Google Trends for a Time Series Marketing Analysis: the LVMH case

Thu Ha Mai
Mariana Montserrat Cerra
Juan Esteban Cruz

Aix Marseille School of Economics
2022

Contents

1 Introduction

Given that 70% of luxury purchases are estimated to be influenced by online interactions (D'Arpizio and Levato 2017), Google Trends could be essential in the growth of big data in the luxury industry. Consumers of luxury items are actively using Google to get inspiration, select the best designs, and make purchases with a single tap (Boone 2016). In 2016, Google alone processed 20 petabytes of data per day and received more than 4 million search inquiries every minute from 2.4 billion internet users (Wedel and Kannan 2016). Google can identify the next major luxury trend using data analytics on this kind of online behaviour (Bain, 2016).

Given its capacity to serve as a leading indicator for forecasting important variables of interest, it is no secret that Google Trends are increasingly influencing business decision-making in a variety of industries (see, for instance, Yu et al. 2019; Siliverstovs and Wochner 2018; Zhang et al. 2018). The luxury business can also profit from using Google Trends to forecast luxury-related variables, such as future purchase decisions, the efficacy of marketing initiatives, and online consumer brand engagement.

Since Google accounts for more than 75% of all internet searches worldwide, we agree with Gordon's (2017) assertion that Google Trends can serve as a gauge for consumer behaviour online (Net Market Share 2019). As a result, we think the luxury sector should take into account relevant Google Trends, often known as "fashion consumer Google Trends," which, according to McDowell (2019), can help firms recognize consumer trends and profit from them.

This report, therefore, illustrates the use of Google Trends data to extract brand search volume, either to assess advertising or predict sales. The choice of different LVMH brands as the data set of interest for this study was influenced by several factors. Firstly, it is the only group present in all five major sectors of the luxury market: Wines and Spirits, Fashion and Leather Goods, Perfumes and Cosmetics, Watches and Jewelry and Selective Retailing. However, in this study, we only focus on the three first sectors for the sake of simplicity. Secondly, the fact that the LVMH Group today comprises more than 75 exceptional Maisons in several countries enables us to track different brands' search volumes worldwide. This is a source for further analysis and forecasting of future online behavioural trends for LVMH so that the management can put in place a series of actions for improving its online footprint and be better positioned to compete with other luxury fashion brands. Thirdly, LVMH believes in the importance of digital innovation and sees 'online' as the first access point to its brand. Therefore, Google Trends-based analytics has the potential to help the brand ensure its online content remains highly relevant to its consumers.

The remainder of this report is organized as follows. Section 2 introduces the characteristics of Google Trends data and the overall process of data collection. Section 3 then focuses on outliers detection and treatment, which is one of the most difficult parts of time series data preprocessing, and the outliers' possible explanations based on real events. Section 4 concentrates entirely on clustering techniques, algorithms and results while section 5 is dedicated to the forecast application and evaluation for each series and clustering. The report concludes in Section 6 by pointing out the key marketing insights given in our analyses.

2 Dataset

2.1 Google Trends Data

According to Google Support website[1], Google Trends gives users access to a vast, unbiased sample of actual Google search requests. It is anonymized (no one is personally identified), categorized (determining the topic for a search query), and aggregated (grouped together). This enables us to show interest in a specific subject from all over the world or even a particular city's geography.

There are two samples of Google Trends data that can be accessed:

- Real-time data is a sample covering the previous seven days
- Non-realtime data, which includes samples from 2004 and the 72 hours prior to your search, is a different set from real-time data

For this project, we use non-real-time data since we aim to study the search queries in a timeframe of 5 years from now.

What is most useful for storytelling with Google Trends data is its normalized data. This means that when we examine the evolution of interest in a topic's search volume, we explore that interest as a percentage of all searches on all topics on Google at that particular moment. The interest in a topic as a percentage of all searches on all topics on Google at the same location and time is what we mean when we talk about a topic's regional search interest.

That normalization is crucial: the number of people searching on Google constantly changes — in 2004, search volume was much smaller than it is today, so raw search numbers will not give us any way to compare searches then and now. By normalizing our data, we can make deeper insights: by comparing different dates, different countries, or cities. The context of our numbers also matters. Google Trends index its data to 100, where 100 is the maximum search interest for the time and location selected.

2.2 Data Collection

2.2.1 The desired dataset

As explained above, the dataset needed for the analysis contains the Google Trends Interest Over Time of:

- Different brands of LVMH
- Various categories/ business groups: perfumes and cosmetics, fashion and leather goods, wines and spirits
  - Perfumes and Cosmetics: 'Sephora', 'Make Up For Ever', 'Benefit Cosmetics', 'Fenty Beauty', 'Guerlain'
  - Fashion and Leather goods: 'Louis Vuitton', 'Christian Dior, S. A.', 'Givenchy', 'Marc Jacobs', 'Kenzo', 'Loewe', 'Fendi', 'Celine'

[1] https://support.google.com/trends/answer/4365533?hl=en&ref_topic=6248052

– Wines and Spirits: 'Dom Pérignon', 'Veuve Clicquot Ponsardin', 'Chandon', 'Ruinart', 'Château Cheval Blanc', 'Hennessy', 'Moët Chandon'

- In a timeframe of 5 years from 27th November 2017 to 27th November 2022
- In different countries: USA, France, Japan since these are the three most prominent markets of LVMH worldwide

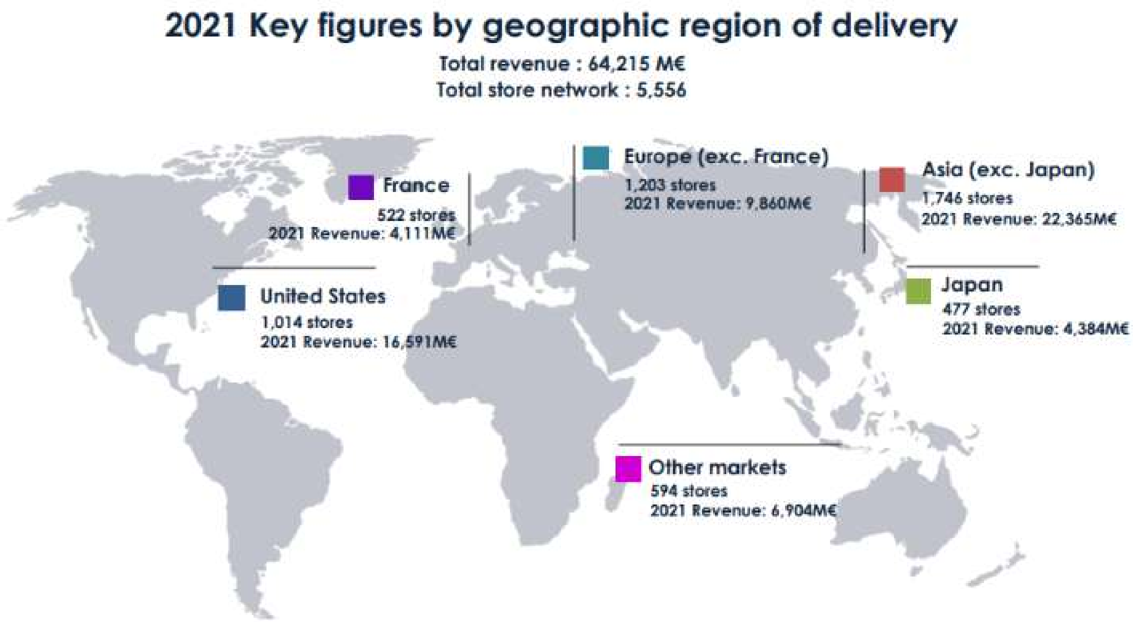

Figure 1: USA, France, and Japan as three biggest country-level markets[2]

2.2.2 The overal process

To pull data from Google Trends, we used an unofficial API called Pytrends developed and maintained by GitHub community[3]. This Python package can automate various tasks, such as retrieving large amounts of data from Google Trends, which is essential for large-scale analysis. However, there are two main challenges when collecting data using Pytrends.

- The current automated Python approaches are inaccurate since they do not query exact keywords: Python offers a variety of ways to extract Google Trends data. None of them, however, creates automatic scripts that can extract exact keywords. The fact that words frequently have many meanings is not rare. For instance, "Chandon" might refer to an LVMH brand of wine and spirits or a place in France. Due to the fact that "Chandon" incorporates both search phrases, a simple search will produce confused results. We have to take this into consideration to use the proper terms when searching.
- It takes time to pull each term one at a time manually: Although Google Trends provides the "Compare" function to compare keywords, the disadvantage is that it scales the results from

[3]https://github.com/GeneralMills/pytrends

0 to 100 based on the most popular term entered. When compared to a popular keyword, the less popular one will quickly lose sensitivity. For instance, when we compare the ”Christian Dior” and the ”Kenzo” brands, ”Kenzo” will essentially show a flat line. In this case, reporting the increasing trend of ”Kenzo” searches will result in significant mistakes. Therefore, it is advised to independently pull the ”Christian Dior” trend and the ”Kenzo” trend.

Firstly, to avoid ambiguity, the keywords must be specific. The pytrend.suggestions method in Pytrends can produce several keyword suggestions. The initial recommendation is typically the most popular one. We can search for those specific keywords in the ”mid” column.

	mid	title	type
0	/m/05nn45	Sephora	Retail
1	/g/1ywtjp667	MAKE UP FOR EVER	Topic
2	/m/0cmwj9	Benefit Cosmetics	Cosmetics company
3	/g/11f53qtpm	Fenty Beauty	Topic
4	/m/0_w4m7v	Guerlain	Topic
5	/m/03h90x	Louis Vuitton	Fashion company
6	/m/04s4g6	Dior	Fashion company
7	/m/0hgt7_w	Givenchy	Fashion label
8	/m/063xvrw	Marc Jacobs	Fashion label
9	/m/0x20kx_	Kenzo S.A.	Fashion company
10	/m/02rf9lb	Loewe	Fashion company
11	/m/063y0r6	Fendi	Fashion label
12	/m/0463vt1	CELINE	Fashion company
13	/m/05lzhs	Dom Pérignon	Topic
14	/m/09dp6h	Veuve Clicquot	Drink company
15	/m/02r5c01	Chandon	Winery in Napa County, California
16	/m/027fswc	Ruinart	Champagne
17	/m/05hh6j	Château Cheval Blanc	Winery in Saint-Émilion, France
18	/m/03hz33	Hennessy	Drink company
19	/m/03hz6q	Moët & Chandon	Winery

Figure 2: The exact keywords for LVMH brands

Then, we pulled Google Trends data by exact keywords by country using the following parameters.

