



The University of Liverpool

The Management School

Business Analytics and Big Data

MSc Project

Predictive Analytics for Identifying Product Opportunities Through Search Trends and Unstructured Data Analysis

Project No: Çağatay Iris-05

Final Report

by Abel Hernández García

15th, September, 2023

Supervisor: Çağatay Iris

Summary

This dissertation investigates the potential of internet search trends and unstructured data to discover consumer behaviour patterns and needs that often go unnoticed. Using information from Google Trends and search suggestions, various analytical methods are implemented: from time series forecasting with the Prophet algorithm, to text mining, semantic networks and data visualisation techniques. The study's findings show the effectiveness of these strategies in identifying emerging trends and consumer preferences. The Prophet algorithm stood out for its ability to predict seasonal fluctuations and long-term trends in Google Trends data, providing a potentially useful decision-making tool. Text mining and semantic networks enrich our understanding of the consumer and what they are looking for. Although the methodologies employed are robust, they face certain limitations, in particular the lack of actual sales data to support our qualitative conclusions. Nevertheless, the study demonstrates the vast potential of these analytical approaches in fields as diverse as market research, product development and strategic planning. This dissertation thus demonstrates that data analysis tools and analytical methods, while not without limitations, offer considerable potential for uncovering insights into consumer behaviour. The techniques presented here are flexible and could be applied in a variety of settings and for multiple purposes, from product development to market entry strategies. If used intelligently and in combination with other forms of data and analysis, they have the potential to become a useful tool for businesses.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Çağatay Iris, for his patience and constant support throughout this academic process. To my parents Abel and Adriana, for being my source of inspiration and my strongest pillar throughout my life. Your faith in me has been the foundation on which I have built each of my achievements. Finally, I want to thank my siblings Alejandra, Elizabeth and Isaac for always being there, my achievements are also theirs.

Contents

| | |
|---|----|
| 1. Introduction | 6 |
| 1.1 Objective..... | 6 |
| 2. Literature Review | 7 |
| 3. Methodology | 10 |
| 3.1 Data Sources..... | 11 |
| 3.2 Instruments and Materials..... | 11 |
| 3.3 Procedures..... | 11 |
| 3.3.1 Data Extraction from Google Trends..... | 11 |
| 3.3.2 Trend Modelling with Prophet and Granular Data Retrieval..... | 12 |
| 3.3.3 Search Term Collection and Method Comparison..... | 12 |
| 3.3.4 Text Mining: Filtering and Cleaning of Unstructured Data..... | 12 |
| 3.3.5 Co-occurrence Matrices..... | 12 |
| 3.3.6 Semantic Networks and Clusters..... | 13 |
| 3.3.7 Validation and Result Visualisation..... | 13 |
| 3.4 Limitations..... | 14 |
| 4. Computational Experiments and Managerial Insights | 14 |
| 4.1 Google Trends: Overview..... | 14 |
| 4.1.1 Understanding Google Trends..... | 14 |
| 4.2 Data Collection from Google Trends..... | 17 |
| 4.2.1 Retrieving Data from Google Trends using gtrendsR..... | 17 |
| 4.2.2 Retrieving Data from Google Manually..... | 19 |
| 4.2.3 Visualizing Interest Over Time from Google Trends..... | 20 |
| 4.3 Forecasting Interest Over Time..... | 22 |
| 4.3.1 Data Frame Preparation..... | 23 |
| 4.3.2 Model Building..... | 24 |
| 4.3.3 Visualising Predictions..... | 25 |
| 4.3.4 Interpretation of Obtained Forecasts..... | 27 |
| 4.4 Granular Data Analysis with Trendecon: In Search of Weekly Seasonality..... | 28 |
| 4.4.1 Experimentation: Forecasting Granular Data..... | 30 |

| | |
|---|-----------|
| 4.4.2 Interpretation of Forecasts Obtained with Granular Data | 31 |
| 4.5 Unstructured Data Analysis: Search Suggestions..... | 34 |
| 4.5.1 What is Unstructured Data? | 34 |
| 4.5.2 Obtaining Search Suggestions: Vector Method and Alphabetical Levels Method..... | 35 |
| 4.5.3 Experimentation with Word Clouds..... | 39 |
| 4.5.4 Comparison of Methods Used for Collecting Search Suggestions..... | 41 |
| 4.6 Using Related Queries from Google Trends as a Vector for Obtaining Search Suggestions..... | 42 |
| 4.7 Co-occurrence Matrix and Semantic Networks..... | 43 |
| 4.7.1 Creating the Co-occurrence Matrix..... | 44 |
| 4.7.2 Creating a Semantic Network..... | 45 |
| 4.7.2.1 Algorithm for Semantic Network: The Louvain Algorithm..... | 45 |
| 4.7.2.2 Visualization of the Semantic Network..... | 46 |
| 4.8 Additional Insights and Visualizations..... | 48 |
| 5. Conclusion..... | 55 |
| 6. Recommendations..... | 57 |
| 7. References..... | 59 |

1. Introduction

In today's digital era, businesses are faced with the constant challenge of keeping up with shifting market demands and trends. Early recognition of these trends can provide a competitive edge and facilitate more informed decisions regarding new product launches. According to Chen, Chiang, and Storey (2012), leveraging business intelligence and analytics, including monitoring market trends, can lead to significant impacts on business performance and decision-making. Search trends, particularly data from Google Trends, have emerged as a tool for understanding and predicting market dynamics (Duwe et al., 2018; Silva et al., 2019; Pekar, 2018). Meanwhile, unstructured data encompassing everything from social media to online news articles, provides valuable insights into public sentiment and interest trends (Lenz & Winker, 2020).

In this context, this study aims to investigate one such approach, employing the power of online search trends and unstructured data analysis to identify potential opportunities for product launches. The challenge lies not only in gathering and processing this data but also in deciphering meaningful insights that can guide the strategic decisions.

This study proposes to develop a predictive model that uses unstructured data analysis along with Google Trends data to predict product opportunities. In doing so, this study hopes to provide a methodology that businesses can use to inform their product launch strategies and make more data-informed decisions to achieve market success.

1.1 Objective

The primary objective of this dissertation is to develop a comprehensive approach to identify product opportunities by exploiting the power of online search trends and unstructured data analysis. In addition, this project also aims to create a scalable and adaptable model that can serve as a predictive tool for companies looking to expand their product offerings into new markets or launch successful products from one country into another.

To achieve this the project aims to:

- Understand and evaluate the potential of tools such as Google Trends in predicting market trends and identifying product launch opportunities across various sectors.
- Develop and implement machine learning and data mining techniques to analyse unstructured data, with a focus on identifying trends and insights that can inform product launch strategies.
- Conduct a comparative analysis of various data sources and methodologies highlighted in the literature, including search query data, and topic modelling from unstructured text, to determine the most effective combination for predicting product opportunities.
- Develop a model that integrates findings from the above objectives to provide a robust, data-driven framework for identifying product launch opportunities in various sectors.
- Develop a model that can adapt to different market conditions and incorporate real-time changes in consumer search trends to predict successful product launch opportunities.
- Identify key indicators in search trends and unstructured data that signal a market's readiness or receptiveness for a new product or a successful product from another country.

A future objective will be to create a user-friendly interface for this model that enables companies to easily input data, interpret results, and make informed decisions about product launches in new markets.

2. Literature Review

Identifying product launch opportunities in today's hyper-competitive market requires data-driven insights and dynamic tools for analysis. In this regard, the use of search trends and unstructured data analysis has gained an increasing attention. (Chen et al., 2012). One such tool that has been extensively studied is Google Trends, a tool that tracks search queries over time, providing insights into public interest in various topics (Duwe et al., 2018; Silva et al.,

2019; Hanna & Hanna, 2019; Pekar, 2018; Wu & Brynjolfsson, 2015). Google Trends has been acknowledged for its ability to reflect real-time changes in public interest based on search behaviour, making it a valuable tool for understanding and predicting trends (Nutti et al., 2014).

Existing studies have explored how online search trends and unstructured data can be used to anticipate market trends and even inform decisions regarding product launches. This body of literature, encompassing fields as diverse as health, fashion, economics, and technology, provides insights into the multifaceted applications of online search data, which contributes to the achievement of the objective of this dissertation.

The relevance and applicability of Google Trends in predicting market trends have been examined across several fields. Duwe et al. (2018) concluded that despite its limitations in detecting early signs of product or technology development, Google Trends showed potential in predicting market trends due to its correlation with market diffusion processes. This underscores the complexities of predicting the diffusion of innovative products and technologies and highlights the importance of diversifying data sources. Similarly, in the pharmaceutical domain, Hanna and Hanna in 2019 utilized Google Trends and Google data to understand patient needs, demonstrating the potential of such tools in informing marketing strategies and patient advice.

In contrast to these findings, other researchers have pointed out the potential of Google Trends as an early predictor in specific sectors. Silva et al. delved in 2019 into the potential of Google Trends for fashion consumer analytics and forecasting, revealing that it could identify seasonal patterns, conduct competitor analysis, and identify brand extension opportunities. This notion was further supported by Wu and Brynjolfsson (2015), who utilized Google Trends to predict housing market trends and employed a seasonal autoregressive (AR) model to argue that search queries could provide a more accurate and timely picture of market demand and supply. The authors found that search indices are not only valuable in predicting contemporaneous sales but also exhibit superior predictive power for future sales compared to expert forecasts from the National Association of Realtors.

The integration of unstructured data in identifying product launch opportunities has also been addressed in the literature. Lenz and Winker introduced in 2020 an innovative approach that combines neural embedding methods to generate meaningful and coherent topics from unstructured news articles, proposing it as a suitable proxy for measuring the diffusion of innovations, instead of focusing on patents, their study examines news articles as a means to measure the diffusion of innovations. Pekar in 2018 also highlighted the potential of user-generated content from Google Trends and Twitter in forecasting consumer expenditure, which could potentially be extended to predict product launch opportunities. The study emphasizes the potential of integrating online search data with social media sentiments for more effective prediction, an approach also advocated by D'Avanzo and Pilato in 2017 for comprehensive decision making. The incorporation of sentiments and emotions analysis from social media into Google Trends analysis could be a critical aspect of predicting consumer behaviour and identifying new product launch opportunities.

Despite the potential of Google Trends and unstructured data analysis, a key theme in these studies is the need for caution. Duwe et al. (2018), Pekar (2018), and Jun et al. (2018) highlighted the limitations and variabilities of using search information alone and emphasized the need for understanding the characteristics of big data and integrating Google Trends data with other data sources for a comprehensive analysis. In developing a predictive model, the challenges and limitations of using these data sources should be recognised. For example, Google Trends data is limited in its time resolution, with daily data only available for a query period shorter than 9 months (Tse, 2021). Despite this, methods have been developed to reconstruct Google Trends daily data for extended periods, which could be utilised in this research (Tse, 2021).

On the other hand, in the context of analysing trends for successful product launches the potential uses of autocomplete can provide valuable insights for understanding user behaviour and preferences during the research process. In a study conducted by Ward, Hahn, and Feist (2012), the potential uses of autocomplete as a research tool were explored, providing valuable insights into user behaviour and preferences during the information search process. This research is particularly relevant to the objective of analysing trends for successful product launches. The findings revealed that autocomplete can be utilized to locate

known items with partial citations, helping users discover specific products or related information even with limited details. This aspect is also crucial in understanding user preferences and interests during the pre-launch phase of a product. Furthermore autocomplete plays a role in building user confidence and validating research topics, offering insights into the popularity and relevance of different product features or variations. (Ward, Hahn, and Feist; 2012).

By exploring how users utilize autocomplete in their search queries, the study sheds light on the factors that influence its adoption and effectiveness. Incorporating the insights from this study on autocomplete into the analysis of trends for product launches can provide businesses with a competitive edge and ensure successful introductions of their products.

These studies reflect the potential of Google Trends and unstructured data analysis in predicting market trends and identifying product launch opportunities. They underscore the value of integrating different data sources and adopting innovative methodologies to gain forward-looking insights into market developments. This literature review provides a foundation for the development of a predictive model integrating online search trends and unstructured data analysis for identifying product launch opportunities (Duwe et al., 2018; Silva et al., 2019; Pekar, 2018; D'Avanzo & Pilato, 2017; Wu & Brynjolfsson, 2015; Hanna & Hanna, 2019; Jun et al., 2018; Lenz & Winker, 2020; Yoo et al., 2018).

3. Methodology

The present study adopts a methodological approach of quantitative experimentation with an interpretative component for data analysis. Predictive models for time series were used. We used text mining techniques, clustering algorithms and data visualisation to examine patterns in the search terms and their relationships. We also included a qualitative analysis of the results to interpret the meaning and relevance of the insights identified.

3.1 Data Sources

The sample for this study consists of data obtained from the Google Trends platform and Google search suggestions. These terms were also collected and filtered according to various criteria in order to ensure the relevance and applicability to the context of the experimentation presented in this dissertation. Subsequently, interpretations of the identified patterns were made with external sources such as news, articles, as well as data from Amazon Seller's Product Opportunity Explorer platform.

3.2 Instruments and Materials

The following tools were used for data collection, processing and visualisation:

R software: Statistical programming language that was used for data analysis

R libraries:

- igraph: Used to implement Louvain's algorithm in the construction of semantic networks.
- wordcloud: To visualise frequent keywords in a graphical format.
- tm: Used for pre-processing and cleaning of text.
- gtrendsR and trendecon: For extracting and managing data from Google Trends.
- visNetwork: Used in the visualisation of networks. Software de Visualización: R Studio.

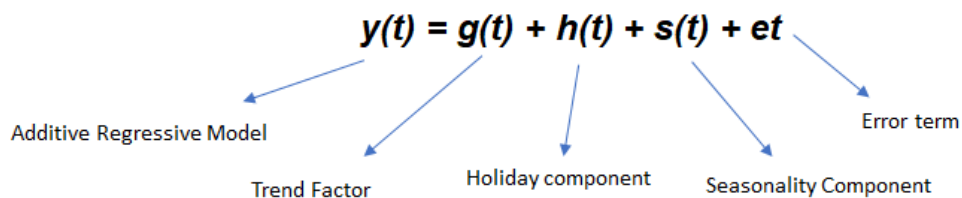
3.3 Procedures

3.3.1 Data Extraction from Google Trends

Data like Google Trends time series with their respective relative frequencies of search terms will be extracted and examined. This data will be subjected to visual analysis to identify patterns and trends that can provide an effective representation of market behaviour and consumer interest.

3.3.2 Trend Modelling with Prophet and Granular Data Retrieval

Applying the Prophet forecasting algorithm to model future trends in relative interest of search terms as well as to decompose the time series into various components, including trend, seasonality and holidays. The formula is the following:



(Facebook, no date)

3.3.3 Search Term Collection and Method Comparison

Collection of Google search terms by two methods: by vectors and by alphabetical levels, allowing us to have two points of comparison to evaluate the most convenient method according to our objectives.

3.3.4 Text Mining: Filtering and Cleaning of Unstructured Data

A process of data cleaning and filtering is performed once the terms have been collected, this to remove redundancies, discard the irrelevant terms and improve the quality of the dataset for further analysis.

3.3.5 Co-occurrence Matrices

Generation of co-occurrence matrices to analyse the relationship between search terms and their frequency. This step is important to understand how certain terms are interrelated, which provides the basis for the semantic network analysis in the next step.

3.3.6 Semantic Networks and Clusters

Semantic network creation based on the co-occurrence matrix using Louvain's algorithm to identify clusters. This algorithm optimises the modularity of the network to identify clusters of words or nodes that are highly interconnected (Murniyati, et al., 2023).

In this algorithm, the aim is to maximise Q until it can no longer be improved:

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j),$$

Where:

- A_{ij} is the link weight between nodes i and j .
- k_i and k_j are the degrees of nodes i and j , respectively.
- m is the total sum of the link weights in the network.
- $\delta(c_i, c_j)$ is 1 if nodes i and j are in the same community and 0 otherwise.

3.3.7 Validation and Visualisation of Results

To validate the insights identified we implemented a process that incorporates multiple external data sources. This process is conducted using supporting information extracted from reliable media such as news and research articles, as well as analytical tools such as the time series provided by Google Trends. Additionally, we cross-check the findings with search patterns and consumer behaviour on Amazon's Seller platform.

The validation strategy consists of comparing the most recurrent keywords and topics identified in our semantic network with the emerging and predominant trends in Google Trends, the news and with the most consulted search terms on Amazon. This approach can provide us with a more solid basis for evaluating not only the relevance but also the practical applicability of the insights found.

3.4 Limitations

1. **Geographical Bias in Search Suggestions:** The keywords or search terms suggested by Google can be influenced by the geographic location from which the query is made. This can introduce a geographic bias that limits the generalisability of our findings to a more global context or, conversely, can be an advantage for identifying more local insights (Google, no date).
2. **Absence of Actual Sales Data:** This study does not include an actual sales database to correlate with our results. While this represents a limitation in the external validation of the results, it also opens an opportunity for future research that could use real sales data sets for a more robust comparison.
3. **Data granularity in Google Trends:** The granularity at which Google Trends presents data can be a limiting factor, especially when more detailed analysis is required. However, this problem was partially mitigated by the use of the R library `trendecon` which allows for a finer disaggregation of the data.
4. **Scalability and Ethical Considerations:** Although Louvain's algorithm is efficient for handling large volumes of data, ethical considerations related to large-scale data collection, especially from Google may limit the scalability of the project.
5. **Use of Third-Party Data:** As much of the data used in this study is drawn from third-party platforms such as Google, it is subject to the data usage policies of these platforms, which could affect the replicability and long-term validity of the study.

By taking these limitations into consideration, the present dissertation provides a baseline for future research that can address these challenges more extensively.

4. Computational Experiments and Managerial Insights

4.1 Google Trends: Overview

In the age of technology and the proliferation of internet usage, there is a large amount of data that reflects public interest, sentiment and behavioural patterns over time (Escobar and Mercado, 2017). As mentioned in the literature review, there are great

opportunities for companies to analyse these trends. One of the fundamental principles of this research is based on discerning patterns in this vast sea of data in order to predict product identify opportunities in different markets. Google Trends, a tool developed by Google, emerges as a vital source in this exploration.

Like a digital footprint, search queries reflect the collective interests, concerns and curiosities of the public. Google Trends offers a macroscopic view of these digital footprints with data going back to 2004 potentially allowing us to unearth patterns, detect anomalies or even predicting future events such as sales increases (Rogers, 2021; Camusso and Jorge, 2019; Marrs, 2021).

