FoodVis

Visualising the interest in nutrition in Portugal

Adriana Gonçalves Domingos

Thesis to obtain the Master of Science Degree in

Information Systems and Computer Engineering

Supervisors: Prof. Sandra Pereira Gama

Examination Committee

Chairperson: Prof. Miguel Nuno Dias Alves Pupo Correia
Supervisor: Prof. Sandra Pereira Gama
Members of the Committee: Prof. Maria Beatriz Carmo

October 2017
Acknowledgments

First and foremost, I would like to thank and express my gratitude to my thesis supervisor Professor Sandra Gama, who steered me in the right direction whenever she felt I needed it. For always being available to help in every question I had about my research or writing, for her advice and guidance in all the phases of this thesis. She made it possible to have a higher quality work and I am very grateful about that.

I must express my acknowledgment to the team of PNPAS from DGS who were involved in the development of this thesis, especially to Doctor Pedro Graça and Dr. Sofia Sousa constantly giving me insights on what has already done and how it can be improved according to their needs. Also, thanks to Dr. Pedro Pinto and Dr. Rui da Silva, from Order of Nutritionists who helped me in the beginning of the development helping me to find the best path to follow.

I would like to express my great gratitude towards my family and my boyfriend José Cândido, for providing me with unfailing support, love, patience and continuous encouragement throughout my years of study and through the process of researching and writing this thesis.

Finally, I would like to thank my friends and colleagues in particular to Manuel Alves and Jessica Ribeiro, for their support, motivation and contributions that made my research richer. A special thanks to my friend Pedro Saraiva for all its availability, support and critical spirit especially in the process of writing this thesis.

This thesis would not have been possible without them. Thank you.
Abstract

Nowadays, people are growing more concerned about what they eat and the desire to know further information about food and healthy lifestyle has increased, being the web the main used way to search. The National Program for the Promotion of Healthy Eating (PNPAS) from the Directorate-General of Health in Portugal (DGS) has two websites related to nutrition: Alimentação Saudável and Nutrimento. To aid their team in exploring, understanding and analysing data collected about the users’ views and searches on these websites, we developed an Information Visualisation system called FoodVis. All the process and decisions taken during the development of FoodVis are explained in this thesis. An iterative and incremental paradigm was followed during development, with the goal to allow them to understand the interest in nutrition in Portugal and provide to the population even better content, adapted to their needs. After achieving the final version of our system, we perform usability tests in order to understand if the system achieved is useful, functional and has a good usability taking into account users’ needs. The results of those tests are presented in this document and allowed us to confirm that FoodVis was successfully accepted among users.

Keywords

Information Visualisation, Websites’ metrics, Nutrition, Health
Resumo

Nos dias de hoje as pessoas estão cada vez mais preocupadas com o que comem e o desejo de quererem saber mais informações sobre alimentação e estilo de vida saudável tornou-se maior, sendo a internet a principal forma utilizada para procurar por informação. O Programa Nacional para a Promoção da Alimentação Saudável (PNPAS) da Direcção-Geral de Saúde (DGS) possui duas páginas relacionados com nutrição: Alimentação Saudável e Nutriamento. Para ajudar a sua equipa na exploração, compreensão e análise de dados recolhidos sobre as visitas e procuras dos utilizadores destes sites, desenvolvemos um sistema de Visualização de Informação chamado FoodVis. Todo o processo e as decisões tomadas durante o desenvolvimento do FoodVis são explicados nesta dissertação. Foi seguido um paradigma iterativo e incremental durante o desenvolvimento, com o objetivo de que fosse possível entender o interesse pela nutrição em Portugal e proporcionar à população conteúdo ainda melhor, adaptado às suas necessidades. Depois de alcançar a versão final do nosso sistema, realizamos testes de usabilidade para perceber se o sistema alcançado é útil, funcional e tem boa usabilidade, tendo em conta as necessidades dos utilizadores. Os resultados desses testes são apresentados neste documento e permitiram-nos confirmar que o FoodVis foi aceite com sucesso entre os utilizadores.

Palavras Chave

Visualização de Informação, Métricas de Websites, Nutrição, Saúde
# Contents

1 Introduction ................................. 1  
  1.1 Goals ................................. 4  
  1.2 Document Structure ....................... 5  

2 Related Work .............................. 7  
  2.1 Web Analytics Tools ......................... 9  
  2.2 Nutritional Information ....................... 12  
  2.3 Food Similarity and Correlation ................. 19  
  2.4 Nutrition and Diet tools ....................... 24  
  2.5 Discussion ............................ 29  

3 Visualising the interest in nutrition in Portugal .... 31  
  3.1 Initial Idea .......................... 33  
  3.2 Understanding the requirements ................. 34  
    3.2.1 PNPAS .......................... 34  
    3.2.2 Websites ........................ 35  
      3.2.2.A Nutrimento .................. 35  
      3.2.2.B Alimentação Saudável .......... 35  
    3.2.3 Google Analytics .................. 36  
    3.2.4 Problem Domain .................... 36  
    3.2.5 Tasks and Questions ................. 37  
  3.3 Data Gathering ........................ 38  
    3.3.1 How ............................ 38  
    3.3.2 Add-on .......................... 39  
    3.3.3 Parsing .......................... 41  
  3.4 Low Fidelity Prototyping (LFP) ............... 41  
    3.4.1 Brainstorm Session with DGS .......... 41  
    3.4.2 Preliminary Validation ............... 43  
  3.5 Functional Prototype (FP) ................. 45
3.5.1 First version .......................................................... 45
3.5.2 Second version ..................................................... 52
3.5.3 Final version: FoodVis ......................................... 57
3.6 Architecture .......................................................... 61

4 Evaluation ............................................................... 63
4.1 Usability Tests ......................................................... 65
  4.1.1 Protocol ........................................................... 65
  4.1.2 Results ............................................................ 67
    4.1.2.A Response Time ............................................. 67
    4.1.2.B Errors ....................................................... 69
    4.1.2.C System Usability Scale (SUS) ......................... 71
  4.1.3 Discussion ....................................................... 72
4.2 Case Studies ........................................................ 73
  4.2.1 User 1 ............................................................ 73
  4.2.2 User 2 ............................................................ 74
  4.2.3 Discussion ....................................................... 75
4.3 Discussion .......................................................... 75

5 Conclusions and Future Work ..................................... 77
List of Figures

2.1 Example of Google Analytics Dashboard. .......................................................... 9
2.2 Example of Watson Customer Experience Analytics dashboard, from IBM. ........ 10
2.3 Example of Flurry Analytics dashboard. ......................................................... 11
2.4 Example of Adobe Analytics Ultimate dashboard. ......................................... 12
2.5 Example of Piwik dashboard. ................................................................. 12
2.6 NUT [1] interface. From left to right: Food List view, Overview, Nutrient view and Detail view. 14
2.7 Newtrition [2] interface. Left side from top to bottom: graph display method, calories and calculated hours of exercise for the meal displayed, nutrient details and warnings panel. · 15
2.8 Fingerprint [3]. Comparison of two select food products. ................................. 16
2.9 ManyLists [4] interface. Here four coffee drinks are compared. ....................... 16
2.10 Interface used by Hariadi et al. [5]. Left side before recognising the image and right side after recognising it. ................................................................. 17
2.11 SELFNutritionData 10 page example. ............................................................ 18
2.12 Sage 12 page example. · ........................................................................... 19
2.13 The networks of foods and nutrients developed by Kim et. al [6]. ....................... 20
2.14 FoodMicrobionet [7]. · ............................................................................ 21
2.15 Dust and Magnet, the old version on the left and the updated version to multitouch displays on the right. ................................................................. 21
2.16 Method developed by Kusmierczyk et al. [8] ................................................. 22
2.17 FoodMood [9] interface. · ........................................................................ 23
2.18 Visualizing Nutrition17 page example. ......................................................... 24
2.19 FridgeNet [10], nutrition information visualisation. · .................................... 25
2.20 Bayu et al. [11], example of Apple Nutritional Information. ......................... 25
2.21 Riehmann et al. [12], attribute manifestations of characters. · ...................... 26
2.22 On the left hGraph a visualisation designed in 2010, on the right the improved version of it, the hFigures developed in 2016. · ........................................... 28
2.23 Pratt et al. [13]. Graphical presentation of food nutrient content per calory relative to recommendations. In this example, fiber (x axis) and protein (y axis) are selected as nutrients to plot.

3.1 Direção-Geral da Saúde (DGS) websites from which we collected data.

3.2 Menu of Google Analytics Spreadsheet Add-on.

3.3 Create a new report at Google Analytics Spreadsheet Add-on.

3.4 Initial data sample related to Users, taken from Google Analytics Spreadsheet Add-on.

3.5 Initial data sample related to Searches, taken from Google Analytics Spreadsheet Add-on.

3.6 Low Fidelity Prototype - Two different options to represent number of views.

3.7 Low Fidelity Prototype - Three options to represent most searched keywords.

3.8 Low Fidelity Prototype - The dashboard.

3.9 Low Fidelity Prototype - Stacked bars when districts are selected.

3.10 Colour blending example with chosen colours.

3.11 Functional Prototype - The beginning.

3.12 Functional Prototype - Selected Districts.

3.13 Functional Prototype - What happens when we click on a bar.

3.14 Functional Prototype - First version of the graphs.

3.15 Functional Prototype - Topbar.

3.16 Functional Prototype - Calendars to select begin and end date.

3.17 Functional Prototype - Sidebar.

3.18 Functional Prototype - Charts with updated green, the same of from DGS logo.

3.19 Functional Prototype - Search dashboard, first version.

3.20 Functional Prototype - Trying to include results from Inquérito Alimentar Nacional e de Atividade Física (IAN-AF).

3.21 Functional Prototype - Wordcloud with data from website "Alimentação Saudável".

3.22 Functional Prototype - Search view with a selected word in Wordcloud.

3.23 Functional Prototype - Error when user try to select more than one word.

3.24 Functional Prototype - Title and Download button from Heatmap.

3.25 Functional Prototype - Warning about changing the date since the previous date was invalid.

3.26 FoodVis architecture.

4.1 Box Plot for time duration to answer the question.

4.2 Box Plot for errors performed to answer the question.

4.3 FoodVis with the keyword “Snack” selected.

4.4 FoodVis with the keyword “Vegetarian” selected.
List of Tables

2.1 Visualisation studied approaches overall characteristics ........................................ 30

4.1 Users tests duration for each question. ................................................................. 68
4.2 Users number of errors made for each question .................................................... 70
4.3 System Usability Scale (SUS) - score from each user. ............................................ 72
### Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>IAN-AF</td>
<td>Inquérito Alimentar Nacional e de Atividade Física</td>
</tr>
<tr>
<td>CSS</td>
<td>Cascading Style Sheets</td>
</tr>
<tr>
<td>CSV</td>
<td>Comma-Separated Values</td>
</tr>
<tr>
<td>D3</td>
<td>Data-Driven Documents</td>
</tr>
<tr>
<td>DGS</td>
<td>Direção-Geral da Saúde</td>
</tr>
<tr>
<td>HTML</td>
<td>HyperText Markup Language</td>
</tr>
<tr>
<td>InfoVis</td>
<td>Information Visualisation</td>
</tr>
<tr>
<td>JSON</td>
<td>JavaScript Object Notation</td>
</tr>
<tr>
<td>LFP</td>
<td>Low-Fidelity Prototype</td>
</tr>
<tr>
<td>NFKD</td>
<td>Normalization Form Compatibility Decomposition</td>
</tr>
<tr>
<td>PNPAS</td>
<td>Programa Nacional para a Promoção da Alimentação Saudável</td>
</tr>
<tr>
<td>SUS</td>
<td>System Usability Scale</td>
</tr>
<tr>
<td>SVG</td>
<td>Scalable Vector Graphics</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
</tbody>
</table>
Introduction

Contents

1.1 Goals ......................................................... 4
1.2 Document Structure ........................................ 5
Nowadays, people are growing more concerned about what they eat and the desire to know further information about food and healthy lifestyle has increased. Food is at the core of our daily survival, the consequences of not having a balanced diet are numerous and will determine how healthy we will be in the long run. There are tips on how to have a healthy lifestyle and nutrition available everywhere, on the Internet, radio, television, newspapers, magazines and books. Also, everyone has an opinion about that, the fitness instructors, doctors, neighbours, friends, and even grocery clerks will offer all kinds of information.

It is imperative not to believe and act upon everything we read or hear about nutrition. But how could we separate credible good nutrition information from the rest? It is important to make an extra effort to find out if the information we are reading is credible and supported by government regulated health experts before acting upon it. To solve this problem we contacted Doctor Pedro Graça, the director of the Programa Nacional para a Promoção da Alimentação Saudável (PNPAS) that means National Program for the Promotion of Healthy Eating from DGS.

Food-based chronic diseases are already the leading cause of death and disease in Western societies [14]. To counter this situation, the DGS is trying to make available to the Portuguese population tools that allow them to improve their health. But how do the population nowadays feed their growing interest in health, wellness and nutrition? They search for information on the web. Taking this into account the PNPAS created two websites that frequently update with credible information which could promote better daily choices. However, they have difficulty in realizing if they are achieving their goals and reaching the population.

For our work, we chose to focus on understanding the interest in nutrition in Portugal taking into account the data collected from two websites from the DGS, more specifically from the PNPAS which is a national strategy in the field of food and nutrition. The strategy is based on guidelines proposed by the World Health Organization (WHO), the European Commission, and also guidelines derived from experiences in other countries and in Portugal [14].