- KEYWORDS: We specified the keywords we need by entering different brand names
- COUNTRY: Companies with an international footprint could use this function to get trends across countries
- DATE INTERVAL: We chose the time range of the trend as five years from now
- SEARCH TYPE: We selected the source of the search as Google Search

After pulling the data, we convert the data from weekly to monthly for the ease of analysis.

	date	Sephora-US	Make Up For Ever-US	Benefit Cosmetics-US	Fenty Beauty-US	Guerlain-US	Louis Vuitton-US	Christian Dior-US	Givenchy-US	Marc Jacobs-US	...	Loewe-JP	Fendi-JP	Celine-JP	Dom Pérignon-JP	Veuve Clicquot-JP	Chandon-JP	Ruinart-JP
0	2017-12-03	68	80	84	65	66	60	20	49	36	...	8	47	30	55	66	0	30
1	2017-12-10	78	76	81	82	89	67	23	50	42	...	10	45	31	79	61	0	54
2	2017-12-17	82	76	77	76	83	75	24	55	47	...	8	56	35	69	54	36	0
3	2017-12-24	92	100	89	100	82	84	19	57	43	...	8	52	33	83	51	52	0
4	2017-12-31	58	76	70	60	81	59	16	45	27	...	9	49	33	80	41	0	0

Figure 3: The final dataset with brands by country

Below is an example of Keyword Web Search Interest Over Time for the brand Louis Vuitton in the three aforementioned markets.

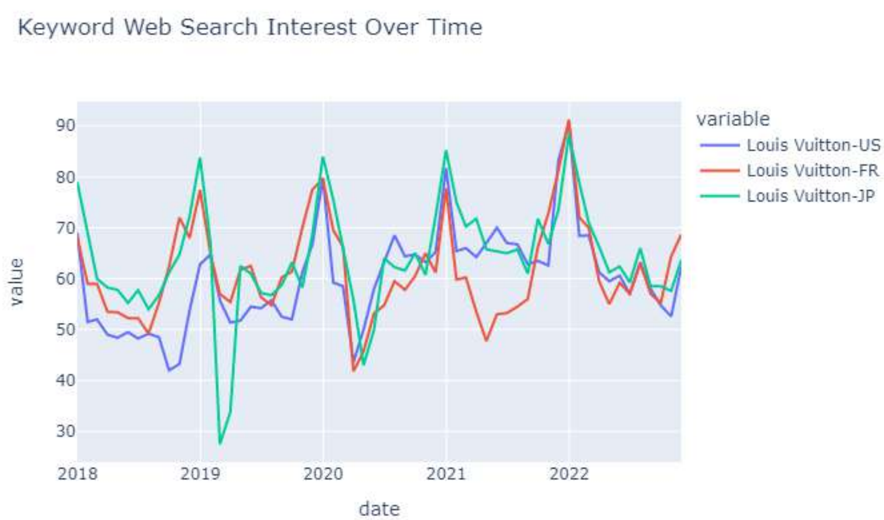

Figure 4: ”Louis Vuitton” Web Search Interest Over Time in the USA, France and Japan

3 Dealing with Outliers

It is necessary to identify the outliers and deal with these observations as they are values that can generate noise in the identification of clusters and even more in forecasting the time trends we are using. In the dataset we are analyzing, outliers could be characterized as observations that have a sudden increase in Google searches and can exist due to the appearance of news and events that put the brands in the public eye and impact Google searches for short periods. As we have different time series, manual outlier identification and correction would take much time; therefore, a good understanding and implementation of an existing package would improve the outlier analysis.

For the outlier analysis, we based on the library Kats, a toolkit created by Facebook’s Infrastructure Data Science team that aims to analyze time series quickly. Kats has been created thanks to the great importance of time series analysis for industries and Data Science. This package focuses on understanding the data's characteristics, detecting regressions and anomalies, and forecasting the future behavior of the time series. Punctually, we use the Kats "OutlierDetector module" to identify and deal with the outliers' values in all the time series.

Initially, the algorithm makes a seasonal additive decomposition of the time series for detecting outliers. An additive decomposition implies that the time series consists of three parts that added constitute the complete behavior of the data through the analyzed period. Each time series is decomposed into its trend, seasonal behavior, and residual. The trend component will be able to capture the long-term trajectory; this long-term trajectory is helpful as it will capture long-lasting circumstances that affect the trend. On the other hand, seasonality can catch periodic and generally predictable behaviors over the years, such as the logical increase in Google searches (and sales) of products and fashion houses during December due to the Christmas season. For the detection of outliers, the focus is placed on the residual of the decomposition of the series. Kats will generate the residual by removing the trend and the seasonality if it is strong enough. All the points that are considered outliers are those that lie outside three times the interquartile range of the residual.

To show how exactly we used the Kats package to detect outliers, the trend in Google of the fashion house "Kenzo" will be an example. The following plots show the time series for the trend "Kenzo" in the three countries we are considering.

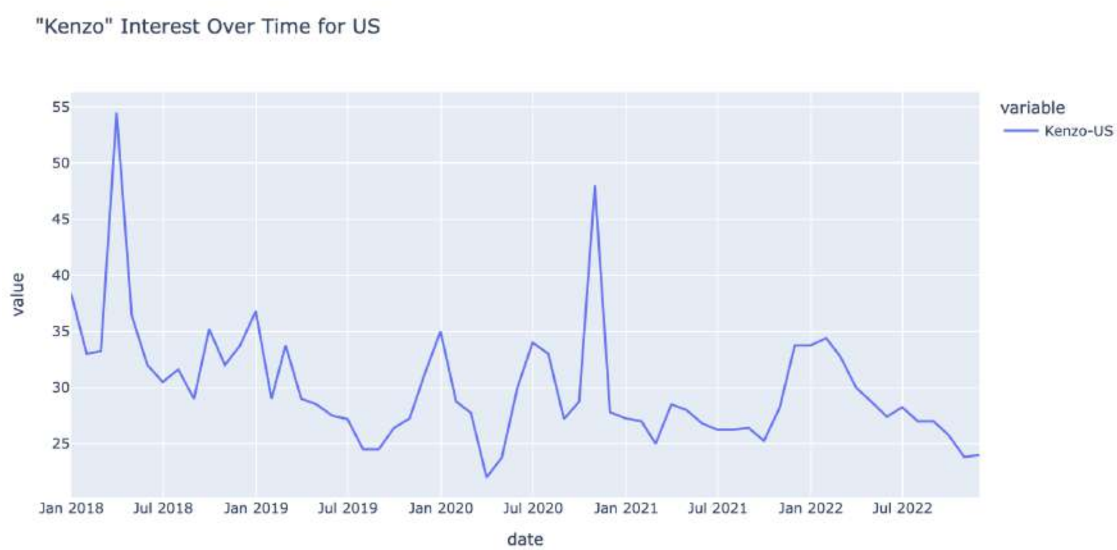

Figure 5: "Kenzo" Interest over time for the US

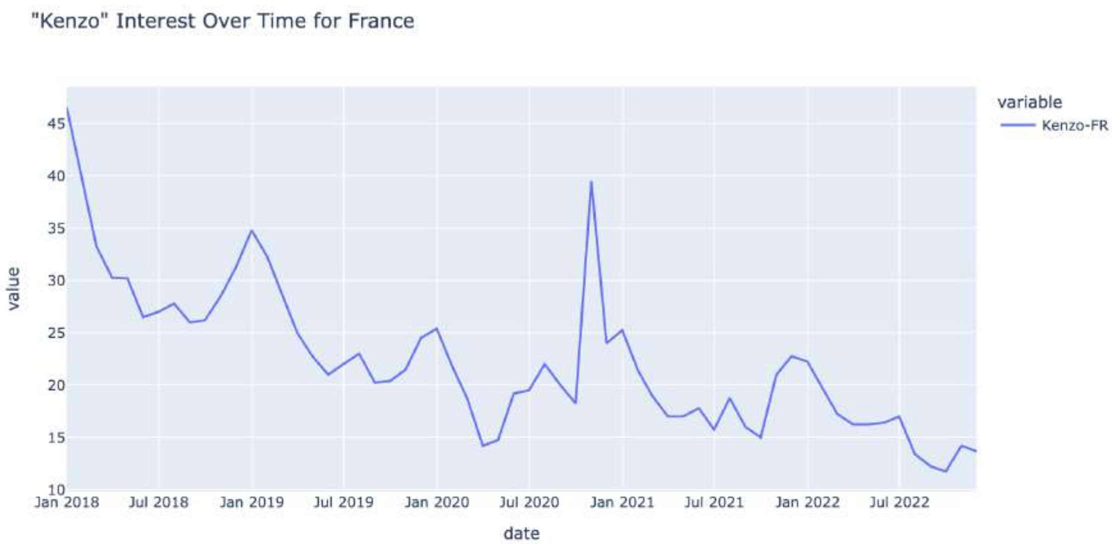

Figure 6: ”Kenzo” Interest over time for France

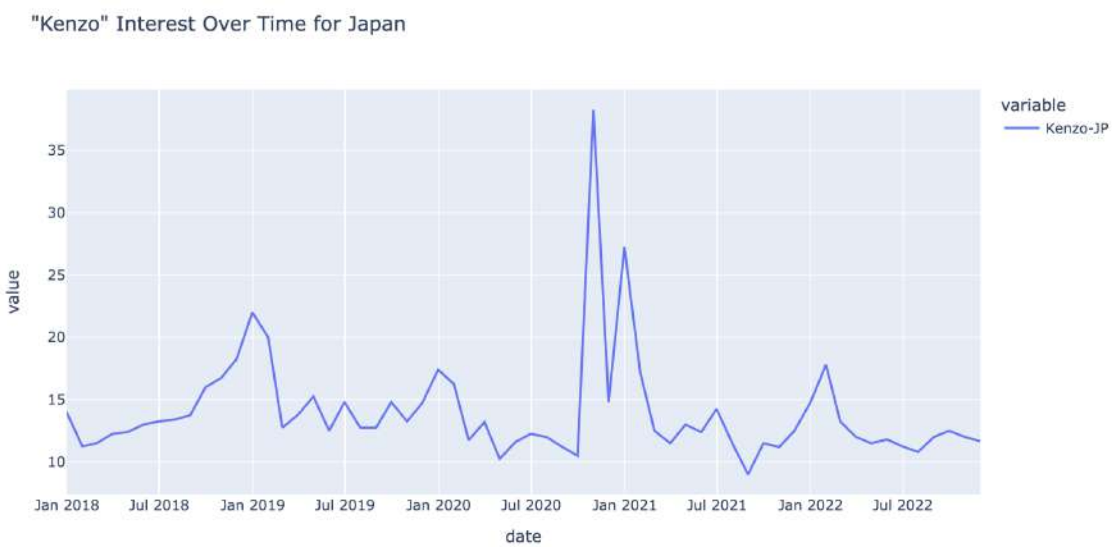

Figure 7: ”Kenzo” Interest over time for Japan

Based on the plots of the time series for the ”Kenzo” trend, we can believe in the existence of two outliers for the United States trend, one at the beginning of 2018 and the other in October 2020. For the France ”Kenzo” trend plot, the observations for December 2017 and October 2020 can be considered outliers. In the case of Japan, we can believe outliers in the values of October and December 2020. The Kats package will generate the series decomposition shown in the following graph for ”Kenzo” US. Here the trend is identified with the orange line in the first graph, the seasonality in the second graph, and the residual in the third one.