4.1.2 Understanding Google Trends

Google Trends is a free online tool provided by Google that provides information on the relative volume of searches performed in different time periods and geographic regions, the data it provides is relative in that it does not show the absolute number of searches per x time period but rather shows values and fluctuations on a scale of 100. (Rogers, 2021) Despite this, as mentioned earlier in the literature review, for the market researchers, marketers, academics and the analysts, it serves as a treasure trove of information, capturing the mood of users around the world. (Silva, et al., 2019).

Given its large user base, Google's search engine provides a microcosm of the world's digital behaviour. Google Trends data essentially quantifies this behaviour charting the ebb and flow of the interest in particular topics, keywords, products or people's names over time (Pool and Hernández-von, 2017). This is especially important in market research, as it allows for the assessment of product popularity, brand awareness and even emerging market needs (Silva, et al., 2019). Among the parameters that the platform allows us to enter are the possibility of adding up to 5 search terms, if we want to make a comparison, which is quite useful for any type of research; enter the time span from which we want the results of the trends to be returned and which gives us the possibility of retrieving data from the year 2004 to the day before our query; In addition, the tool allows us to select if we want it to show us the trends

of a specific country or worldwide as well as if we want the results to reflect the interest of people registered in the web search, image search, news search, Google Shopping or Youtube Search (Google, n. d).

In addition to showing the interest of the search term queried over the time, Google Trends exposes other sections such as "Related Queries" and "Related Topics", which are the sections that provide additional information about the search terms that have been entered (Google, no date; Rogers, 2021). These sections allows for further exploration of the interests and trends associated with a specific term, which function as follows:

1. Related Queries

This section displays search queries that are related in some way to the term you enter.

These queries can be similar terms, variants, related queries, or terms that are often searched for in conjunction with the main term. By examining related queries, you can get a more complete picture of how people are approaching the same topic from different angles. (Google, n.d.)

2. Related Topics

This section highlights broader topics that are related to the original search term. Related topics, may include broader concepts and categories that encompass the topic being explored. For example, if you are searching for information on "football", related topics might include "sports", "football league" and "football teams". This section allows you to contextualise the search term within a broader context. (Google, n.d.).

In addition, in Google Trends these two sections show their results with the terms "rising" and "top" which refer to specific functions:

Rising: When a search term shows a "rising" trend in Google Trends, it means that its popularity is increasing over a specific period of time compared to its previous history. This could be due to current events, relevant news or any other factor that is generating an increase in searche related to that term (Google, n.d.).

Top: The terms or topics that appear in the "Top" section of Google Trends are the most popular at a specific time and place. These are the search terms that are being queried most often at that time. The "Top" section provides a snapshot of the hottest and most relevant trends in a given place and time (Google, not date).

In terms of data retrieval, however, Google Trends offers two main avenues. The first is directly from the Google Trends website, ideal for quick ad hoc queries and quick insights. In this way a person can easily enter the site and make quick comparisons; enter the time range, the search term(s) if a comparison is desired, as well as see the search terms or topics related to the search term we are querying, this data is presented in a very intuitive and visual way; additionally it gives us the option to download the results of the query in .csv format. The second, more complete method is through the Google Trends API, which allows for automated and more detailed data collection. While both methods have their advantages, it is critical to understand the difference between them. Manual collection may be simpler and more straightforward, but may not be feasible for large-scale data collection. On the other hand, the API offers more flexibility, but requires a more technical approach, which will be discussed in more detail in the following sections.

4.2 Data Collection from Google Trends

4.2.1 Retrieving Data from Google Trends using gtrendsR

Technological progress and the demand for more efficient processes have resulted in the development of APIs by technology companies which allow for the automated collection or transmission of data.

In the case of Google Trends, the gtrendsR package for R has been widely used by the scientific community to access Google Trends data because of its ease of use (Raubenheimer, 2021; Silva, et al., 2019; Balouli and Chine, 2022.). Therefore, this study used the R package gtrendsR, which makes it easy to obtain data without having to download it manually.

The aim of the first part of this study focused on retrieving data directly from Google Trends and visualising it in the Rstudio software by means of the previously mentioned package, `gtrendsR`. To do this, it was necessary to install and load all the necessary packages, after which the collection process is initiated by specifying the search terms of interest. As allowed by the Google Trends interface, there is the possibility to add a maximum of 5 search terms. In this first part of the study it was decided to make a trend analysis of different technological devices as search terms: "laptop", "desktop" and "tablet". In addition, the search was restricted to Great Britain and data was requested for the last 5 years, as shown in the code below:

```
device <- gtrends(  
  keyword = c("laptop", "desktop", "tablet"),  
  geo = "GB",  
  time = "today+5-y",  
  tz = 0  
)
```

The query returns a list of different components representing different datasets and metadata related to the query, including:

interest_over_time: Provides search trends for the query term or terms over time. This is basically the number of relative searches in a specific period of time.

interest_by_region / interest_by_country / interest_by_city: It shows the popularity of a search term by region or city or if the query was made for worldwide data, it will also display the popularity by country. This information is useful to see where a term is most popular geographically.

related_topics: As mentioned above, the related topics section identifies topics or categories of interest related to the original keyword or query. These are broader and may not exactly be searches that people enter into Google, but rather more general categories or topics that relate to the original query. In our study, particularly for the search term "laptop", among the resulting related topics were

mainly "HP Laptop", "Dell", "Microsoft Windows", among others. These are general topics that are related to "laptop".

related_queries: It identifies exact search variants or additional queries that people are likely to make in relation to the original keyword. Continuing with this example of "laptop", related queries may include "laptop price", "best laptop", "laptop UK", among others and these are specific searches that people may be performing. These search terms are more detailed and directly reflect what users are typing into the Google search box in relation to the original query.

In addition, it is worth noting that `related_topics` and `related_queries` present their results under the categories "rising" and "top", which were mentioned above. On the one hand, "rising" refers to terms or topics that have recently gained popularity, showing emerging trends, while "top" highlights those that are consistently popular or have the biggest search volume in a given period.

4.2.2 Retrieving Data from Google Manually

There are times when manual data collection from the Google Trends website is still a viable and sometimes necessary option when you can't or simply don't want to use third-party packages such as `gtrendsR`. While automation via an API may offer a more direct, official and systematic extraction of data, not having such an option may lead one to opt for the traditional method. However, it is important to note that this method has its own challenges such as the need to clean up the `.csv` file downloaded from the site before use.

The code for this method also starts by loading the relevant packages such as `readr` which is part of the tidyverse suite of packages and which allows us to read `.csv` files and has a number of functions for reading different types of data, making this very versatile for various data import tasks (Wickham, Hester and Bryan, n.d.).

To read our .csv file, the file needs to be manually cleaned up, as it contains headers and a different structure than necessary, and the date column needs to be formatted correctly.

```
data <- read.csv("laptop_interest_over_time.csv", skip = 2, header = TRUE)
data$Week <- as.Date(data$Week, format = "%Y-%m-%d")
```

It is important to note that for the following sections of this document, the gtrendsR package was used instead of parsing .csv files.

4.2.3 Visualizing Interest Over Time from Google Trends

Before diving into analysis and predictions, getting a clear picture of what the data looks like is essential, as it helps us understand the nature and structure of the data. In addition, by making visualisations we can observe any anomalies or quickly identify patterns.

As mentioned above, interest_over_time in Google Trends reflects how interest in the search terms (in this case, "laptop", "desktop" and "tablet") has changed over time. Recall that Google Trends provides this data on a relative basis for each term, with 100 being the point of highest popularity during the selected time period.

After retrieving the data from Google Trends using the gtrendsR package, essential pre-processing was performed to ensure that the dates were in the correct format and to remove any "noise" from the last month.

With the data cleaned, we proceeded to plot the trend of interest over time for the three search terms using the ggplot2 library.

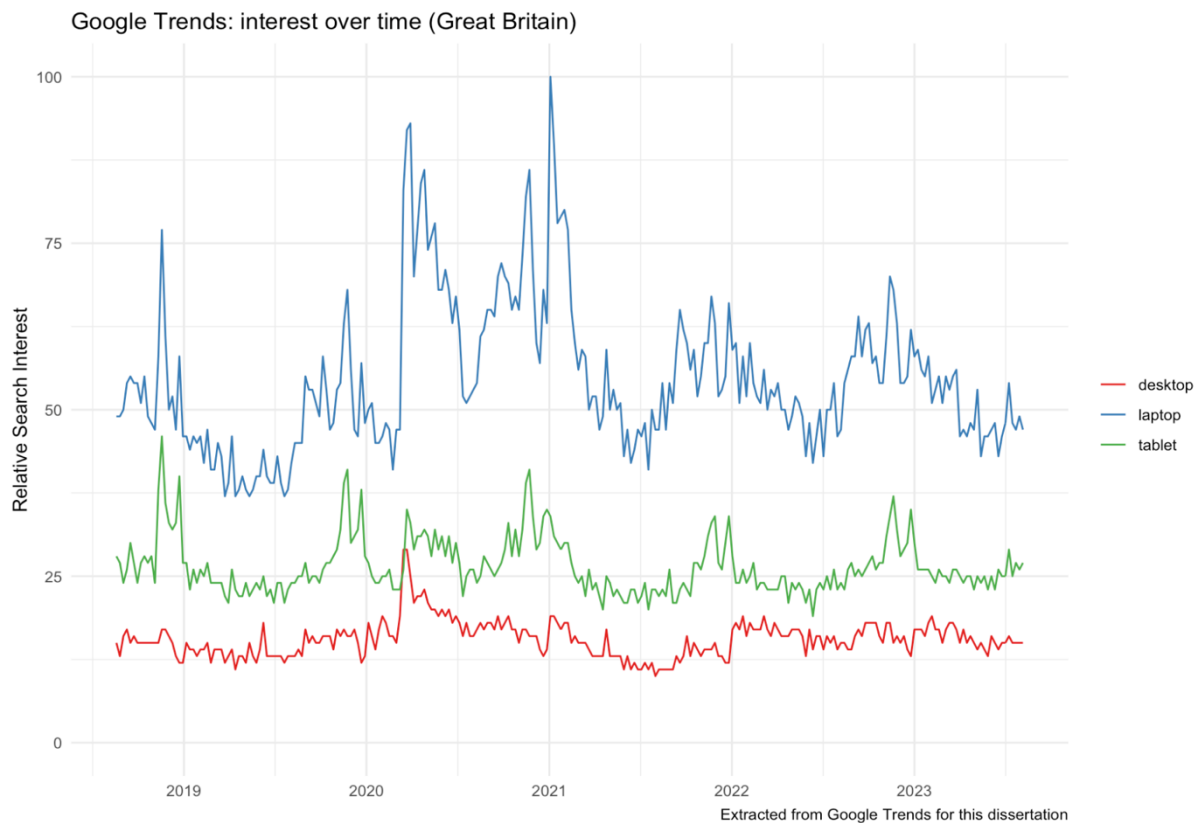


Figure 1. Interest Over Time

Components of the Graph shown in Figure 1:

X-axis: Represents time, showing how interest in search terms has evolved over time, in this case over the last 5 years.

Y-axis: Shows the relative interest index, with a range from 0 to 100.

Trend lines: There are three lines, each representing one of the search terms ("laptop", "desktop", "tablet"). These lines show the evolution of interest in these terms over time.

Looking at the graph in figure 1, we can discern patterns of interest over time, mainly in the terms "laptop" and "tablet", where their respective lines at first glance show some seasonality, making evident an increase in interest in the last quarter of each year; on the other hand, it is almost impossible to detect at first glance a seasonality in the search term "desktop". Another interesting fact is how "laptop" had a considerable growth during the first year of the COVID19 pandemic, which may coincide with the fact that as a result of this, laptop sales in 2020 had a 12.7% and 79.2 million units sold in the third quarter, its highest growth since 2014, due to the implementation of remote classes and remote work among other

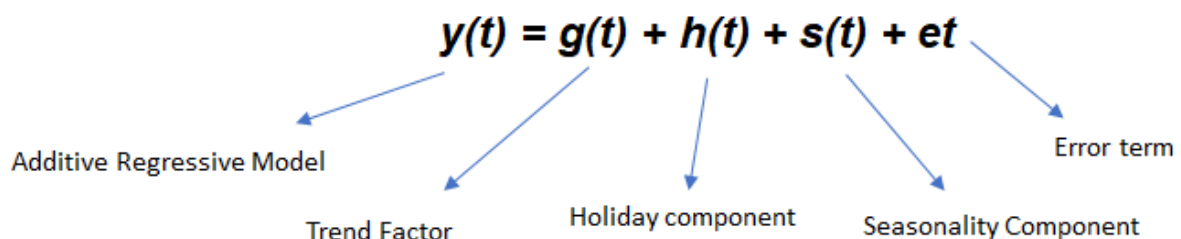
factors (Page, 2020; Pressman, 2021). A fact that does not seem to be a coincidence at all and that is directly related to the objective of this dissertation.

4.3 Forecasting Interest Over Time

Product search trends on platforms such as Google offer a unique window into consumer behaviour and preferences (Pool and Hernández-von, 2017). The question therefore arises: can the analysis of these trends predict future demand, changes in consumer preferences and the adaptability of the market to those changes?

As mentioned at the beginning the main objective of this dissertation is to design an approach to identify business opportunities in relation to product launch, marketing or further product strategies by harnessing the power of online search trends and unstructured data analysis. Furthermore, this dissertation seeks to eventually create a scalable and adaptable model that functions as a predictive tool for companies looking to diversify their product offerings into new markets or launch successful products from one country to another. In this context, it is essential to use tools that allow accurate and understandable prediction of these trends to perform our analysis.

The model developed by Facebook, prophet, which is a tool that facilitates the task of predicting time series data, is based on an additive model where non-linear trends are fitted with daily, weekly and yearly seasonalities, plus the effects of holidays. (Facebook, n.d.)



This mathematical approach, which combines both automatic and manual components, is attractive for addressing the challenges presented by Google Trends data, which often show

seasonal patterns and are influenced by specific events. Given the consistency and flexibility of prophet, I have decided to use it as my first forecasting model in this analysis.

4.3.1 Data Frame Preparation

It is worth noting that the following experimentation was made possible by parts of code in R by Christopher Yee in 2019, in which he explains the use of gtrendsR in R Studio and the possibility of using prediction models.

Before diving into predictions, it is essential to properly prepare our data. The format and structure of the data can significantly influence the quality of the predictions.

For this transformation process, the 'date' column was converted to a suitable date format, and it was decided to remove data from the last 15 days to avoid any noise or incompleteness, thus ensuring that the dataset is in an optimal format for modelling.

```
device_timeseries <- as_tibble(device$interest_over_time) %>%  
  mutate(date = ymd(date)) %>% # Convert to date format.  
  filter(date < Sys.Date() - 15) # Here we remove "noisy" data from the last 15 days
```

Likewise, only data related to the term "laptop" was filtered out, since of the three search terms, it is the one that shows a clear seasonality, which is feasible for our model. Finally, prophet requires the DataFrame to contain specific columns: ds for dates and y for values, so we proceeded with their renaming.

```
laptop_timeseries <- device_timeseries %>%  
  filter(keyword == 'laptop') %>%  
  select(date, hits) %>%  
  mutate(date = ymd(date)) %>%  
  rename(ds = date, y = hits) %>%  
  arrange(ds) #ARRANGE BY DATE
```

4.3.2 Model Building

With the data ready, we moved on to model building using prophet. The beauty of prophet lies in its ability to decompose the time series into its constituent components and model each component individually (Taylor and Letham, 2018; Yee, 2019). This allows for more accurate forecasting, taking into account seasonal variations and long-term trends.

The model is trained with historical laptop data and then a future time frame was created to make predictions for the next two years (730 days). The capabilities of prophet to automatically detect and model trends and seasonalities in the data make this process efficient and robust.

```
laptop_m <- prophet(laptop_timeseries)
laptop_future <- make_future_dataframe(laptop_m, periods = 730, freq = "day", include_history = TRUE)
laptop_ftdata <- as_tibble(predict(laptop_m, laptop_future))
```

We build a model using the prophet function applied to the laptop_timeseries time series. Essentially, prophet fits a model to the historical search data for "laptop" and creates an object that can be used to make future predictions. The result is stored in the laptop_m variable.

Subsequently, the make_future_dataframe function from the prophet package is used to create a DataFrame that extends into the future from the historical data from which the model was built. This is necessary to then make predictions on this extended set of dates.

Finally, the predict function was used to make predictions with the laptop_m model on the laptop_future DataFrame and the as_tibble function that converts the result of the predictions into a modern version of the data frame for R that is usually more user-friendly in terms of data manipulation and visualisation.

This is how these lines of code first build a model based on historical data from Google Trends, then create a set of dates that extends into the future to make predictions, and finally make and store those predictions.

4.3.3 Visualising predictions

Once the model has been built and the predictions have been generated, it is essential to visualise these results. It is important to understand that visualisation not only provides us with a graphical representation of the predictions but is a really important part of helping us to interpret and communicate the findings (Unwin, 2002).

Visualisation with ggplot2: This The first technique for visualising the predictions was with the use of the R library, ggplot2, which provides control and flexibility as we can customise every aspect of the graph, from colours to titles, as well as shading and line thickness. Figure 2 shows the result obtained with R using this method, in which the predictions were configured to be displayed in green.