Having the data how could we get an overview of the interest of Portuguese people about nutrition or even find relevant data patterns on their visits and searches on these websites? The solution for this problem may be to apply the techniques of Information Visualisation (InfoVis). InfoVis is a research area that aims to communicate big data in a clear and efficient way using graphics. It is characterised as the area that, through the application of interactive computer graphics techniques, helps in the analysis and understanding of a significant set of data [15]. It is defined by Munzner et al. [16] as a computer-based visualisation system that provides visual representations of datasets designed to help people carry out tasks more effectively. So, basically it aids users in exploring, understanding and analysing data through progressive, iterative visual exploration [17].

InfoVis may ease the processes related to the interpretation of information in our context, contributing
to people making inadequate diet decisions leading to obesity and many other health problems that can arise later in life. If DGS understands what people are looking for and if they are reaching the people or not, they could adapt their contents to give the users the information they need or change their strategy to reach more people. That way people could have better information about nutrition, leading to better food choices that if they are made, especially early in life, can help reverse some of these growing trends. With the study of other approaches, its advantages and disadvantages, it is possible to gather and explore an integrated solution that harnesses the best of these methodologies and overcomes current research gaps.

1.1 Goals

The main objective of this work is to study ways to visually identify the relevant data patterns in data collected from two websites from the DGS, focusing on users’ views and searches, allowing the DGS to understand the interest in nutrition in Portugal and helping them to improve their strategy.

In order to fulfil our goals, we created a visualisation which makes complex data easy to digest and accessible to all, more concretely the data related to statistics of those websites, collected since their creation. That way we are able to provide accurate information in an understandable way. This visualisation will let users break out specific metrics, from specific dates or regions, that can be adjusted to more easily present all the information the user wants to understand. For example, by selecting some period of time, users may see an overview by year, month, day or hour of the webpage views or searches, and they can also see from which districts they had come. This information must be contained on a single page and accessible among the several existing devices and operating systems.

It should also be possible to compare the metrics from different districts or even compare the metrics from both websites to understand which one had better results. That way users can have new means of learning about the results of their websites, understanding if they are following the right strategy leading to better results, reaching more people and get population better informed about nutrition. Those interactions happen at the same time that the user interacts with the corresponding data, by starting with a global picture it is possible to explore and discover deeper information. The dashboard must use multiple views, that dynamically change and focus the information according to the user inputs.

During the development, an iterative and incremental paradigm was followed receiving feedback from DGS in each stage, with the goal that in the final they are allowed to understand the interest in nutrition in Portugal and provide to the population even better content, adapted to their needs. After the final visualisation was achieved, it was validated by a set of users to ensure that our visualisation has a good usability and meet users’ needs.
With our project, we attempt to solve the above problems and help DGS to engage the population on their websites making them think more seriously about their diet choices by providing a smart, simple, and personalized visualisation of the visitors’ behaviour and interests.

1.2 Document Structure

In this document, we present in Section 2 the results of the research done to some of the literature related to this thesis context, information visualisation in a nutritional context, followed by its discussion. Then in Section 3, the detailed description of the implementation process will be described. The evaluation process is available in Section 4 including all the results and conclusions from Usability Tests and Case Studies performed to validate our work. The document ends up with the main conclusions regarding all the work done, the results achieved and suggestions for future research, in Section 5.
2 Related Work

Contents

2.1 Web Analytics Tools ................................................... 9
2.2 Nutritional Information ............................................... 12
2.3 Food Similarity and Correlation .................................... 19
2.4 Nutrition and Diet tools .............................................. 24
2.5 Discussion ............................................................. 29
Much work has been done regarding the visualisation of nutritional data, but not so much work has been done regarding the visualisation of metrics related to nutritional websites. Our work will focus on those metrics but since they are related to nutritional websites with nutritional information we decided to research about this area. In this section, we present the results of the research done to some of the literature related to web analytics and nutrition. A brief description of some Web Analytics Tools and some of their shortcomings will also be presented. After that, we present the literature related to nutrition where we will focus on three essential concepts, *Nutritional Information*, *Food Similarity and Correlation* and *Nutrition and Diet tools*, giving insights on what has already done and how it can be improved.

After that, we present a table (Table 2.1) that focuses on works’ main points of comparison and discussion purposes.

### 2.1 Web Analytics Tools

Nowadays, every company need to use tools to understand the performance of its websites. That is why there are a large number of web analytics tools available. Most of these web analytics tools are part of hosted web analytics services offered by several companies such as Google, IBM, and Yahoo.

Google Analytics\(^1\) is a free service offered by Google that provides digital analytics tools to analyse data from websites for a better understanding of the user experience as we can see in the example of Figure 2.1.

![Figure 2.1: Example of Google Analytics Dashboard.](https://www.google.com/analytics/)

It has many features and is the tool currently used by DGS to collect and analyse the data from their websites but the main problem identified is that since it has soo many functionalities it becomes complex.

\(^1\)https://www.google.com/analytics/
to understand how to use. That is why they provide online courses to learn how to use it\(^2\) however, for a team which is focused on health and nutrition, like DGS that do not have anyone specialised in this area neither with this formation, it becomes complex to understand.

One of the identified problems is for example in analysing the user views from Portugal. We may see the Sessions by each district during one week, but we can not find a way to see information about the islands of Azores and Madeira and we can not also explore how those Sessions varies in a specific district or compare multiple districts. With our work, we aim to allow to see the information about the Islands and Continental Portugal, using the same approach as Google Analytics regarding the use of darker colours to present districts with higher values and lighter colours for districts with lower values. We intended also to allow the user to get the detailed information about one district or compare them in an easy way.

![Figure 2.2: Example of Watson Customer Experience Analytics dashboard, from IBM.](image)

In order to understand every step of the customer journeys, IBM has a product named Watson Customer Experience Analytics\(^3\) which is a software as a service (SaaS) solution that provides essentially four functionalities (as in Figure 2.2): role-based dashboards and analysis of mindset, journey and behaviour of users. One of the main disadvantages is that it does not have a Free Version.

Flurry Analytics\(^4\) is a software that is integrated into the Yahoo Developer Network suite of products. As a web analytics tool, it provides resources for the user to gain a deep level of understanding about your users’ behaviour in their apps. Flurry Analytics has focused Dashboards, in fullscreen mode, that

---

\(^2\)https://analytics.google.com/analytics/academy/

\(^3\)https://www.ibm.com/us-en/marketplace/customer-experience-analytics

\(^4\)https://developer.yahoo.com/analytics/
let the user view all the metrics from a given area on a single page (as in Figure 2.3). The main problem is the flexibility to look at multiple metrics simultaneously and understand multiple apps as a whole.

![Example of Flurry Analytics dashboard.](image)

**Figure 2.3**: Example of Flurry Analytics dashboard.

Also Adobe has an Analytic Tool called Adobe Analytics⁵, that solution uses the data to get a real-time understanding of the business and it is specially good for large enterprises that rely on driving large numbers of end users. Adobe Analytics does not have a Free Version, it has three main subscriptions which are from the lower to the higher: Select helps to gather and analyse relevant and real-time data so we can make the best decisions possible, Prime supports to go beyond basic analytics and dive into true customer intelligence gathered from all the channels and Ultimate that pairs the analytics expertise with the power of machine learning (Fig.2.4).

The problem is that it has a great complexity in implementation and maintenance, so it is not simple to use by anyone, and for smaller companies, it is likely to be overkill, because the setup, for example, can take many weeks. Another reason is that however Adobe Analytics has three different subscriptions, it is not cheap at all, so it is not accessible for everyone. The users ended up paying a lot for the features that, unless they are experts, they will unlikely use.

Another solution is Piwik⁶, a tool used by individuals, big and small companies. That tool (Fig.2.5) can tack every users’ move right is and it is very similar to Google Analytics, although unlike Google’s tool, you have to host the analytics on your own server. Piwik is a free and open source software but to expand their functionalities we have to add plugins from the Piwik Marketplace and most of them have a cost per year. It is not intended for use by every user since it is a PHP MySQL software which the user has to download and install on his own webserver so it requires some knowledge about development.

---

⁵http://www.adobe.com/uk/data-analytics-cloud/analytics.html/
⁶https://piwik.org/
2.2 Nutritional Information

All packaged foods include a label with details of the food content and composition. According to European Parliament and the Council of the European Union [18], the mandatory nutrition information should refer to 100 g or 100 ml amounts and, if appropriate, provide additional portion-based declarations. The nutrition information per portion or per consumption unit shall include the energy value and the amounts of protein, carbohydrate, sugars, fat, saturates and salt. Nutrients such as cholesterol, vitamin A, vitamin C, calcium, and iron, are optional to include.

That information can affect individual and public lifestyle decisions and health status, but it is not
easy to deal with it because we eat so many things during the day that in the end, we have a significant amount of information to analyse. Another problem is that we also consume non-packaged food and that food may not have labels. To solve those problems and to better understand what is in our food we may apply some visualisation techniques so we can explore the nutrient content of common foods through the visualisation of a nutritional database.

Nutrition Understanding Tool (NUT) by Dawson et al. [1], is an example of a web application that uses bar charts to help users in understanding the food through exploration of both nutrients and foods, allowing comparisons between two foods (Fig. 2.6b).

The columns of the bar chart represent each nutrient and food groups are represented by colour-coded lines stacked within the column in relative order, giving information about the types of foods that tend to contain more or less of a nutrient (Fig. 2.6a). The NUT interface has 11 colours: 8 to encode food groups, 1 for the background and 2 colours to highlight foods and comparison.

This application uses the dataset of National Nutrient Database for Standard Reference (USDA NNDSR). There are other visualisations using this dataset, but they don’t support high-level exploration.

For it to be approachable to everyday users, data has been reduced by collapsing some nutrients together, as in the case of Sodium and Calcium into the quantity minerals and removing others that are less interesting. The overview consists of 16 nutrient dimensions, from the full dataset of 46.

The same happens with food: in the end it has 6014 foods out of a total of 7906. Foods are ordered by food group, using the title provided in the dataset and organised first in alphabetical order and then by similarity (e.g. fresh or frozen).

Selecting a food highlights the relative position of that food within each nutrient bar in the overview through a black bar that is positioned over two thirds of the column keeping the information below visible. The name of selected food is also shown with the amounts of each detailed nutrient per 100g. Positive points are the context-keeping when selecting the food name and the information about the types of food that tend to contain more or less of a nutrient.

On the other hand, the interface presents some weaknesses: the eight colours used to encode food groups are not easily distinguishable from each other [16], no explanation about the meaning of the colour coding used, several duplicates food names which leave users confused about what food he should select, an excessive quantity of text and non-user friendly attribute selection.

Another approach that also uses bar charts is Newtrition by Bush et al. [2], a visualisation which attempts to engage users to think more seriously about their diet choices by providing a more accessible visualisation of what they eat. The chosen representation is a bar chart (Fig. 2.7a) when comparing an item to itself and a and stacked bar chart (Fig. 2.7b) when comparing foods. The database used in this project was the USDA database.
It is a multi-tiered lookup system. For easy data access, it first searches for exact matches, then prefix matches, then any substring matches. To enable users to simply explore this application they also provide at the bottom immediate access links to some foods from the most popular fast foods. The graph displays fat, protein and carbs and it could be presented on grams or based on daily value percentage but we can also see nutrient details.

They map calories to the sorts of physical activities that are equivalent to them, although calory calculations may have up to 20 minutes of error, as the estimates are based on calculations for a 160-pound person. The application also has a warning panel for abnormally high sugar, trans fat, saturated fat, cholesterol, and sodium so that users may learn about the potential health risks of these ingredients.

Good parts of this system that differentiate it from the previous are: multiple views of graph displays that could be in grams or daily values, warning panel to alert for risk situations and amount of physical activity needed to spend food calories.

However some limitations arise: a stacked bar chart is not the best approach when trying to make nutrient values easier to relate because comparing subcomponents across different bars is considerably more difficult [16], no explanation about the meaning of the colour coding used and other components of visualisation should be improved for better understanding, because even features like amount of physical activity and warnings are purely textual.

A different solution is proposed by Mah et al. [3], fingerprint is a visualisation designed to compare multiple aspects of two products, specifically it is applied to the nutritional data in food items (Fig. 2.8a).

The problem is that data between packages is presented in inconsistent ways, so they chose multivitamins as their data sources and converted the data to share the same units and serving sizes between items.

The fingerprints are represented as a collection of B-spline curves so each product can then have a
distinct fingerprint associated with it. The final design splits the fingerprint into two vertical halves, each representing a product, with zero beginning at the bottom of the centre line. To measure the length of these curves they used relative amounts, the larger of two values is a curve that extends all the way to the vertical line at the top, while smaller values have a curve with length proportional to the larger value.

To get more information about the products we can move the mouse over the fingerprint and it presents an information summary (Fig. 2.8b). This solution makes it very simple to understand what is the best product, although the colours may conduct users in error because they have no meaning.

A Spatial Layouts with Animated Transitions is designed by Liu et al. [4]. In ManyLists\(^8\) we can compare at least four products, each with dozens of attributes (Fig. 2.9).

It features a stepwise animation to guide the user through complex visualisation steps. So, identical features with the same values are placed on top of the screen, features that are unique are listed at the bottom, identical features, but with different values across two or more products are aligned and in the end, the display is compacted. For each feature, ManyLists uses a light green colour to show the best value among the products (Fig. 2.9b).

Users can adjust threshold or similarity (e.g. they can specify that products that vary in calory count by less than 10% should be considered as identical) and they can also change the default settings of the goodness of the feature.

---

\(^8\)http://www.cs.umd.edu/hcil/manylists/
Comparing two foods.

Figure 2.8: Fingerprint [3]. Comparison of two select food products.

A usability study with fourteen participants provided evidence that this visualisation is easy to learn and animated transitions are helpful. On a 1-9 scale, where 1 is failing and 9 is exemplary, they gave it an average rating of 7.7.