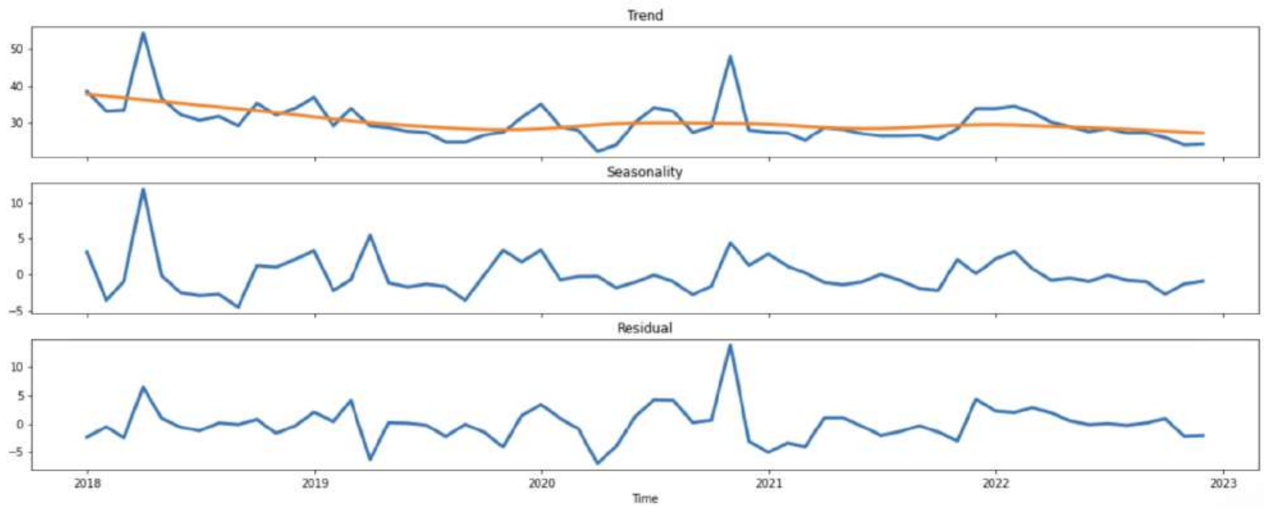

Figure 8: Time series decomposition for "Kenzo US"

When applying the detection algorithm, it detects two outliers for The United States trend (3/2018 and 10/2020), one outlier for the trend in France (10/2022), and two outliers for Japan (10/2020 and 12/2020). Kats allows the removal of the outliers but also interpolates them for not having missing values and creating by a linear interpolation a value that fits the data better. The removal of outliers and their interpolation can be seen in the following graphs shown for the "Kenzo" trend.

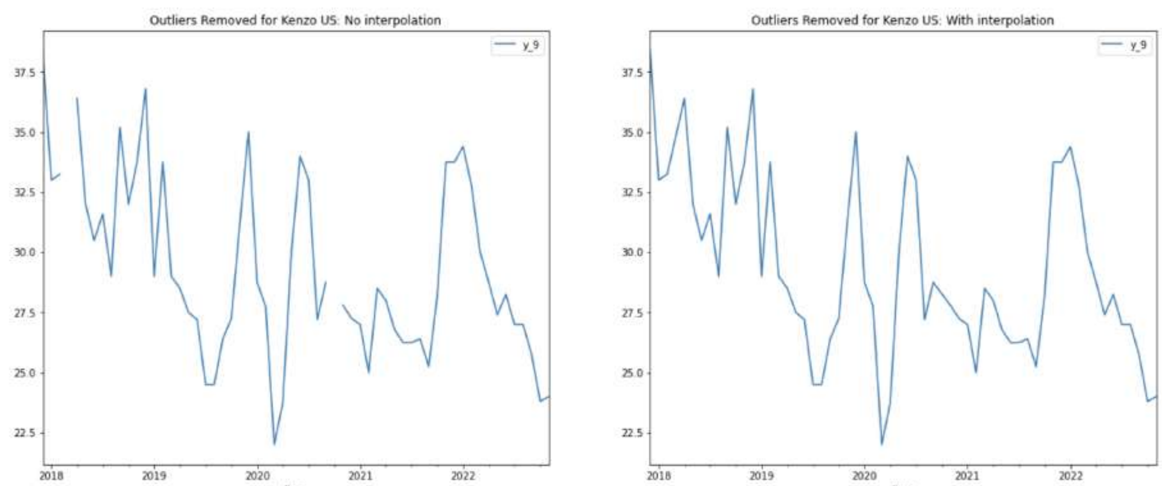

Figure 9: Outliers removed for "Kenzo" in the US. Left side: without interpolation. Right side: With interpolation.

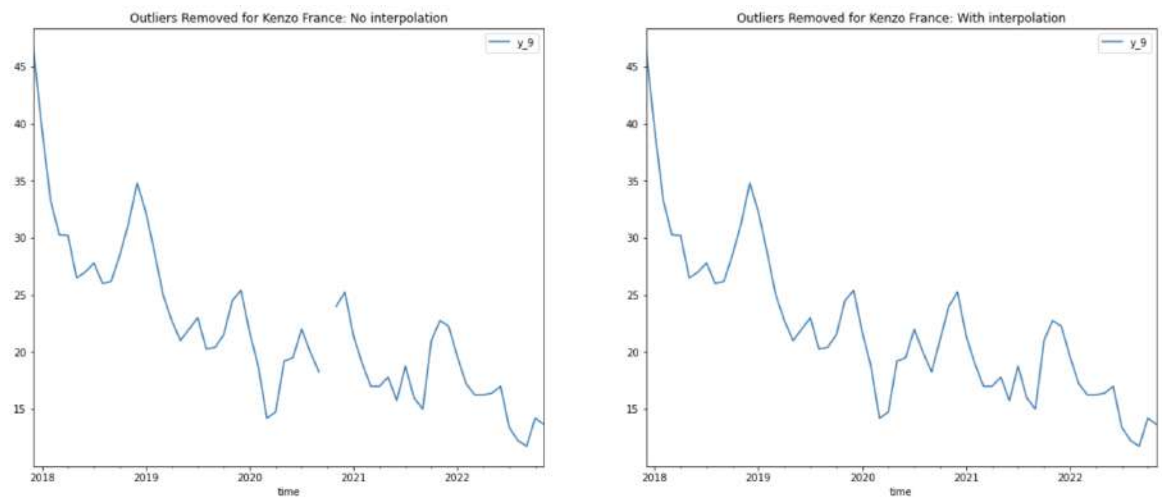

Figure 10: Outliers removed for "Kenzo" in France. Left side: without interpolation. Right side: With interpolation.

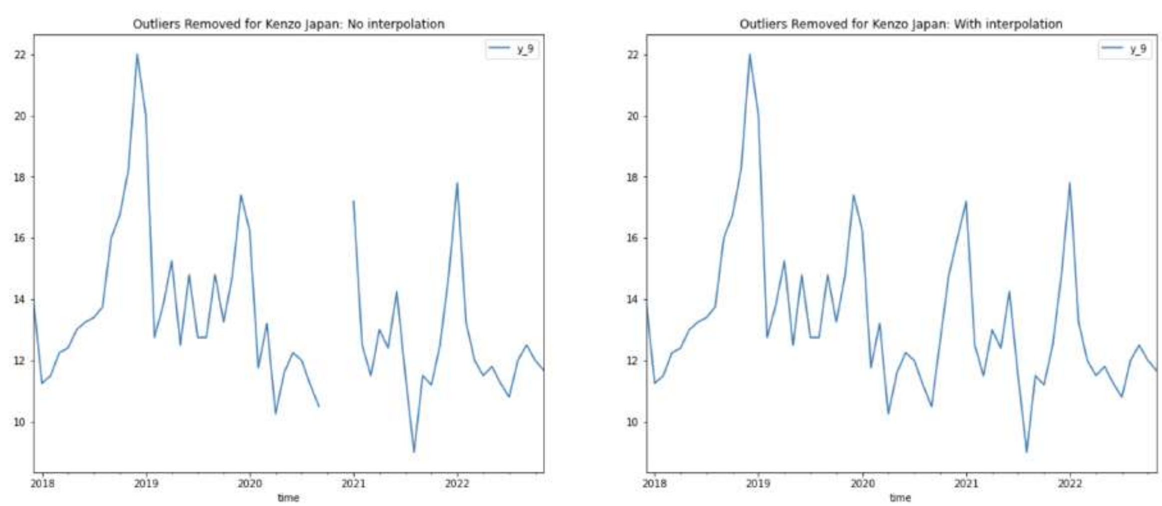

Figure 11: Outliers removed for "Kenzo" in Japan. Left side: without interpolation. Right side: With interpolation.

Beyond dealing with outliers as observations that can put noise in the results, it's essential to understand that these outliers are related to events that atypically increased Google searches. In the case of "Kenzo", which we are taking as an example, the presence of the same atypical data in the three countries is evident. In October 2020, Google searches for "Kenzo" increased significantly in all three countries thanks to the fact that the brand's Japanese fashion designer Kenzo Takada, the founder of the brand, died on the 4 of October at the age of 81 in Paris after contracting Covid-19. This process of identification and interpolation of outliers was carried out for all series. Likewise, possible causes that could generate such atypical increases in Google searches were sought. The results of the identified outliers and their possible causes are shown in the following table.