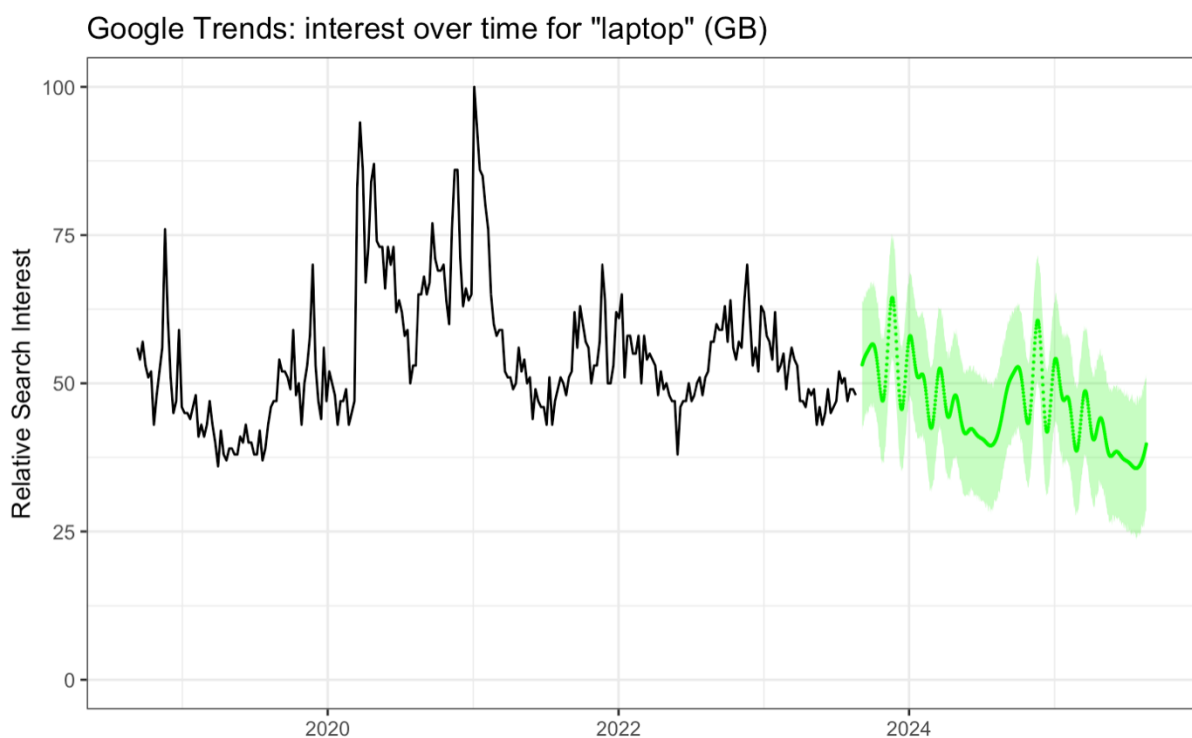


Figure 2. Interest Over Time and Forecast Using Prophet and visualising with ggplot2

The next way of visualisation is with the tool included in the prophet package, which is also customisable and much more interactive. To visualise the results, we are mainly interested in 2 things: looking at the graph with the forecast compared to the actual data, as well as any seasonality. For this, we will use two functions included in the prophet package;

dyplot.prophet for the time series plot and prophet_plot_components, which will show seasonality and trend.

```
dyplot.prophet(laptop_m , laptop_ftdata , uncertainty = TRUE) #interactive graphic
prophet_plot_components(laptop_m , laptop_ftdata , uncertainty = TRUE) #trend and seasonality
```

The time series graph shown in Figure 3 shows the historical data, in this case of interest for the search term "laptop", which are presented as black dots, on the other hand, the prediction is represented by a purple line, this is the prediction of the model based on the historical data, the visualisation is accompanied by a shaded area which represents the confidence intervals, both upper and lower and which represent the expected variability.

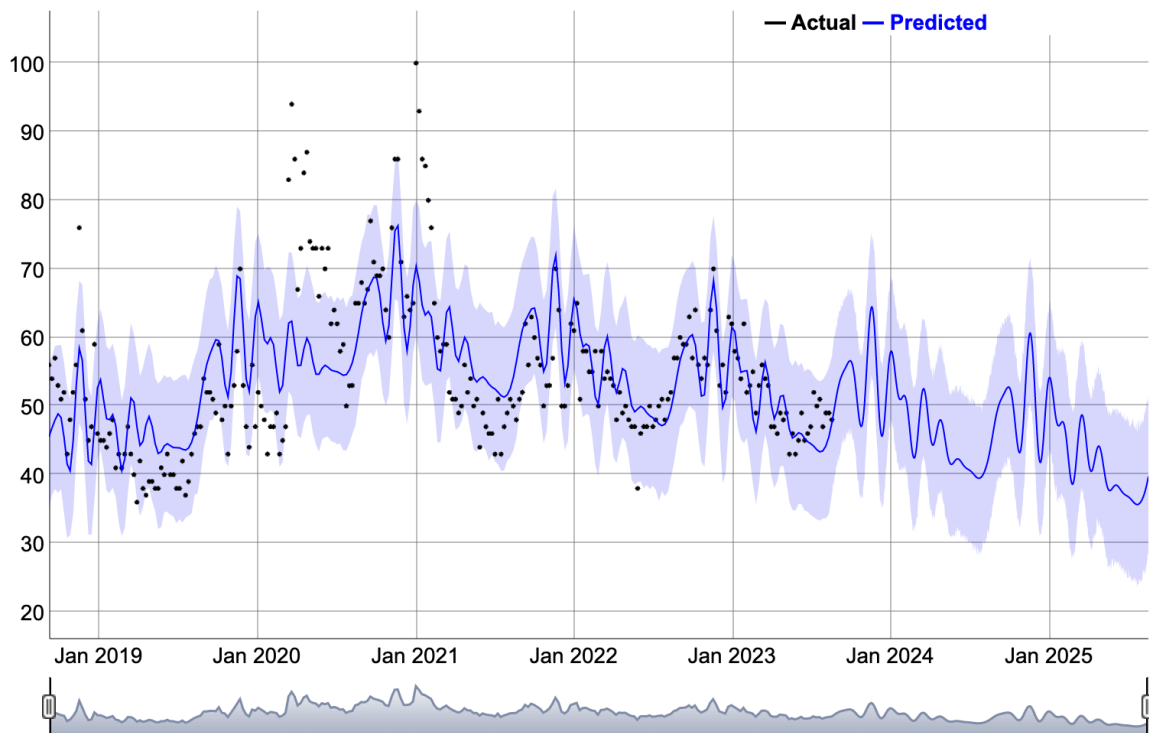


Figure 3. Interest Over Time and Forecast and Visualisation Using Prophet

The difference between using ggplot (Figure 2) and the prophet visualiser (Figure 3) lies mainly in the fact that with the former we need to combine the actual data with the predicted data with code, as well as configure the appearance of the graph, which prophet makes much simpler. Both methods have their own strengths and the choice will depend on the context

and the needs of the analysis, however, for this part of the dissertation we will focus on using the interface provided by prophet.

4.3.4 Interpretation of Obtained Forecasts

As could be observed in the historical data, a certain seasonality in the number of searches for the word 'laptop' is easily discernible. This seasonal pattern observed in Figure 4 manifests itself most prominently during the last four months of each year, peaking in the month of December. This phenomenon could be related to several variables, such as the Christmas shopping period, the end-of-year offers, or the need to purchase new equipment for the next academic or fiscal year.

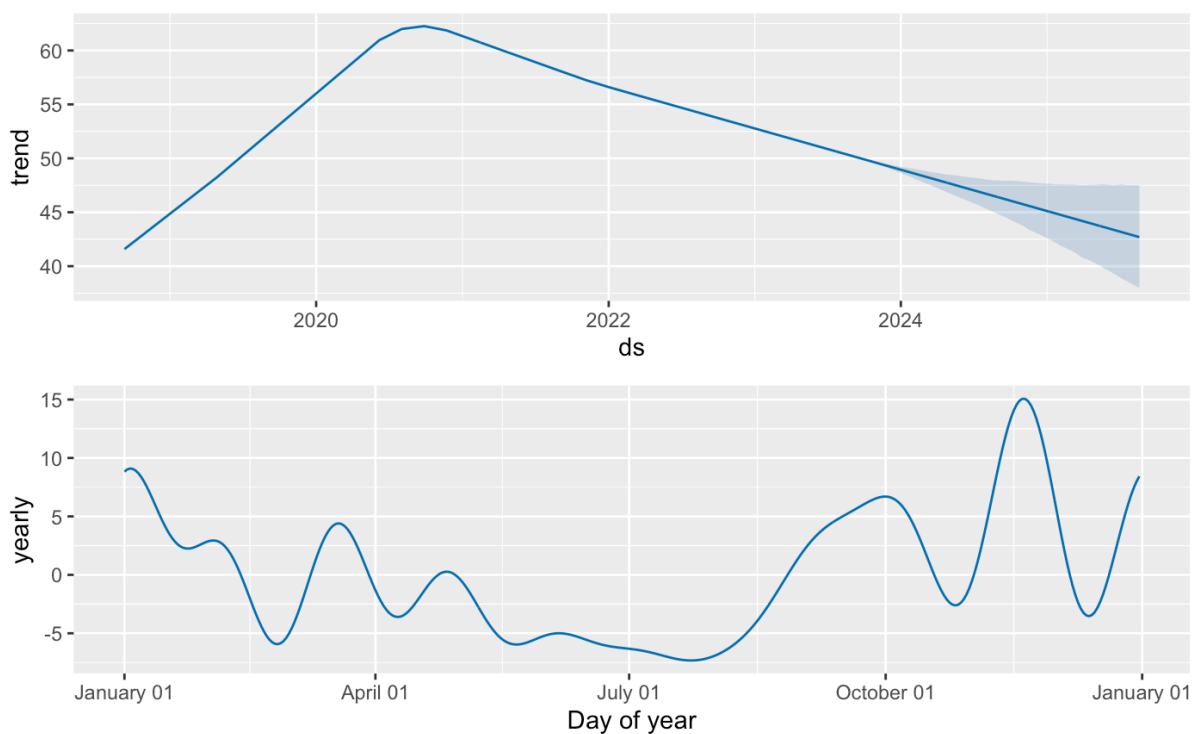


Figure 4. Seasonability with Weekly Data

Projected with the 'prophet' model for the next 2 years, or 730 days, we can again observe seasonality, which suggests that the model was able to understand the periodicity of the historical data. However, it is also important to note that we can notice a decreasing trend over time in the volume of searches. This downward trend could be interpreted as a gradual

change in consumer preferences, or if the peak of the rise in searches in 2020 was caused by the COVID19 pandemic, perhaps due to the end of the pandemic the values are gradually returning to normal. Another possible cause could be an increase in the diversification of technology-related searches, which could include more specific terms or new categories of devices, which we do not know at this point.

The `prophet` model decomposes the elements individually, which facilitates interpretation and more granular analysis of the time series, as well as including the ability to visualise the uncertainty interval and providing us with a range in which future values are likely to lie. This could be particularly useful for strategic decision making, as it could provide a degree of certainty.

At this point we can see that the forecast generated by the 'prophet' model not only validates patterns observed historically with the naked eye, but also offers valuable insights into possible future trends which, if there is a correlation between search interest and sales, could make it an extremely useful tool for any business strategy or market analysis.

With the above process it was possible to observe the forecast, as well as the trend and seasonality of the year making it possible to identify the months of the year in which users are most interested in a specific product according to search engine queries for that term. However, in a marketing strategy it may be convenient not only to know the seasonality during the year, but it would also be very useful to know which day or days of the week users are most interested in the search term in question. Hence the need for the following experimentation in this dissertation.

4.4 Granular Data Analysis with Trendecon: In Search of Weekly Seasonality

To identify whether there is a weekly seasonality in a time series, a higher resolution of data is necessary, and in this context, daily data are really valuable. By having data that show people's interest on a daily basis, it is possible to observe more precisely and in more detail the fluctuations and trends that occur over the course of a week.

For example, with daily data we could detect specific patterns related to consumer behaviour on different days of the week. We could observe whether searches or sales of a product increase on weekends compared to weekdays, or if there are peaks of activity on certain days such as Mondays or Fridays. These insights could be crucial for making informed decisions related to marketing strategy, inventory management or even human resource planning.

In contrast, with lower resolution, such as weekly or monthly data, we would lose the ability to analyse patterns within each week. We could identify trends over time, identify seasonality within the year as demonstrated in the previous experiment, but we would not have the level of detail needed to understand what is driving those trends at a more granular level.

However, this presents a challenge when the objective is to analyse data from Google Trends, due to the limitation that the platform presents when providing query data. The limitation in question basically lies in the fact that the data is presented on a weekly basis when queries are made for more than 90 days, and on a monthly basis when queries are made for data of more than 5 years, complicating experimentation. While we could analyse weekly seasonality based on data from 90 days ago, it is still a limitation, as the more data we have, the greater the robustness and accuracy of our analysis, allowing us to identify patterns and trends with greater confidence. Therefore, the availability of more historical data would be helpful for more robust and comprehensive analysis.

To obtain granular information from the daily data, the most viable option is to retrieve the data for each 90 days individually, and then integrate these time series into a single time series. While past publications have shown various methods for obtaining daily data from Google Trends for periods longer than 90 days, for this experimentation it was decided to use *trendecon*, which is an R package created in March 2020 by economists to analyse the impact of the Covid-19 pandemic on the Swiss economy (Trendecon, n.d.; Eichenauer, et al. 2020). In addition to being set up as a project among economists consisting mainly of Angelica Becerra, Vera Z. Eichenauer and Alexander Götz, it received the support of the German State Secretariat as well as the University of St. Gallen. This package essentially aims to create daily time series from Google Trends data. It does this by querying the data several times and then harmonising the data and merging it into a time series (Trendecon, n.d.; Eichenauer, et al., 2020).

4.4.1 Experimentation: Forecasting Granular Data

To obtain the data on a daily basis, the function "ts_gtrends_mwd" from the "trendecon" package was used. This is the function that allows querying Google Trends on a daily basis and obtaining detailed online search time series. In the following, the process is described based on the example provided:

As performed in the previous experimentation with the gtrendsR package, first the search keyword is defined, in this case the word "laptop". In addition, the Google Trends search category is specified as "0", indicating a general search. The "geo" parameter is set to "GB" to limit the search to the UK. Unlike the gtrends function in the gtrendsR package, here we need to specify the exact start and end dates of the query. In this example and to homologate our experimentation with the dates of the previous experiment, the start date was set to "2019-09-05" and the end date to the last date of the available data.

```
robust_laptop <- ts_gtrends_mwd(  
  keyword = "laptop",  
  category = "0",  
  geo = "GB",  
  from = "2019-09-05", #The starting date.  
  to = Sys.Date() #the ending date. Sys.Date() is for the current day.  
)
```

Once the parameters have been defined, the query is executed using the 'ts_gtrends_mwd' function. This function performs daily queries to Google Trends during the specified period and collects the search data for the keyword "laptop" in the UK.

Once this command is executed, the query results are stored in an object we create called 'robust_laptop'. This object will have the detailed time series of the popularity of the keyword "laptop" in the UK, with daily data from the given date to the current date.

As in the previous experiment, where we used weekly data, the first step is to ensure that the data is in a format that the 'prophet' model can easily understand. In this case, we rename the columns containing the dates and values to "ds" and "y", respectively, to ensure that the model can interpret the data correctly.

```
robust_laptop_timeseries <- robust_laptop %>%  
  select(time, value) %>%  
  mutate(time = ymd(time)) %>%  
  rename(ds = time, y = value) %>%  
  arrange(ds) #ARRANGE BY DATE
```

Once the data have been properly formatted, we move on to the model building stage. In this experiment, as in the previous one, the objective is to make a prediction over 730 days. To do this we create a future dataframe using the `make_future_dataframe` function. This function is used to generate a dataframe containing both the history of dates and the future dates for which we want to make predictions.

```
r_laptop_m <- prophet(robust_laptop_timeseries)  
r_laptop_future <- make_future_dataframe(r_laptop_m, periods = 730, freq = "day", include_history = TRUE)  
r_laptop_ftdata <- as_tibble(predict(r_laptop_m, r_laptop_future))
```

Note that we specify the frequency as "day" (`freq = "day"`) and also include the data history (`include_history = TRUE`). By doing so, we ensure that the model takes historical data into account when making future predictions.

Finally, we run the 'predict' function to get the predictions based on the model and the future dataframe we have created. We convert the result into a tibble for easy manipulation and further analysis.

4.4.2 Interpretation of Forecasts Obtained with Granular Data

By generating the plot with `dyplot.prophet` function, we obtain a plot similar to the one obtained in the previous experiment, which can be seen in the upper half of the image below, however a closer inspection reveals a higher density of data points in this version. This higher data density is due to the facts that we are using daily data instead of weekly data, which provides a much higher resolution and detail for trend analysis.

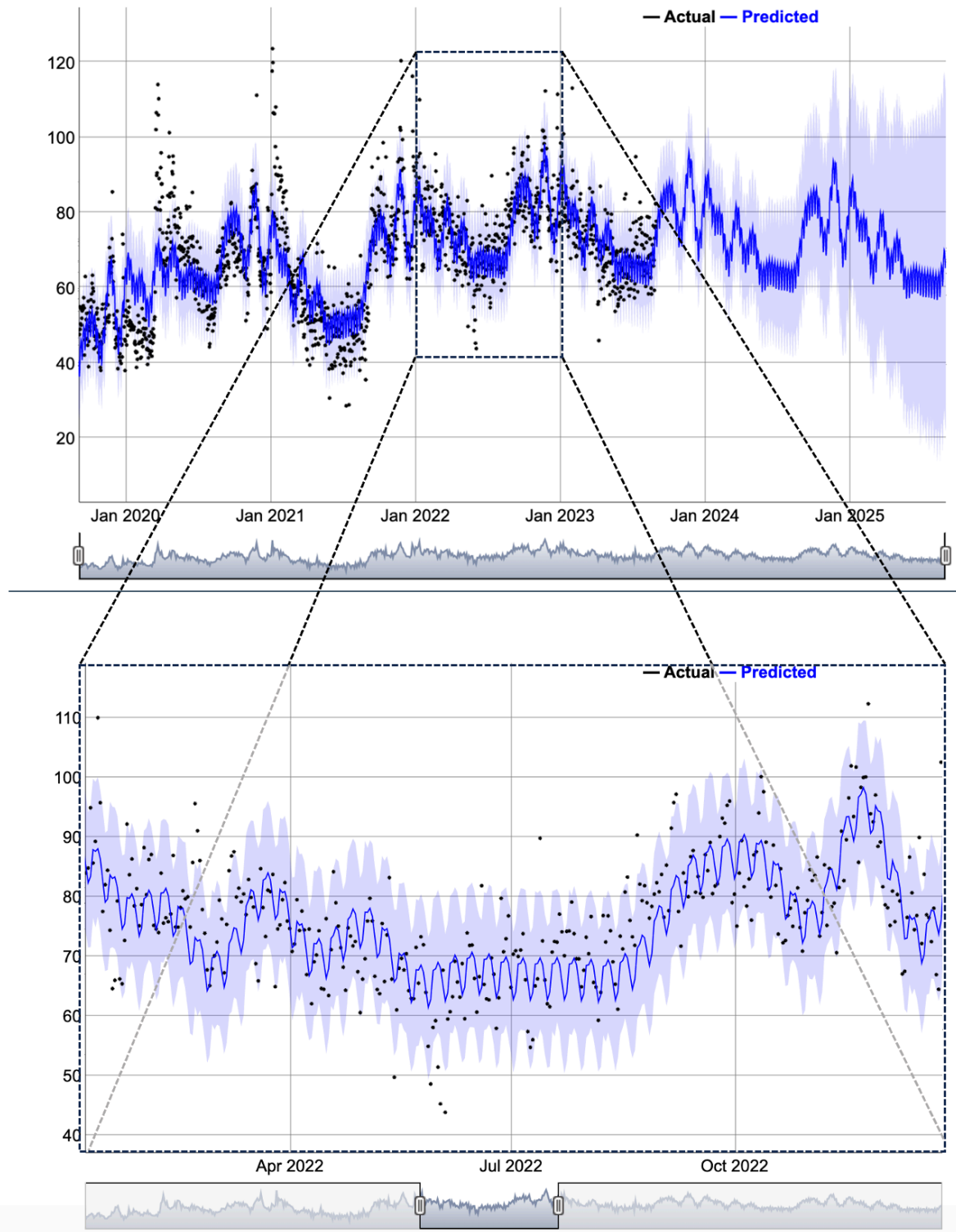


Figure 5. Observing Granular Data

As we zoom in on the interactive graph, we begin to notice patterns that are recurring and consistently manifest themselves each week as seen in the lower section of Figure 5. This repetitive behaviour is a strong indicator that there is some form of weekly seasonality in the data which could be vital to better understand short-term fluctuations in search interest and,

in keeping with the aim of this paper, could have significant implications for marketing or product inventory strategies, for example. It is worth noting that with daily data, greater variability is obtained in the upper and lower bounds of the predicted data going forward. As mentioned above, this variability is represented in the graph as a shadow that is much longer than with weekly forecasts. This result is expected, as the daily data has larger fluctuations, as the daily behaviour within a week is recorded, unlike the weekly data which is smoothed.