Some issues were found in this work, such as scalability and reading problems because we have to read everything to discover the information that we need.

Regarding the concept of getting nutritional information of food, Hariadi et al. [5] developed an Android application that shows the nutrition information by just taking a picture of the food.

First, it sends the picture to a server, which job is to recognize the image using SURF algorithm [19] and identifies the food's name. Then it connects to the FatSecret® server to get the nutrition information. In the end, the application shows the calories, fat, carbohydrate and protein per serving to the user (Fig. 2.10).

9https://www.fatsecret.pt/
There are also websites trying to help in the understanding of nutritional information. SELFNutritionData\textsuperscript{10} is one of these. This page aims at providing the most accurate and comprehensive nutrition analysis available, and make it accessible and understandable to all. The database used comes from the USDA’s\textsuperscript{11} National Nutrient Database for Standard Reference.

For example, if we select black eyed peas (“feijão-frade”) we can see Food Summary (Fig. 2.11a) which contains nutrition facts table, a Nutritional Target Map, a Caloric Ratio Pyramid, the Estimated Glycemic Load and NutritionData’s Opinion. A Nutrient Balance is also available (Fig. 2.11b), Protein Quality and a detailed nutrition information that displays values for more than 130 different nutrients.

The Nutritional Target Map (Fig. 2.11a) allows us to see how foods line up with our nutritional and weight-management goals. The closer a food is to the right edge of the map, the more essential nutrients per calories it contains. Otherwise, the closer a food is to the top edge, the more likely it is to fill up with fewer calories. The Caloric Ratio Pyramid (Fig. 2.11a) shows what percentage of the calories in a food come from carbohydrates, fats, proteins, and alcohol. Carbs are coloured green, fats red and protein blue. The Estimated Glycemic Load (Fig. 2.11a) is related to the food’s effect on blood sugar, it gives an indication of how much a serving of a food is likely to increase blood-sugar levels, which can be useful for diabetic control and weight loss.

Nutrient Balance Indicator (Fig. 2.11b) lets us see nutritional strengths and weaknesses of a food. Each spoke in the wheel represents a different nutrient. The spoke for dietary fiber is coloured green, protein is blue, vitamins are purple, minerals are white, and yellow represents a group of commonly overconsumed nutrients—saturated fat, cholesterol, and sodium. The density of each nutrient is indicated by how far that spoke extends towards the edge of the graph. It also shows a Completeness Score

\textsuperscript{10}\url{http://nutritiondata.self.com/}
\textsuperscript{11}\url{http://www.usda.gov/}
between 0 and 100 that summarizes how complete the food is with respect to 23 essential nutrients.

A problem with this approach is that the used colours don’t have any meaning and as it is not explained anywhere, it may lead the user to invalid conclusions.

![Nutrition Facts](image)

Figure 2.11: SELFNutritionData page example.

Another example of a website with the mission of turning food labeling less confusing, more personalized and interactive is Sage. Sage is a data visualisation platform and it is basically a database of food products where users can search to get clearer information about particular food products and individual food ingredients. The platform shows visualisations of food product data (Fig. 2.12), breaking out specific ingredients to provider clearer indications of when individual products might be a cause for concern (or vice versa). Serving sizes can be adjusted and personal factors such as allergies can be taken into account.

The platform can also map where food and its ingredients come from geographically and also allows users to create visual collections of food products. On the other hand, it is not possible to compare multiple foods or track your cumulative food intake. Food searches can be done by keyword or brand.

---

12[https://sageproject.com/](https://sageproject.com/)
2.3 Food Similarity and Correlation

Another studied paradigm allows us to understand how identical are certain kinds of food and also their mutual relation.

An example that studies the similarity between nutrient content of foods and also the correlation between nutrients was developed by Kim et al. [6]. It is a network-based approach that can be applied to food and nutrition, which must be studied to design healthy diets. The main idea was that two individual nutrients alone may have no impact on our system, but those two nutrients can synergistically affect nutritional fitness.

It is composed by 654 raw foods, and other foods whose nutrient contents have been minimally modified, connected by weighted links. Nutritionally similar foods are recursively grouped into a hierarchical structure and we can see the network divided into two parts, the animal-derived and the plant-derived.

The networks of foods and nutrients offer a global and unbiased view of the organization of nutritional connections and enables the discovery of unexpected knowledge regarding associations between foods and nutrients.

In the food-food network (Fig. 2.13a) each node represents a food, and nodes are connected through links that reflect the similarities between the nutrient contents of foods. Colour is chosen according to food category, and the size corresponds to the nutritional fitness (NF) measure of food. NF increases with the number of irreducible food sets that include food. Curiously, foods with high NF’s tend to be more expensive to purchase than food with low NFs, but there is no correlation between a food’s NF and price.

Then in the nutrient-nutrient network (Fig. 2.13b) each node represents a nutrient, the nodes are connected through correlations between the abundances of nutrients across all foods and colours of
the nodes represent the nutrient type. Shape indicates the hierarchical level of a nutrient, from highest (a general class of nutrients) to lowest (a specific nutrient). Regarding the links, colour thickness correspond to the sign and magnitude of the correlation, respectively.

(a) The food-food network. Different overviews of the network. (b) The nutrient-nutrient network.

Figure 2.13: The networks of foods and nutrients developed by Kim et. al \[6\].

Other study based on network analysis is FoodMicrobionet by Parente et al. \[7\]. A database which results from seventeen studies investigating the structure of bacterial communities in dairy, meat sour-dough, and fermented vegetable products. It was used to analyse nodes and network properties and to build an interactive web-based visualisation. The main purpose is to provide a user-friendly tool to explore multiple datasets of studies of food bacterial communities.

This visualisation\[13\] also provides additional search tools and hyperlinks for the rapid selection of food groups and Operational Taxonomic Units (OTUs) and quick access to external resources. The network has two types of nodes, 552 sample nodes and 964 OUT nodes, and 18,115 edge connections which occur only between samples and associated OTUs.

The colour of the node (Fig. 2.14) was attributed based on a custom field containing families for OTUs and Food subgroup for samples. The size of the nodes was made proportional to the weighted degree of the node and the edge thickness was made proportional to the weight of the connection.

A different visualisation proposed by Dai et al. \[20\], Hands-On is a visualisation that is an updated version of the Dust and Magnet visualisation technique \[21\] for large, multitouch displays that enables the users to simultaneously manipulate multiple magnets\[14\]. The main goal was to allow multiple users to “wade through” the data, manipulating both data items and attribute strengths by hand.

\[13\]http://www.foodmicrobionet.org/fmbn1_0_3web/
\[14\]https://youtu.be/wLXwL38xeK0
Data items are represented as iron dust particles (small circles) and data attributes as magnets. The basic idea is that magnets attract data items with higher values of their attributes more strongly.

One of the possible applications that the authors presented for this system is the visualisation of food items, choosing a cereal for breakfast, for example, has many variables to consider and compare, and this visualisation will help to analyse and take a decision. All different cereals that we want to compare are represented by dust particles and the nutrients are represented by magnets.

As the magnet is dragged around on the screen, the items with higher value of that attribute move faster, which means they have a higher attraction with the magnet. With multiple magnets, the particles are attracted to all magnet simultaneously. That allows us to understand which one has more sugar or protein for example.

The application has three different windows (Fig. 2.15). The main one is where the dust and magnet interaction takes place. The control window is on the top right where we can manipulate features like colour or size encoding, filtering or changing the magnitude of magnet attraction. On the bottom right is the detail window where the details of selected data are shown.

Kusmierczyk et al. [8] developed a method to better understand associations between words and
nutritional values in the context of online food communities, using data from the online food platform which contains more than 240 thousand recipes. How some nutrient values and the most relevant facts (i.e., kilocalories, fat, carbohydrates, proteins, sugars, sodium, cholesterol) were missing for the majority recipes, were presented only 58 thousand recipes. The recipe titles were pre-processed and cleaned, resulting in a vocabulary containing 4679 single words.

Sugars have almost no correlation to fat and cholesterol and are negatively correlated to proteins and sodium (Fig. 2.16a). The associations between carbohydrates and fat, proteins and cholesterol are significantly weaker than for other nutrients. Some of these discoveries are surprising and demonstrate that online food content may be biased in an unexpected way.

To measure the influence of words on nutritional values they applied information gain, calculating for each word information gain to all nutrients (Fig. 2.16b). Observed correlations are very high, showing that the same words are important for all nutrient facts. Based on this, they conclude that it is possible to model text topics in conjunction with related nutrient facts to provide an interpretable and low-dimensional representation of such content.

The empirical analysis reveals a strong correlation between text topics and nutrient facts, so it is possible to predict nutrient facts from meal name.

![Correlations between recipe nutrients.](image1)

![Compares recipe title topics found by their model along with associated nutrient facts weights.](image2)

**Figure 2.16:** Method developed by Kusmierczyk et al. [8].

A treemap visualisation called FoodMood was designed by Dixon et al. [9]. The treemap made sentiment analysis based on English language tweets about food, each block is an entity (foodstuffs), the size of the block represents the number of food tweets from a certain country and the colour of the block represents the average “happy sentiment” (the happiness percentage, red being more happy and blue being less happy) (Fig. 2.17). In FoodMood stick men and money bags are also used to represent country obesity level and Gross Domestic Product per capita, respectively.

The first step is data processing, the system continuously gathers live data from Twitter by querying

---

15 http://allrecipes.com/
16 http://www.foodmood.in/
the Twitter API with terms such as “for dinner”, “for lunch”, “for breakfast”, “I ate” and “I’m eating”. The gathered tweets are analysed to determine whether they contain food types. For each relevant tweet, the FoodMood system consults the geolocation assigned to the tweet, if the tweet is not geotagged it is assumed that it is made from the location that the user has provided in their Twitter profile and in the case the user hasn’t provided location, the tweet is disregarded. At the end of this step, it normalizes the naming of locations that means countries will be referred to with a single naming convention. Then each tweet is sent to the Sentiment Analysis system, an application of natural language processing, computational linguistics, and text analytics, which determine the overall sentiment orientation of the tweet.

FoodMood reveals that 58% of the 50 most tweeted foods have globally contributed to obesity, the sentiment expressed by collecting tweets is positive with certain foods showing peaks at certain times of the year (eg. Chocolate over Easter), users have high sentiment for foods contributing to obesity, the popularity of fast food stands out as a major trend and meat enjoys a high sentiment rating.

A website that also has a visualisation related to this paradigm is Visualizing Nutrition\(^7\). It is a data visualisation developed in 2011 that allows exploring over 150 nutrients, including vitamin D, for over 5000 different food items. We can select to view nutrients by common, where we have many words and the most common are represented with a higher size or by a-z where we have a list of all available nutrients (Fig. 2.18).

Then we can select a nutrient, for example, Potassium and we can see where it could be found, the top 8 foods that contain this nutrient and above we an ordered, size encoded view of all of the food groups and foods by nutrient type. If we pass the mouse over the bars we can get more information about the selected item.

\(^7\)http://viz.lonelydatum.com/nutrition/
2.4 Nutrition and Diet tools

In nutrition, diet is defined by the sum of food consumed by a person or other organism and dietary habits are the habitual decisions an individual or culture makes when choosing what foods to eat [22]. There are several tools to help guide our food consumption, and we will present some of those who try to encode the information to help the user.

FridgeNet was created by Lee et al. [10] and it is similar to existing social networking websites where the main goal is promoting communication and social activity among senior citizens, encouraging the sharing of dietary information.

Some reduced food intake or an unbalanced diet tends to result in vitamin and mineral deficiencies in older people. Also, in people aged 65 years or older, vitamin D intake is far below what is recommended. Based on these facts FridgeNet focuses on only calcium, iron, vitamin C, and vitamin D. It basically allow users to record personal food intake information, post comments, pictures, and voice messages. When a user opens the fridge, a message appears instructing the user to scan a grocery receipt or select a food item name, and then take a picture of the food item removed from the fridge.

Users mark and record consumed food in their stored dietary history, based on that FridgeNet compares the differences in diets between users and their peers and shows a chart split into four sections that represent the four nutrition elements, each colour based on the level of nutrition that the user requires: green represents “sufficient”, yellow “might be insufficient” and red “insufficient” (Fig. 2.19). When a user selects one of the four sections, the comparative results appear in the centre of the screen. The right side of the chart displays recommended foods for the user.

This platform also encourages users to meet face to face using Buy2+gether, a service that enables users to send shopping invitations to neighbours to meet and purchase food together. A study made with a group of 15 older people during three months showed that the diet of the participants improved.
during the study period this demonstrates that online discussion about food nutrition can enhance understanding and familiarity among senior adults. The Buy2-gether service wasn’t used very often.

Figure 2.19: FridgeNet [10], nutrition information visualisation.

Another study that allows the user to understand nutritional information visually is done by Bayu et al. [11]. It is a mobile application using Augmented Reality (AR) technology, that could be useful for diabetes’ patients who need to control calories.

The process is described in four steps: first, a camera scans an available image, then the image is tracked using vision-based tracking to determine the object id contained and if it fails, the user needs to set the camera position and try again. The third step displays a 3D object contained in the image. In the end, it displays the generated information about carbohydrate represent by green, fat coloured in red and protein with yellow colour, in the form of gauge meter (Fig. 2.20).

Figure 2.20: Bayu et al. [11], example of Apple Nutritional Information.

Thinking about children, Riehmann et al. [12] created a different approach for visualising food ingredients. It is glyph-based, depicting two comic-like characters whose shape and features depend on the ingredients contained in a food product. The first character represents a kid that is intended to act as a mirror for the children (Fig. 2.21a). The second is a hamster that acts as a metaphor for an animal or a toy a child cares about (Fig. 2.21b).

