that REI tends to focus more on adventurous activities which results in a higher representation of bicycles suited for that purpose.

Size of dataset:29690 rows
Time range: 01-01-2019 to 30-06-2023
Variables: Class, Sub Class, Style, Color, WTD RTL SLS, WTD AUR, WTD UNITS SLS, TY MTD UNITS SLS, TY STD UNITS SLS, TY WTD WOS - 4 WKS, TY WTD UNITS OH, TY WTD COST OH, Date, Week #, REI Model, Model #, PH1 - Class, PH2 - Brand, PH3, PH4, Category, Platform, Model, Strategic Category

Table 5: Sell-Out Data Variables

Class	Sub Class	Style	Color	WTD RTL SLS
WTD AUR	WTD UNITS SLS	TY MTD UNITS SLS	TY STD UNITS SLS	TY WTD WOS - 4 WKS
TY WTD UNITS OH	TY WTD COST OH	Date	Week #	REI Model
Model #	PH1 - Class	PH2 - Brand	PH3	PH4
Category	Platform	Model	Strategic Category	

16.6 Stock

Description: The stock data contains a monthly snapshot of the inventory on the last day of the month. The data contains the inventory for all storage locations worldwide and it shows the availability on a material level.

Size of dataset: 2,127,267 rows
Time range: 01-01-2019 to 30-06-2023
Variables: Material, Material description, Plant, Plant description, Class Brand Fiscal Period, Stock Qty, id

Table 6: Stock Data Variables

Material	Material description	Plant	Plant Description	Description
Description	Fiscal Period	Stock Qty	id	

16.7 Google Trends

Description: Contains the relative amount of searches over the last 10 years. The search activity is put on a scale ranging from 0 to 100. 0 being no activity and 100 the highest monthly activity in the last 10 years. To acquire the data, a similar proces is followed for every different platform. The trend was found for the search term "Cannondale [PLATFORM NAME]". For example "Cannondale Topstone. This was done for the last 10 years and for both the worldwide and the American trend.

Size of dataset: 482 rows
Time range: 01-07-2013 to 30-06-2023
Variables: Month, Search Activity, Platform, Country

Table 7: Google Trends Data Variables

Month	Search Activity	Platform	Country