Universitat de Lleida

TREBALL FINAL DE GRAU

ESCOLA POLITÈCNICA SUPERIOR
UNIVERSITAT DE LLEIDA
INSPIRING THE FUTURE

Estudiant: Oriol Aguilar Larruy

Titulació: Doble Grau en Enginyeria Informàtica i ADE

Títol de Treball Final de Grau: **Visualization and Analysis of Household Consumption Data**

Director/a: **Roberto García and Mariona Farré**

Presentació

Mes: Juny

Any: 2022

Acknowledgements

I would like to thank all the people that helped me during this project. From the University, I would like to thank my directors, Roberto García, for all the support and advice related with the Rhizomer software that he has given to me during these months, as well as Mariona Farré, for his help in analyzing the Consumption data.

I also want to thank my parents, my siblings and my girlfriend, who have listened all my explanations about the project without exactly understanding what I was talking about.

Contents

List of Figures

List of Tables

1. Introduction

1.1 Introduction

Nowadays many experts qualify data as the new oil. Extracting real knowledge from huge data sets of raw data is a challenging activity and one of the most demanded jobs currently. When having high loads of numerical data stored as rows, the best way to understand the whole picture is by depicting the dataset in charts.

In addition, one of the mission of economists is to translate the reality we live in to data so the situations can be analyzed, compared and studied. Household Consumption is one of the most important macroeconomic indexes because it represents the source of the economic activity of a country by its most important actor, families.

By combining the coding skills required for developing a web application that can manage data and display it in charts and the knowledge related with Household Consumption's determinants and the theory of consumption, useful conclusions can be extracted from a dataset. Without the named skills it would be impossible to extract any important idea because the quantity of information is massive.

The final thesis of a degree is not only an assignment but the opportunity to learn something new and this thesis follows this principle.

In the first part of the thesis, the development of the application is explained. The development has been divided in iterations and each iteration consists on a set of tasks to be done. It includes all the considerations and decisions that have been taken in order to make the most usable and useful application.

This application has been done on top of another project called Rhizomer. Rhizomer is a faceted browser for exploring semantic data from SPARQL endpoints in an interactive manner it is maintained by Roberto García, one of the directors of this thesis. However, the system lacked of a usable way of exploring numerical data so the development of the application is focused on building a feature of Rhizomer that works as a chart visualizer.

The second part focuses on the analysis of consumption from a theoretical point of view and then using empirical data from the European Union (EU). The goal of the part is to demonstrate the theoretical knowledge of consumption with the data from the European Union using the different features developed.

1.2 Aims and Objectives

1.2.1 Objectives of the development of the application

Extracting knowledge from any dataset should be the main challenge and goal of any data-related project. Secondly, when designing an interface the most important challenge is developing something that is usable for the users.

Moreover, the other big challenge is developing a chart visualizer that complements Rhizomer and does not look as something it was not planned to do.

The concise objectives that work as the guide of the thesis are the following ones:

- Transform a set of traditional rows and columns datasets into the graph data format compatible with the standard that Rhizomer uses.
- Represent this numerical data into Rhizomer allowing the user to look up the data in the form of different charts.
- Allow the users to go one step further and facilitate them the extraction of conclusions by offering them some tools like filtering, combining different metrics,...
- Allow the user to use Rhizomer with datasets that contain both numerical and semantic data.

1.2.2 Objectives of the analysis of Consumption Data

The section of this thesis dedicated to the Analysis of Consumption Data from EU countries will be focused on the usage of the chart visualizer tool and the analysis of the data, the main goal is not to forecast the consumption expenditure of the EU in a future nor to detect an unknown conclusion about the macroeconomics of these countries. The main aim will be to be able to explain, justify and understand the situation and some causes of the economic behaviour of Europe in the last years with the help of charts.

The concise objectives that work as the guide of the thesis are the following ones:

- Study the different determinants of consumption form both a microeconomic and macroeconomic approach.
- Extract valuable information like correlations or patterns with the help of the developed Rhizomer Chart Visualizer.

- Extract more precise information with the help of the Rhizomer Chart Visualizer and third data sources and tools in order to understand the situation of household consumption in the last years during the COVID-19 pandemic and the subsequent inflation period.