This result is expected since we have daily data, which captures shorter and more numerous fluctuations by documenting day-to-day search behaviour. In contrast, weekly data provides a smoothed and aggregated view of user behaviour, which minimises the occurrence of sharp peaks or valleys. In addition, the greater variability in predictions based on daily data is a direct consequence of the higher resolution and sensitivity of these data to short-term changes. (Clements and Hendry, 1998).

Having observed these initial characteristics, the next step is to decompose the patterns further using the 'prophet_plot_components' function which will allow us to break down the different components that influence the time series: general trend, annual seasonality and weekly seasonality, among other possible components which can be seen in Figure 6.

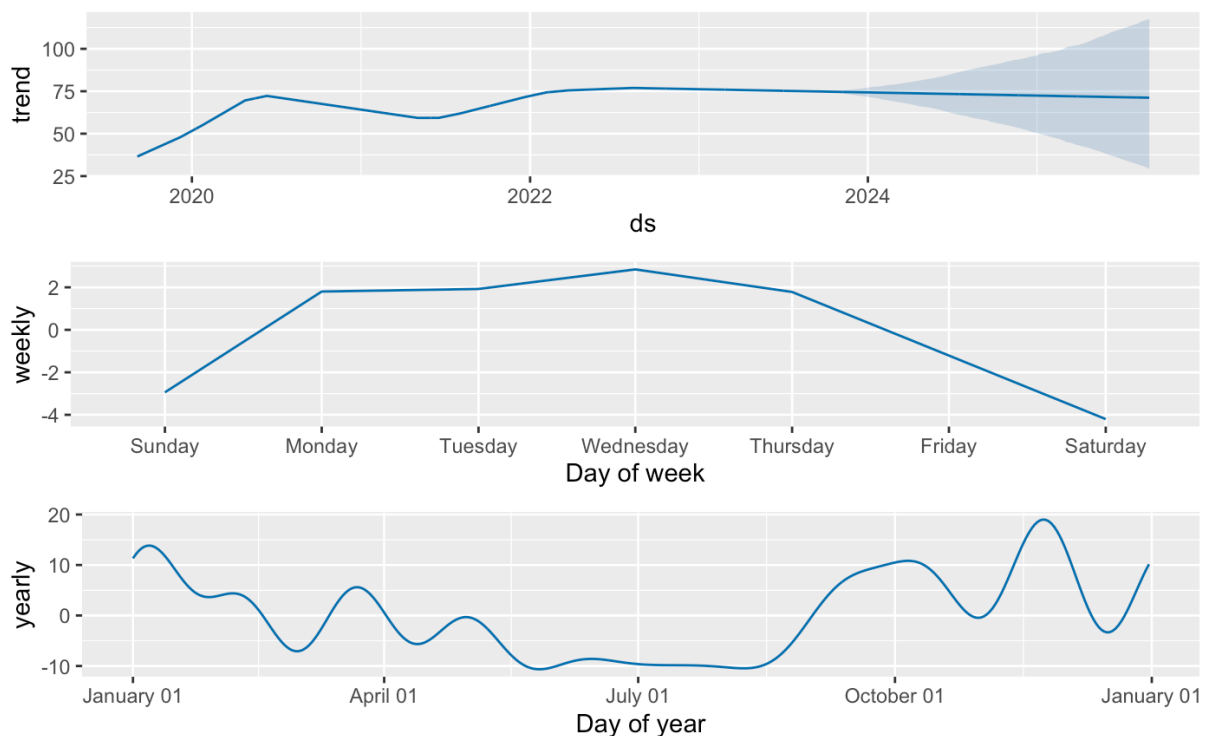


Figure 6. Seasonality with Daily Data

The result is interesting, we can see how there is a higher amount of searches during the week, with Wednesday having the highest amount of searches and a lower volume during the weekend, with Saturday being the day with the lowest volume. In addition, when analysing this granular data, we can see that the yearly trend was also impacted when compared to the previous analysis of weekly data. In this result with daily data and unlike the result with weekly data, we cannot observe a downward trend after the year 2021.

4.5 Unstructured Data Analysis: Search Suggestions

As we continue to explore ways to understand user behaviour and predict market trends based on internet searches, the next area of focus for this dissertation is the analysis of search engine suggestions. It is well known that in this digital age where the internet is the first point of contact for consumer research, understanding how users conduct their search queries can provide valuable insights into audience sentiment and emerging trends.

4.5.1 What is Unstructured Data?

Before we get into the next experiments in this document, we need to understand what "unstructured data" is. Unstructured data is data that does not adhere to a predefined data model or structure, such as tables. And so unlike structured data, which are usually numerical and fit into tables, unstructured data can take the form of text, images, sound, video, etc. In fact most of the data generated on the planet is unstructured data with a 80% of global data, in addition, 90% of unstructured data will have been generated in the last two years (IBM, 2021; Columbus, 2021). It is therefore clear that this type of data requires a different treatment to convert it into a structured form so that it is easier to analyse. Search engine suggestions fall into this category, as they are presented in the form of free text without pre-set formatting, these search suggestions can cover a wide variety of topics and terms where we could even identify user sentiments and intentions.

Search suggestions are not the only phrases or keywords we can analyse in this section, but it is also of great interest to analyse from Google trends the `related_topics`, which as mentioned above identify topics or categories of interest related to the original keyword or query, as well as the `related_queries` which identifies exact variants of searches or additional queries that people often make in relation to the original keyword.

But why is it important to integrate unstructured data into the analysis, the main objective of this dissertation is to develop a comprehensive approach in which product-related opportunities can be identified, thus leveraging search trends and unstructured data. Since search engine suggestions as well as related topics and queries from Google Trends are phrases or words, they are therefore a form of unstructured data and thus their analysis could provide a source of qualitative information. This could of course be complemented by the above experimentation, adding another layer of depth to our predictive and visualisation capabilities. Through this analysis, the aim is to better understand evolving trends, consumer intentions and potential "gaps" in the market that could be opportunities for product launches.

By integrating structured data analysis, such as Google Trends, with unstructured data analysis, we aim to build a more multi-dimensional toolset that can serve companies looking to make decisions in the marketplace.

4.5.2 Obtaining Search Suggestions: Vector Method and Alphabetical Levels Method

Having mentioned the above, this section will explore the functionality to retrieve Google search suggestions in R. The aim is to provide a list of suggested search terms, either from a single search term or from a vector containing several search terms. By doing so, we will get a more complete picture of the users' search behaviour and the keywords that are semantically connected to our initial search term. In the first instance this kind of data could be very useful to understand how the user searches the internet, which would be very useful for the SEO of a website or to identify related products and trends (Ward, Hahn and Feist, 2012; Shenoy and Prabhu, 2016).

The following experiments were made possible thanks to functions for R created by François Joly (no date) that interact directly with the Google search API to obtain search suggestions. The functions in question are: `getGSQueries` and `suggestGSQueries`, which are included in the appendix of this document.

To set up this experiment the first step was to make sure that the XML package is installed, which is essential for parsing the XML-based response we will get from Google's search suggestions.

The `suggestGSQueries` function takes three parameters:

- `search_query` - the initial search keyword(s).
- `code_lang` - the language code for results.
- `level` - the depth of suggestions we're interested in.

```
kwd <- suggestGSQueries('keyword(s)', 'en', 1)
```

As François Joly mentions in his document, there are two main methods to collect Google search suggestions using the function:

By alphabetical levels: in this approach depicted in Figure 7, we use the parameter `level` to define how deep we want to explore the search suggestions. A level 1 simply retrieves the first list of keyword suggestions; level 2 adds another letter to the search term ('keyword a...', 'keyword b...', 'keyword c...', ... 'keyword z...'), and level 3 adds two letters to each search term ('keyword aa', 'keyword ab', ... 'keyword zz').

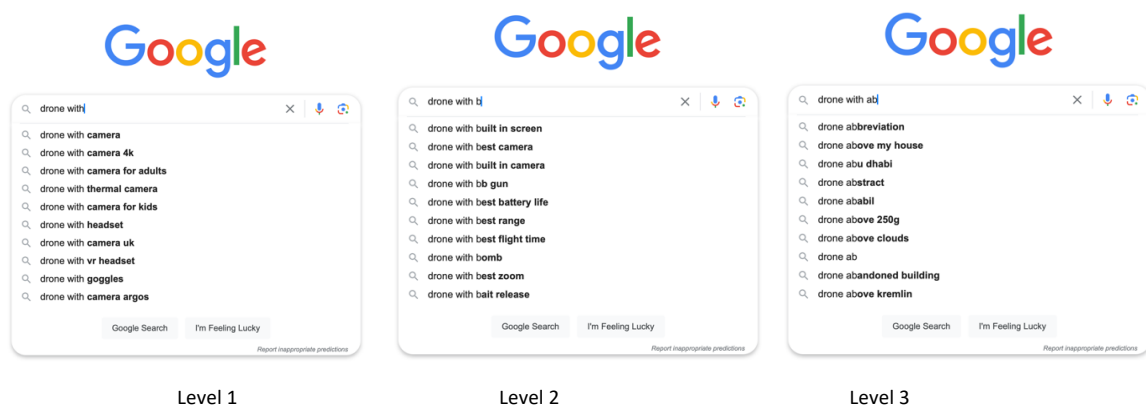


Figure 7 – Alphabetic Levels

By Vectors: With this method there is more flexibility as it allows us to explore further into the suggestions in a nested way. For example, as we can see in Figure 8, if in level 1 we get ten suggestions, which is the maximum amount that Google returns per search term, we can use those ten terms as a vector to pass them through the function and get another 10 suggestions related to each of the original terms, giving us a total of 100 search terms.

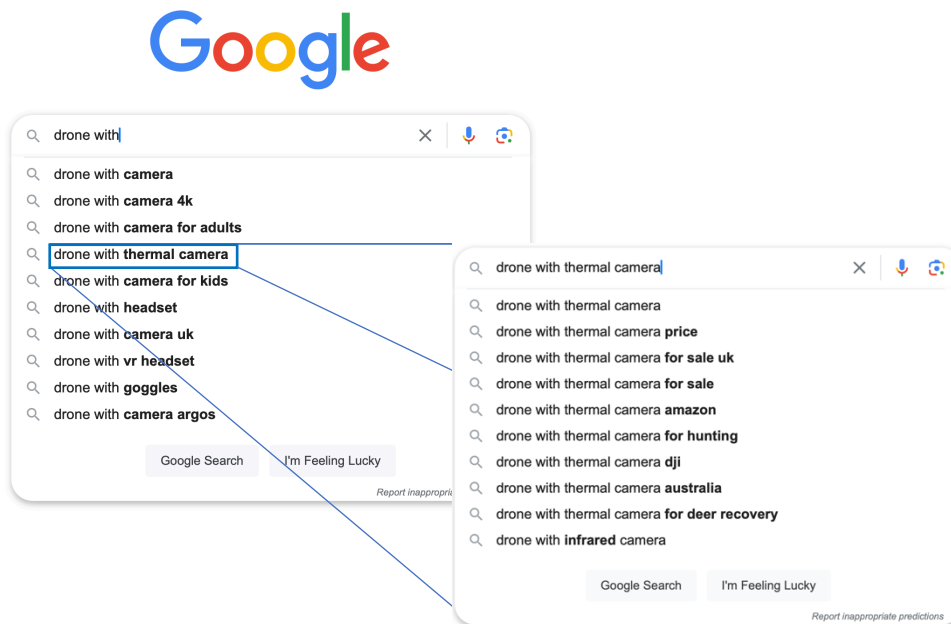


Figure 8 -Requests using a vector.

Experimentation A: with vectors

Resuming our experimentation, in this first stage we employ a vector approach to deepen the analysis of search suggestions. Initially, we use the query 'where can I buy', a phrase commonly typed in the search engine when we are looking for a specific product or service that we do not know where to find.

```
kwd <- suggestGSQueries('where can I buy', 'en', 1)
```

```
where can i buy an oyster card
where can i buy prime drink
where can i buy clomid
where can i buy stamps
where can i buy wegovy uk
where can i buy john lewis vouchers
where can i buy ozempic uk
where can i buy covid tests
where can i buy ozempic
where can i buy naproxen
```

1-10 of 10 rows

After obtaining the first 10 suggestions for this query, we decided to go one step further and explore secondary suggestions for each of the initially suggested terms. This approach allows us to get a more complete understanding of the spectrum of related searches that users might be performing.

```
deeper_kwd <- suggestGSQueries(kwd, 'en', 1)
print(as.data.frame(unlist(deeper_kwd)))
```

The screenshot shows two panels of R console output. The left panel displays the first 10 rows of suggestions for the query 'where can i buy clomid'. The right panel shows a zoomed-in view of suggestions for 'where can i buy clomid' from rows 21 to 30. The suggestions include various related terms and locations.

| Row | Suggestion |
|-----|--|
| 1 | where can i buy an oyster card |
| 2 | where can i buy prime drink |
| 3 | where can i buy clomid |
| 4 | where can i buy stamps |
| 5 | where can i buy wegovy uk |
| 6 | where can i buy john lewis vouchers |
| 7 | where can i buy ozempic uk |
| 8 | where can i buy covid tests |
| 9 | where can i buy ozempic |
| 10 | where can i buy naproxen |
| 21 | where can i buy clomid |
| 22 | where can i buy clomid over the counter uk |
| 23 | where can i buy clomid online |
| 24 | where can i buy clomid in the uk |
| 25 | where can i buy clomid over the counter |
| 26 | where can i buy clomid online uk |
| 27 | where can i buy clomid tablets |
| 28 | where can i buy clomid in south africa |
| 29 | where can i buy clomid pills |
| 30 | where can i buy clomid in nigeria |

As a result of this methodology, we ended up with a set of 89 suggested search terms.

Experimentation B: with alphabetical levels

In this part, the experimentation was carried out, but now with alphabetical level 2, i.e. our search term followed by a space and one by one the letters of the alphabet.

```
kwdlev2 <- suggestGSQueries('where can I buy', 'en', 2)
print(as.data.frame(unlist(kwd)))
```

The screenshot shows a list of search suggestions for the query 'where can i buy acetone'. The suggestions include various related terms and locations.

| Row | Suggestion |
|-----|---------------------------------------|
| 11 | where can i buy acetone |
| 12 | where can i buy a covid test |
| 13 | where can i buy amazon gift cards |
| 14 | where can i buy activ8 joint complete |
| 15 | where can i buy a tortoise |
| 16 | where can i buy amoxicillin |
| 17 | where can i buy apple cider vinegar |
| 18 | where can i buy aldi perfume |
| 19 | where can i buy a monkey uk |
| 20 | where can i buy baking soda |

In this way we get 260 suggested search terms, a much larger number than with the use of vectors. However, here the question arises, which method would add more value? Although more than twice as many suggestions were obtained with the alphabetical method than with the vector method, this does not ensure that they are more useful. In order to interpret the

data in a more visual way, word clouds were created to provide a representation of the most frequently mentioned words in a corpus of text.

4.5.3 Experimentation with Word Clouds

After a data cleaning process in which we removed redundant or irrelevant terms, such as the query phrase 'where can I buy' itself, we generated a wordcloud for both experiments. This visualisation was done with the aim of providing us with a clearer understanding of the most common and relevant search queries by users.