US Trends	Outliers	Possible reason
Fenty Beauty	12/2017	It is still a new brand as it was founded in September 2017 and they launched the first matte lipstick collection.
Christian Dior	12/2021 08/2022 09/2022 20/2022	They launched the Christmas atelier of dreams collection. In September they relaunched the new Miss Dior Eau de parfum which increased the Google searches for the following months.
Givenchy	12/2019 12/2021	For Christmas, they launched the ”Red Line” holiday makeup collection. The brand launched a new collection with Bstroy, it is a mix of physical and digital (NFTs) art.
Kenzo	03/2018 10/2020	Momento fall collection. Designer Kenzo Takada dies.
Loewe	01/2022	The brand had one of the most viral fashion shows and collections.
Celine	09/2018 02/2022 11/2022	First show under Hedi Slimane’s direction.
Dom Pérignon	12/2020	High alcohol consumption for holidays, this is associated with the first Christmas of the pandemic.
Veuve Clicquot	12/2020	
Ruinart	12/2020	
Moët & Chandon	12/2020	

France Trends	Outliers	Possible reason
Givenchy	03/2018	They launched the limited edition Couture collection.
Kenzo	10/2020	Designer Kenzo Takada dies.
Loewe	01/2022	The brand had one of the most viral fashion shows and collections.
Celine	10/2018 11/2018 12/2018 02/2019 03/2019 04/2019 05/2019 06/2019 07/2019	The first show under Hedi Slimane’s direction and ready-to-wear collection in January that could have a long-lasting effect.
Dom Pérignon	12/2020	High alcohol consumption for holidays.
Ruinart	12/2017	
Veuve Cliquot	04/2020	Instagram campaign with Didier Mariotti’s virtual tasting lesson.
Chandon	02/2019	Saint Valentine’s limited edition box for its Moët Impérial Rosé champagne.

Japan Trends	Outliers	Possible reason
Make up Forever	02/2019	Conversation of the LVMH Japan K.K. president Norbert Laurent. Festive for the setsubun plus an increase in searches of various Google trends for February 2019.
Benefit Cosmetics	02/2019 3/2019	
Guerlain	02/2019	Increase in searches of various Google trends for February 2019 and Guerlain's success with Abeille Royale skincare and Rouge G makeup.
Louis Vuitton	02/2019	Increase in searches of various Google trends for February 2019 and Louis Vuitton opened a pop-up store in Tokyo.
Marc Jacobs	04/2020	Marc Jacobs fashion show and make-up campaign through Instagram.
Kenzo	10/2020 12/2020	Designer Kenzo Takada dies.
Loewe	12/2021 01/2022	Christmas and the brand had one of the most viral fashion shows and collections.
Veuve Clicquot	12/2020 12/2021 06/2022	High alcohol consumption for holidays. 250 anniversary of one champagne with a new international campaign and a new collection.
Chandon	11/2018	
Moët & Chandon	02/2019	Increase in searches of various Google trends for February 2019.

These outliers can be generalized to the existence of new collections and fashion shows associated with the leading fashion events that occur during the year, such as the fashion weeks in September and February, the week of haute couture in January, and men's fashion week in January. Finally, from the identification of the outliers, it is evident that the pandemic -beyond generating hasty reductions in Google searches- caused long-term effects that were translated into the trend of the time series. The fact that Covid-19 is part of the trend is due to the long duration of the pandemic and lockdowns.

4 Clustering

Clustering is a data mining approach that divides homogenous data into uniform groups (clusters) where we do not have important information about those groups. Its primary purpose is to detect comparable time series patterns and reveal underlying information in the data.

In this section, we will begin by presenting the methods for selecting the number of clusters, which were the Elbow method and the Silhouette method. Then, we will refer to the two clustering algorithms: k-means and hierarchical clustering, about their drawbacks and advantages, and the motivation for why we decided to continue with one of them. We will also introduce the DTW Barycenter Averaging (DBA) algorithm, which was used for computing the average of the series within a cluster. Following, we will move to the results and present the cluster assignment for the trends in each country and draw some insights about how they relate to each other. Finally, we will plot each cluster together with its average cluster, which will be used later in the forecasting section.

4.1 Techniques to determine the number of clusters

4.1.1 Elbow method

The ideal number of clusters into which the data may be grouped is a critical stage for any unsupervised technique.

The Elbow Method is one of the most prominent approaches for determining the ideal value of k. For it, we run k-means clustering for a range of clusters k and calculate the sum of squared distances from each point to its assigned centre for each value (distortions). When the distortions are plotted and the plot resembles an arm, the optimal value of k is the "elbow" (the point of inflexion on the curve).

4.1.2 Silhouette Method

Because there might not be a distinct point where the curve begins to flatten in the Elbow method, it might be difficult to select a reasonable number of clusters to employ. Therefore, for a more confident choice, the silhouette Method is employed in conjunction with the Elbow Method.

The silhouette score quantifies how similar an object is to its own cluster (cohesion) when compared to other clusters (separation). The silhouette has a value between -1 and +1, with a high value indicating that the object is well-matched to its own cluster but poorly matched to nearby clusters.

4.2 Clustering Algorithms

4.2.1 K-means Clustering

K-Means clustering is a method for unsupervised learning. It is based on distances or dissimilarity metrics between multivariate data and cluster centroids. The method generates data clusters by dividing samples into k groups and minimizing the sum of squares in each group. The k-means method is designed to reduce the average distance of components from the cluster center.

The distance measure is often used to determine the similarity/dissimilarity of two time series. The Euclidean distance metric and the DTW distance measure are well-known distance measurements.

The Euclidean distance metric is inadequate for time series since it is insensitive to time shifts and ignores the data's time dimension. If two time series are highly correlated, but even a single time step displaces one, the Euclidean distance incorrectly measures their distance.

It is preferable to implement dynamic time warping (DTW) to compare time series. DTW is a technique for determining the similarity of two time series that do not entirely match in time or length. Given this difference, we adapted the k-means clustering to time series using dynamic time warping.

4.2.2 Hierarchical Clustering

Hierarchical clustering is an alternative to k-means clustering for detecting groups in data collection. The Agglomerative Clustering method is an unsupervised learning technique that uses a bottom-up strategy to produce hierarchical clustering: each observation begins in its own cluster, and clusters are sequentially aggregated into bigger superclusters.

Unlike k-means, hierarchical clustering creates a hierarchy of clusters and hence does not require us to define the number of clusters beforehand. It is also faster since it acts on a matrix of pairwise distances between observations rather than directly on the data. Furthermore, hierarchical clustering results may be easily displayed using an appealing tree-based representation known as a dendrogram. Given these advantages, we decided to use this second clustering approach as the base for the forecasting section.

4.3 DTW Barycenter Averaging (DBA) algorithm

To forecast each cluster, we first needed to compute the series' average within the clusters. We applied the DTW Barycenter Averaging (DBA) algorithm, which minimizes the sum of squared DTW distances between the barycenter and the cluster's series. A barycenter, also known as a cluster centroid, is the average sequence of a set of time series in DTW space. As a result, regardless of where temporal changes occur among the series, the centroids have an average form that reflects the shape of the cluster's series.

4.4 Results for clustering

4.4.1 Selection of number of clusters

We first determine the ideal number of clusters before employing the clustering algorithms. The plots and analysis will be shown in this report for the case of the US. To keep the final results comparable, we will equally use the number of clusters k selected by the methods for the three countries. In figure 12, we have the results for the Elbow method for the US. From it, we could infer that the optimal k is equal to 4, given that it seems to be the value from which the curve begins to flatten, but this is not so evident. Given this ambiguity, we show in figure 13 the results of the Silhouette method. According to it, the optimal number of clusters would be 4, given that the silhouette score is the highest at this point. Now that we have selected four clusters for the three countries, we can apply the clustering algorithms in the next section.

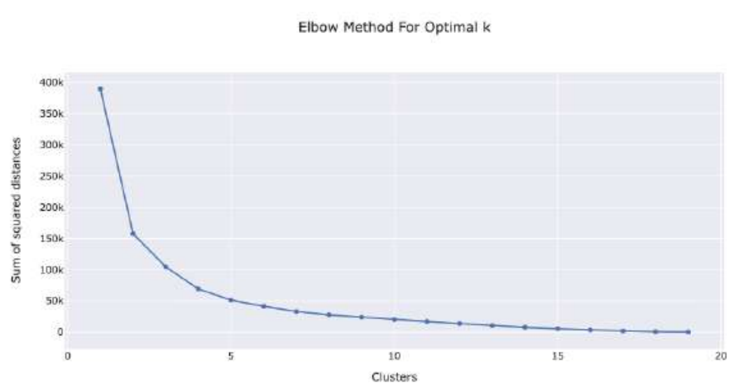

Figure 12: Elbow method for optimal K

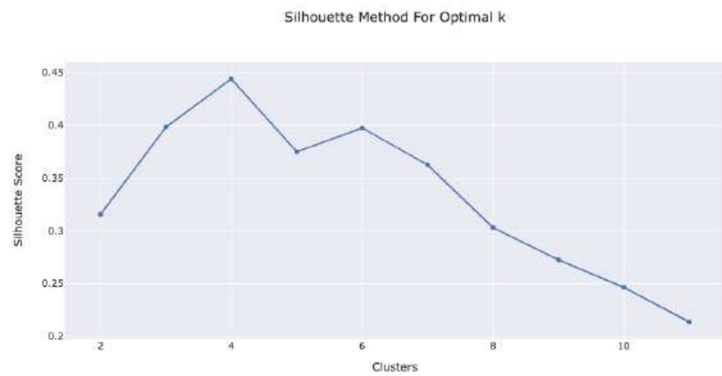

Figure 13: Silhouette method for optimal K

4.4.2 Hierarchical Clustering Assignment

This report shows only the results for Hierarchical Clustering, given its advantages compared to K-means clustering. Nevertheless, the results for the K-means clustering for the US can be seen in the code attached to this report.

We used Scikit-learn to perform the Agglomerative Clustering algorithm, a free software machine learning library for the Python programming language. The number of clusters defined for each country was 4, and the cluster assignment of the algorithm was used as labels for plotting the dendrogram. The hierarchical clustering results are displayed using a dendrogram for each country. In figures 14, 15, and 16, the vertical axis represents the twenty series for the US, France, and Japan. The horizontal scale on the dendrogram represents the distance or dissimilarity between the series. Each joining of two clusters is represented on the diagram by splitting a horizontal line into two horizontal lines. The horizontal position of the split, shown by a short bar, gives the distance between the two clusters.

Some insights we could take from the cluster's assignment are that,

- In the case of the US, the cluster denoted by the green color is exclusively made of wine brands, clearly stating similarities between their time series. The red group comprises fashion brands, while the others mix cosmetics and fashion brands.