2. State of the Art

2.1 Semantic Web

Web technologies have evolved since the apparition of the World Wide Web for the first time back in the 90s. The first versions of the web, known as Web 1.0 was unidirectional and its content was static, it was mainly used as an informative tool used by cultural and educational institutions.[20]

Web 2.0 started at the beginning of the 21st century, web-related technologies improved and Internet users increased so the web became a social affair. The web wasn't anymore static and users could be part by contributing content. It was the time when social networks and internet forums became popular.[20]

However, it is the third stage of the web, Web 3.0, the one that interests us. It is also called Semantic Web or Data Web and its main goal is to make Web content machine-readable. The following quote from Tim Berners-Lee, the father of the World Wide Web, is the best way to explain not only the concept itself but the direction where Semantic Web is intended to go:

> *"I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web – the content, links, and transactions between people and computers. A "Semantic Web", which makes this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The "intelligent agents" people have touted for ages will finally materialize."[19]*

2.1.1 URI's, RDF and Vocabularies

For understanding much better the concept of how a machine can understand Web content, which is generally conceived as natural language, we must explain different concepts that are key to understand how the Semantic Web works:

If we want that machines understanding human language, we cannot express it as simple array of characters. For example, if a human bilingual in English and Spanish reads "Germany" and "Alemania" he will understand the concept behind these two words even though they are spelled completely differently. For a computer, it is much more complex to recognize the concept behind these words. Because of this, it is why URI's (Uniform Resource Identifier) exists. It is a universally unique identifier used to represent the underlying concepts. In the prior case, every time that we would want to represent the state of Germany, in an unambiguous and machine-readable way, we can use the following URI: https://dbpedia.org/page/Germany.

However, language is not represented only with independent words. Sentences imply the connection of different concepts that alltogether make more sense. In this project, we will use RDF (Resource Data Framework) to represent these statements that we have mentioned before.[21] RDF is a directed graph composed of triple statements. An RDF graph statement is represented by: 1) a node for the subject, 2) an arc that goes from a subject to an object for the predicate and 3) a node for the object. Each of the three parts of the statement can be identified by a URI. An example of the representation of a sentence in RDF would be the following one:

"Berlin and Munich are cities in Germany, a country with 83 million inhabitants"

```
<https://dbpedia.org/page/Munich>
    foaf:citiesOf <https://dbpedia.org/page/Germany>.
<https://dbpedia.org/page/Berlin>
    foaf:citiesOf <https://dbpedia.org/page/Germany>.
<https://dbpedia.org/page/Germany>
    foaf:inhabitans "83000000" .
```

And the following graph representation.

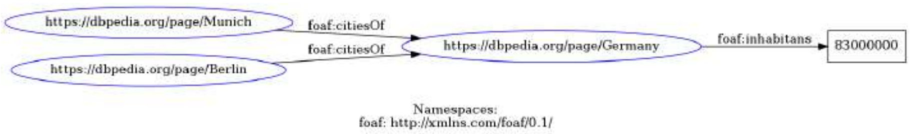

Figure 2.1: RDF example

As it can be seen in the last figure, URIs can be both objects and subjects and arcs represent the relationship between them.

To further assist computers when processing data, we can group similar concepts into classes. For example, in the previous example, Munich and Berlin's URIs could be classified as Cities and Germany as Country. By doing this, the graph is enriched and its interpretation is easier because, instead of representing the individual instances, the classes provide additional context that helps the computer process the data.

Vocabularies consist on the collections of Classes and Properties, the name that receives the relations between the subject and the predicate. And it is recommended to use them as a way of standardizing the RDF dataset, facilitating data processing and integration.

2.1.2 SPARQL

SPARQL is an RDF query language capable of retrieving and manipulating data stored in RDF format. The following is an example of a SPARQL query to show country capitals in Africa, using a fictional ontology:

```
PREFIX ex: <http://example.com/exampleOntology#>
SELECT ?capital ?country
WHERE {
  ?x ex:cityname ?capital ;
     ex:isCapitalOf ?y .
  ?y ex:countryname ?country ;
     ex:isInContinent ex:Africa .
}
```

2.2 Rhizomer and Data Visualization

Rhizomer follows the Shneiderman's Mantra [23] when helping users explore data, consisting in three steps.

- **Overview:** get the full picture of the data set at hand.
- **Zoom & Filter:** zoom in on items of interest and filter out uninteresting items.
- **Details:** after zooming and filtering the user arrives at the resources of interest and gets their details.