The application is intended for children between the ages four and eight. It aims at supporting parents
explaining if some food product is healthy or not. To start, the user only needs to take a picture of the barcode and the application will detect the European Article Number (EAN) and based on the ingredients contained, the values are ranked in three categories: (1) recommended amount, (2) too much and (3) excessive concentration, for every visual attribute.

Normally the nutritional values are visualised using red if the amount is high, yellow if the amount is moderate and green if the amount is low. In this case, the represented concepts are general caloric intake, sugar, fat, and salt. Calories are mapped to the size of the belly, fat by the roundness of the cheeks, sugar is linked to tooth decay and salt can be seen through the figure is sweating or not.

Every child recognized and considers 32 attributes with 3 possible manifestations each, they have a rate of 93.2% of positive matches over all visual attributes of all products and participants. The most errors occurred when considering the belly of both and the second was when considering the cheeks.

![The hamster character.](image1)

![The kid character.](image2)

Figure 2.21: Riehmann et al. [12], attribute manifestations of characters.

Another visualisation technique is called hGraph [23] and was originally designed in 2010 with the purpose of providing a single source representation of a person’s overall health state, being particularly well-suited for viewing complex healthcare data. It is a visualisation technique that normalizes the scales used for depicting the various health care metrics. The green circle represents the good area for a metric, while the area either inside or outside the circle represents the bad, either too high or too low (Fig. 2.22a). hGraph can be used in a variety of scenarios, can accommodate both small and large amounts of data, makes it easy to identify metrics that exist in a normal range, versus those that may be too high or low, allowing the identification of condition patterns and provides a healthy score.

Based on hGraph, Ledesma et al. [24] created hFigures where the key improvement is the addition of multiple graphs as a mechanism to compare values of health measurements over time. The aim of the library is to provide tools to assist experts and non-experts in the decision-making process of assessing the situation of a patient and its evolution over time.
The dataset is structured as a set of measurements where each has its collection of samples. All
the measurements have the same numerical value, and they are represented using a circular layout
divided into sectors, a pie chart. At the end of the group an extra value to leave a blank space between
the circular areas was inserted. The array has the same constant number multiple times, the number
of measurements plus an additional number for each group. After the measurements are plotted, the
labels are added to increase readability, and their position is defined by a given range to avoid overlaps
and clutter (Fig. 2.22b).

To colour-coded entries, a traffic light-based approach is used, following user feedback. The green
colour means that the values are within the recommended, yellow suggest a warning or follow-up action
needed and the red indicated a critical threshold was passed.

To compare the evolution of these measurements, the hFigures library allows the graphical represen-
tation of any number samples resulting in a set of graphs or polygons overlapping or stacking with each
other. To differentiate them, and the users can see the difference between two points in time, it is used
a lighter set of colours. To test this library, they designed an application for visualising the health situ-
tion of a modelled patient and how this has changed over time within a health coaching program. This
application has three components: activity timeline, the hFigures visualisation library and longitudinal
measurements.

The activity timeline represents the health interventions that the modelled patient has done during
the health coaching program. The hFigures library is utilized to display the set of measurements taken
during the health coaching program, giving an overview of the health situation of the patient. Finally, the
longitudinal measurements also display the same set of measurements using the longitudinal temporal
representation.

In the end, the heuristics evaluation’s average response was 6.3 out of 7 points, and the Cognitive
Walkthrough done by usability experts indicated no design or mismatch errors. In the Computer System
Usability Questionnaire (CSUQ) the system obtained an average score of 6.13 out of 7, and in After
Scenario Questionnaire (ASQ) the overall satisfaction score was 6.64 out of 7. The results indicate that
the library was helpful in assisting users in understanding health data and its evolution over time. A
user evaluation study with 14 participants demonstrated high satisfaction with the usability of the system
and also show that successful visualisation can assist individuals to understand their holistic health and
wellness data. [25]

The last study is from Pratt et al. [13] and the goal was to determine the efficacy of using a two-
dimensional plot to present quantitative values of two selected nutrients together with a target box rep-
resenting recommendations for those nutrients.

Two studies were conducted to test this hypothesis. The objective of the first study was to determine if
showing nutrition information graphically instead of numerically improves the ability of users to process
and recall that information. The purpose of the second study was to identify if graphically presenting nutrition information allows it to be more efficiently used for decision making.

The developed graphical display method presents food content per calory of nutrient one on the y-axis and nutrient two would be plotted on the x-axis. It also has a target representing a range of dietary recommendations for these nutrients near the centre of the plot (Fig. 2.23).

Nutrients are plotted on a calory baseline, so individual foods as well as combinations of foods can be visualised in a single chart. Fiber and protein were selected as the two nutrients to be used because they can support a weight management focus.

The chart is divided into colour-coded regions to help users identify how easily foods fit into the target eating pattern. The amount of calories can be presented by the circle sizes of data points on the plot and/or can be presented numerically next to their respective data point.

Studies suggest that although colour-coding using a traffic light-style approach is most favoured by consumers, it does not lead to behaviour change when shown for more than one nutrient. Based on their findings, the authors believe that this happens because processing information for many nutrients may be too overwhelming for consumers to make a decision under a time constraint, particularly when ratings of various nutrients conflict with each other.

Although the study showed a significant improvement in protein per calory purchased, but it has a weaker effect on fiber. Graphically presenting nutrient information two-dimensionally improved the ability of individuals to process and recall the information compared with numeric information. Graphical signposting effectively changed purchasing behaviour toward the intended direction of reducing calories without reducing protein, although numerical signposting had no effect on calory purchased.
2.5 Discussion

Information visualisation techniques provide researchers with an opportunity to take insights in an interactive and better way than ever before. With Web Analytics Tools, we have some ideas on how we should organize our system. Through the visualisation of metrics related to nutritional websites, we are able to visualise the most significant trends and interesting shifts in behaviour that can lead the website owners to understand the best approach to follow based on their own visitors’ behaviour.

Taking into account the analysis Web Analytics Tools, we conclude that the best approach will be to use a Dashboard, with the graphs that we need, placed in different boxes since all of them also do it well and it fits in our context. That Dashboard should contain an expandable menu at the left side and a top bar with information since this is what all Dashboards that we studied have and will be more intuitive to the user. One problem identified on the Web Apps was that all of them are focused on analysing data from one source at a time, so we will try to solve this issue by providing a system that easily allows to compare data from two websites and understand which one has better results in a brief view.

The different cited approaches related to nutrition, according to the paradigms they present, can meet the characteristics we desire to have in a useful InfoVis, since our metrics are about websites with nutrition context. In order to compare important features from the studied approaches, we summarized these characteristics in Table 2.1 and the selected features are described below.

The Detailed view details whether it is possible to have detailed information about the information presented. Compare means that we can compare at least two kinds of foods. Multiple views, as the name says itself, describes whether the visualisation has multiple views of the food data. Aggregation is when it is possible to merge elements together to get information about a group of items for example “dairy products”. Interactive means that is possible to interact with the visualisation.

Taking into account all features that we wish to have in our context, the most complete approach is...
Table 2.1: Visualisation studied approaches overall characteristics

<table>
<thead>
<tr>
<th>Studied approaches</th>
<th>Detailed view</th>
<th>Compare</th>
<th>Multiple views</th>
<th>Aggregation</th>
<th>Interactive</th>
</tr>
</thead>
<tbody>
<tr>
<td>NUT [1]</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Newtrition [2]</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Fingerprint [3]</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>ManyLists [4]</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Hariadi et al. [5]</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>SELFNutritionData [10]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Sage [12]</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>✓</td>
</tr>
<tr>
<td>Kim et al. [6]</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
</tr>
<tr>
<td>FoodMicrobionet [7]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>x</td>
</tr>
<tr>
<td>Dust and Magnet [20]</td>
<td>x</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>✓</td>
</tr>
<tr>
<td>Kusmicrycz et al. [8]</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>FoodMood [9]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Visualizing Nutrition [17]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>FridgeNet [10]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Bayu et al. [11]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Glyph-Based Characters [12]</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>hGraph [23]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>hFigures [24]</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
<td>✓</td>
</tr>
<tr>
<td>Pratt et al. [13]</td>
<td>x</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

NUT [1]. Although it doesn’t have multiple views and although it is possible to compare, it is not easy to do it. The worst results came from the paradigm of *Nutrition and Diet tools* which don’t have almost any of the desired characteristics. The less supported feature is aggregation.

With our work, we aim to solve the problems found and create a visualisation with all of these characteristics, providing a Detailed View of our metrics, allowing to Compare metrics from both websites, providing Multiple Views of the same data to give different overviews, having Aggregation and also being of course Interactive for the user to explore freely.

To do that, we will take advantage some of NUT [1] principles, but it will be completed with relevant aspects from other studies, like the simple and clean layout used in Sage [12] even allowing multiple views, the simple overview to understand if the best results come from one website or other like we have in Fingerprint [3], and the interactivity allowed in FoodMood [9]. We will also improve those visualisations by providing new features like *General Trends* and *Pattern Discovery* methods. Our work will follow some principles of *Google Analytics* since it was one tool that users already knew, allowing to select a period of time and choose if we want to see this time period by Day, Month or Year, adding a new feature to also see by the Hour. Our software will be free, only with the important information for the users and easy to use by them, that aims to solve the problems found in the other Web Analytics Tools such as being too expensive and too complex.
Visualising the interest in nutrition in Portugal

<table>
<thead>
<tr>
<th>Contents</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Initial Idea</td>
<td>33</td>
</tr>
<tr>
<td>3.2 Understanding the requirements</td>
<td>34</td>
</tr>
<tr>
<td>3.3 Data Gathering</td>
<td>38</td>
</tr>
<tr>
<td>3.4 Low Fidelity Prototyping (LFP)</td>
<td>41</td>
</tr>
<tr>
<td>3.5 Functional Prototype (FP)</td>
<td>45</td>
</tr>
<tr>
<td>3.6 Architecture</td>
<td>61</td>
</tr>
</tbody>
</table>
Visualising and understanding data isn’t an easy task, especially when we have large amounts of data. Taking that into account, for our work we chose to focus on helping Directorate-General of Health in Portugal (DGS)\(^1\) more specifically the National Program for the Promotion of Healthy Eating (PNPAS). They have two different websites, and our goal was to develop a tool who could help them understand the data collected from both because they can’t do it easily with the tools they have available.

FoodVis applies the techniques of InfoVis and provide a set of ways to explore the data from both websites individually or together, we can have information about users, searches, search keywords, the region of origin, date and time. That will help them to understand, for example, what people are looking for and then provide them what they need based on that information.

3.1 Initial Idea

As we know, food is an essential part of everyone’s life and the choices we make each day affect our health not only today but in the future. One of the big problems is understanding the nutritional information of what we eat because there are many variables to take into account and many different foods.

So, in the early stage of this thesis, the main idea was the development of a visualisation using nutritional data to aid users in exploring, understanding, and analysing food data, contributing to avoid risk situations, and look for possible alternatives to singular foods.

However, together with the DGS and the Order of Nutritionists, we realised that nutrient visualisation was something that already had been done by several researchers, and we identified as being more important, and with a higher level of priority, to understand the population’s interest patterns in order to support future decisions of the responsible health entities. But let’s explain step by step how we found what would be the focus of our thesis.

The first entity we contacted was the Order of Nutritionists, the regulatory and licensing body for the Nutritionists profession in Portugal. Nutritionists are qualified to provide information about food and healthy eating, so we thought that they are the right people to help us. We contacted them via email, and after a few days we ended up phoning to see if they could help us. They were available and very interested in the idea, but they were not sure how they could help us right because they also needed to understand better what is InfoVis. So we went to Porto, where their headquarters are, to have a meeting with Dr. Pedro Pinto and Dr. Rui da Silva, two of the leaders of the Order of Nutritionists.

Based on this meeting we realized that they were not the right entity to help us, since their mission is to regulate access to the profession of a nutritionist and the exercise of the professions of nutritionist and dietitian. They could not provide us with the data we needed because all the data they had was related

\(^1\)A public body of the Ministry of Health that positions itself as a reference for all those who think and operate in the healthcare field.
to the professionals and not to nutrition. However, they gave us some contacts of people who could help us as they were linked to healthy eating projects at the national level. And this is how we found Doctor Pedro Graça, the director of the PNPAS from DGS.

We contacted Doctor Pedro Graça who was also very interested in working with us and scheduled a meeting at the headquarters of the DGS in Lisbon. In this session, the concept of InfoVis was presented and based on that we brainstormed about the different things that would be interesting to explore.

After analysing the various possibilities in more detail, it was important to converge on a more concrete theme. For that, we would have to consider the data that was available, which was not many, and also what would be more important in the long run for the DGS. It was at this point that we realised that would be important to look at the systems of online communication of the ministry of health. This would allow realising if the PNPAS, which was created five years ago, was reaching the goals of the programme and it was something very important for them because they had never looked at their own results in this viewpoint.

Taking all those things into account, together we found that the most promising approach would be to analyse the data of the two websites that they have on healthy eating and understand the interest in nutrition in Portugal. It would be helpful to realize if they are taking the right approach, whether they are reaching out to people, what topics they should explore the most, and what times of the year they should do it.

3.2 Understanding the requirements

Once we decided that the most promising approach would be to analyse the data from DGS websites, it was necessary to see what information we could get and what information would be the most interesting and important. We met again at DGS headquarters in Lisbon, this time to know more about the websites and to see what parameters were possible to extract. With the insight of Doctor Pedro Graça the director of the PNPAS from DGS and Dr. Sofia Sousa, also from the team of the PNPAS, we identified concrete questions and tasks that we expect our visualisation to tackle.