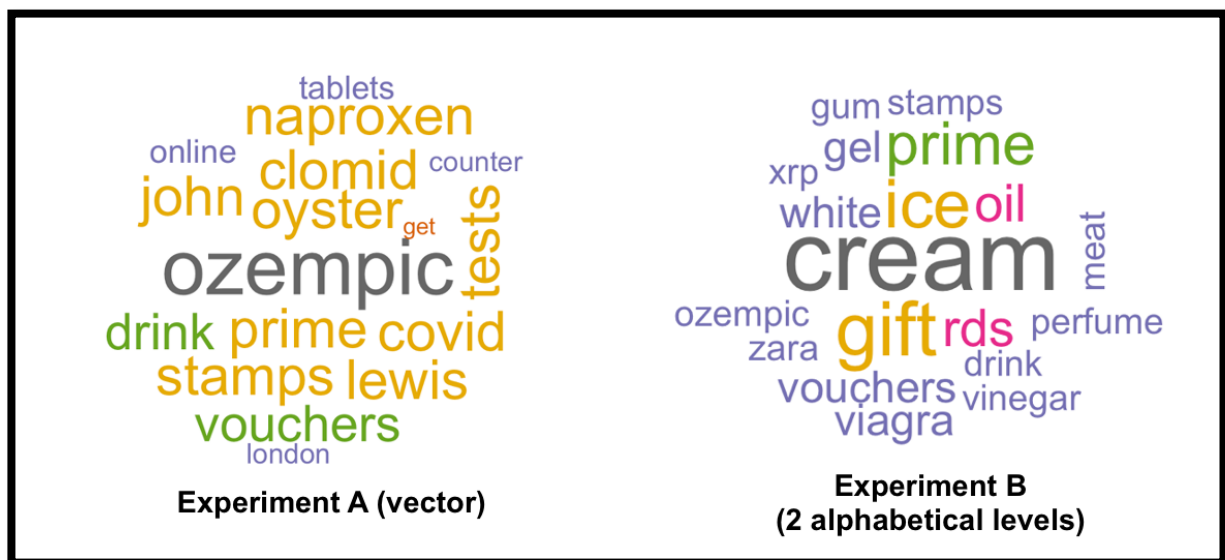


Figure 9 – Word clouds

As we can see in Figure 9, the results are different, although we can identify some products in both word clouds and this could be interesting to see what users are looking for when completing the sentence "where can I buy..." we can not really get any significant insight yet between the differences that exist between the vector method and the alphabetic level method, so we proceed to expand the experimentation by adding another vector in experiment A. For this we used the following code:

```
kwd <- suggestGSQueries('where can I buy ', 'en', 1)
kwd2 <- suggestGSQueries(kwd, 'en', 1)
kwd3 <- suggestGSQueries(kwd2, 'en', 1)
```

Likewise, in the method of alphabetical levels, level 3 was used instead of the level 2 that had been used previously, as follows:

```
kwd_lev3 <- suggestGSQueries('where can I buy ', 'en', 3)
```

By its nature, this code generates queries as follows:

- Where can I buy a
- Where can I buy aa
- Where can I buy ab
- Where can I buy ac
- ...
- Where can I buy zz

As a result, the number of search terms multiplied from 89 suggestions to 580 suggestions in the vector method and from 260 to an extraordinary 6,138 suggestions in the alphabetical method to generate the following word clouds contained in Figure 10.

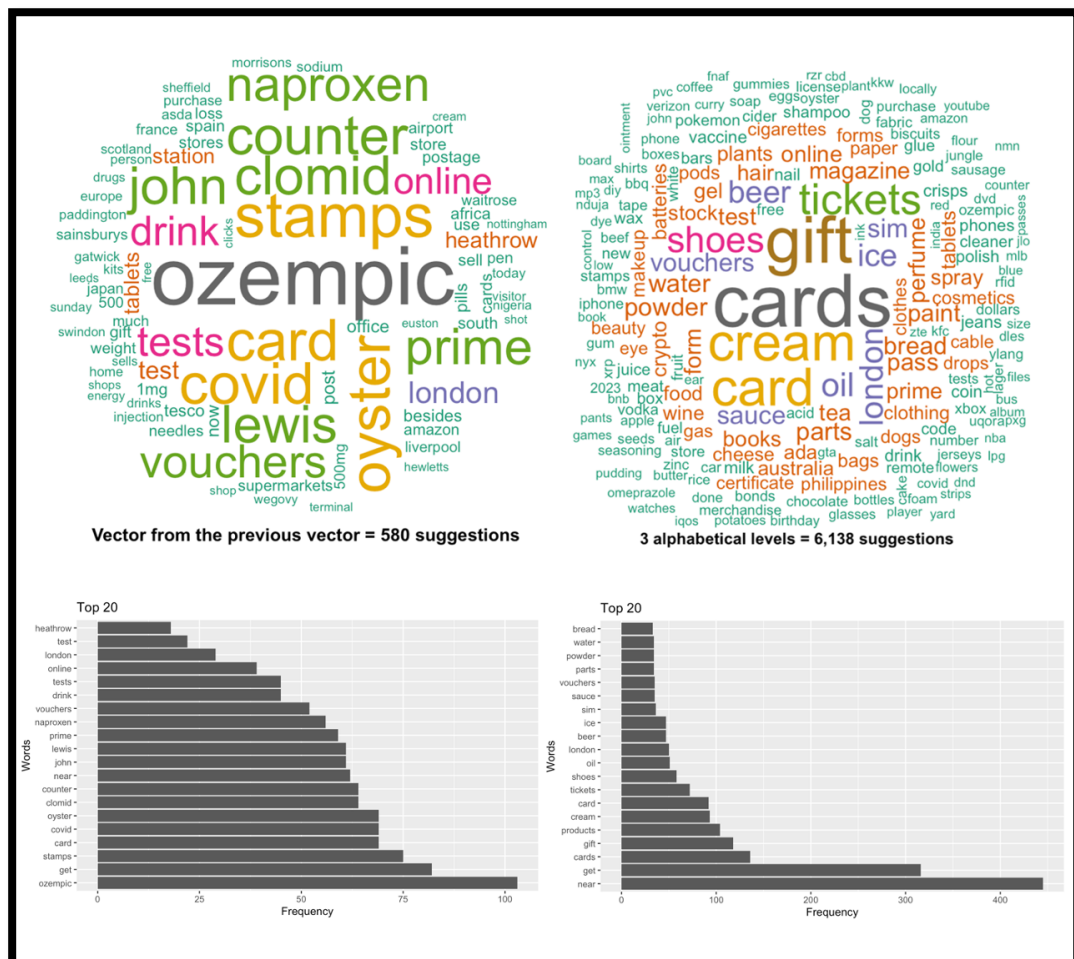


Figure 10 – Second set of word clouds

4.5.4 Comparison of Methods Used for Collecting Search Suggestions

Looking at the word cloud on the top left in Figure 10, we can see that with the vector method the sentences obtained are more specific. Moreover this method seems to have a multifaceted coverage, as it is more useful when we want to explore different aspects of an initial query, which could provide a better understanding of the users' specific intentions. For example, for the phrase "where can I buy an oyster card," multiple extensions are generated, covering everything from geographical proximity ("near me") to more specific locations ("at Heathrow Airport").

As the vector method is a method in which search suggestions are queried from a previous set of search suggestions, it could be very useful for generating visual networks in which the co-occurrence of words in search terms is observed, which we will explore later. Interestingly enough, this method tends to generate terms that are specific to certain locations, such as train stations, cities and countries, or in this case, the place where people expect to find a certain product.

On the other hand, a relevant aspect of the alphabetic levels method, which can be seen on the right hand side of Figure 10, is that this method has the ability to generate a significantly higher volume of search suggestions. However, this increase in quantity may come at the cost of query quality and specificity. In other words, while it does provide a broader view of overall search trends, it may also lack the depth and detail that is provided by the vector-based method. This trade-off between quantity and quality can be crucial when deciding which method to use depending on whether you are looking for an overview or a more focused analysis.

If the goal is to perform an in-depth analysis to understand how queries are modified to express more specific intentions, especially in relation to geolocation, the vector method is more appropriate. On the other hand, if what is sought is a more general and broader view of the varied search intentions that users might have, regardless of how much weight they have in relation to each other, the alphabetical method might be more appropriate.

At this point we can conclude that both methods have their own advantages and limitations. Which makes it really attractive to employ both approaches separately if our goal is to create a comprehensive tool that helps companies to comprehensively understand user search behaviour. While the vector-based method is better suited to explore variability and specificity in queries, especially in terms of location and detail, the alphabetical method is more effective in providing a quicker overview of general areas of interest around a keyword. By integrating these two approaches, companies could benefit from both a detailed analysis and an overview, allowing for more informed and strategic decision-making for product launches, or to improve their marketing campaigns and SEO strategies.

4.6 Using Related Queries from Google Trends as a Vector for Obtaining Search Suggestions

The use of the "rising" and "top" related_queries in Google Trends, which we explained in section 4.1.2 and 4.2.1, presents a good opportunity to enrich and contextualise the search suggestions generated by the vector method. Rising terms reflect queries that have gained relevance over a specific period of time, while top terms denote the most popular queries in general and that have remained in good standing over time. By incorporating this data into the initial query vector as shown in Figure 11, more relevant and opportunistic search suggestions can be generated. This not only captures current and emerging trends, but also allows us to gain a deeper understanding of dynamic changes in user preferences and needs. This strategy can be very useful to identify products or services that are gaining traction or to uncover areas of need that might have been overlooked by companies or other providers. By combining these data sources, you get a more robust and adaptive tool that can effectively guide decision-making in a variety of contexts, from product development to adapting to rapidly evolving market trends.

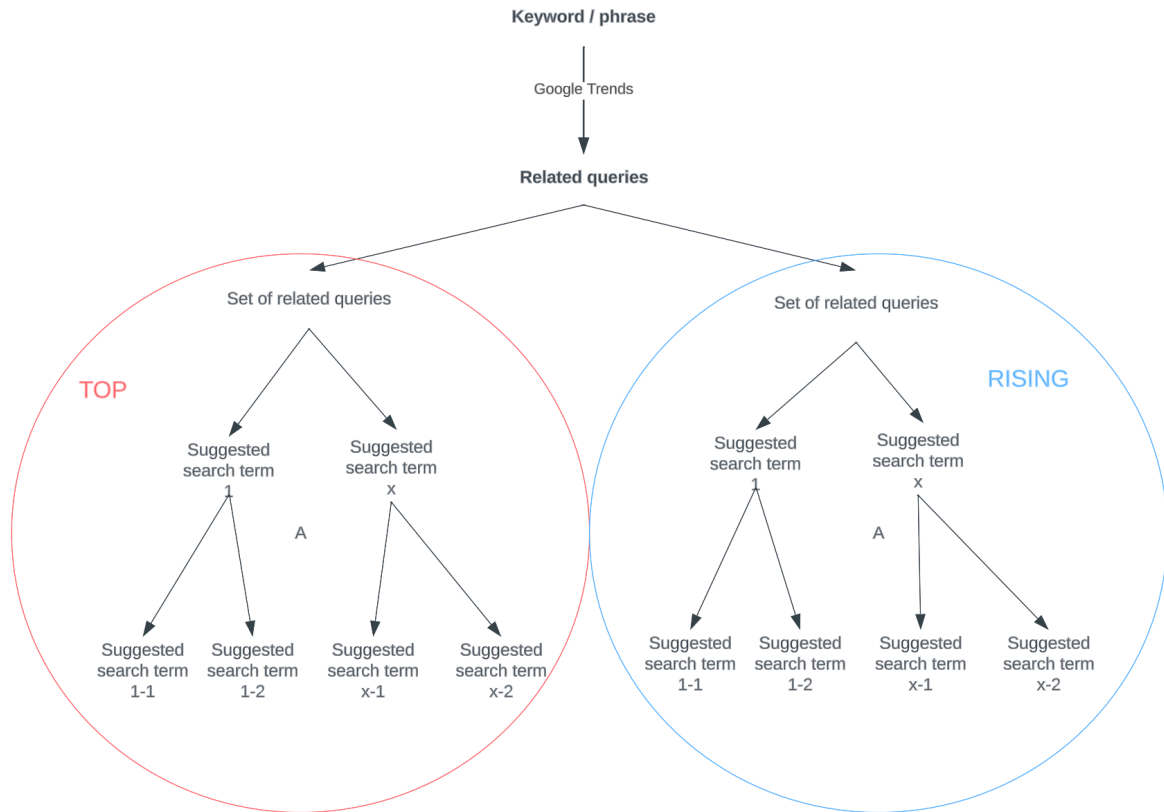


Figure 11 – Proposed methodology for obtaining suggested search terms

4.7 Co-occurrence Matrix and Semantic Networks

Given the multifaceted nature of the vector method, which expands the initial query in multiple directions to encompass a wider range of terms related to the initial query, we decided to employ co-occurrence matrices to subsequently create semantic networks and deepen our analysis of the search suggestions generated by this method.

Semantic networks provide an intuitive visual representation of how search terms are interconnected, allowing us to better understand the underlying relationships and patterns that exist between various search queries (Driegera, 2013). This additional exploration can provide us with much more detailed and easily visualised insights into specific user intentions, which can be invaluable for fine-tuning strategies for substitute product introduction, marketing, search engine optimisation (SEO), etc.

4.7.1 Creating the Co-occurrence Matrix

For the following experiments we used the `igraph` package which is a tool for the generation of the co-occurrence matrix and which was installed and run in R Studio (Csárdi and Nepusz, 2006). Subsequently, we obtained the `related_queries` from Google Trends for the search term "laptop" for the last 5 years in the UK, our first experimentation in this section, as can be seen in the first code block in Figure 12. These data were stored in two separate variables, one for the top `related_queries` and one for the rising `related_queries`. The latter can be seen in the second code block of the same Figure.

In the third code block shown in the same figure, the `suggestGSQueries` function, as we did in the previous experiments, was used to generate Google search suggestions, in this case based on the terms stored in the `rq_rising` and `rq_top` vectors. These suggestions will be aimed at further refining the related queries, providing a more detailed understanding of the sub-themes around the search term.

```
```{r}
keyword <- gtrends(
 keyword = c("laptop"),
 geo = "GB",
 time = "today+5-y",
 tz = 0
)
```

```{r}
#RISING
keyword_related_queries <- keyword$related_queries #Getting the related queries from Google Trends

rising_keyword <- keyword_related_queries %>% #Filtering just rising related queries
 filter(related_queries == 'rising')

Getting the 'value' column from them and storing it in a vector
rq_rising <- rising_keyword$value
```

```{r}
kwd_RIS <- suggestGSQueries(rq_rising, 'en', 1)
```
```

Figure 12 – Obtaining suggestions from rising `related_queries`.

After this process, the suggested search terms are processed to eliminate irrelevant, common words, or if what we want is, as in this case, to have a clearer and easier to read representation, we can also eliminate the initial keyword "laptop", as well as the so-called "stop words" of the English language which are recommended to be eliminated in any text-mining process (Silva and Ribeiro, 2003). Once cleaned, these words are used to construct a co-occurrence matrix, which will record which words appear together most frequently as and how.

4.7.2 Creating a Semantic Network

After obtaining our co-occurrence matrix, we proceeded to create a semantic network from it, where the nodes will represent single words and the edges will represent the co-occurrences that exist between them. In addition, the degree of each vertex was calculated, and the size of the vertices was adjusted according to their degree in order to visualise more intuitively which words are more central in the network.

4.7.2.1 Algorithm for the semantic network: The Louvain Algorithm

For this section, a Louvain design algorithm found in the `igraph` package that detects clusters in the matrix was used. This algorithm is efficient and effective in finding cluster structures in large network datasets. The use of the Louvain algorithm allowed the identification of subgroups of words that tend to co-occur together more frequently, which makes it much easier to interpret the overall pattern of the semantic network, and the same algorithm was used to assign different colours to each cluster found, to help with the interpretation of the data (Csárdi and Nepusz, no date; Junker, 2020; Murniyati, et al. 2023).

It was decided to use Louvain's algorithm due to several reasons that make it suitable for our study. First, the algorithm does not require the number of clusters to be specified, which is an advantage when analysing new or unstructured datasets as in this case, where the optimal

number of clusters is not obvious (Junker, 2020; Murniyati, et al. 2023; Pradana, Kusumawardani and Permanasari, 2020).

Furthermore, compared to other clustering methods such as K-means and hierarchical techniques, the Louvain algorithm has the ability to split data more evenly and accurately than others. This is due to its focus on optimising modularity, which results in more efficient identification of communities or clusters that are strongly interconnected (Pradana, Kusumawardani and Permanasari, 2020). This is especially useful in this case, where co-occurrence and connectivity between different search terms can be indicative of deeper semantic relationships or consumer behaviour patterns.

In the Louvain logarithm, the aim is to maximise Q until it can no longer be improved:

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j),$$

Where:

- A_{ij} is the link weight between nodes i and j .
- k_i and k_j are the degrees of nodes i and j , respectively.
- m is the total sum of the link weights in the network.
- $\delta(c_i, c_j)$ is 1 if nodes i and j are in the same community and 0 otherwise.

4.7.2.2 Visualization of the Semantic Network

Understanding the above, we finally proceed with the visualisation of the semantic network by means of the `visNetwork` library, which generates a representation that is interactive and allows a better understanding of the network structure, incorporating elements such as changes in node size based on degree, as well as the opportunity to visualise a specific node and its respective connections (Thieurmel, 2022). Furthermore, for a better understanding of the semantic network, the clusters that were previously calculated were visualised with

colours. The network also allows for interactive manipulation, such as dragging or rearranging the nodes, allowing for a clearer visualisation and deeper exploration.

In Figure 13 we can see the result of a semantic network using the search suggestions obtained from the rising related_queries of the word "laptop". Note that the words laptop and laptops themselves were previously filtered for a clearer view.