These three steps give the user maximum freedom to explore what they're interested in, while still maintaining context of the overall data set and the ability to point users in the right direction with particular zoomed and filtered features [23]. These steps are already used in the current version of Rhizomer and will be as well the guideline for this project.

2.2.1 How Rhizomer Works

When users configure a dataset in Rhizomer and open it, they will see the first step of Shneiderman's Mantra, the overview. Depending on the configuration chosen when setting up the dataset, the information will be shown as a word-cloud or a graph.

The words and circles we can see are the classes of the dataset. The bigger the circle or the word means the bigger its presence in the dataset. In the case of the graph overview, the most frequent relations between classes are displayed.

Figure 2.2: Overview of DBPedia data in Word Cloud format

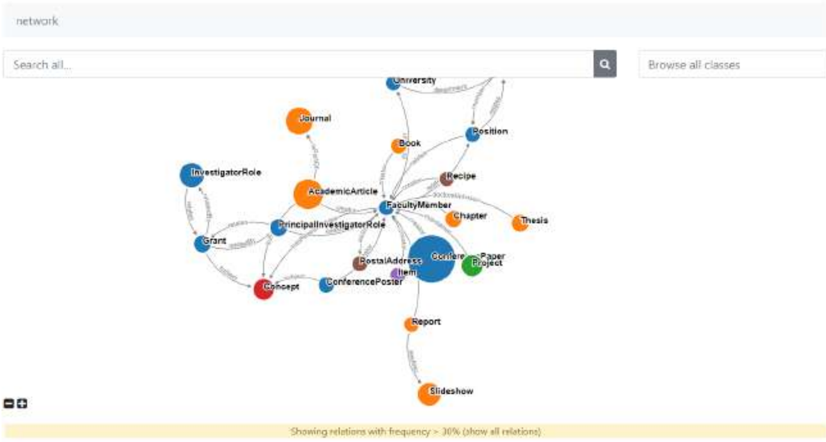

Figure 2.3: Overview of academic data in Graph format

By clicking on any class the following view appears. It's the one related with the second step, the Zoom and Filter. The user can see all the members or instances of a class and can filter them using their facets. The facets are all the relations that an object has. They can be another object or a simple scalar that represents an attribute of the object. This view is organized by pages for being less greedy when working with large datasets.

All the instances can be clicked and the user can see on demand all its attributes and other classes that are linked with that instance. This view consist on the third step of the mantra, the visualization of the details.

2.2.2 Rhizomer's Architecture

Similar to many web apps, Rhizomer is divided in two layers, the front-end and the back-end. This thesis is going to be focused in the development of the front-end, but it is necessary

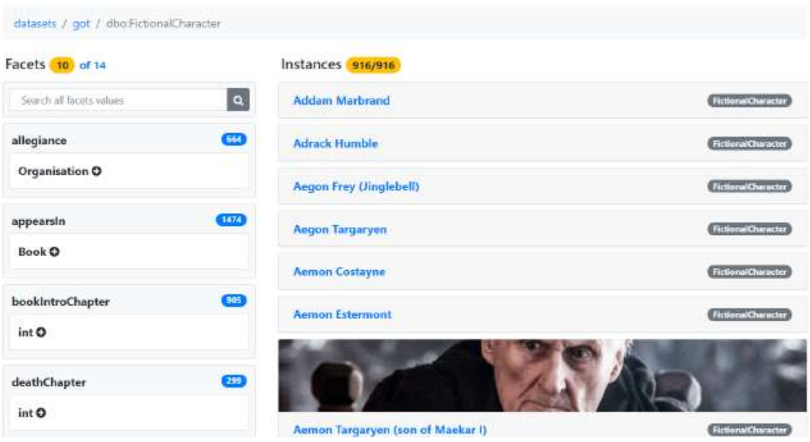

Figure 2.4: Faceted view of the Fictional Characters class of Game of Thrones dataset

to explain the overall architecture of the app, as shown in Figure 2.5, to understand better how the API calls work.