3.2.1 PNPAS

PNPAS means “National Program for the Promotion of Healthy Eating”, it began in 2012 and is the first national strategy in the field of food and nutrition. It was created because food based chronic diseases are already the leading cause of death and disease in Western societies and Portugal already counted with a million of adult obese in the year that the program began. The strategy is based on guidelines proposed by the WHO, the European Commission, derived from experiences in other countries and also the retrospective analysis of previous initiatives in Portugal [14]. This set of actions aimed
at guaranteeing and encouraging the access and consumption of certain types of food with the objective of improving the nutritional state and health of the population

### 3.2.2 Websites

The PNPAS has two sites related to healthy eating, and it was from these two that we extracted the data for our visualisation. They were released in different years, but both were on 16th of October because it is the "World Food Day". The first one is called Nutrimento (Fig. 3.1a) and the second Alimentação Saudável (Fig. 3.1b) and each one of them will be introduced in the following subsections.

![Figure 3.1: DGS websites from which we collected data.](http://nutrimento.pt/)

#### 3.2.2.A Nutrimento

The word Nutrimento is described by Emílio Peres as "a nutrient or nutritive substance or principle, element useful to the functioning of the organism, which is proper to food" [26].

The site was launched on October 16, 2014 and it is the blog of PNPAS. It aims to keep the population abreast of the current relevance of the program; actions developed and relevant articles on nutrition in general. It has a daily update and has responded to the ability to interact with readers as it allows them to comment on articles.

#### 3.2.2.B Alimentação Saudável

Alimentação Saudável which means healthy diet, was launched on October 16, 2015 and unlike the blog, the site is somewhat more static and less interactive since it does not allow comments. It functions as an information repository, more institutional where any professional can obtain information. It appeared as an attempt to house all the documents made available in a more organized way. There

---

you can find information about the various programs they have, partnerships, projects, healthy eating, health and illness, food and recipes, information for professionals and other tools.

### 3.2.3 Google Analytics

The data was collected through the Google Analytics tool which is a web analytics service provided by Google that tracks and reports website traffic. They already tried to use this tool but it’s very complicated to get the answers they need because it captures a lot of data points so it makes complex to get simple things.

We explored together Google Analytics and joined in an Excel file some examples of data that could be collected and would be more important to analyse. For example:

- **Sessions**: simply means the amount of times that site has been visited.
- **Users**: amount of people who have visited the site (both new and returning).
- **Average Session Duration**: how long visitors are staying on the site.
- **Pageviews**: how many pages have been viewed.
- **Total Unique Searches**: the total number of times that the site search was used. This excludes multiple searches on the same keyword during the same session.
- **Search Keyword**: keywords that visitors search.
- **Geographical Data**: provides geographical dimensions, such as City, Country, Continent, etc. The values for these dimensions are automatically derived from the IP address of the hit.
- **Date and hour**: when things happened.

After that, it was time to look closely at the data, see how we could relate them and which would give us more knowledge to improve these websites.

### 3.2.4 Problem Domain

We still did not have the data, but at this point, we knew what we could get from Google Analytics so we discussed what would be the most important aspects to focus on.

We concluded that the most important thing would be to see:

- **Search Keywords**: view the most searched terms.
- **Users**: number of views, which could give us for example cyclical information.
They could be seen in two different granularities:

- **Time:** Year, Month, Day, Hour.
- **Space:** Country, Region.

### 3.2.5 Tasks and Questions

After choosing the aspects that we were going to focus on, we had to decide which concrete questions and tasks we expect our visualisation to answer and validate them with the team of PNPAS. The selected tasks to be supported were:

1. **Compare the statistics from both websites to understand if the trends of views are the same and identify the most viewed periods of the year.**
   - **Description:** This task will allow users to understand how the number of views on both pages varies. The user can choose to view the pages alone or together. It will be easy to identify the most visited periods that can have year, month, day or hour granularity.
   - **Example questions:** What is the month that most people visit the webpages? Was the month that one of the sites had the most visits the same one the other had? Do the visits occur mostly at a specific time of the month? Do the visits occur mostly at any specific time of the day?

2. **Realize from what main regions in Portugal are the visitors of the webpages.**
   - **Description:** To understand which regions of Portugal are most accessible to both sites, this can be seen in any chosen period of time.
   - **Example questions:** Which region of the country has visited the websites since its existence? Which city most visited them during December 2016?

3. **Present the most searched keywords.**
   - **Description:** This task will allow us to perceive which keywords the Portuguese are most looking. By filtering for time, we will be able to perceive what was most searched in a particular period.
   - **Example questions:** What are the most searched keywords? Is there any time of year that a keyword is more searched? What is the most searched keyword in the last year?

4. **Understand if there is some time of the year that one keyword is more searched.**
(a) **Description**: It will allow you to see how the number of searches varied over time for each of the keywords. That also gives us an insight about the districts who done more searches about that keyword.

(b) **Example questions**: Is there any time of year that people search more about "salt"? Which region of the country is searching more about "vegan"?

5. **View what are the most searched keywords in each region.**

(a) **Description**: By selecting one district we can see what are the most searched keywords in these district.

(b) **Example questions**: What are the most ten searched keywords in the district of "Faro"?

### 3.3 Data Gathering

The next step was get the data, cleaning and processing it. We had to analyse which data supported our tasks, so we had to decide which information from the original dataset (or datasets) we would use. Then we parsed the original data into .csv format to use with Data-Driven Documents (D3), chose a strategy for dealing with missing or erroneous data entries, and finally decided an appropriate data abstraction.

#### 3.3.1 How

The available data presented on the current solution is available at Google Analytics so we had to understand how could we get all the data since the begining of the websites with all the parameters we needed.

We went with Dr. Sofía Sousa, one of the team members PNPAS, to meet with the company responsible for their websites, ActiveMedia. The meeting took place in its offices in Lisbon and we had the opportunity to speak with the project manager. We started by explaining our idea, what we wanted to do with the data and what data we would need. We also asked for an opinion to see if there was more data available that we had not reached, so we explored Google Analytics to see different options but we concluded that the ones we had chosen were the best option.

To extract the data it was necessary to have an authorisation to access the data of google analytics since it is not public and it requires authentication. They did not know exactly how best to extract all the information from the site with the parameters we needed, although they gave us some tips, and the access to the platform so we could explore freely.

---

4 [https://activemedia.pt/](https://activemedia.pt/)
3.3.2 Add-on

After searching for different ways to get the data we needed, we realised that the best way to get such data was through an Add-on available for Google Sheets. We had to install the Add-on, and there we could create reports (Fig. 3.2) specifying what parameters we wanted to get and after that, we ran the reports to generate the spreadsheets with the required information.

Figure 3.2: Menu of Google Analytics Spreadsheet Add-on

When we create a new report we had to choose which account we want to use and which property. In our case we want data from account "DGSaude" and properties "Alimentação Saudável" and "Nutrimiento" so we have to create different reports for each one. We also had to specify the metrics and dimensions that we want to get so we can create it.

One of the problems here is that we have more than two hundred thousand entries but each report has a maximum of ten thousands rows, and each spreadsheet has two million cells limit. To overcome this challenge we had to do different reports for each ten thousands rows. As we can see in Figure 3.3, that means the first report has the start index zero (empty) and max results are ten thousands because it is the maximum rows, the second report of the same data has the start index ten thousands and one, and max results are two thousand... This is done until we have all available rows, taking into account that each spreadsheet can only have seven reports. For example, to get all data from all users we have three spreadsheets with seven reports each one.

In Figure 3.3 we can see the configuration option for users and for searches reports. To get the data about users (Fig. 3.3(a)) we use the metric users and the dimensions country, region, date and hour. On the other hand to get the data about searches (Fig. 3.3(b)) we use the metric search uniques and the dimensions country, region, search keyword, date and hour.

After all configurations we had to run all reports. A sample of the displayed information about users on the spreadsheets is depicted in Figure 3.4 and in the Figure 3.5 we can see the displayed information.

---

5 https://developers.google.com/analytics/solutions/google-analytics-spreadsheet-add-on
6 https://www.google.com/sheets/about/
Figure 3.3: Create a new report at Google Analytics Spreadsheet Add-on about searches. To get the data from the other site, all these steps must be repeated only by changing the properties that will generate a new ID.

Figure 3.4: Initial data sample related to Users, taken from Google Analytics Spreadsheet Add-on

Figure 3.5: Initial data sample related to Searches, taken from Google Analytics Spreadsheet Add-on

After that we finally could join all reports and we get four different Excel files with informations about:

1. Users from "Nutriimento"
2. Users from "Alimentação Saudável"
3. Searches from "Nutriimento"
4. Searches from "Alimentação Saudável"
3.3.3 Parsing

After gathering all relevant data, it was necessary to parse the data into JavaScript Object Notation (JSON) or Comma-Separated Values (CSV) format to use with D3. We chose to parse to a JSON file because it keeps a clear and structured format and it doesn’t repeat data to reference it because one label could contain multiple data. To parse the data we used a script written in Node.js\(^7\) that reads a .xlsx file and saves it to a JSON file.

To obtain only information considered relevant and necessary within the context of our work, we made a Python\(^9\) script. The original file contains information about different countries and regions, and some items that don’t have information had the value of "(not set)", as we want to view the interest in nutrition in Portugal we only want the items that have “Portugal” as the country. To be consistent with all the data, when for example we are viewing something about a specific regions we decided to remove all regions that have the value not set. We also decided to split the date into the day, month and year to be easier to process.

After treating the data, the data sets became much cleaner, containing only relevant information. The datasets about users contain in each line information about the country (in this case there are only data from Portugal), region, year, month, day, hour and number of users. The datasets about searches have the country (once again only about Portugal), region, search keyword, year, month, day, hour and number of search unique.

3.4 Low Fidelity Prototyping (LFP)

Since the FoodVis’s development process followed an incremental and iterative approach the next phase uses a Low-Fidelity Prototype (LFP) to sketch our visualisation, taking into account the tasks and questions that FoodVis must answer. This type of prototypes is useful because they tend to be simple, cheap and quick to produce and test [27].

After sketching the prototype, we had to validate it with the PNPAS team before we skip to the next phase.

3.4.1 Brainstorm Session with DGS

We started by outlining different options that we considered good to present the data, so we could then discuss what would be the best approach, comparing them with each other.

\(^7\) https://nodejs.org/en/  
\(^8\) file extension for an open XML spreadsheet file format used by Microsoft Excel  
\(^9\) https://www.python.org/
Regarding the number of users who visited the page, the idea was to understand aspects such as: how trends varied over time if the number of visits increased or decreased if there was a month with many more views, the months in which they decreased and the months in which they increased. To apply that idea we presented to DGS two options, a Bar Chart and a Line Chart, as we can see in Figure 3.6. Both of them could represent the number of views for each month with time and are easy to analyse. Regardless of the idiom chosen, the number of searches could be represented in the same way as the number of users, in this case instead of view the total views we would see the total of searches per month.

Concerning search data, the idea was to present also the top of searched keywords to understand what Portuguese people are looking for. To achieve such goal, we explored three options: a Wordcloud, a Heatmap and a TreeMap as we can see in Figure 3.7. One advantage of using the Wordcloud or the TreeMap is that they allow an overview of most searched keywords through a hierarchical structure where the ones with a higher size represent in this case the most searched. The advantage of using the Heatmap is that we can get information on which months people searched the most for the keyword.

To present the data related to different regions, we chose to use a map of Portugal not forgetting to include the Azores and Madeira Islands.
3.4.2 Preliminary Validation

Our sketches were done, so we went again to DGS to present our prototypes and discuss what would be the best way to present our data.

Regarding the prototypes of Figure 3.6 we chose to use the Bar Chart. Then if we wanted to see the results of both websites at the same time we could use Stacked Bar Charts that are good for showing a total. This chart would have on the x axis the selected period and on the y axis the number of views or searches. Each bar could represent one year, one month, one day or one hour, according to the type of view that we choose.

About the way to see the most searched keywords we compared the options of Figure 3.7 and they preferred the Wordcloud because it is easy to understand the most searched terms. The size of the keyword is given according to the number of searches that means an oversized word has more searches otherwise an undersized word has fewer searches.

The Heatmap exactly as it was at the Figure 3.7 was not used, but we felt that this could be seen in another way. We decided to include a Heatmap not just for keywords, but on the main screen to complete the Bar Chart, presenting the same data, but in a different language. In this way, we can have a different overview, cyclically visualise the time and thus give the possibility to identify patterns. If each bar in the bar chart represents a year, in the Heatmap we will have the years on the x-axis. If each bar is
a month then we have the months on the x-axis and the years on the y-axis. On the other hand, if each bar represents a day, we will have the days on the x-axis and the months on the y-axis. Finally, if each bar represents one hour, in Heatmap we will have the hours on the x-axis and the days on the y-axis.

In the heatmap, we could also use the same colour scheme that we would use on the map, where the darkest colour represents more views or searches, and a lighter colour means fewer views or searches.

Related to the map we decide to present as we described before, and it would stay side by side with the other graphs.

To resume all the ideas we made another sketch that is presented at Figure 3.8. On the left side we also put one lateral menu that expands only when clicked and where the options to change website and between views or searches will be placed.

When we were discussing the prototypes, we also thought how could we compare two districts. We concluded that a good way would be to click on the map on the district we want and that action should change the colour of the path of this district, a pill with the selected district name should be added and the Bar Chart and the HeatMap should be updated to present data related to the region. When we select another district it’s path also change the colour, another pill is added, and we get the stacked bars for both districts as we can see in Figure 3.9.
3.5 Functional Prototype (FP)

Having a general notion of the graphical appearance of each design visualisation and also the data that we need, the process of creating the first functional prototype for each one took place. That should be as close as possible to the final representation of the system. In this phase, the members of the PNPAS evaluated the system once again since we are following an iterative and incremental model. This evaluation was being done through direct observation and think-aloud technique followed by a briefing about FoodVis.