- For France, the red cluster merges wine and fashion brands. The purple group contains the three categories of wine, fashion, and cosmetics.

- In Japan, we notice that the green group is of fashion and cosmetics brands, while the orange group is wine and fashion brands.

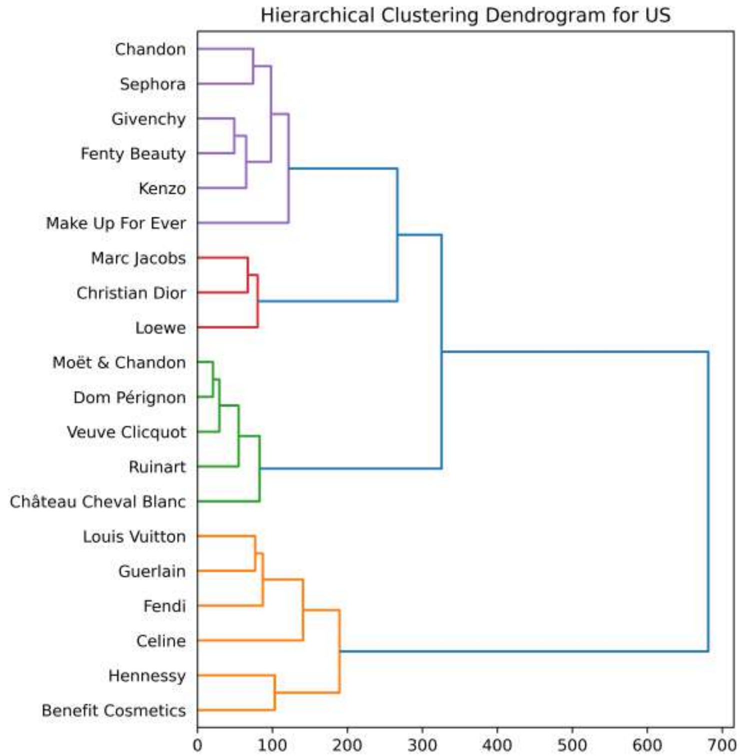

Figure 14: Hierarchical Clustering Dendrogram for the US

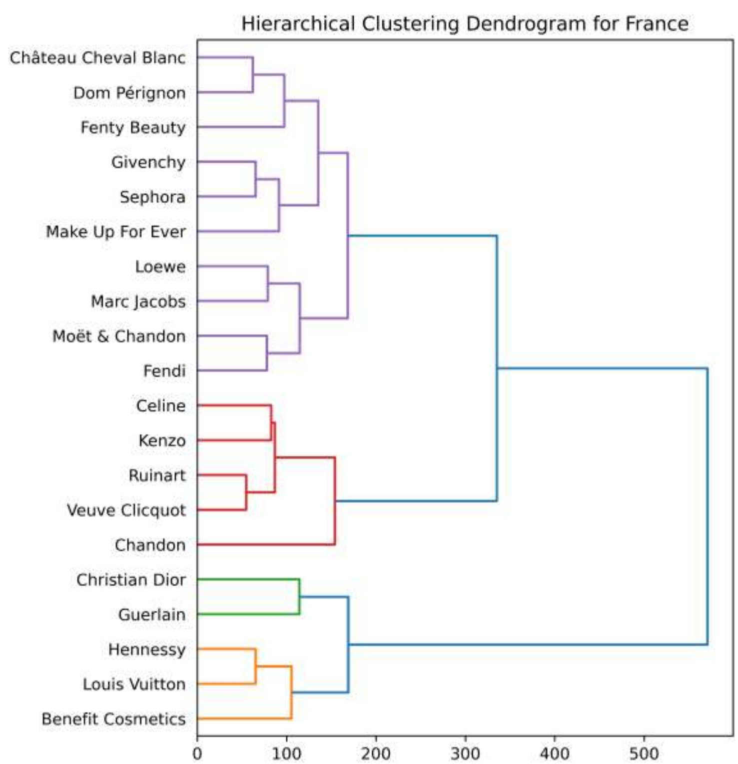

Figure 15: Hierarchical Clustering Dendrogram for France

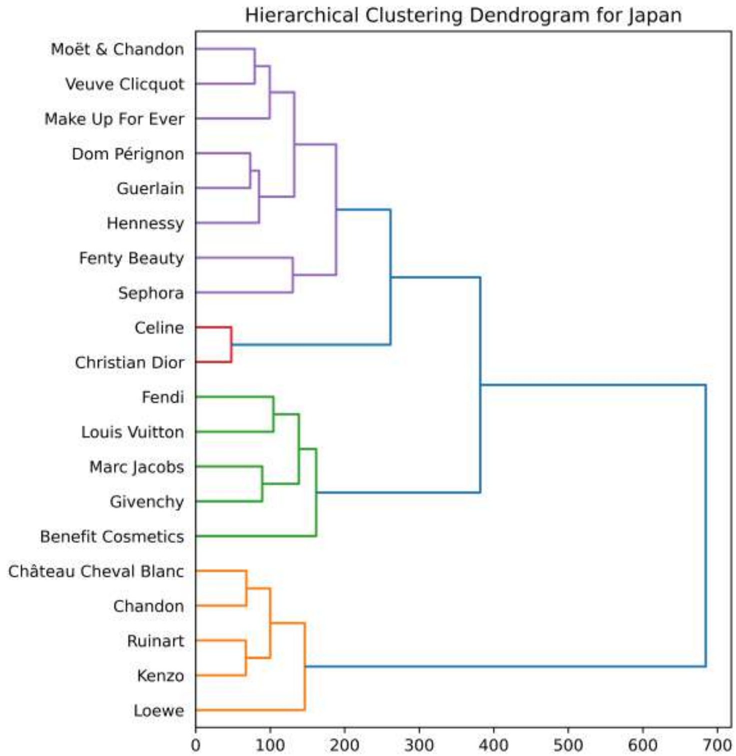

Figure 16: Hierarchical Clustering Dendrogram for Japan

Now that we have the cluster assignment, we can calculate the average of each cluster that will be used in the next section for forecasting.

We have the countries' clusters in figures 17, 18, and 19. For each cluster, we plotted all the series in gray. Then, to see the group's movement or shape, we took the average of the clusters using the DTW Barycenter Averaging (DBA) algorithm, and we plotted it in red.

- In figure 17, we can see that cluster 3 of the US has an upward trend. This cluster is composed of Loewe, Christian Dior, and Marc Jacobs. Meanwhile, cluster 2, consisting of only wine brands, has a relatively stable trend with peaks that might be related to the seasonality of the end of the year.

- Figure 18, which corresponds to France, shows the average clusters in red with peaks and stable trends, except for cluster 1, with a slightly upward trend.

- Figure 19 plots the clusters for Japan. In this case, the third group has an increasing trend, containing the time series for Christian Dior and Celine.

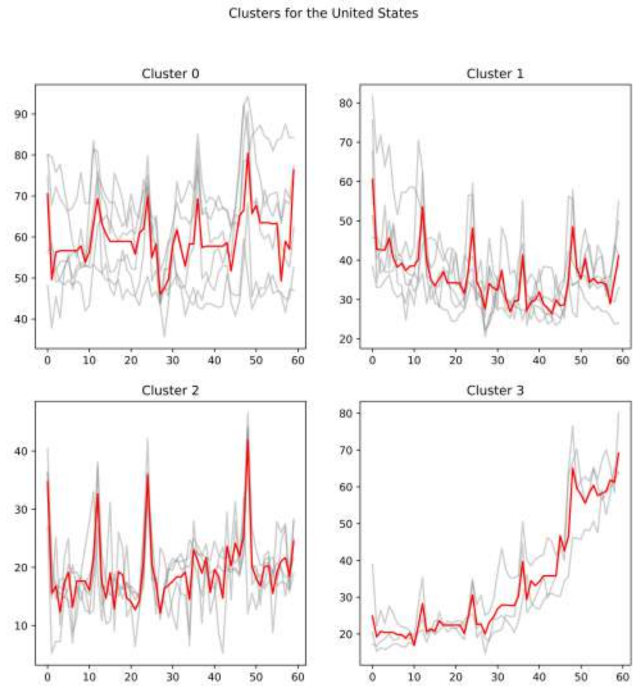

Figure 17: Clusters for the US

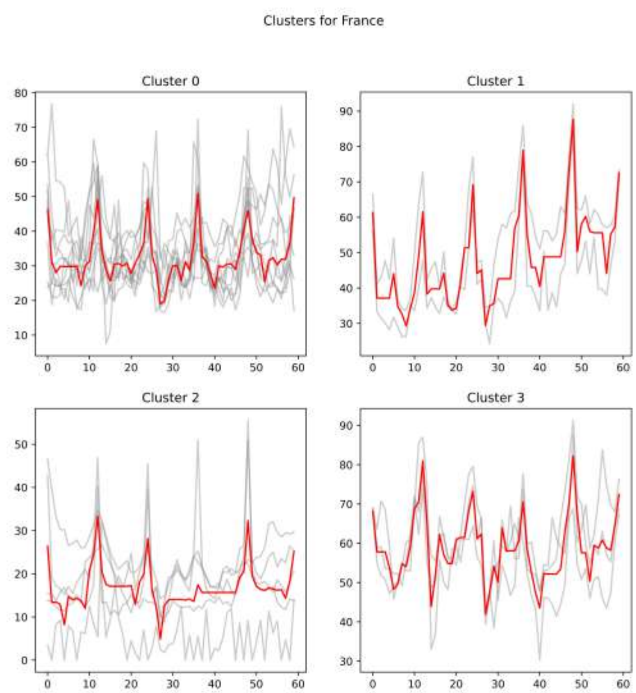

Figure 18: Clusters for France

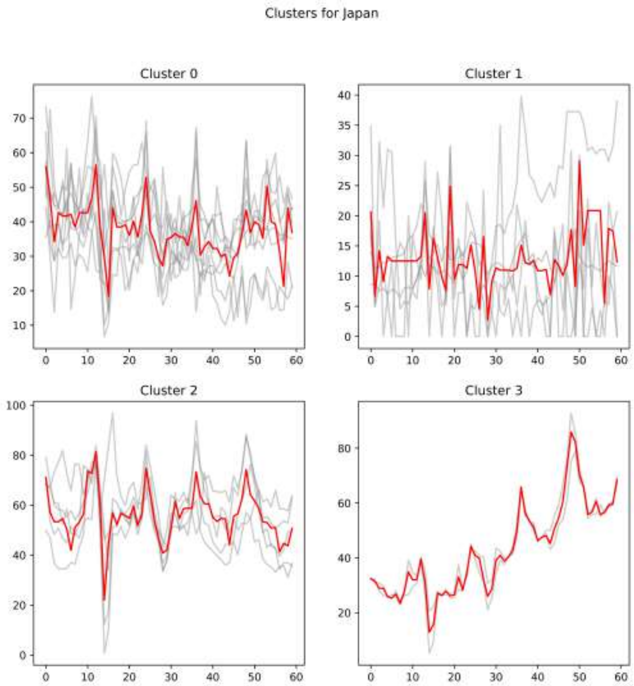

Figure 19: Clusters for Japan

5 Forecast

5.1 Forecast for each series

To predict the Google Trends data for each LVMH brand in each country, we used the open-source software Prophet[4] released by Facebook's Core Data Science team. Prophet is a method for predicting time series data that uses an additive model to suit non-linear trends with seasonality that occurs annually, monthly, daily, and on weekends as well as during holidays. Strongly seasonal time series and multiple seasons of historical data are ideal for it. Prophet typically manages outliers well and is robust to missing data and changes in the trend.