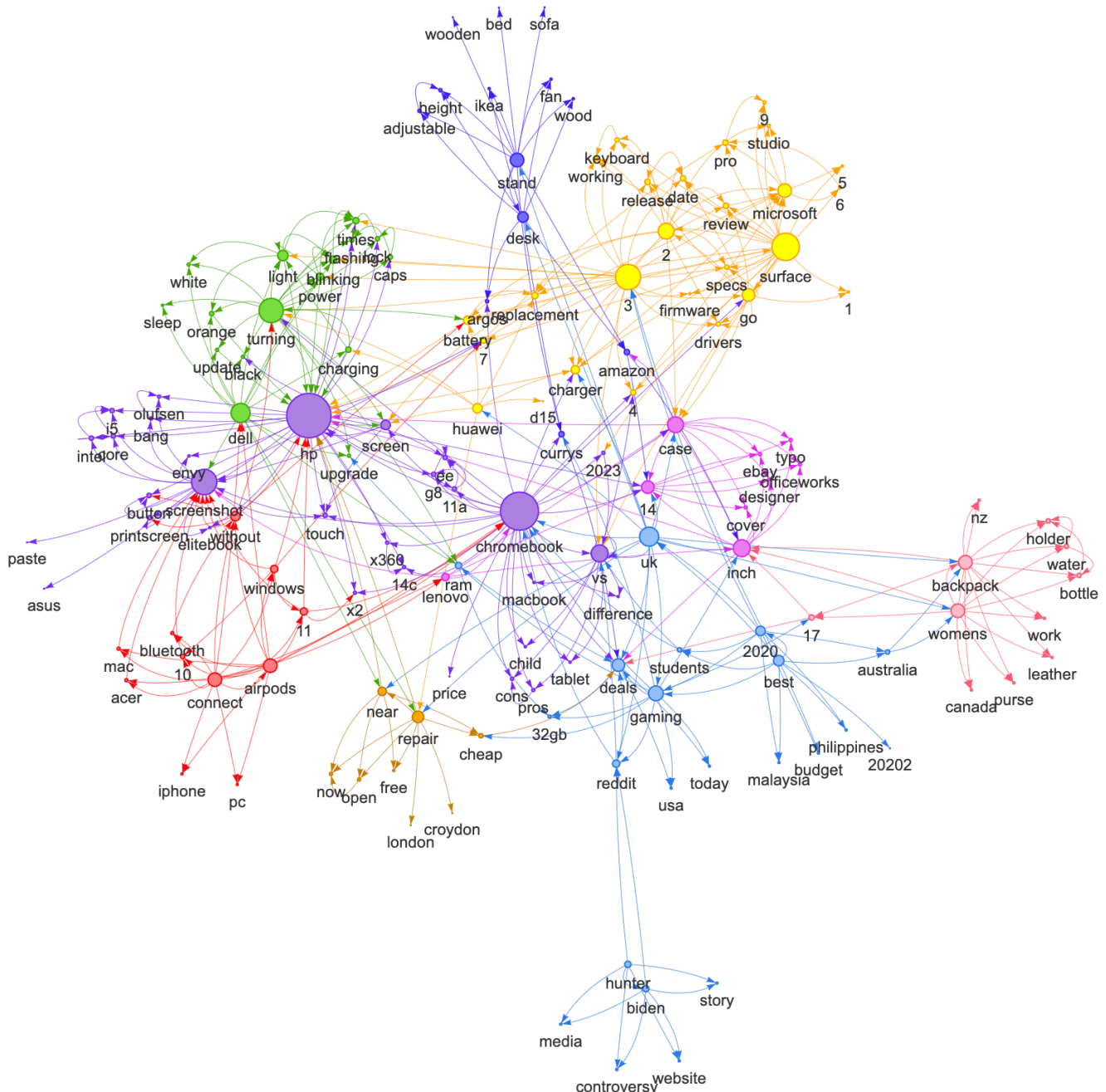


Figure 13 – Semantic Network. Rising related_queries. Original keyword: laptop

However the network in Figure 13 is somewhat complicated to interpret the nodes and their connections, so we then use the interactive capability of the 'visNetwork' library to highlight the node connections as can be seen in the example in Figure 14.

4.8 More Insights and Visualisations

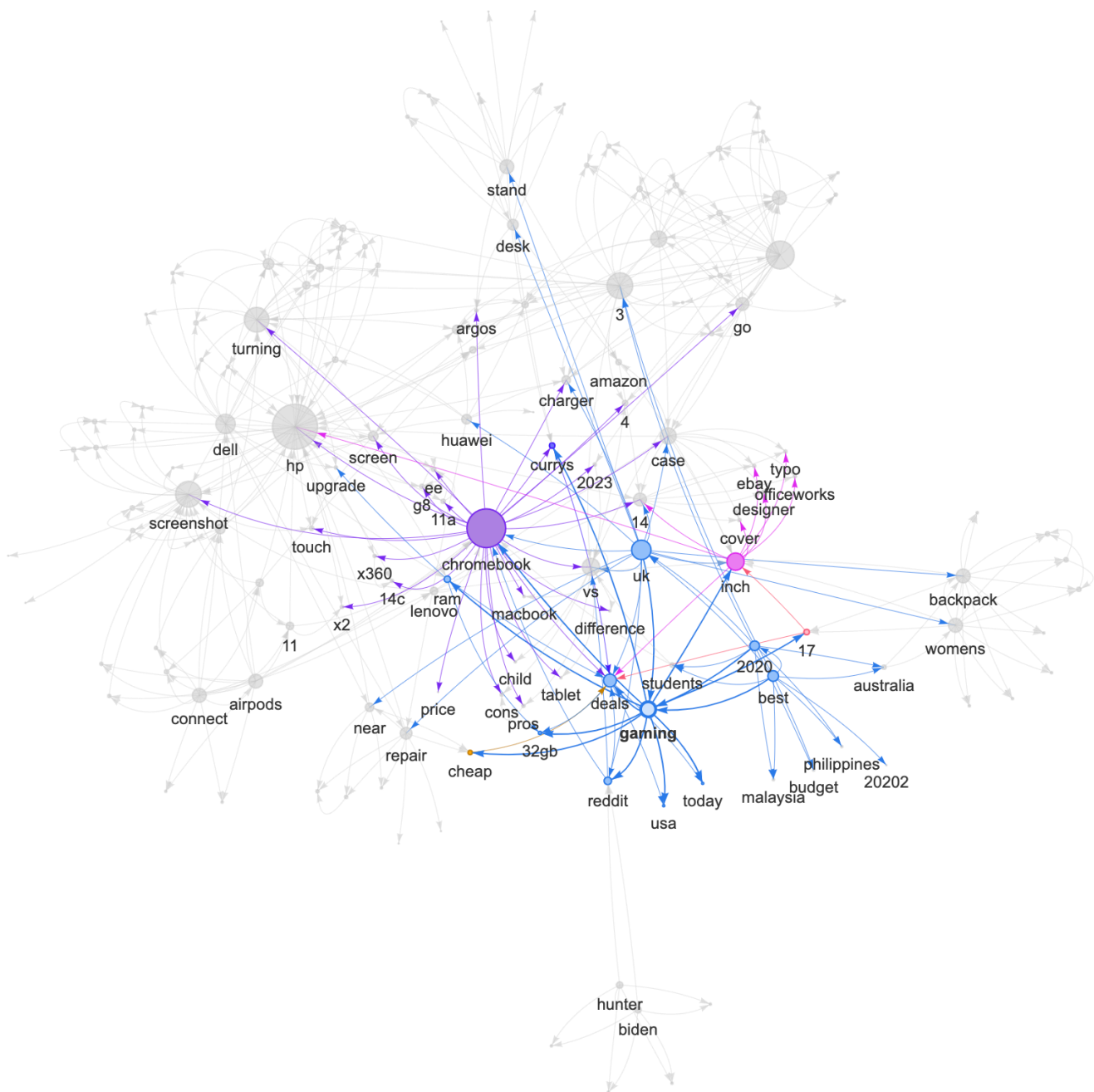


Figure 14 – Semantic Network: rising related_queries. Original keyword: laptop.

When examining the semantic network of suggested search terms from the "Rising Related Queries" for the word laptop shown in Figure 14, it is interesting to note the Chromebook in gaming as having the highest co-occurrence in terms of user interest in search engines. It is important to remember that "Rising Related Queries" represent terms that have experienced significant growth in searches over a given period of time, suggesting emerging or renewed interest from the public. This possible increase in search interest was further corroborated by looking at the Google Trends time series shown in Figure 16. Also, upon further investigation, this increase in search interest in Chromebooks for gaming coincides with the growth in recent months of cloud computing services for gaming, with an estimated growth of \$1.619 billion dollars in this market for the period 2022 to 2027 alone (PR Newswire, 2023). But what does this have to do with Chromebooks? This finding in turn coincides with the announcement that the new Chromebooks will be compatible with the GeForce Now subscription service, Nvidia's new cloud computing service for this purpose (Fear, 2020). It should be remembered that Chromebooks have distinguished themselves over time as not exactly expensive, which can make it very appealing to be able to purchase an affordable Chromebook along with a monthly subscription service that provides the computing capabilities of an expensive computer with large processing power. As an additional piece of information, Google, the company behind Chromebooks, recently announced that all of its Chromebooks will come with a 3-month free subscription to the service offered by Nvidia (Shanklin, 2023).

In contrast when we focus our attention on the semantic network analysis generated from the "Top Related Queries" represented in Figure 15, we notice that Asus and desktop computers have maintained a dominant position in terms of gaming searches over the last five years. Here it is also worth remembering that "Top Related Queries" are those terms that have been consistently popular over a period of time, indicating sustained demand and strong market interest. This suggests that while Chromebook is an emerging brand, Asus has managed to maintain a solid and consistent presence in the gaming market.

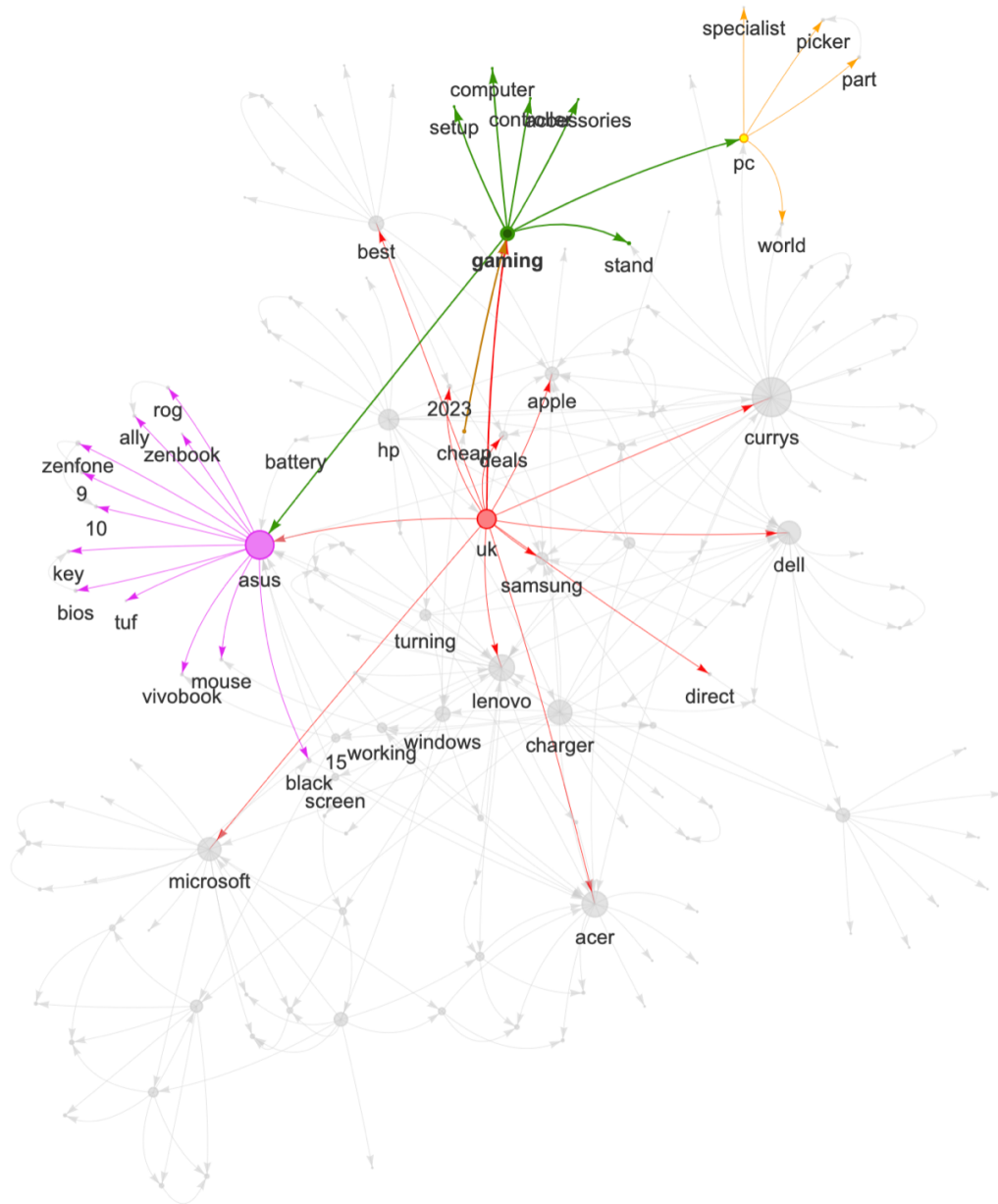


Figure 15 – Semantic Network: top related_queries. Original keyword: laptop.

Although no actual company sales database was available for this paper to compare with these findings, Google Trends time series analysis was performed as shown in Figure 16 for different gaming related brands, and indeed, we can find that the Asus brand has consistently remained at the top of the search results over the last year well above other brands such as Dell or HP.

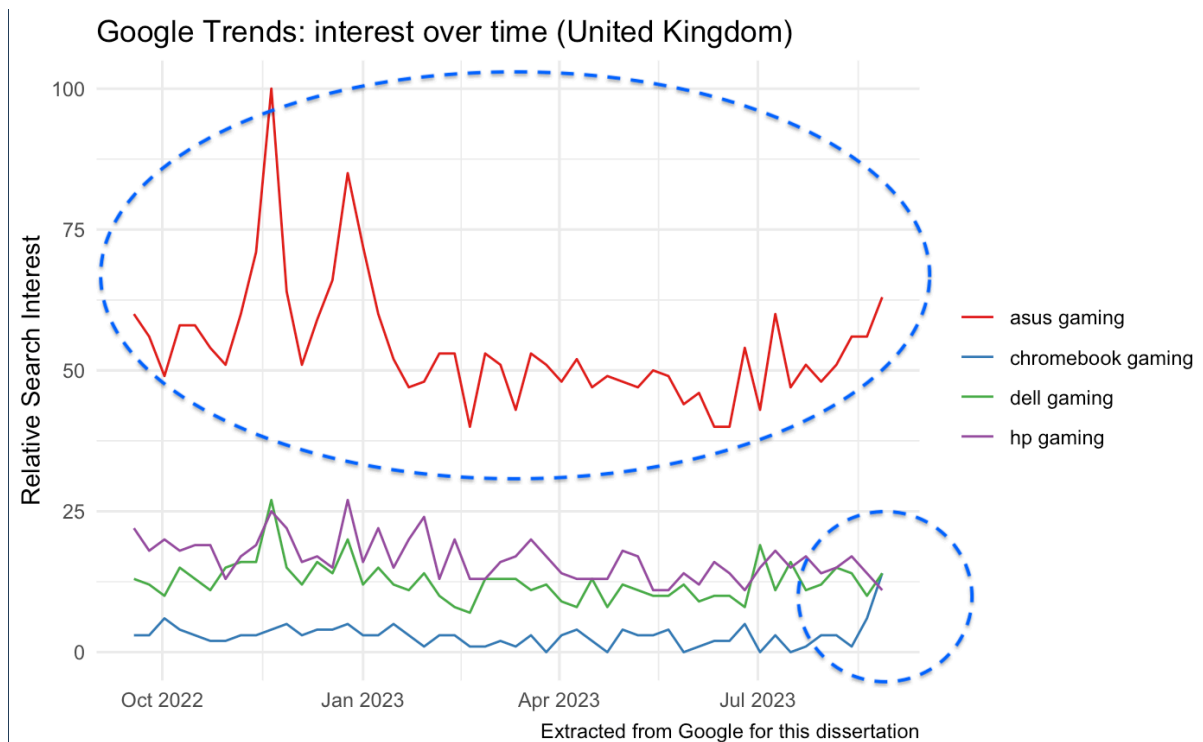


Figure 16 – Interest Over Time for Laptop Brands

While this does not necessarily represent that Asus is the sales leader in the gaming market, it could provide an avenue for future research based on comparison with actual sales data sets. In addition, an analysis in Amazon's Product Opportunity Explorer seller tool, shown in Figure 17, which is a tool that analyses click-through patterns and online shop purchases, showed that indeed, the two most clicked products in the UK after searching for "gaming laptops" belong to the Asus brand, with the FX506HC model at the top with a count of 110,478 clicks in the last 365 days. In total three Asus laptops appear in the "top 7" most clicked products within this category in the online shop, which supports our hypothesis and underlines Asus' leadership in the gaming arena by reinforcing the idea that this brand enjoys a high level of consumer interest, which also partially validates our observations based on Google Trends metrics.

Niche view ASIN view **New!**

Matching Products for "gaming laptop" in United Kingdom

Showing 1 to 25 of 1000

Filter results

| Product Name | Brand | Category | Launch Date | Click Count (Past 360 days) | Average (Past 360 days) |
|---|--------------|-----------------------------|-------------|-----------------------------|-------------------------|
| ASUS TUF F15 FX506HC 15.6 144Hz Full HD Gaming Laptop (Intel i5-11400H NVIDIA Ge... | ASUS | Laptops/Traditional Laptops | 11/2022 | 110,478 | |
| ASUS TUF Gaming FX506HCB 15.6 144Hz Gaming Laptop (Intel i5-11400H Nvidia GeForc... | ASUS | Laptops/Traditional Laptops | 1/2022 | 48,168 | |
| MSI Katana GF76 Gaming Laptop (12UB-027UK) 12th Gen i7-12700H 17.3 Inch FHD 144 ... | MSI Notebook | Laptops/Traditional Laptops | 2/2022 | 41,060 | |
| Lenovo IdeaPad Gaming 3 15.6 Inch FHD Laptop - (Ryzen 5 5600H NVIDIA GeForce RTX... | Lenovo | Laptops/Traditional Laptops | 8/2022 | 37,686 | |
| HP Victus Gaming Laptop 15-fb0020na Bundle with HyperX Pulsefire Mouse & HyperX ... | HP | Laptops/Traditional Laptops | 3/2023 | 34,370 | |
| Acer Nitro 5 AN515-57 15.6 Inch Gaming Laptop - (Intel Core i5-11400H 16GB 512GB... | Acer | Laptops/Traditional Laptops | 2/2022 | 32,187 | |
| ASUS TUF Gaming FX506LH 15.6 Full HD 144Hz Gaming Laptop (Intel i5-10300H Nvidia... | ASUS | Laptops/Traditional Laptops | 1/2022 | 31,136 | |

Figure 17 - Product Opportunity Explorer, Amazon Seller.

Since we experimented with the word clouds where we used the phrase "where can I buy..." for experimentation and which is a term that people use in the search engine when they cannot easily find a product or service, we noticed that "Ozempic" was a term that appeared very frequently in the search suggestions. By analysing this word further and in order to find out what is going on with it and why it is so popular, in Figure 19 we can first observe that indeed this term shows connections with the nodes "near" and "counter" which may suggest that it is something that is not easily found and people are looking for an outlet to get it, furthermore, by visualising the interconnected nodes of the same cluster, we can realise that the terms are related to medical words as well as words related to weight loss. By searching the internet, we can corroborate that it is indeed a controlled drug in the UK and although it is used for people with type 2 diabetes, there is a strong trend among people to purchase this drug for weight loss, which has consequently led to a shortage that has not been able to meet the demand for patients with diabetes (Blum, 2022; Fairclough, 2023). The increased interest in this medicine can also be corroborated in Google Trends as shown in Figure 20.