The last step was to construct a second version of a functional prototype. Its testing consists of performance user tests. Since the approach is iterative, prototypes generated during one session they were reused or revisited in another with the same or a different set of stakeholders. When problems were found in user testing, they were fixed and then more tests and observations were conducted to see the effects of the fixes.

3.5.1 First version

Taking into account the tasks supported by our visualisation, the questions our visualisation must answer and the data abstraction selected, each visualisation was implemented. At a first stage individually developed and later on incorporated together on a single page. We started to focus on questions related to Users and after with Searches.

The technologies used at this stage of development were then: HyperText Markup Language (HTML) for the page's construction, Cascading Style Sheets (CSS) for its styling, JavaScript and jQuery for
overall interactivity across the page, Bootstrap\textsuperscript{10} to build the responsive page layout, gridstack.js\textsuperscript{11} a plugin to allow the widget layout with drag-and-drop and resize options which also works with touch devices, SweetAlert2\textsuperscript{12} a responsive and visually appealing way to do popup boxes, and finally, D3.js a library for manipulating documents based on data for constructing the visualisations.

When we were going to start designing the visualisations one of the aspects discussed was what colour to use. We started by thinking of the following: regardless of the graphs presented, we would always have three different views, one with data from the website "Alimentação Saudável", another with data related to the website "Nutrimento" and a third with data from both.

As we had two dimensions (two websites) and one of the interests pointed out by the DGS was to understand if people went to one site or another, we chose to use the colour blending. In colour blending each different data is represented in a distinct colour and then when we have both of them at the same time, we could use a colour that consists of blending its properties' colours. However, the amount of each colour could not be evident with some colours used so, based on the study made by Gama et al. [28] that performed a user-study to verify to which colour pairs humans can perceive the amount of each component in a particular colour, we choose to blend green and yellow.

![Figure 3.10: Colour blending example with chosen colours.](image)

While we were implementing the first version of the Map, we faced some problems. The first is that we didn’t know exactly how we could draw the map so after a small search we found that TopoJSON was the solution. It is an extension of GeoJSON that encodes topology, so the first step was to find TopoJSON from Portugal and the islands of Açores and Madeira. Here appeared the second problem: we could not find any specification where both of them were together. To solve this, we searched for individual files, and then we created three different Scalable Vector Graphics (SVG) for each one of them: Portugal Continental, Açores and Madeira. To quickly identify the islands we added a legend over it saying only "Açores" and "Madeira".

Another problem here was that regions’ names of our dataset were wrong so they don’t matched with the names of TopoJSON file when we wanted to fill them with the colour. In the JSON file, we had

\textsuperscript{10}http://getbootstrap.com/
\textsuperscript{11}http://gridstackjs.com/
\textsuperscript{12}https://limonte.github.io/sweetalert2/
names like “Aveiro District” instead of “Aveiro”, so we had to fix the names of the file before we used them to draw the graph and that way everything work.

In the Heatmap Chart and the Map Chart, we used a lighter colour for smaller values and a darker colour for higher values, so we chose to blend the main colours with grey. In Figure 3.10 we can see the result of the selected colours. When we have data only from “Alimentação Saudável” the colour used was green when data is only from “Nutriimento” the colour was yellow, and when we see data from both, we blend those colours according to the percentage of each one.

When we finished implementing the Heatmap and the Portugal Map, we realised that we couldn’t see the differences between the colours. We understood that this was happening because we were using a linear scale and as we have big differences between values we couldn’t spot the differences between colours. To solve this problem, we tried to change to a logarithmic scale and it worked, so we kept it that way.

We needed to have a coordinated view that would allow us to create highly interactive methods such as filter, integrate data between different visualisations, obtain information with more or less granularity, change between Users and Searches view, change between websites, and so forth. As such we chose an interactive dashboard with several views, where each of these views (alone or together) gave us the answers to our questions.

In our dashboard, we decided to have a top bar with the title of our system and one expandable menu on the left side where we could change the options. After individually develop the three graphs of Users view we incorporated them together as we can see in Figure 3.11. Each one of them is in a different box that we could drag and drop, resize or minimise and the dashboard is also responsive to the screen sizes.

We added a feature on the Bar Chart to order the bars from the higher value to the lowest, which allows the user to know which months have the highest values easily. When we move the mouse pointer over the bars of Bar Chart, the squares of Heat Map or the districts of the Map, details on data points are displayed next to the view.

After that, we start trying to implement the option of compare districts by clicking in some region on the map which should update the Bar Chart to have stacked bars for each selected district and a list at the bottom of the map with the districts’ names. At the top of this list we have the option to remove all selected districts because it is easier when we select many of them but we can also remove one at a time, by clicking again on it.

Here we had to choose between two options because they would generate two different graphs. The first option was to keep the same scale in y-axis in the Bar Chart and the second was to change the scale of the y-axis according to the new maximum value. We tried this two options, and we chose to modify the scale because if we kept it and chose some district with low values, we were not able to
observe the bars.

The advantage of the way we chose, is that the bars from Bar Chart don’t get invisible, and we don’t lose the context. That allows to see how a district with low values varies through the time. We also took into account the fact that this could confuse the user because he may not realise that the scale has changed and to solve this we add a transition effect that happens each time that scale change.

Our initial idea was to use the same colour on the path of the selected region, the pill with the name and the bar of the bar chart as in Figure 3.12(a), but we did not like the final result of only changing the path and we decided to change to colour of the fill too as in Figure 3.12(b), because it is much more readable.

Another problem regarding this feature was the colours to use on selected districts because we have twenty districts and in the worst case the user may choose them all. We need to have twenty different
colours that can be distinguished on the map that is showing the data from one website (green or yellow) or data from both sites (blending between green and yellow). To achieve all the colours we needed we took into account a palette of 12 colours suggested by Ware [29] trying each one of them and excluding the ones that were not readable in our context, so we focused on: Red, Green, Blue, Pink, Cyan, Orange and Brown. As we need twenty colours, we get them by variating the saturation, or saturation and lightness of the previous selected colours. The purple is also a colour that fits well in our context, but we use it to highlight another thing that we will explain later in this document.

Also regarding the colours of the selected districts, after testing a few times and discussing whether it made sense, we decided to change something else. The colour of each selected region in a first approach was to be assigned in sequential order regardless of which district it was. This way the user could be confused with the fact that in a first interaction when he selects a district one colour is assigned to it and, in a second interaction, when he selects the same district again, it has another colour assigned. So we have decided to change this aspect so that each district has always the same colour associated with itself.

After that, we began to think about what interaction mechanisms we could add to our system and help us answer to our questions. To understand how things varied in detail in a month or a day we decided to add a new interaction.

By clicking on a bar of the bar graph as in as in Figure 3.13(a), the user will see the data only for the clicked period, the bar graph itself is updated for that period and also the Heatmap and the Map of Portugal. That is, if we are in an overview of the different months and click for example in the bar for September 2016, the charts are updated so that the days of that month are presented. In this case, then we have one bar in the bar graph for each day of that month, as well as a square in the heatmap for each of those days and the map of Portugal is also updated for that period as in Figure 3.13(b).

If we click on a bar again, for example referring to the day September 16, 2016, we will see represented in the charts the hours from a day. Each bar in the bar graph and each square of the heatmap will now serve one hour, and the map of Portugal will also be updated with the values of the new period as in Figure 3.13(c).

To remove all these filters, we added a button that says "Remove all filters" by clicking on it, the dashboard changes to present all the available data since the beginning date until the ending date. The default type of view is by month, so if we were in a different type it will be changed too.

After doing that we started to notice that our system was too slow when it had to update the data because it had a lot of calculations to do, since the original data is described for each hour of each day.

We started thinking about what we could change to improve this, and we ended up changing the initial Python script to get the data that we wanted. As we could have data presented as months, days or hours what we did was sum all the entries of the same month into one object only and save that file.
After that, we did the same but summing by days and saving it to a new file. That way our system loads much faster because when we want to see the general view by month, it uses the file with the sums by month directly and doesn’t have to calculate anything. When the view is by day, the behavior is the same using the file with the sums by days. If we are viewing the hours it doesn’t change because it’s the lowest granularity that we can have.

Regarding the map, since we only need to know the sum of the values between the initial and final selected date we also made a Python script to get a new file with these values and reduce the loading time.

After all of this was done for the website “Alimentação Saudável” we decided to add in our side menu the option to switch between the different data and update the different graphics accordingly.

At the end of this phase, in addition to the options to order the bars, click on the bars and click on the map described above, when “Alimentação Saudável” is selected, we viewed our dashboard as it is in Figure 3.13(a) but we could open the lateral menu as we can see in Figure 3.14(a) and change the look to “Nutrimento” and all the data is updated as we know. Or we also could change the data to both then we would see as in Figure 3.14(b).

Once we were following the incremental, iterative methodology, we felt it was time to schedule a new meeting with the DGS to test this cycle of development. Despite in this version we only had the Dashboard related to Users data and the Dashboard with data about searches had not yet been developed,
since it will include also these three graphs we thought that was the right time to validate what we have
done before we continue our work, introducing to Doctor Pedro Graça and Dr. Sofia Sousa this first
version of the functional prototype.

The purpose of this session was to get feedback and detect problems on this phase of the system to
allow us to redesign it on the next phase taking this feedback into account. We also wanted to discover if
there were some interesting data, related to the area of nutrition, that we could use to compare with the
data we already had. That was once again done through direct observation and think aloud technique
followed by a briefing, writing notes while we were watching users using the system.

The main suggestions/problems detected after this meeting were:

1. Have something on the main screen that tells us what data we are viewing because we only knew
   which one were if we opened the side menu;

2. Make the filters to switch between months, days or hours more intuitive;

3. Use the logos provided by DGS (sent by email after the meeting) in order to customize the system
to their style;

4. Increase the contrast between colours because in most of the graphics we could not see the
difference between the tones.

Regarding the issue of other nutrition-related data that we could cross with ours, the possible options
would be the results from the National Food Survey conducted by IAN-AF, where we can see, for exam-
ple, adherence to the Mediterranean dietary pattern or the prevalence of very high risk categories waist
circumference and waist-hip perimeter.
3.5.2 Second version

Given the feedback received we began to make changes and develop the features that were still missing.

To solve problem number 1, a bar with information was added at the top of the main page. There we have information about what data we are viewing, whether we are seeing the number of Users or the number of Searches, the time period that is being presented and still how is the data grouped (by year, month, day or hour).

![Functional Prototype - Topbar.](image)

The Figure 3.15 shows how the topbar looks like when we are seeing data from both websites, from 01-03-2017 until 22-03-2017 and the information presented is grouped by days. If we change to see only information of one website the colour of the other will change to grey, meaning that is disabled. And for example if we change from days (that means "Dia" that is highlighted in figure below the date) for data grouped by Month (in the figure it is "Mês"), the word "Dia" will change his colour to grey and is also removed the bold and the word "Mês" will change to black and bold, showing that it is the selected. This solves problem number 2.

![Functional Prototype - Calendars to select begin and end date.](image)

Now the time can be filtered by clicking on the bars or by choosing the start and end date to view through a calendar, available at the top of the page. The date can be entered via a text input or by clicking on the desired day in the calendar that expands when we click on the text box with a start date as in Figure 3.16(a) or end date as in Figure 3.16(b). When we choose the start date, we can see the
present selected date highlighted in the dark blue, the time until the end date highlighted in light blue and the end date highlighted with the colour grey. When we are selecting the end date, the behaviour is similar, highlighting the end date in this case with dark blue, because it is the one we are choosing, and the start date in grey.

This calendar has a fast navigation since if we click on the arrows on the side of the month name we go to the next or the previous month, but if we click on the month name, we saw all the months of the respective year. After that we can click on some month name and see the days of the month or click on the year, we will see all the years. It does not allow you to select days lower than the first day we have in our data or days higher than the last day for which data exists. As we can see in Figure 3.16 for example, the days before October 15 are scratched because they are disabled since we only have data since October 15, 2015. Another rule is that the end date cannot be lower than the starting date.

Below this calendar, we could see how the bar graph and the heatmap are grouped (hour, day, month or year). This could be changed two ways: by clicking on the sidebar or directly clicking on the top bar. We also added the option to group the data per year, where the user can see the total of each year, but always according to the selected period. Means that if we are seeing data referring to only one month of 2016 and we change to data grouped per year, we will only see one bar for the year 2016 and this bar will have the value of the total of that selected days.

Due to the fact that we added new options and received the feedback that was confusing how to modify the options we decided to redesign our side menu as we can see in Figure 3.17. On our new
sidebar (Figure 3.17(b)) we have three main options that could be expandable to select what we want: Dashboard where we can select which dashboard we want to see, if we want the view related to the number of Users who visited ("Visitas") or the dashboard of the Searches made ("Pesquisas"); Dataset ("Dados") where we may select which data we want to view, if we want only the website "Alimentação Saudável", only "Nutrimento" or both ("Ambos"); Group by ("Agrupar por") is where we can also select the presentation type of the Bar Chart and Heatmap, if we want data grouped by Year ("Ano"), Month ("Mês"), Day ("Dia") or Hour ("Hora"). And as we can see in the figure, the options selected are highlighted from the others with a darker grey.

Regarding the suggestion number 3, besides adding the logos on the top bar of the websites that we are analysing, as we can see on Figure 3.16, we also added a bottom bar with all requested logos: Portuguese Republic\textsuperscript{13}, National Health Service\textsuperscript{14}, DGS\textsuperscript{15}, PNPAS\textsuperscript{16} and Nutrimento\textsuperscript{17}. Each logo has a link to the respective website that we can open a new page by clicking on it.