The first step is to transform the dataset from long to wide format as Prophet requires at least two pre-defined columns as inputs, a *ds* column, and a *y* column.

- The *ds* column has the time information. The column *date* is renamed to *ds*
- The *y* column has the time series values. In this analysis, *y* represents the Interest Over Time series of each brand name in each country

After grouping the wide-format dataset by brand name, we defined a specific function to forecast the Google Trends time series.

- The input data is individual time series data for a group.

[4]https://facebook.github.io/prophet/

- *Prophet()* initiates the time series model with the default hyperparameters, and we gave the model the name m
- *m.fit(group)* fits the prophet model on the individual time series data, which is the Interest Over Time data for each brand in a country.
- *make_future_dataframe* creates a new dataframe called future for the forecasting. *periods=12* and *freq=MS* means that we will forecast for 12 months of data
- After predicting on the future dataframe, prophet produces a long list of outputs. We only kept *ds*, *yhat*, *yhat_lower* and *yhat_upper*. *yhat* is the predicted value. *yhat_lower* and *yhat_upper* are the lower and upper bound of the confidence interval.
- A new column called brand is created in the forecast dataframe to indicate the brand name in a specific country for the predictions.

The next step is to make multiple time series forecasting using a for-loop. Down below is an illustration of one time series, ”Christian Dior - US” pulled from the aforementioned process.

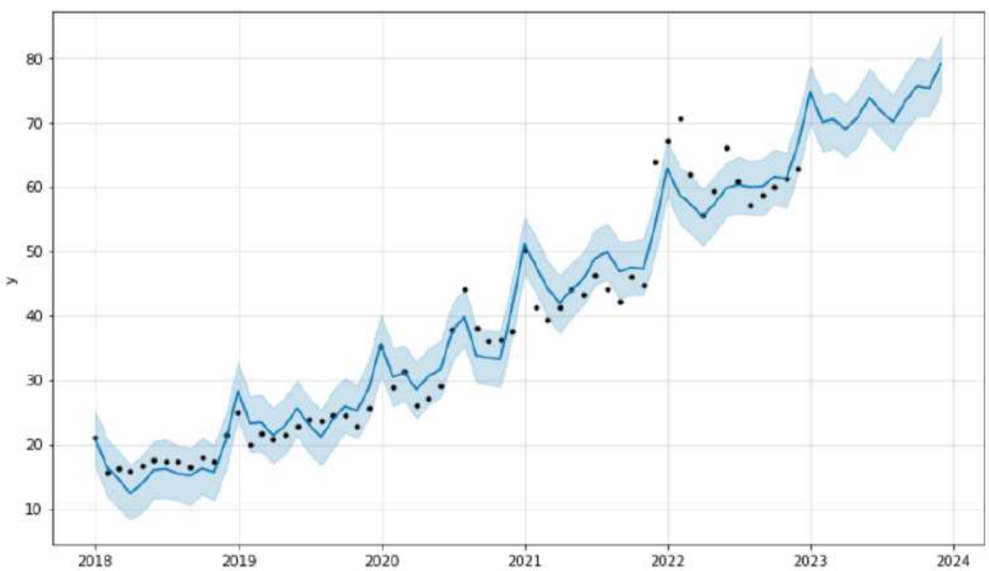

Figure 20: Forecast for Christian Dior US in the next 12 months

The original data are the black dots, and the light blue area is the confidence interval while the blue line is the forecasting result. As can be observed from the graph, the Search Interest Over Time for ”Christian Dior - US” is predicted to increase in the next 12 months with its 02 peaks falling at the end of 2022 and 2023. This is expected in reality since this period is a holiday season when consumers tend to purchase luxurious items as holiday presents. The trend will decrease slightly in the next 3-4 months and again increase around May or June 2023 when the fashion summer seasons start. The same pattern is witnessed until the end of 2023.

5.2 Forecast for each cluster

For forecasting the clusters, we performed a similar process to the one described for forecasting each series. We also used the software Prophet, and we changed data from the wide format to the long format, ending with 720 rows, corresponding to 60 months (5 years) for 4 clusters and three countries. We then grouped the dataset by the 12 clusters we have, and by following the same steps as for forecasting the time series, we defined a function to forecast the clusters to then applied it into a loop for doing it for all the 12 clusters. In the figure below, we present the results for the forecast of cluster nº 3 of the US. As seen in figure 17, this cluster comprises Loewe, Christian Dior, and Marc Jacobs, and it presents an increasing trend. Therefore, it would be interesting to predict if this upward trend in Google's search is likely to continue. This suspicion can be confirmed by looking at the graph. The cluster's prediction peaked at the end of the year, followed by a minor fall, but continuing with an increasing trend for 2023.

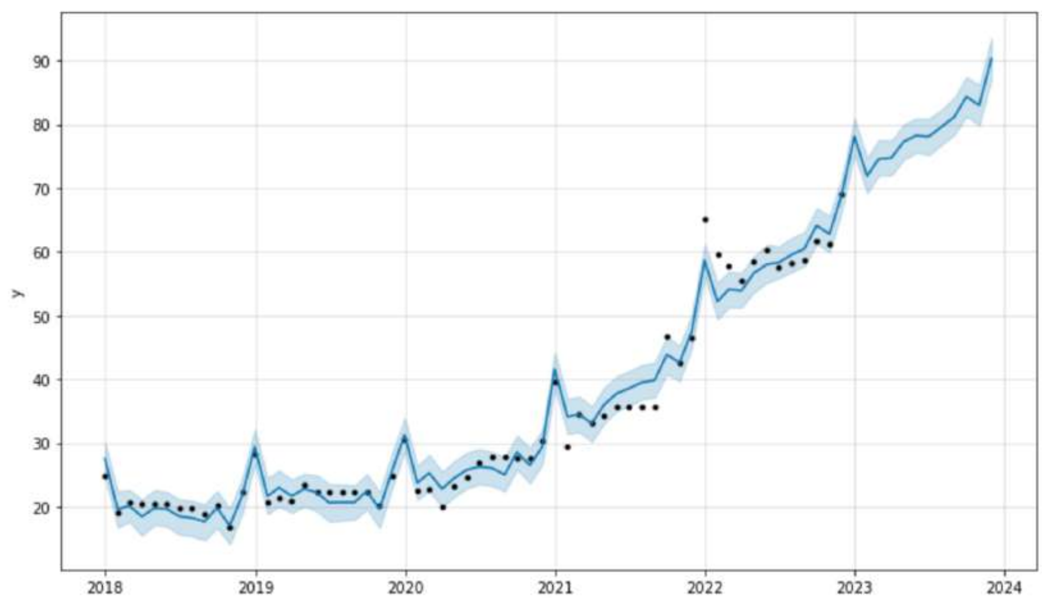

Figure 21: Forecast for cluster nº 3 in the US for the next 12 months

5.3 Evaluation of the forecasting models

In this section, we briefly showed two metrics, R2 score and Root Mean Squared Error (RMSE), to evaluate the accuracy of two forecasting models, one used for each series and one for each cluster. Following the examples above of forecasting the series 'Christian Dior - US' and cluster no. 3 in the US, the metrics for them are shown in the table below.

Forecasting metrics	Trend Christian Dior - US	Cluster 3 of the US
R2 score	0.959	0.979
RMSE	3.346	2.175

As shown in the table, for the series model, the R-squared value is pretty high (95.9%) while the

RMSE remains low (3.346). It means that our forecasting model for each series performs pretty well and is not exposed to overfitting problem. The model of each cluster performs even better with a higher R-squared value (97.9%) and lower RMSE (2.175) compared to that of the series forecasting model.

6 Marketing insights

6.1 Identifying Seasonal Patterns in Demand

It is reasonable to assume that Google Trends could serve as a potential indicator for changes in future sales if a luxury company could find evidence of a significant positive correlation between its historical sales for a product and consumer interest in the said product, as indicated by luxury consumer Google Trends. Notably, Boone et al. (2017) discovered evidence that Google Trends enhances sales estimates. Therefore, it would be beneficial if luxury brands could accurately predict future seasonality movements in Google Trends for luxury consumers, enabling better decisions on what to stock, when to stock it, and how much to stock, as well as using consumer trends to determine when to reduce and increase the price points from a marketing perspective. For example, our illustration for the brand search ”Christian Dior - US” implies that LVMH should invest more resources in marketing strategy, and stock preparation in the US market as this search term is predicted to increase in the next 12 months, especially in the summer and winter seasons. This strategy could also be applied to the other series that were clustered together with ”Christian Dior - US”, which are Loewe and Marc Jacobs, given that this cluster is also predicted to increase in the trend, as can be seen in figure 21.

6.2 Identifying Better Marketing Terms

Consumer demands for acceptable language and ethical advertising have made fashion marketing and advertising more challenging than ever (Bae et al. 2015). According to Forni (2018), it might be harmful to fail to recognize the necessity for regional language variations while promoting fashion online. Google Trends’ data on luxury consumers can assist luxury firms to make sure they are utilizing the right keywords for online marketing and search engine optimization in the nations where they do business. The ability to analyze Google Trends for luxury consumers by nation offers further value because it helps marketers to tailor their websites to the search preferences of consumers in any given country around the world. Having the ability to predict these future luxury consumer Google Trends with accuracy can enable improve search engine optimization for luxury brands. Our part of Outliers’ possible explanations is a great illustration of this implication.