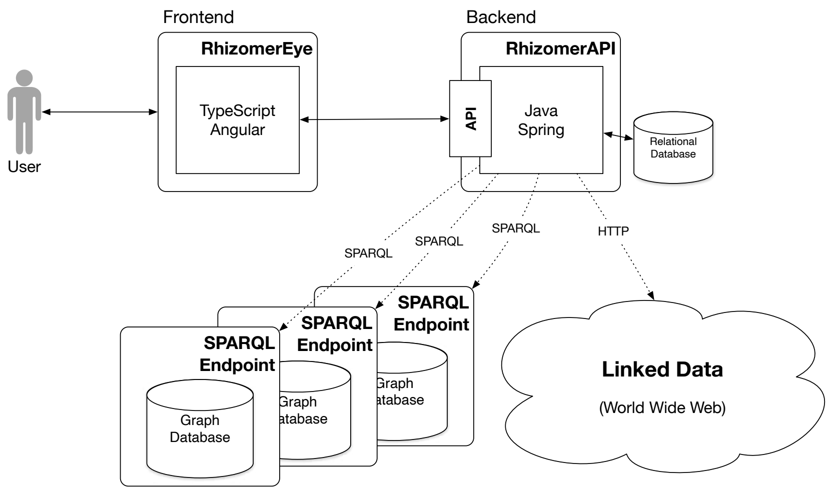

Figure 2.5: Rhizomer's Architecture

The back-end is connected with the graph data stores using SPARQL to extract it for the app purposes. The back-end is also responsible for retrieving the right data when filters are applied. The front-end is the responsible of making the calls, sorting, managing and displaying the data. The information is returned to the front-end in a JSON format, specifically JSON-LD. Additionally, to make the front-end more efficient and reduce response time, the API provides results pagination.

2.2.3 Angular and Google Charts

Angular is a web application framework based on Typescript, open-source and maintained by Google, used by Rhizomer for its front-end. It is based on a MVC (Model-View-Controller) model, a software design pattern used for user interfaces that divides the logic of the program into layers.[24]

- **Model:** Manages the data and the logic of the application.
- **View:** Represents the model in a visual way. It is represented by the HTML file. Multiple views can be controlled by the same model.
- **Controller:** Accepts the inputs of the user and commands it to the model.

Different frameworks like Google Charts and D3 have been considered to display the charts. D3 is a Javascript library that can work for Angular. Its strength is its powerful customization and it let users create their charts with customized features like interactivity and so. It was developed by Mike Bostock and it was widely used in the NYTimes for their interactive-data stories. However, its main drawback is its steep learning curve. The charts that are designed for this thesis have to be as simple as possible so there is no need to use such a powerful and complex library.

On contrast, Google Charts is much simple. It has a rich library with all the necessary charts and it allows some customization at a certain point. Additionally, it has a special library for Angular (angular-google-charts[1]) that will facilitate the development.

[1] More information about this library in the following repository: https://github.com/FERNman/angular-google-charts

3. Development

The development of the system has been planned as iterative development, by dividing the whole process into small chunks called iterations. With each iteration, additional features can be designed, developed and tested until there is a fully functional software application ready to be deployed.

The main benefit of iterative development is the flexibility that gives to the developer compared with the traditional development. Sometimes there can be unforeseen problems that don't surface until development begins and by dividing the work to do, those can be detected easily and solved in future iterations.

3.1 Iteration 1: Data Gathering

In order to be able to use Rhizomer web platform, you need to have your data in RDF format. Usually, the most common sources of economic data don't have the data in this format so a prior effort in order to convert the data had to be done before being able to start developing the new features of the system.

3.1.1 Data Source and Data Merging

The data source of all the economic indexes represented is Eurostat. The page offers Tab Separated Values (TSV) files of a large number of indexes of all the EU members during the last years, one file per index. The data in those files is represented in a tabular way, where each column represents a year and each row a country. The problem with this disposition of the data as separate files is that it makes more difficult to transform it into RDF.

Because of this, a Python script using pandas library was developed in order to be able to merge all the files into a single one. The returned dataframe has as identifier the concatenation of both the country code and the year. The other columns represent each index, as Figure 3.1 illustrates.

GDP.csv			
Country	2018	2019	2020
AT	11516	11517	11518
...			

HICP.csv			
Country	2018	2019	2020
AT	1.03	1.04	1.035
...			

output.csv				
Id	Country	Year	GDP	HICP
AT-2018	Austria	2018	11516	1.03
AT-2019	Austria	2019	11517	1.04
AT-2020	Austria	2020	11518	1.035
...				

Figure 3.1: How the different tabular files were merged into a single one

3.1.2 RDF creation and SPARQL Query

Once the tabular file with all the chosen data is prepared, it is necessary to transform it into RDF. The software used for this is OpenRefine, a desktop application that allows data transformation and data cleaning as well and it is widely used for creating RDF files.