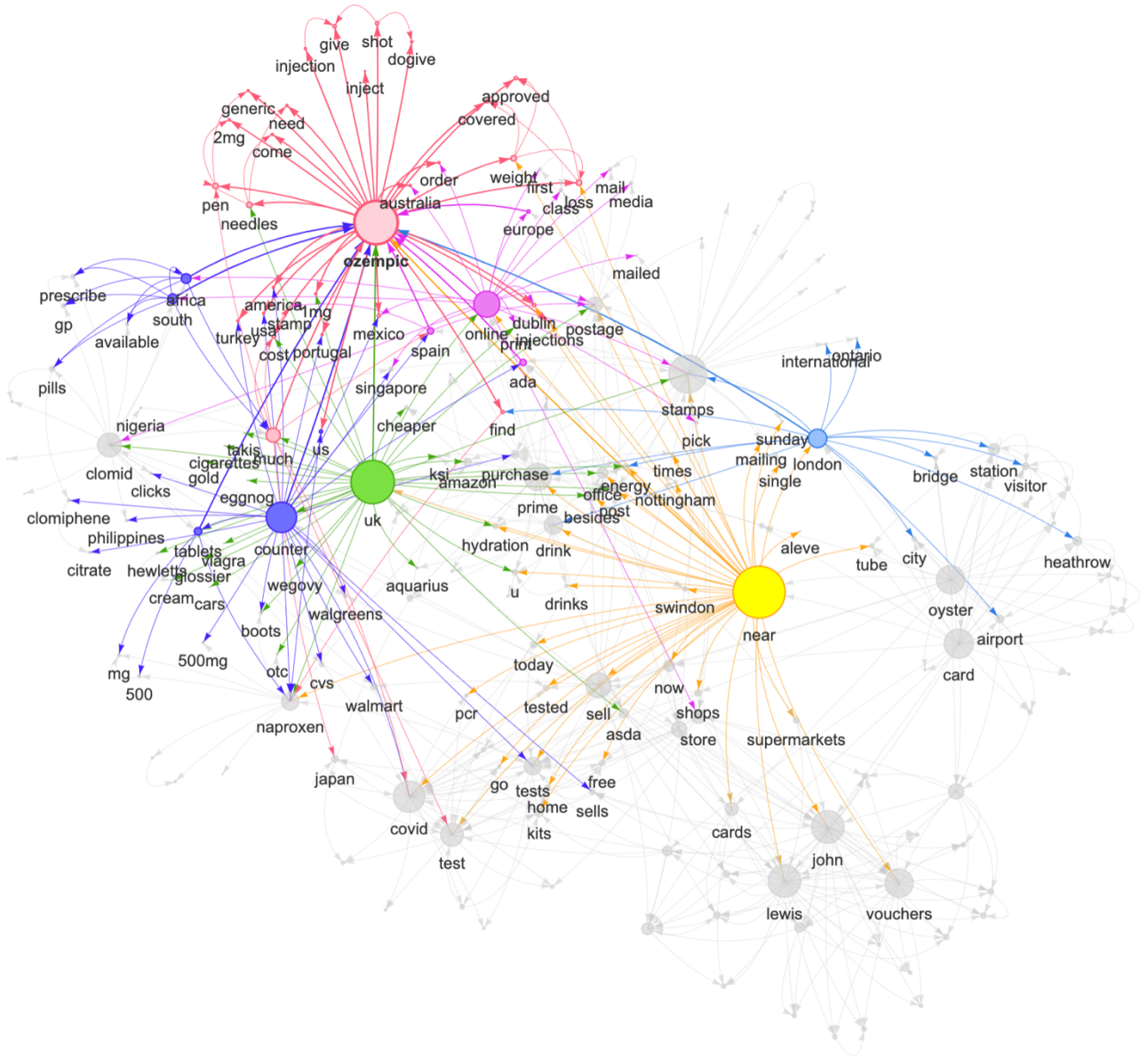


Figure 19 - Semantic Network - "Where can I buy". Highlighting the word "ozempic"

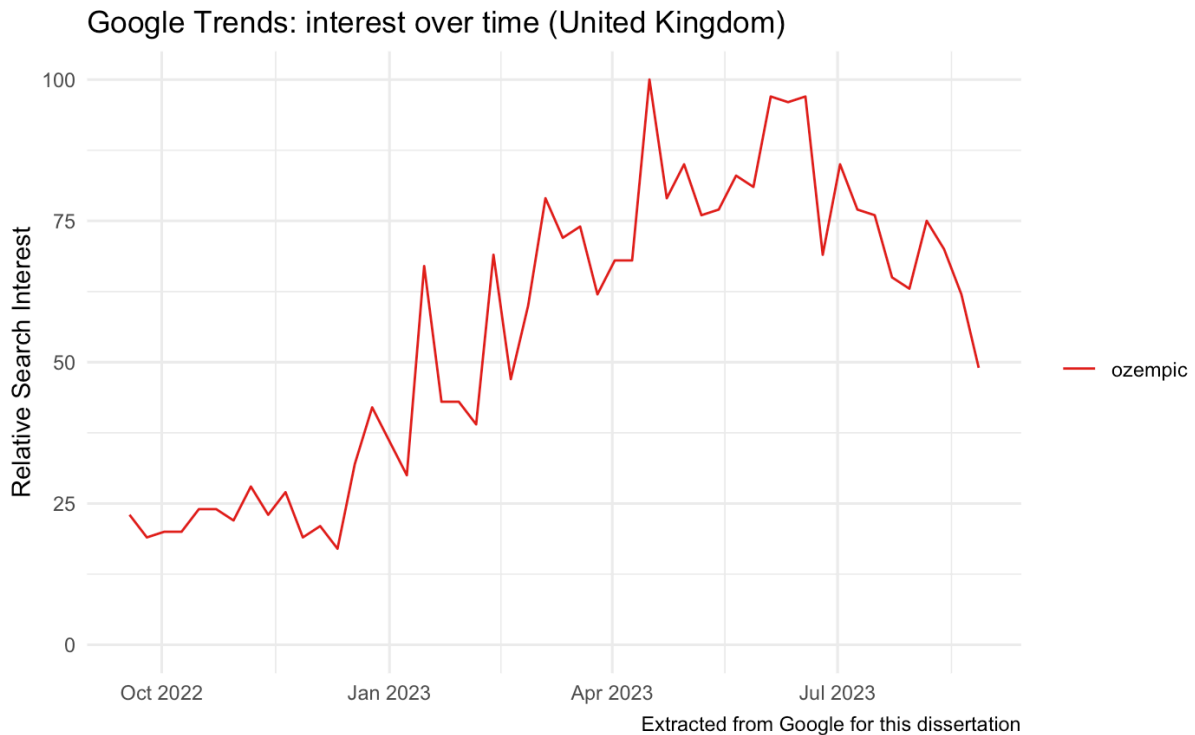


Figure 20 – Interest Over Time: "Ozempic". Data from Google Trends.

As we can see from the above experiments, semantic networks of co-occurrence generated from search suggestions become a useful tool for discovering and understanding users' specific needs and preferences. By analysing how certain search terms co-occur, it is possible to identify areas of user interest or need that may not be immediately apparent. Knowing this can be key to decision making in a variety of areas, from developing new products to adapting existing services to more efficiently meet consumer demands and expectations. In general, these tools allow us to gain insight into consumer psychology, which is invaluable for any initiative that seeks to respond accurately to market needs. Based on the various experiments that were carried out, the model shown in Figure 21 could be further developed in the future to generate a comprehensive and intuitive tool to help companies find insights for the development, marketing and positioning of products in the market.

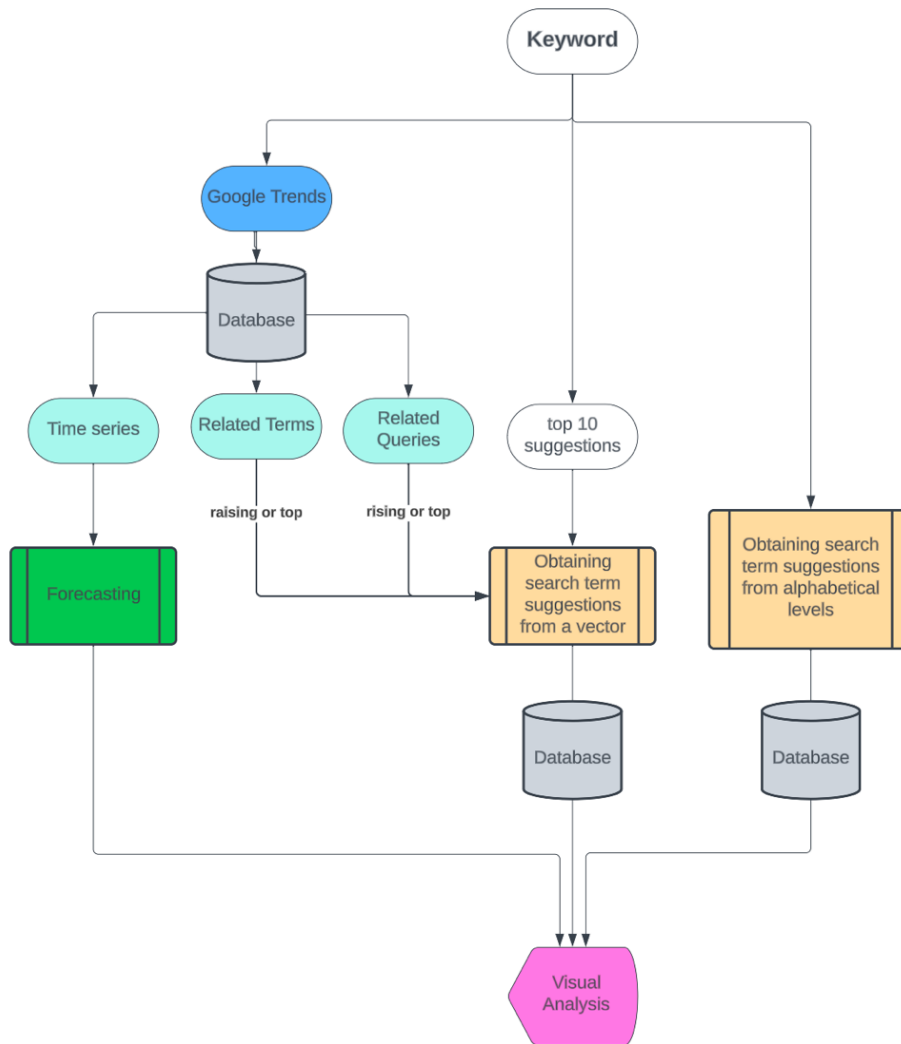


Figure 21 – A suggested model that can be improved in the future to generate an intuitive tool for finding ideas for product development, marketing and positioning in the market.

5. Conclusion

This study was started with the objective of exploring how trends as well as unstructured data can be used in conjunction with various data analysis methods to identify areas of interest or need that may not be immediately apparent in consumer behaviour. To do this we used data from Google Trends as well as search suggestions provided by Google's algorithm, we analysed this data with a variety of techniques including forecasting interest over time using

the Prophet algorithm, Text Mining techniques to analyse search suggestions, semantic networking and various tools to visualise this data.

The first question asked was whether we could identify areas of interest or need that were not immediately obvious. The answer to this, according to the results obtained in the experimentation, is yes. Prophet's algorithm, for example, has proven to be highly effective at capturing patterns in the Google Trends time series, allowing us to predict future interest in products and detect seasonal fluctuations and trends. In combination with more detailed analysis of search suggestions and co-occurrence matrices, we have been able to unravel both emerging trends and areas that may have been overlooked in more conventional analysis approaches.

When the question of whether we can gain meaningful insights through these methods arises, the evidence suggests that we can. Text mining and semantic networks provide rich insights into the "landscape" of terms and topics that are relevant to consumers or target groups. Word clouds and visualisations help to identify not only which terms are most prominent, but also how these terms relate to each other, offering valuable insights for decision makers in areas such as marketing, product development and strategic fit.

Now, can the analysis of these trends forecast future demand, a company's sales, changes in consumer preferences and the adaptability of the market to those changes? Based on the experimentation of this study and current data, the answer is "maybe". For, without a direct numerical correlation with actual sales sets, which is a limitation of this study, we cannot say for certain that these predicted trends would translate into purchasing behaviour. However, it is suggested that future research could close that gap by correlating qualitative findings with actual sales data to provide a more robust model.

This dissertation thus demonstrates that data analysis tools and analytical methods, while not without limitations, offer considerable potential for uncovering insights into consumer behaviour. The techniques presented here are flexible and could be applied in a variety of settings and for multiple purposes, from product development to market entry strategies. If

used intelligently and in combination with other forms of data and analysis, they have the potential to become a useful tool for businesses.

6. Recommendations

Finally, what next? This exploratory dissertation has laid some preliminary groundwork for the further development of a comprehensive tool to help companies identify product-related opportunities, but there is still much to be done. To begin with, the inclusion of actual sales data would allow for a much more rigorous and quantifiable analysis of how online search trends translate into purchase behaviour. In addition, the adaptability of these methods to different industries and geographic contexts could be the subject of further research, such as the inclusion of these tools in the public sector.

For Future Research

As was mentioned before, one of the limitations of this study was the lack of a set of actual sales. It is therefore recommended that future studies incorporate this data for more in-depth analyses that can correlate search trends with actual purchase behaviour. The following would be additional future research recommendations:

- **Studies in Other Industries:** This study focused specifically on consumer products, however it would be useful and interesting to apply the same techniques and eventually generate a tool that would be useful for other industries such as government, for example, where these techniques could be applied to public health or safety issues.
- **Temporal Update:** Given that consumer trends are dynamic, it would be interesting to update this analysis of unstructured data on a regular basis to capture changes in behaviour over time.
- **More Advanced Text Mining Methodologies:** While the text mining techniques used in this study were useful, there are more advanced methods that could certainly offer deeper insights.

For Business

Adoption of Analytical Tools: Business decision makers should consider adopting similar analytical tools to monitor consumer trends. Also, companies could have continuous monitoring of online search trends, as this can be an early indicator of changes in consumer preferences, and they could be conducting validations through A/B experiments to ensure that the strategies taken from these tools are actually effective.

Another aspect to consider is data ethics and privacy. Since this dissertation touched upon the consideration of ethical aspects related to data collection and use, companies should also always be aware of this and act in an ethical and responsible manner. Furthermore, these considerations should be evaluated before generating any tool that can be offered as a service to other companies.

7. Reference:

1. Balouli, H.E. and Chine, L. (2022) Google Trends Analysis Using R: Application on Algerian Tourism. First Online: 03 December. Springer. Available at: https://link.springer.com/chapter/10.1007/978-3-031-06971-0_25 (Accessed: 20 August 2023).
2. Blum, D. (2022). What Is Ozempic and Why Is It Getting So Much Attention? The New York Times. Available at <https://www.nytimes.com/2022/11/22/well/ozempic-diabetes-weight-loss.html> (Accessed: 05 September 2023).
3. Camusso, M.F. and Jorge, R.E. (2019) Google Correlate y Google Trends como herramientas para realizar un nowcast de las ventas minoristas. Centro de Estudios y Servicios de la Bolsa de Comercio de Santa Fe. Facultad de Ciencias Económicas de la Universidad Nacional del Litoral. Available at: https://www.bcsf.com.ar/ces/pdf/2019_Camusso-Jorge.pdf (Accessed: 17 July 2023).
4. Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 36(4), 1165-1188. Available at: [https://www.scirp.org/\(S\(lz5mqp453edsnp55rrgjt55\)\)/reference/referencespapers.aspx?referenceid=2097965](https://www.scirp.org/(S(lz5mqp453edsnp55rrgjt55))/reference/referencespapers.aspx?referenceid=2097965) (Accessed: 21 June 2023)
5. Clements, M., & Hendry, D. (1998). *Forecasting Economic Time Series*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511599286
6. Columbus, L. (2021) 'Roundup Of Machine Learning Forecasts And Market Estimates', Forbes. Available at: <https://www.ibm.com/cloud/blog/structured-vs-unstructured-data> (Accessed: 07 August 2023).
7. Cordón, O. (no date) Algoritmos de Poda y Visualización de Redes. Dpto. Ciencias de la Computación e Inteligencia Artificial, *Universidad de Granada*. Available at:

<https://sci2s.ugr.es/sites/default/files/files/Teaching/GraduatesCourses/RedesSistemasCompejos/Tema04-2-VisualizaciondeRedes-13-14.pdf> (Accessed: 20 August 2023).