7 References

Bae, Su Yun, Nancy Rudd, and Anil Bilgihan. 2015. Offensive advertising in the fashion industry: Sexual objectification and ethical judgments of consumers. Journal of Global Fashion Marketing 6: 236–49.

Bain, Marc. 2016. Google Already Knows the Fashion Trends that Will Be Popular this Fall. Available online: https://qz.com/767643/google-already-knows-the-fashion-trends-thatll-be-popular-this-fall/ (accessed on 07 December 2022).

Boone, Tonya, Ram Ganeshan, Robert L. Hicks, and Nada R. Sanders. 2017. Can Google Trends Improve your Sales Forecast? Production and Operations Management 27: 1770–74.

Boone, Torrence. 2016. Fashion Trends 2016: Google Data Shows What Shoppers Want. Available online: https://www.thinkwithgoogle.com/advertising-channels/search/fashion-trends-2016-google-data-consumer-insights/ (accessed on 07 December 2022).

D'Arpizio, Claudia, and Federica Levato. 2017. The Millennial State of Mind. Available online: https://farfetch.prezly.com/bain–company-media-pack (accessed on 07 December 2022).

Forni, Christine. 2018. Using Google Trends to Get an Edge in Digital Marketing. Available online: https://growthpilots.com/using-google-trends-edge-digital-marketing/ (accessed on 07 December 2022).

Gordon, Steven. 2017. Google Trends: A Metric of Consumer Behavior. White Paper. Sphere Quantitative Insights. Available online: https://spherei.com/wp-content/uploads/2017/09/whitepaper-6-30-17-Using-Google-Trends.pdf?pdf=GoogleTrends-A-Metric-Of-Consumer-Behavior (accessed on 07 December 2022).

McDowell, Maghan. 2019. What Google Search can Reveal about Trends. Available online: https://www.voguebusiness.com/technology/big-data-trends-spate-google (accessed on 07 December 2022).

Net Market Share. 2019. Desktop Search Engine Market Share. Available online: https://goo.gl/ESpbRp (accessed on 07 December 2022).

Siliverstovs, Boriss, and Daniel S. Wochner. 2018. Google Trends and reality: Do the proportions match?: Appraising the informational value of online search behavior: Evidence from Swiss tourism regions. Journal of Economic Behavious Organization 145: 1–23.

Wedel, Michel, and P. K. Kannan. 2016. Marketing analytics for data-rich environments. Journal of Marketing 80: 97–121.

Yu, Lean, Yaqing Zhao, Ling Tang, and Zebin Yang. 2019. Online big data-driven oil consumption forecasting with Google trends. International Journal of Forecasting 35: 213–23.

Zhang, Zuochao, Yongjie Zhang, Dehua Shen, and Wei Zhang. 2018. The cross-correlations between online sentiment proxies: Evidence from Google Trends and Twitter. Physica A: Statistical Mechanics and Its Applications 508: 67–75.

8 Appendix

In this appendix we will present five graphs, each showing a brand with the fitted model for the three markets we are analyzing, the US, France, and Japan. This is done to have a visual comparison for each brand along the three countries. For example, in the case of Sephora, the trends and the predictions in the US and France are quite similar, with high peaks at the end of each year. Therefore, they could follow similar marketing strategies. On the other side, for Sephora in Japan, we notice a decreasing trend with small peaks each December. With this, we identify a problem of consumer engagement with the brand in Japan, indicating that it might be advisable to focus on marketing campaigns that would target this issue.

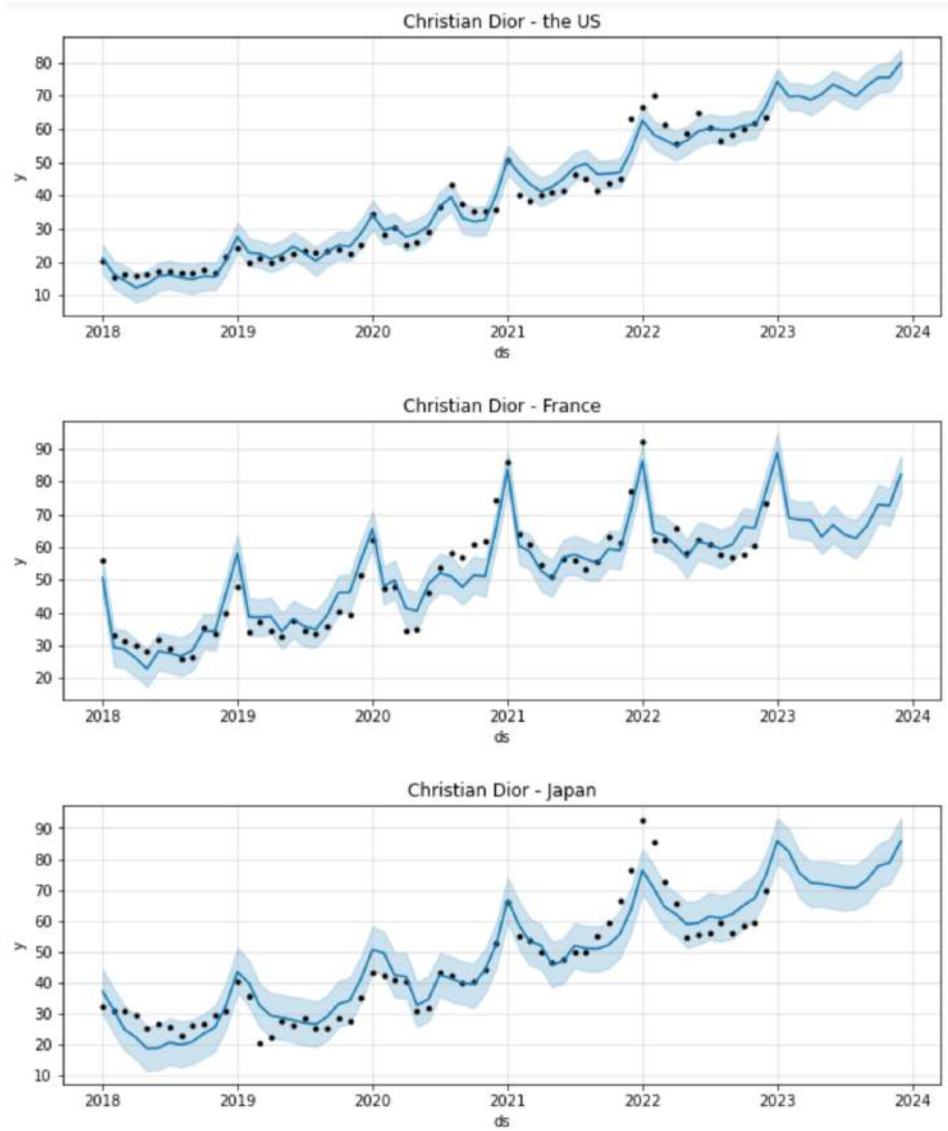

Figure 22: Forecast for Christian Dior in the US, France, and Japan for the next 12 months

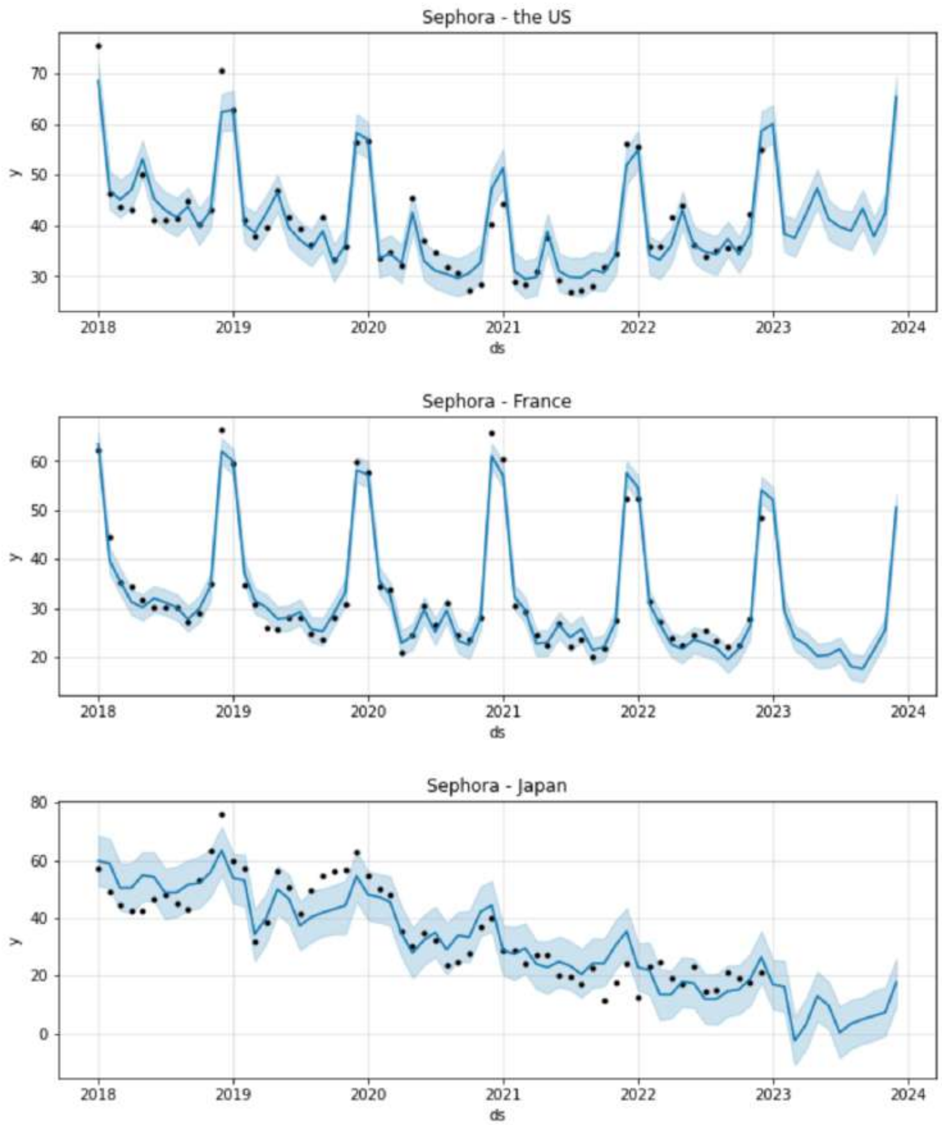

Figure 23: Forecast for Sephora in the US, France, and Japan for the next 12 months