Another thing we changed was the green colour we were previously using. Since we wanted to customize the page according to the DGS image and the logo has the dark green colour in its composition we changed our green to the green used in the DGS logo as we can see in Figure 3.18. We also considered to use the green present in the PNPAS logo, however, after testing it we realise that it was too clear, because it was more difficult to understand the differences between darker and clear tones.

The change of the green colour helped to solve the problem 4, but we also changed another thing. Previously, we were calculating the min and the max values that we could have in general, so it generates something like we have in Figure 3.18. We can see when we were viewing the data from the beginning to the end as in Figure 3.13(a) the green was darker, and we almost did not see differences because all values are high when compared to the lowest value that we can have. When we change to view only one month as in Figure 3.13(b), the values are lower, so in general, we have a lighter green than the previous one. And when we were viewing only one day as in Figure 3.13(c) the values for each hour are even lower, so we have an even lighter green.

Now, we updated to calculate always the min and max values according to the view. That means that we always have some element with the colour of darkest green, the one with the highest value and one element with the grey colour which is the one with the lowest value. That way we can easily spot the differences between the different colours, independently of the view we are.

We also changed the colour of the bars to the main colour: green or yellow instead of grey. This is because when we change to the stacked bars we represent the part of website "Alimentação Saudável" in green and the part of "Nutrimento" with yellow, so we understood that it was confusing for users to

\textsuperscript{13}http://www.portugal.gov.pt/pt.aspx
\textsuperscript{14}https://www.sns.gov.pt/
\textsuperscript{15}https://www.dgs.pt/
\textsuperscript{16}http://www.alimentacaosaudavel.dgs.pt/
\textsuperscript{17}http://nutrimento.pt/
have the bars with the gray colour when we were to see only one of the websites.

Another option was added to the map of Portugal, besides being able to remove all the selected districts ("Remover todos"), now we can also select all the districts ("Selecionar todos") at once. The options of adding/removing only one region by clicking on it remain available, the difference is that we have two quick accesses if we want to remove or select everything.

Since the graphs in the users’ Dashboard were already validated with DGS, we proceeded to implement the Search Dashboard. It will have a similar graphical representation, including just an additional chart of the Wordcloud type where we can see the most searched keywords.

We started by developing the same graphs since the difference was the fact that one represents the total number of users and another the total number of searches performed but the logic applied behind is the same and the functionalities that would be available as well.

Through the side menu, we can then change to the new Research Dashboard, the previous graphs are removed from the screen, and the new graphs with data related to the surveys appear as we can see in the Figure 3.19. In this case, we tried to colourise the path of the map with dark lines to present both possibilities to the DGS in the next meeting and decide which approach we would follow.

After implementing this we realised that we only had data about searches since the day 6 of January of 2016 for both websites. We tried to solve this question but we found out that there was no way to get the data before that date, because it was related to Google Analytics which only started saving those parameters since that day.

We had another scheduled meeting, so before we did the Wordcloud chart, which would involve word processing, we opted to develop the part that would relate to the IAN-AF data that was provided to us.

We focused on the values related to the waist circumference by regions; this is what regions that
have the prevalence of abdominal obesity. To integrate this data into our system, the solution was to create a new Panel where the three different maps could be seen side by side: number of users, number of visits and abdominal obesity, as we can see in Figure 3.20. We use the same encoding colours used on IAN-AF report with all the results to show this data.

After that, we met again with the DGS to show the adjustments made as well as the new functionalities before we proceeded with development. In this meeting, we made another formative evaluation to check if this version of the prototype continued to meet users’ needs and to verify users’ intentions while using the system. For doing this we used essentially the direct observation and think-aloud techniques. We needed to collect and then make sense of the stream of notes made while watching users in a controlled environment [30].
The main suggestions/problems identified were:

1. Remove the description words "from the website" and "from the blog" in the topbar because the logo is enough to them;

2. Regarding the IAN-AF data we concluded that we should not proceed with development of this since they do not give relevant information. Basically we can’t answer any question with it so we have chosen to remove it from the system;

3. What they wanted us to give priority was the wordcloud chart since it would allow to answer many important questions such as the most searched keywords in a specific period or region;

4. Regarding the wordcloud, they asked us to include an option to choose how many keywords they wanted to see, that way they can adjust it to their needs;

5. Add a feature that allows to download each chart so they can include easily the images of those graphs on future articles;

6. They would prefer that when changing between graphs, the same date is always kept, if possible, otherwise a notification should appear;

7. Include some elements that make it easier to see if we are seeing the dashboard of Users or Searches.

3.5.3 Final version: FoodVis

We should notice that the final version of the system includes all the improvements that have been made so far and which are described in the previous sections. In addition, new features have been added in this release, which are then described in this section.

After the meeting with DGS, we started to develop wordcloud since it was the only graph missing and that would answer us to the remaining questions.

However, it’s not only about programming because it involves some pre-processing of the data before we can represent it. This happens because the data we have is exactly what people look for, so they contain the same words in many different ways like singular, plural, with errors, more than a word, and so on, and we intended to standardize this aspect.

We made a script in Python to process the data, and the first step was to normalize the keywords strings using the method `normalise` that returns the Unicode Normalization Form of a given string (if the value isn’t a string, it will be converted to one first) in our case using the Unicode Normalization Form Compatibility Decomposition (NFKD). This removes all numbers, accents and non-unicode characters and we add an option to also convert all uppercase letters to lowercase. In this way, we have all the
words a little more alike, but it is not enough because although it solves the problem of having the same word with upper and lower case, with accents and without accents, for example, it does not solve the singular and plural problem, writing errors and multiple keywords together in one, for example.

So the second step was to find a way to solve our problem, and we thought about two possible solutions: Lemmatization and Stemming. A study made by Ganesh Jivani [31] explains the differences between them, basically both of them reduce a word variant to its ‘stem’ in stemming and ‘lemma’ in lemmatizing. The difference is that in stemming applies a set of rules without bothering about the part of speech (POS) or the context of the word occurrence while lemmatizing deals with in reducing the word forms to its root form after understanding the POS and the context of the word in the given sentence.

After a search we concluded that Lemmatization would be a better option to our case, but we faced another problem: we didn’t find an algorithm to Portuguese words that fits exactly on what we needed because most of them are for other languages. We decided to create our algorithm based on a Lemmatization List, so we created a Dictionary with each correspondence, and our script reads our keyword and finds out in the created Dictionary, what is the Lemma.

We ran our script, we went to see the results, and we detected several keywords that did not have any corresponding lemma since, for example, acronyms, more than one word together and errors of writing were not predicted by the algorithm. For these cases, we have chosen to add the correct correspondence manually. Some of the previous lemmas also had to be changed since there were lemmas that did not make sense in our context like for example the use of verbs like "aguar" instead of the name "water".

After this we could finally move on to programming the wordcloud. We decided to use the same colour scheme we had in the remaining charts. When we are viewing words only from the website Healthy Eating these are represented in green when we are viewing words from the website Nutrition these are represented in yellow and when we are viewing both we see the mixture of the two colours. In all cases the word size is proportional to the number of searches performed, this is a most searched word will have a larger size.

As DGS asked us, we inserted an option for the user to choose the number of words that he wants to see each time. To do this before we list them, we calculate the number of searches of each word, we sort by descending order and we only present the number of words requested. As we can see in Figure 3.21 under the words, we have the option to choose how many words we want to see, and in this case, we select thirty but it could be a value from ten to one hundred, in steps of ten.

When we pass the mouse over a word we can see the information about what word it is and how many occurrences it has but if we click on some word we may see how the searches of the selected word vary throw the time as we can see in Figure 3.22. The colour used to highlight the selected word had to be different from the colours used to highlight the select districts because we don’t want to confuse the
user, taking that into account we choose to use the Purple colour. It is the one missing from the palette of 12 colours suggested by Ware [29] that also fits in our context, as we had mention in Section 3.5.1.

As we can see in Figure 3.22 when a word is selected all the charts are updated, in the Bar Chart and Heatmap we may see the total number of searches of these for a month and the Portugal map is also updated with the total searches in each region. To remove the selected word, the user just has to click on it again, just as it does with the selected districts on the map.

The difference here concerning the map is that we can only select one word at a time. We decided to do this because we do not have much data on this context so when we chose multiple words, the graphs were not understandable, and that too was not very important for what we want to know. So, if the user tries to select more than one word, he will get a message as in Figure 3.23 saying that he can only select one word. If he wants to select another, he has to remove the last one by clicking on it.

As some words have low values of searches and the size is proportional to that they are very small.
To solve that problem we decided not only to put in the tooltip the word itself but also allow the user to zoom in the map of the wordcloud, that way he could analyse it in detail.

So when we are in the Dashboard related to the searches the differences related to the Dashboard of the visits are also that: if we change the period of time the words of the wordcloud are updated to the ones referring to the selected period, otherwise if we select one district, we will see the most searched words in those district. Basically, every graphs are synchronized between them.

At the last meeting, two things that DGS asked us to include were a button to download the graphic, and also to make it more explicit what we were seeing. In this way, we have chosen to redesign the top bar of the boxes where we added an option to download as we see in the Figure 3.24. Another change was that now the title is always visible so we do not lose the context, unlike previously that it was only showed when the chart was minimised.

![Figure 3.24: Functional Prototype - Title and Download button from Heatmap.](image)

So that there is no confusion about the Dashboard we are seeing, aside from letting the title always visible we have also changed our top bar. Now it only says “Visits” or “Searches” and in addition, the font size has been increased and the word “Visits” is accompanied by an eye and the word “Searches” of a magnifying glass.

Finally, regarding the issue of the dates we are seeing, as suggested by the DGS, we maintain the same selected period when we change between the graphs. However, since we have three different starting dates it may not be possible to display one particular graph over the same period as another, so in these situations, an alert appears for the user to notice the change. In Figure 3.25 we may see an example of an alert received when we change from Users to Searches because we only have data about searches since day 6 of January, 2016.
3.6 Architecture

To develop our solution, we used a layered architecture because this approach supports the incremental development of systems, it’s changeable and portable. That solution allows changes to be localised because each layer only relies on the facilities and services offered by the layer immediately beneath it [32]. FoodVis’s visualisation architecture can be divided into three major layers, User Interface (Front-end), Business Logic (Back-End) and Data. In Fig. 3.26 we present our layered architecture with three main layers.

Users only interact with the User Interface component, which is the top layer and provides user interface facilities. It is composed of three layers, the View Layer which is the actual page the user interacts with, the Styling Layer where all CSS and media content is stored and the Scripts Layer, which stores FoodVis’s idioms management and construction logic.
The user starts by accessing FoodVis website. That can be done from any computer connected to the Internet using a standard browser, instead of using an application that has been installed on their local computer. With this web-based application, users access our system via a normal environment, the web browser, so the application itself only needs to be developed for a single operating system, and it is accessible anytime, anywhere, via a computer with an Internet connection.

To be able to construct the visualisations of FoodVis page, each script file needs data to be displayed. For each user’s interaction, the active visualisations make data requests to the Business Logic’s Layer, this layer identifies which data is needed for each request and claims it to the Data Layer, where all JSON files are stored. Those files were previously generated from .xlsx files and processed by a Python script that returns the final JSON file.

After fetching the required data, information flows backwards, from the Data Layer to the User Interface Layer, in which the visualisation is then created, applied the respective styling and finally rendered to the user’s visualisation, where he will be presented with the information described in Section 3.5.
Contents

4.1 Usability Tests .................................................. 65
4.2 Case Studies .................................................... 73
4.3 Discussion ....................................................... 75
After finishing the last cycle of development, a set of users tested the FoodVis in order to gather quantitative and qualitative usability metrics to ensure that our visualisation meet users’ needs. The evaluation consisted of two components: Usability Tests and Case Studies. The usability tests with users were aimed at evaluating the system regarding interactivity and usability, with users without knowledge of the domain. In the case studies, the objective was the evaluation by potential users of DGS realising if it was useful, functional, intuitive and with good usability.

4.1 Usability Tests

When the final visualisation was achieved, a group of twenty users tested our system, FoodVis. This evaluation was done to assess the success of the final prototype and to check that a standard was upheld, it is known as summative evaluation [30]. Users are presented with a list of tasks, and their performance was evaluated through quantitative measures: the time it takes the user to do the task, the number of errors made, if any, and the level of satisfaction while doing such tasks.

The evaluation degrees were as follows: a preparation stage where all necessary materials were designed and created, the actual testing following a well-defined protocol, and after, the analysis and discussion of the gathered results was done.

4.1.1 Protocol

Before the tests were carried out, we need to prepare to ensure that everyone followed the same protocol. For this, it was necessary to develop a test script with the order and description of everything we had to do and questionnaires to be filled out by the users.

As a requirement for the evaluation process, it was established that it should be done with at least 20 users. We used a sampling technique called Convenience sampling, which means that our users were selected because of their convenient accessibility and proximity to us, not being made any restrictions on the basis of age, gender or educational background. Although as we refer, we established that all subjects should have the same conditions of evaluation, we used a controlled and typical use context environment and also the same tools to perform the tests.

All our tests pursued the following order:

1. A profiling survey was used to collect key demographic information about our users such as gender, the range of age, education and origin country.

2. A contextualization about what is FoodVis and the current evaluation goals was given to users. This consisted on following one prepared script to ensure that we did not forget to explain anything
and everyone knew the same. Basically, we gave an overall overview of our system, explaining each visualisation's purpose and features, performing a small demonstration of the system.

3. Each user had a five-minute period to explore FoodVis freely.

4. A set of five questions with a random order was made. They were informed that for each question the same starting point would be used, more concretely, it was the FoodVis' homepage, so we would refresh the page between each one of them. To answer each question, the user had to perform the corresponding task in our system. Always highlighting the fact that what was being tested was the FoodVis system and not the participants to give more confidence and comfort to explore the system, we asked them to inform us when they were ready to do the task so that we could collect the time duration and number of errors made.