8. Csárdi, G., & Nepusz, T. (2006) 'The igraph software package for complex network research.' *InterJournal Complex Systems*, 1695. Available at: <https://doi.org/10.5281/zenodo.3630268> (Accessed: 22 August 2023)
9. Csárdi, G., & Nepusz, T. (2006) *iGraph Reference Manual*. Available at: <https://igraph.org/c/doc/igraph-Introduction.html> (Accessed: 22 August 2023)
10. Csárdi, G., & Nepusz, T. (No date) Cluster Louvain: Finding community structure by multi-level optimization of modularity. The Comprehensive R Archive Network. Available at: https://search.r-project.org/CRAN/refmans/igraph/html/cluster_louvain.html (Accessed: 25 August 2023)
11. D. Duwe, F. Herrmann and D. Spath, "Forecasting the Diffusion of Product and Technology Innovations: Using Google Trends as an Example," 2018 Portland International Conference on Management of Engineering and Technology (PICMET), Honolulu, HI, USA, 2018, pp. 1-7, doi: 10.23919/PICMET.2018.8481971. Available at: <https://ieeexplore.ieee.org/abstract/document/8481971> (Accessed: 01 July 2023)
12. D'Avanzo, E., & Pilato, G. (2017). Using Twitter sentiment and emotions analysis of Google Trends for decisions making. *Program*, 51(3), 322-350. Available at: <https://doi.org/10.1108/PROG-02-2016-0015> (Accessed 02 July 2023)
13. Divakaran, P. K. P., Palmer, A., Søndergaard, H. A., & Matkovskyy, R. (2017). Pre-launch Prediction of Market Performance for Short Lifecycle Products Using Online Community Data. *Journal of Interactive Marketing*, 38, 12-28. ISSN 1094-9968. Available at: <https://doi.org/10.1016/j.intmar.2016.10.004>. (Accessed: 04 July 2023)

14. Driegera, P. (2013) Semantic Network Analysis as a Method for Visual Text Analytics. Elsevier. Pages 4-17. doi: 10.1016/j.sbspro.2013.05.053 (Accessed: [Date]) (Accessed: 20 August 2023).
15. Escobar, M. and Mercado, M. (2017) Big Data: An Analysis of Its Use and Application in the Context of the Digital Era. Available at: <https://heinonline.org/HOL/LandingPage?handle=hein.journals/revpropin28&div=12&id=&page=> (Accessed: 15 August 2023).
16. Facebook (no date) Prophet: Forecasting at Scale. Available at: <https://facebook.github.io/prophet/> (Accessed: 30 July 2023)
17. Fairclough, S. (2023). Ozempic shortage hits diabetes patient after weight loss use. BBC News. Available at <https://www.bbc.co.uk/news/uk-wales-66735012> (Accessed: 07 September 2023).
18. Fear, A. (2020) 'Wherever You Go with Chromebook, GeForce NOW Lets You Bring Your Games with You'. Nvidia. August 18, 2020. Available at: <https://blogs.nvidia.com/blog/2020/08/18/geforce-now-open-a-chromebook/> (Accessed: 02 September 2023).
19. Google News Initiative (no date) Google Trends: Understanding the data. Available at: <https://newsinitiative.withgoogle.com/en-gb/resources/trainings/fundamentals/google-trends-understanding-the-data/> (Accessed: 21 July 2023).
20. Google Search Help (no date) 'How Google autocomplete predictions work'. Available at: <https://support.google.com/websearch/answer/7368877#zippy=%2Cautocomplete-policies> (Accessed: 25 August 2023)

21. H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 36(4), 1165-1188. Available at: [https://www.scirp.org/\(S\(lz5mqp453edsnp55rrgjt55\)\)/reference/referencespapers.aspx?referenceid=2097965](https://www.scirp.org/(S(lz5mqp453edsnp55rrgjt55))/reference/referencespapers.aspx?referenceid=2097965) (Accessed: 21 June 2023)
22. Hanna, A., & Hanna, L-A. (2019). What, where and when? Using Google Trends and Google to investigate patient needs and inform pharmacy practice. *International Journal of Pharmacy Practice*, 27(1), 80–87. Available at: <https://doi.org/10.1111/ijpp.12445> (Accessed 02 July 2023)
23. IBM Cloud Education (2021) 'Structured vs. Unstructured Data: What's the Difference?', IBM. Available at: <https://www.ibm.com/cloud/blog/structured-vs-unstructured-data> (Accessed: 07 August 2023).
24. Joly, F. (no date) Grab Google Suggest Search Queries using R. R for SEO. Available at: <https://www.rforseo.com/apis/grab-google-suggest-search-queries-using-r> (Accessed: 08 August 2023).
25. Jun, S.-P., Yoo, H. S., & Choi, S. (2018). Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. *Technological Forecasting and Social Change*, 130, 69-87. ISSN 0040-1625. <https://doi.org/10.1016/j.techfore.2017.11.009>. (Accessed: 29 June 2023)
26. Junker, N. (2020) 'Community Detection with Louvain and Infomap'. *R bloggers*. Available at: <https://www.r-bloggers.com/2020/03/community-detection-with-louvain-and-infomap/> (Accessed: 22 August 2023)
27. Lenz, D. and Winker, P., 2020. Measuring the diffusion of innovations with paragraph vector topic models. *PLOS ONE*. Available at: <https://doi.org/10.1371/journal.pone.0226685> (Accessed 4 July 2023).

28. Lopes Dias, Á., & Lages, L. F. (2021). Measuring market-sensing capabilities for new product development success. *Journal of Small Business and Enterprise Development*, 28(7), 1012-1034. DOI: 10.1108/JSBED-07-2019-0216. Available at: https://www.researchgate.net/publication/355118292_Measuring_market-sensing_capabilities_for_new_product_development_success (Accessed 04 July 2023)
29. Marrs, M. (2021) Google Correlate: Google Trends in Reverse. *WordStream*, 17 November. Available at: <https://www.wordstream.com/blog/ws/2011/05/25/google-correlate> (Accessed: 17 July 2023).
30. Murniyati, et al. 2023. 'Expanding Louvain Algorithm for Clustering Relationship Formation', *International Journal of Advanced Computer Science and Applications (IJACSA)*, 14(1), pp. 701. Available at: https://thesai.org/Downloads/Volume14No1/Paper_77-Expanding_Louvain_Algorithm_for_Clustering_Relationship.pdf (Accessed: 02 September 2023).
31. Nuti, S. V., Wayda, B., Ranasinghe, I., Wang, S., Dreyer, R. P., Chen, S. I., & Murugiah, K. (2014). The Use of Google Trends in Health Care Research: A Systematic Review. *PLoS ONE*, 9(10), e109583. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4215636/> (Accessed: 21 June 2023)
32. Page, C. (2020) PC Sales Just Broke A 10-Year Record Thanks To The Pandemic. *Forbes*. Available at: <https://www.forbes.com/sites/carlypage/2020/10/10/pc-sales-just-broke-a-10-year-record-thanks-to-the-pandemic/?sh=439208083c2c> (Accessed: 20 August 2023).
33. Pekar, V. (2018). Mining for Signals of Future Consumer Expenditure on Twitter and Google Trends. In 2nd International Conference on Advanced Research Methods and

Analytics (CARMA2018). Universitat Politècnica de València. Available at: <http://dx.doi.org/10.4995/CARMA2018.2018.8337> (Accessed: 03 July 2023)

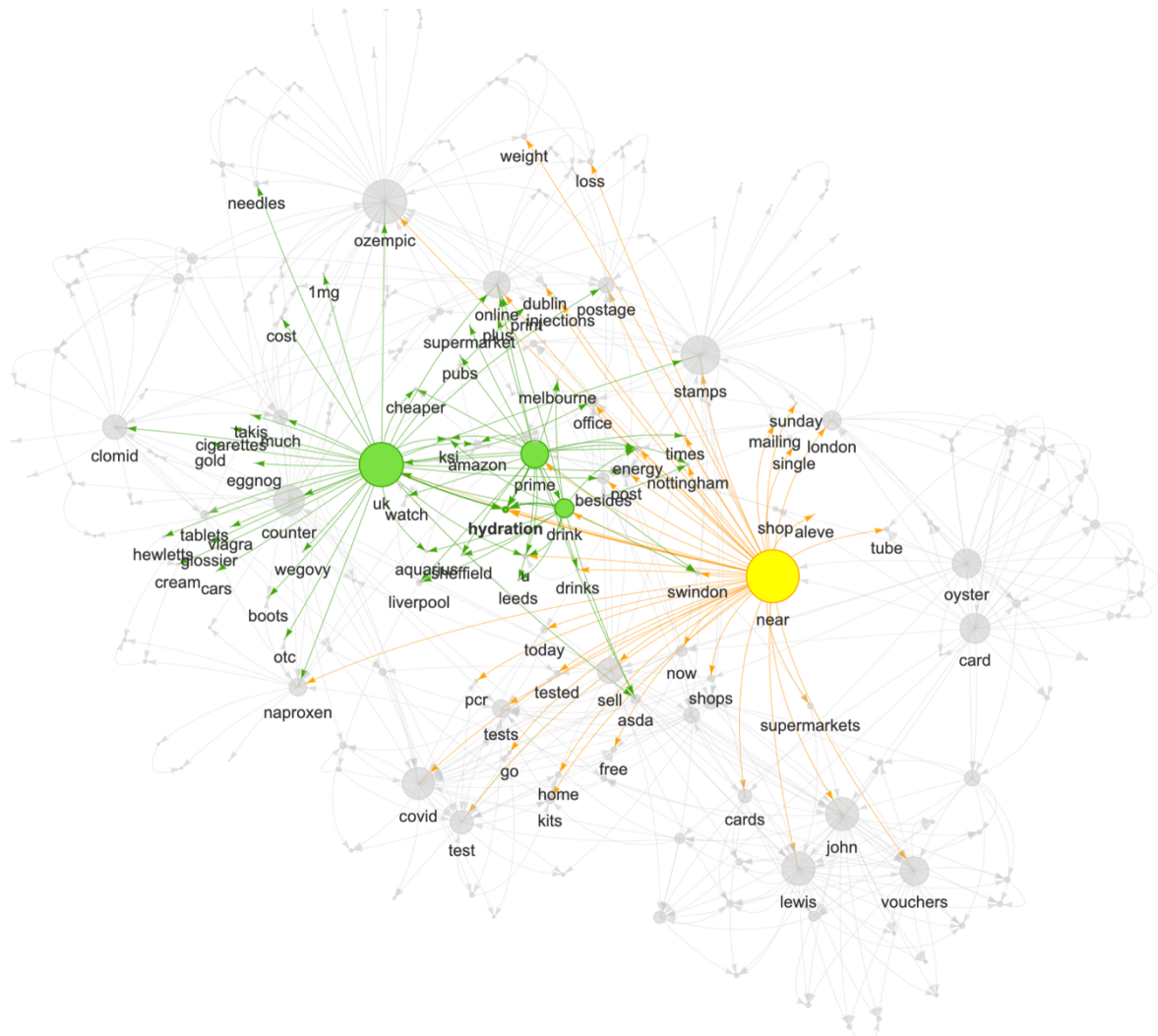
34. Pool E. and Hernández-von W. (2017) Google Trends y Ubersuggest como herramientas de análisis del comportamiento del consumidor. Estudio de caso de la microempresa Yaxkin, Cancún-México. RECAI Revista de Estudios en Contaduría, Administración e Infomática, 6(16), pp. 25-46. Available at: <https://www.redalyc.org/pdf/6379/637967154002.pdf> (Accessed: 20 July 2023).
35. PR Newswire. (2023) 'Cloud gaming market size is set to grow by USD 1,619.32 million from 2022-2027. Increased availability of high-speed Internet to boost the market.' Yahoo Finance. June 5, 2023. Available at: <https://finance.yahoo.com/news/cloud-gaming-market-size-set-061500526.html> (Accessed: 24 August)
36. Pradana, C., Kusumawardani, S. and Permanasari, A. (2020) Comparison Clustering Performance Based on Moodle Log Mining. IOP Conference Series: Materials Science and Engineering. Available at: <http://dx.doi.org/10.1088/1757-899X/722/1/012012> (Accessed: 02 September 2023).
37. Pressman, A. (2021) PC sales have surged for at-home workers and learners during the pandemic. Fortune. Available at: <https://fortune.com/2021/01/11/covid-computer-sales-lenovo-hp-dell-apple/> (Accessed: 20 July 2023).
38. Raubenheimer, J.E. (2021) Google Trends Extraction Tool for Google Trends Extended for Health data. The University of Sydney. Available at: <https://doi.org/10.1016/j.simpa.2021.100060> (Accessed: 27 July 2023).
39. Rogers, S. (2021) 15 tips for getting the most out of Google Trends. Google Blog, 11 August. Available at: <https://blog.google/products/search/15-tips-getting-most-out-google-trends> (Accessed: 15 July 2023).

40. Shanklin, W. (2023) 'Google gives Chromebook owners three months of GeForce Now'. Engadget. Updated Thu, Aug 31, 2023. Available at: <https://www.engadget.com/google-gives-chromebook-owners-three-months-of-geforce-now-cloud-gaming-182734681.html> (Accessed: 25 August 2023)
41. Shenoy, A., & Prabhu, A. (2016) Keyword Research and Strategy. In: [Book Title], Springer. First Online: 27 July 2016. Available at: https://link.springer.com/chapter/10.1007/978-1-4842-1854-9_7 (Accessed: 04 August 2023).
42. Silva, C. and Ribeiro, B. (2003) The importance of stop word removal on recall values in text categorization, Proceedings of the International Joint Conference on Neural Networks, 2003., Portland, OR, USA, 2003, pp. 1661-1666 vol.3, doi: 10.1109/IJCNN.2003.1223656. Available at: https://www.researchgate.net/publication/4030175_The_importance_of_stop_word_removal_on_recall_values_in_text_categorization (Accessed: 24 August 2023)
43. Silva, E., et al. (2019). Googling Fashion: Forecasting Fashion Consumer Behaviour Using Google Trends. Social Sciences, 8(4), 111. Available at: <http://dx.doi.org/10.3390/socsci8040111>. (Accessed: 07 July 2023)
44. Taylor, S.J. and Letham, B. (2018) 'Forecasting at Scale', The American Statistician, 72(1), pp. 37-45. Available at: <https://doi.org/10.1080/00031305.2017.1380080> (Accessed: 05 August 2023)
45. The Comprehensive R Archive Network (2022) Package "gtrendsR", 13 October. Available at: <https://cran.r-project.org/web/packages/gtrendsR/gtrendsR.pdf> (Accessed: 17 August 2023).
46. Thieurmel, B. (2022) Introduction to visNetwork. The Comprehensive R Archive Network. Available at: <https://cran.r->

project.org/web/packages/visNetwork/vignettes/Introduction-to-visNetwork.html
(Accessed: 21 August 2023).

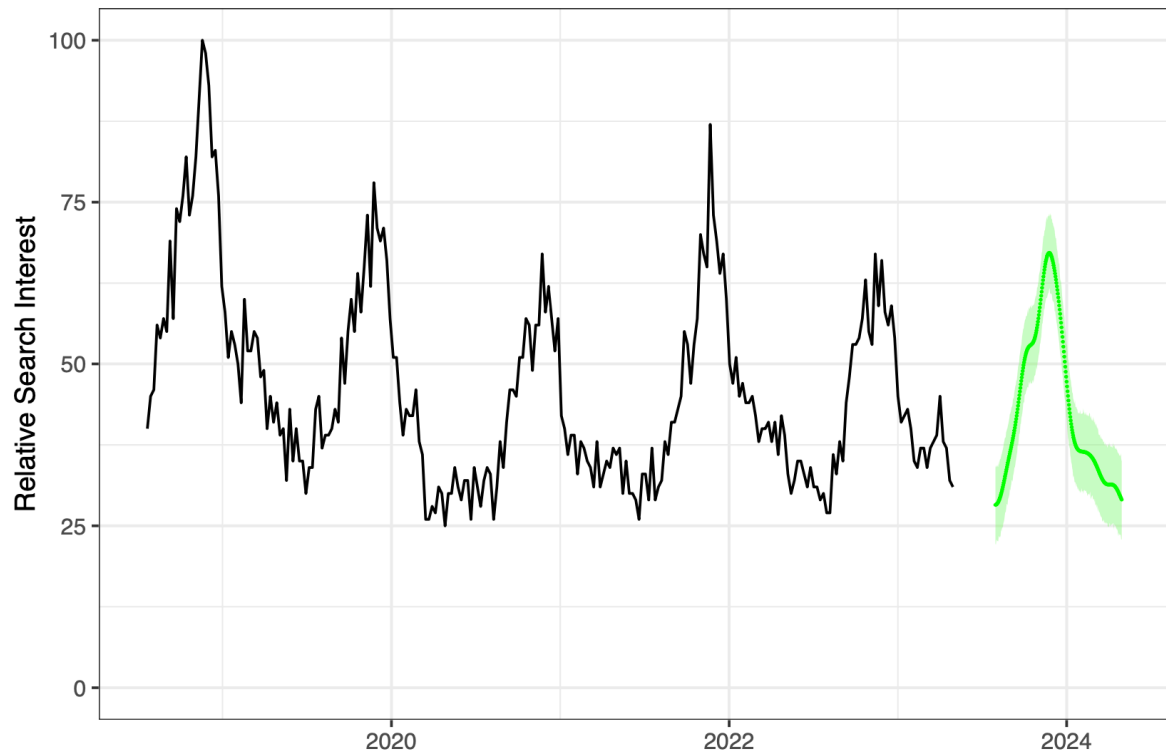
47. Trendecon. (no date) Introduction to Trendecon. Available at:
<https://trendecon.github.io/trendecon/articles/intro.html> (Accessed: 06 August 2023)
48. Eichenauer, et al. (2020) Constructing daily economic sentiment indices based on Google Trends. ETH Zurich. Available at: <https://doi.org/10.3929/ethz-b-000423817> (Accessed: 06 August 2023)
49. Tse, Q. (2021). Reconstruct Google Trends Daily Data for Extended Period. Towards Data Science. Available at: <https://towardsdatascience.com/reconstruct-google-trends-daily-data-for-extended-period-75b6ca1d3420> (Accessed 22 June 2023)
50. Unwin, A. (2002) Visualisation for Data Mining. Department of Computer-Oriented Statistics and Data Analysis, University of Augsburg, 86135 Augsburg, Germany.
Available at:
<https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=0fb89194c2707c4cb3fae0e843702be2aa508bf2> (Accessed: 05 August 2023)
51. Vanani, I. R., et al. (2022). Using text mining algorithms in identifying emerging trends for recommender systems. *Quality & Quantity*, 56(6), 3189-3211. DOI: 10.1007/s11135-021-01177-9. Available at https://www.researchgate.net/publication/352012643_Using_text_mining_algorithms_in_identifying_emerging_trends_for_recommender_systems (Accessed: 03 July 2023)
52. Ward, D., Hahn, J., & Feist, K. (2012). Autocomplete as Research Tool: A Study on Providing Search Suggestions. *Information Technology and Libraries*, 31(4), 6–19. Available at: <https://doi.org/10.6017/ital.v31i4.1930> (Accessed: 03 August 2023)

53. Wickham, H., Hester, J. and Bryan, J. (no date) readr. Available at:
<https://readr.tidyverse.org> (Accessed: 30 July 2023).
54. Wu, L., & Brynjolfsson, E. (2015). The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales. In A. Goldfarb, S. M. Greenstein, & C. E. Tucker (Eds.), *Economic Analysis of the Digital Economy* (pp. 49-75). Chicago, IL: Chicago Scholarship Online. Available at:
<https://doi.org/10.7208/chicago/9780226206981.003.0003> (Accessed: 3 July 2023).
55. Yee, C. (2019) Mining Google Trends data with R. Available at:
<https://www.christopheryee.org/blog/mining-google-trends-data-with-r-featuring-gtrends/> (Accessed: 30 July 2023)
56. Yoo, S., Song, J., & Jeong, O. (2018). Social media contents based sentiment analysis and prediction system. *Expert Systems with Applications*, 105, 102-111. ISSN 0957-4174. Available at: <https://doi.org/10.1016/j.eswa.2018.02.013>. (Accessed: 05 July 2023)



Semantic Network - Suggested Search Terms from Rising Related Queries. Keyword: “Where can I buy...”. Highlighting: Hydration.

Google Trends: interest over time for "jacket" (US)



Forecasting Interest Over Time of Google Trends – Keyword: “Bomber Jacket”. Forecast:
Next 360 days. Model: Prophet.