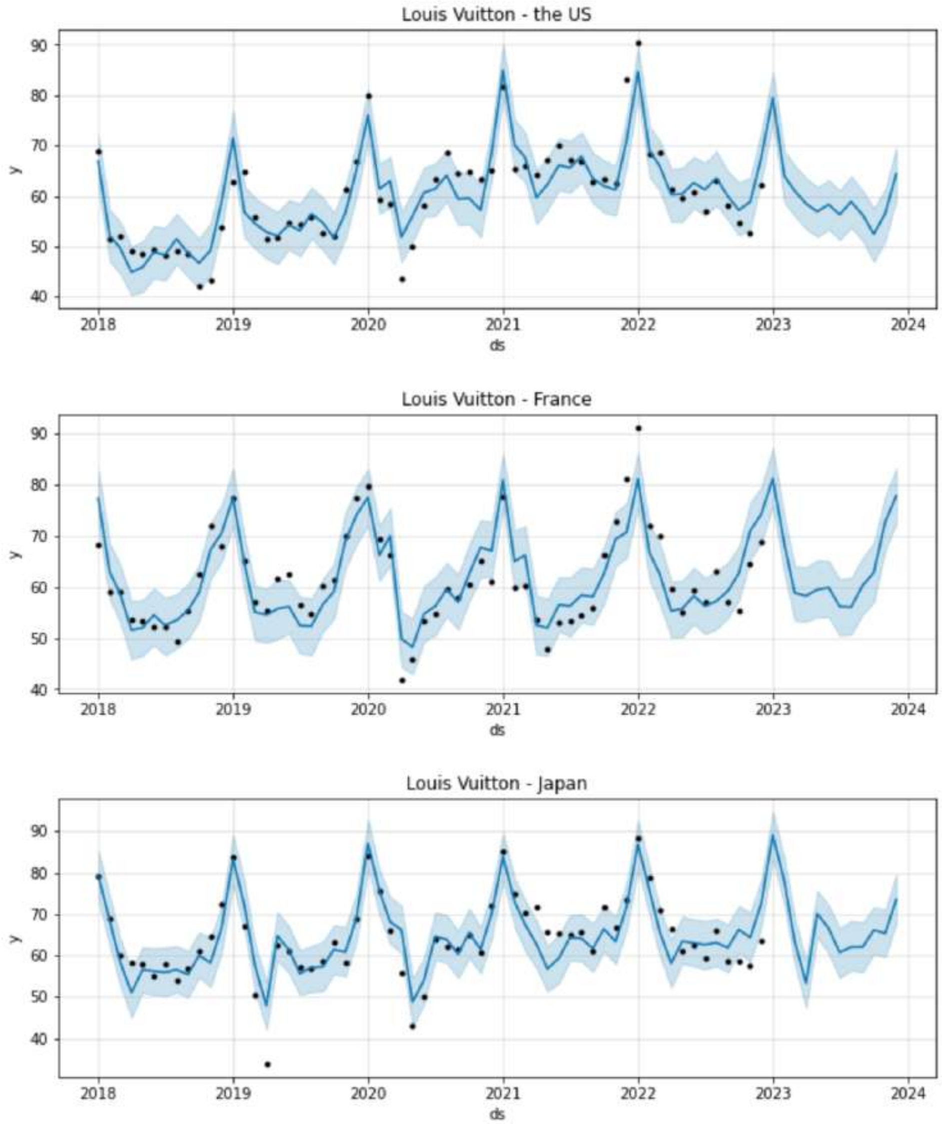

Figure 24: Forecast for Louis Vuitton in the US, France, and Japan for the next 12 months

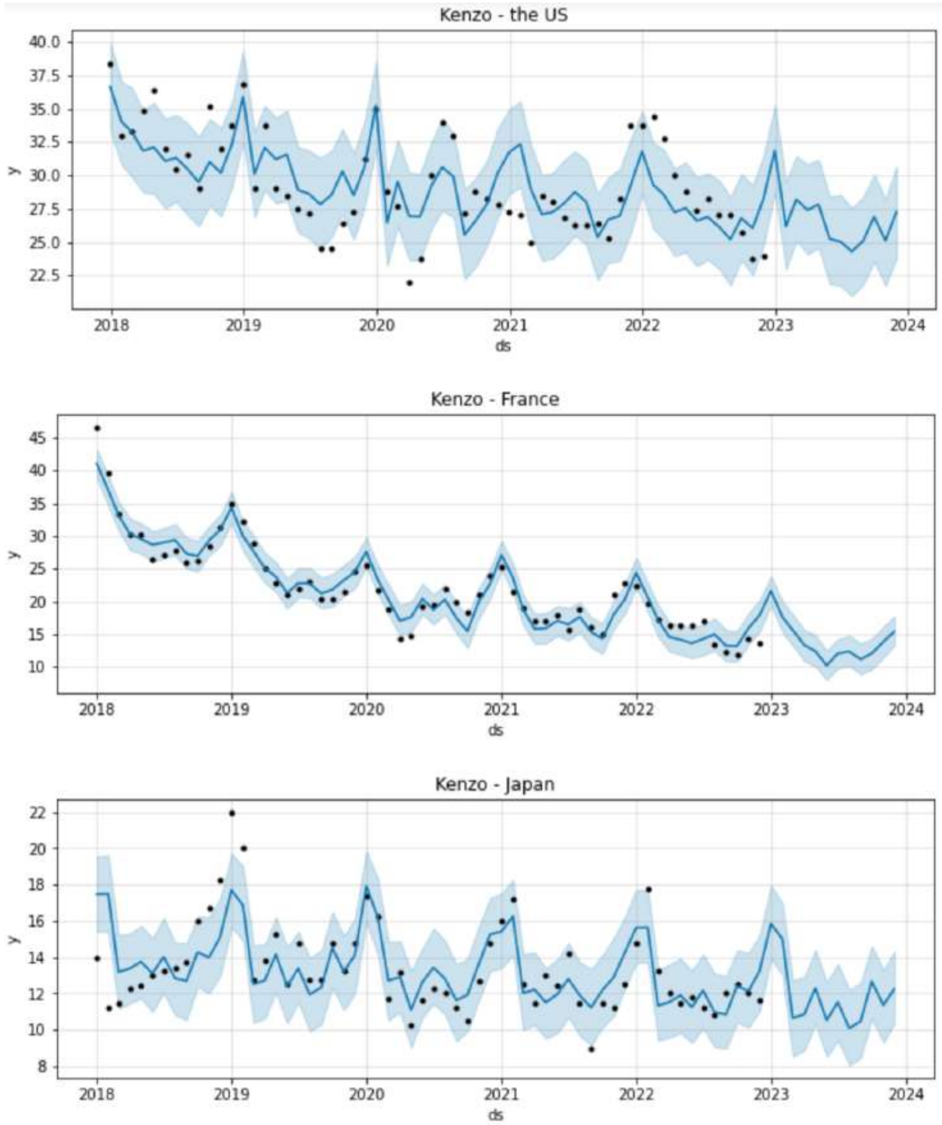

Figure 25: Forecast for Kenzo in the US, France, and Japan for the next 12 months

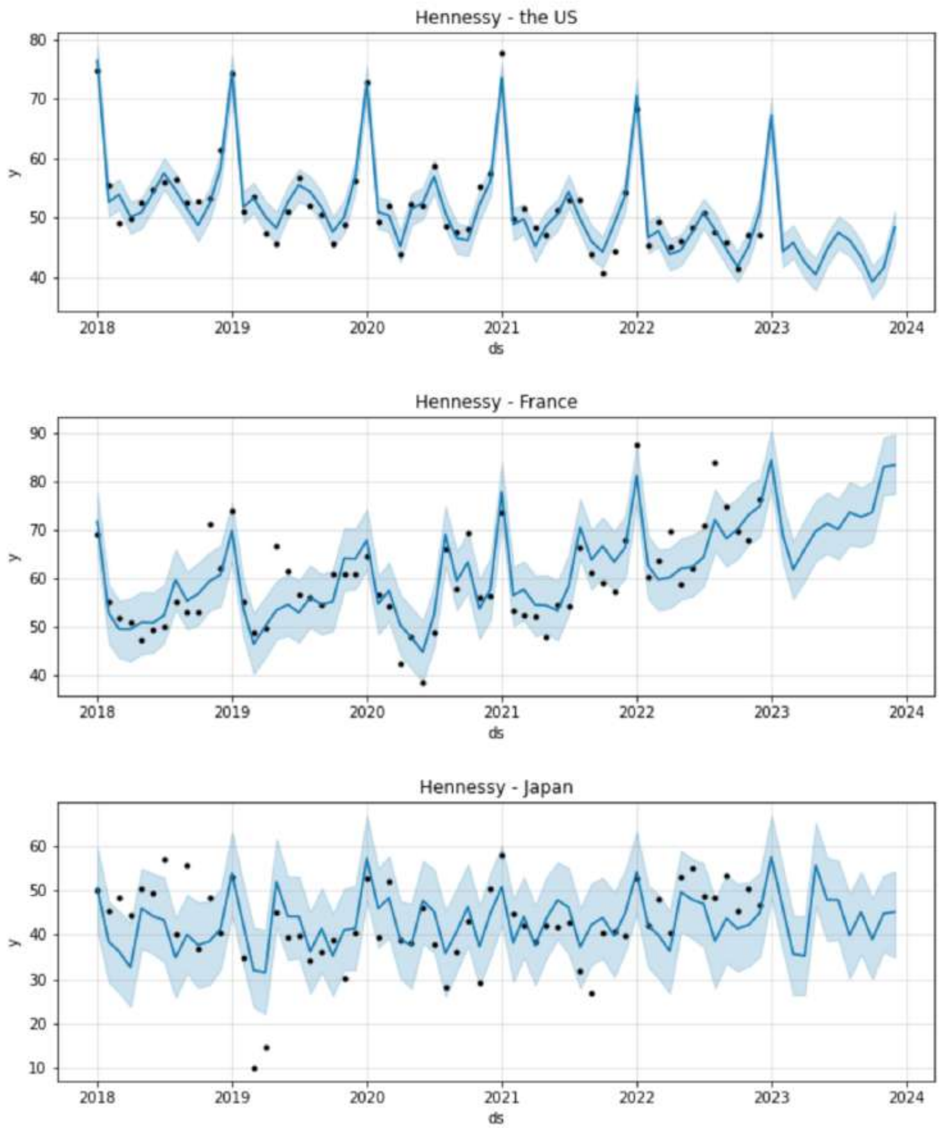

Figure 26: Forecast for Hennessy in the US, France, and Japan for the next 12 months