5. A debriefing about the previous stage.

6. Another survey, now to understand their level of satisfaction while doing the tasks. For this we used the System Usability Scale (SUS) a simple, ten-item scale giving a global view of subjective assessments of usability [33] about the user experience with FoodVis. We followed the guidelines established by Brooke [33], each question had a degree of disagreement or agreement, with a range from Strongly Disagree (1) to Strongly Agree (5) respectively, from which the user could choose. Users were asked to answer each question with their true opinion, but we recommend them to not think too much about it, and if undecided to pick the middle score of the presented scale.

7. We offered a lollipop with a label saying “Thank you” to thank everyone for their participation.

The five questions randomly asked to the users were:

1. During the year of 2016, the month in which the website "Alimentação Saudável" had more visits was the same month in which the website "Nutrimento" had more visits? If so, please indicate which was the month.

2. Compare the visits made by the district of "Lisboa" with the visits made by the district of "Faro" in the website "Alimentação Saudável". In the month that more visits were made by the districts, how many were from the district of Lisbon and how many from the district of Faro?

3. What is the most searched term on both sites?

4. Considering the searches made on the website "Nutrimento", in what month was the keyword "Sport" most searched?

5. What are the ten most searched keywords in the district of "Faro" on the website "Nutrimento"?
Each task was considered to have a correct result when the user rightly answered to the question, not having any help in its accomplishment, unless the user gets lost for a long time.

4.1.2 Results

User evaluation was performed with twenty users. All of them are from Portugal, with the age group from 18 to 30 years old, while 10% (2 users) are female and the other 90% (18 users) are male. Regarding the educational qualifications, 65% (13 users) had a Master's degree, 20% (2 users) had a Bachelor's degree and 15% (3 users) had a High School degree.

As described in the Section 4.1.1, while the user were performing the task to answer each question, the response time and the number of errors made were noted, and at the end, the SUS questionnaire was answered with those results being presented in the next sections.

4.1.2.A Response Time

Regarding the duration time of executing each task, the results for each user and question are presented in Table 4.1.
There we can also see some statistics about those values, that show us the values referring to the minimum, maximum and average time spent executing each task, the value of the standard deviation and the confidence interval with the confidence level of 95%. For a even better understanding of the data, we also created a box-plot that is presented in Figure 4.1 that summarizes the following measures: median, upper and lower quartiles and minimum and maximum data values.

After analysing the data from Table 4.1 and box-plot from Figure 4.1 some conclusions could be made.

The max and mean duration time of questions Q1, Q3 and Q5 are higher than the others, and the question Q1 is the one with the highest values, which was expected because it was the most complex question. This happens because in this task the user had to view the visits from website “Alimentação Saudável” and also the visits from website “Nutrimento” and in each one of them understand which was the most visited month to understand if they were the same. To do this some users used the option...
to sort the bars from the highest to the lowest value, which makes the task faster, but others did not remember that option and stopped to look for the bars and check one by one what was the highest value, loosing more time. Another thing was related to the change between websites because there are two ways to do it, one way implies going to the lateral menu, choose to view the options of Dataset and change to the new dataset that we want, the other option is clicking on the top bar of the website name that we want which is fastest but only some of them choose this way.

Through the confidence interval we are 95% confident that the mean time taken to perform question Q1 on further testing will be between 25.67 and 37.43 seconds, which shows that even for users unfamiliar with the FoodVis system, they are able to understand and execute it fast.

Questions Q2, and Q4 represent the questions with the lowest minimum and mean duration times, with a response time of 9 seconds, proving these answers were quite intuitive for users. Those questions are also the ones with the smallest value dispersion, showing that users have a high level of agreement with each other.

Since question Q2 is performed on the home page, it is the easiest. The user only had to realize that the dashboard presented was already about the visits, the dataset was also the right one and then find and click on the districts. Consequently, as expected, it is also the task with the lowest values of the duration time, maximum time, standard deviation and dispersion.

### 4.1.2.B Errors

While users are interacting with FoodVis, errors during the process were registered. The results are consolidated on Table 4.2.
As in Section 4.1.2.A we present a few statistics about the values, like the values referring to the minimum, maximum and the mean of errors performed executing each task, the value of the standard deviation and the confidence interval with the confidence level of 95%. For a better perception of the results, a box-plot was created too. It is presented in Figure 4.2 and condense the results of: median, upper and lower quartiles and minimum and maximum data values.

The most of the detected errors was due to the fact of users clicking on the wrong menu first to change between dashboard or dataset. Through the feedback received we understand that users felt some difficulties on remembering what is the difference between Dashboard and Dataset, so they did not know in which option they could change between User views and Searches and where they could change between "Alimentação Saudável", "Nutrimento" or both.

After inspect the results from Table 4.2 and box-plot presented in Figure 4.2, we could form some conclusions.
As we could see, 25% (5 users) did not make any error during its evaluations, 55% (11 users) did one or two errors, and the remaining 20% made more than two errors. The maximum number of errors registered for a particular user was four and it was for user U01 and U11. Those users are also the only ones that perform two errors in one single question, the user U01 in question Q1 and the user U11 in question Q5.

It was expected that the question Q1 who is the most complex would have more errors, and this was what happened. On the other hand, the question Q5 also had the same number of errors, which was not expected because it was not so complex. This happens in question Q5 because two main reasons. The first reason was because we asked for the users to show the ten most searched keywords and we wanted that they use the option to see only ten keywords, but some of them kept the default option with twenty keywords answered with that default option. The second reason was what we mentioned before, as they have to change from Users to Searches some users pick the wrong menu (Dataset instead of Dashboard) at the first time.

After all the number of errors is not bad since there is not any question with more than 50% of errors and the reason for most of them is only the confusion between two options on the first time using the system, but the user quickly understands what is the right option and then chooses the right one.

### 4.1.2.C System Usability Scale (SUS)

The final step in the evaluation of FoodVis, was regarding the SUS questionnaire. We grouped the user’s questionnaires and for each one of them, the SUS score was calculated following the scoring guidelines provided on the works of Brooke [33].
The data from the results of SUS is presented in Table 4.3. As we can see, the minimal and maximum rating given were 80 and 100 respectively, on a range between 0 and 100.

The mean rating of our system was 90.13 points, based on Bangor et al. [34] through this average SUS score we could make a comparison with other metrics before established from other systems to understand if our system is considered "Worst Imaginable", "Awful", "Poor", "OK", "Good", "Excellent" or "Best Imaginable". By correlating our system, with the adjacent metrics, we conclude that the achieved score falls into the range of what is considered "Excellent".

### 4.1.3 Discussion

By analysing the test results, we can consider that the system’s objectives were achieved. The previously described user testing session shows that the system allows a quick perception of the data about the population's interest in nutrition in Portugal and that the interaction is intuitive since in the tests performed the number of errors is low.
It is also possible to conclude that the system has a good usability, reaching 91.13 points in the SUS being considered as having an excellent usability. We can not forget the fact that the user only interacted with the system in this session, which means that when he would use the system again, he would perform even better results related to time duration executing each task and number of errors.

4.2 Case Studies

In addition to testing with voluntary users, the system has also been evaluated by two DGS members who are part of PNPAS and who will use the system, Doctor Pedro Graça and Dr. Sofia Sousa. This was made to understand if it was useful, functional and it had a good usability and unlike the previous evaluation, in the case studies we do not measure the time or errors, only comments made aloud were recorded during the free use of the system by the user.

Basically, this test consisted of simulating navigation tasks that will be performed in the use of the system. The task development process consisted of performing the task on the part of the user, following the approach of saying aloud what they are thinking and what they intend to do, commenting on the interactions they were having with the system. This is useful also for us to understand what they thought that is more relevant, and what they found useful in the system.

4.2.1 User 1

The first user was Dr. Sofia Sousa, who helped us a lot throughout the development and already knows the system well. Before this evaluation was made, we made a demonstration of the actual state of the system and explained once again each visualisation’s purpose and features. Dr. Sofia Sousa really liked the fact that we were able to include the option to download the graphics, she is responsible for updating both websites and has already in mind to make publications there with data from our system and include the images of our graphics.

She also thinks it is very useful to be able to see the most searched terms, she said, “This is very good! We can see what people are looking for on the last month for example, then we may analyse if we are responding to these searches.”. She took a look at the FoodVis and said: "For example, we have a lot of searches about snacks (Figure 4.3), and this is a topic without many articles in our websites, we can plan to publish something about snacks to respond to this demand.”

Dr. Sofia Sousa also told us that she found our system much easier than Google Analytics, she said: “Google Analytics is not easy to use, it’s not intuitive, there are lots of courses to learn how to use it exactly for that reason. With FoodVis we can see exactly what we are interested in, we can see only one or both of the websites at the same time, in a simple way, easy to explore and that allows us to draw new conclusions.”
4.2.2 User 2

The second user was Doctor Pedro Graça, the director of the PNPAS who also knows the system and was part of its development. In this case, we did not perform the demonstration again because he already had listened when we made the previous one. Doctor Pedro Graça was also very enthusiastic about the final system, saying: "We have a lot of interesting stuff here! This is like seeing a disease in an organism that is still alive, it is much more interesting to see it *in vivo* that *in vitro*".

They had the idea, for example, that vegetarianism is more a urban phenomenon, so he showed a lot of interest in exploring the most searched keywords by regions to check this. He clicked on the word vegetarian and saw that on the map of Portugal the areas that showed more searches for the term were actually mostly littoral, like we may see in Figure 4.4. He said "This is very interesting! We consider this to be an urban phenomenon and here we can prove that."

Doctor Pedro Graca is specialized in the field of nutrition and told us that our idea of exploring this data this way was very interesting because it is something that is not usually done in the area of nutrition. He showed us a magazine from the world’s largest food and beverage company, which has for the first time dedicated two pages of its paper magazine to also showcase trends for a particular term, search

![Figure 4.3: FoodVis with the keyword “Snack” selected.](image)

![Figure 4.4: FoodVis with the keyword “Vegetarian” selected.](image)
spikes, wordcloud of most wanted terms like as we can see using our system.

4.2.3 Discussion

They both thought the system was really interesting to them since they had never been able to do anything like this and considered it very important these days. Through FoodVis they can adapt the contents of the websites, understand the interests of the population of Portugal about nutrition and respond to their needs. They commended us for taking this initiative because they feel that our work is very valuable because we worked directly with them and responded to their needs and it was not an external analysis. In addition, nutrition is an area that has never been worked out in this way and for them to now have the opportunity to analyse their own results in this way will be very useful.

4.3 Discussion

With the usability tests we can conclude that users may answer all the questions using our system with few errors and without taking too much time. The usability of the system was also classified through the SUS as Excellent reaching 91.13 points on a range between 0 and 100. In general, the feedback received was very good, users said that the system is fast, easy to use and interesting.

Through the tests carried out with the members of the DGS we could verify that our system was in agreement with its objectives, it answered the questions that they found more relevant to answer, being easy to use and intuitive. FoodVis complies with the requirements initially proposed and users are very satisfied with the system developed.

Considering all the described evaluation results, we may confirm that FoodVis was successfully accepted among users, that way we may validate all the work done during FoodVis’s development process. Furthermore, by being able to answer all purposed questions and a high SUS score the evaluation assured this dissertation’s main goal which was: to study ways to visually identify relevant information about the interest of Portuguese population in nutrition, contributing to understanding their needs, and look for possible improvements on the information shared.
Conclusions and Future Work
Food is an essential part of our life and it is not easy to understand the nutritional information and what is the best for our health. Taking this into account our initial idea was to develop a visualisation that helps the users in exploring and understanding food data, however, we realised that there was not enough available data about this subject so together with the DGS we found that the most promising approach was to analyse data from their websites, like user' views and searches, in order to help them to understand the interest in nutrition in Portugal. FoodVis is an information visualisation system that provides information about metrics collected from two websites, Alimentação Saudável and Nutrimento, related with nutrition that belongs to the DGS. It consists in an interactive dashboard with interlinked views that provide to the user different views on the existing data, allowing pattern discovery, compare, get a detailed view and understanding the trends.

In the development of this system, we followed iterative and incremental approach, with cycles of "design, test, measure, and redesign" being repeated as often as necessary in order to understand if the prototype kept meeting users’ needs and to verify the users’ intentions while using the system. FoodVis can be accessed from any computer connected to the Internet only using a standard browser without any previous installation, it has a drag-and-drop mechanism allowing the users to change the default dashboard layout and also minimize or resize some graph. All these mechanisms are responsive, automatically adapt themselves while maintaining the current information context and provide to each user a personal and richer experience.

When the final visualisation was achieved, a group of twenty users tested FoodVis. Users were presented with five questions to be answered by using FoodVis and their performance was evaluated through quantitative measures: the time it takes the user to do the task, the number of errors made, and the level of satisfaction (SUS questionnaire) while doing such tasks. The users were able to answer all questions properly with a low number of errors and not spending much time. The usability of the system was considered as Excellent, achieving a global score of 91.13. Our case studies verified that FoodVis was in agreement with its objectives, it answered all proposed questions and is easy to use.

Future work includes adding new metrics, like the average of new users in the visits, compare posts to understand what are the most read and understand if it was related to some time of the year. The back-end could be improved to have an easier what to gather the data from Google Analytics and keep the system up to date, using other plugins and create a database since the amount of data will grow. Also, new idioms should be considered to be integrated within FoodVis’s dashboard. In the Wordcloud a deeper processing of the keywords could be made and new features like searching and groups of terms could be added. It could as well have new ways of decreasing user proneness to error when they had to change between dashboards, as also, improve the top bar finding another way to clarify which data they are seeing and in which dashboard they are, and later carry out new usability testing to validate such research.
Bibliography