Open Refine allows you to modify the columns of any tabular file with a simple scripting language. Because of this, creating a row with an URI from the information of another column can be done easily. However, the main advantage that OpenRefine offers is the RDF generator. Figure 3.2 illustrates how users create the RDF. It allows the users to generate the triplets that constitute RDF at their convenience. Moreover, its users can decide the classes, the vocabularies and the properties that better describe the data being converted.

Figure 3.2: How to create a RDF with OpenRefine

As it has been explained previously in Figure 3.1, each row describes one country in a year and the values of the economic indexes. Therefore, the subject of all the triplets will be this identifier in the form of a URI. The properties that link the subject with the different objects are derived from name of the objects' columns.

The object can be a literal value or a URI like the subject, in which case the object can be the subject of other relationships. In the case of the values of the different indexes it makes sense to represent them as literals but the country's name, for example, can be replaced with a URI using OpenRefine's tools.

The vocabulary chosen for representing the macroeconomic data is the Data Cube[1], more specifically multi-measure observations. It is intended to work with multi-dimensional data, such as economic indexes, on the web in such a way that it can be linked to related data sets and concepts so it suits perfectly the data we want to represent.

3.2 Iteration 2: First Design of Chart Visualizer

Designing a web application is a complex process because it includes a lot of decisions to be considered. From choosing the colours and the layout to the interaction and the routes that a user has to do in order to get to a certain point of the system.

When trying to expand the functionalities of an already-working web application, things change. The challenge is not thinking about a new layout but trying to incorporate the new things into the old system while maintaining the same order, aesthetic and usability of the previous version. All the visual objects that we add should look as they were planned in the first design.

As it has been previously said, the library chosen for the chart representation is Google Charts for Angular, so the design had to be adapted to the possibilities that it offers. This Angular extension allows the customization of the charts in a certain way: the colour, the tags but the type of charts is limited the collection of charts selected has to be rich enough to translate the data into different valuable information.

3.2.1 Design

The main goal of the chart visualization is to give the user another way of looking at the different classes of the dataset, so the design has to preserve the same layout as the interface that represents the instances.

Regarding the usability and the interactivity of the new feature, it is expected that the chart visualization is accessed with a button in a visible position at the top of the page and the different types of charts have to be able to be selected with a simple button-group. How the chart is created and which attributes of the instances will become the axes will be determined in the next iterations when actually coding and API calls are involved.

The collection of selected charts consists of 4 representations. A table chart, a bar chart, a line chart and a pie chart. The selection of these charts is based on giving the user as much information as possible in a limited number of charts. The evolution of a metric, the comparison between peers and the weight of a metric respect the total, are frequently asked questions when analyzing a dataset so it is necessary to have these charts available for the user. However, Google Charts offers the possibility to edit the charts as they desire so it may be an open door to give to the user.

[1] Handling multiple measures with Data Cubes: `https://www.w3.org/TR/vocab-data-cube/#dsd-mm`

Moreover, the facets filters must affect the charts in real-time at the same way they affect the faceted view results.

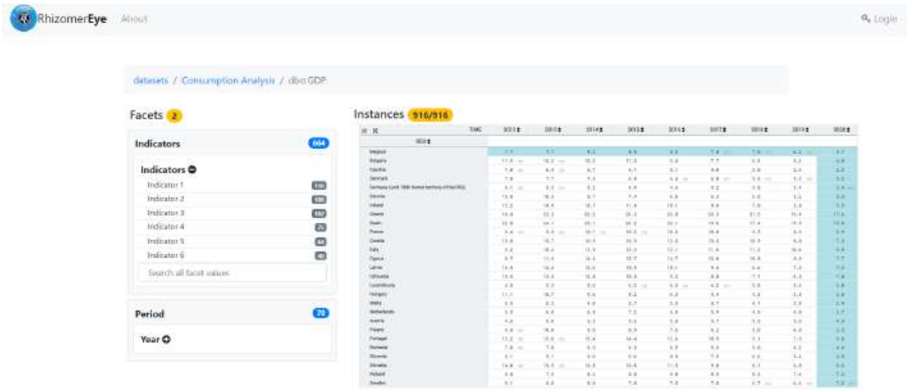

Figure 3.3: Design of the table prototype

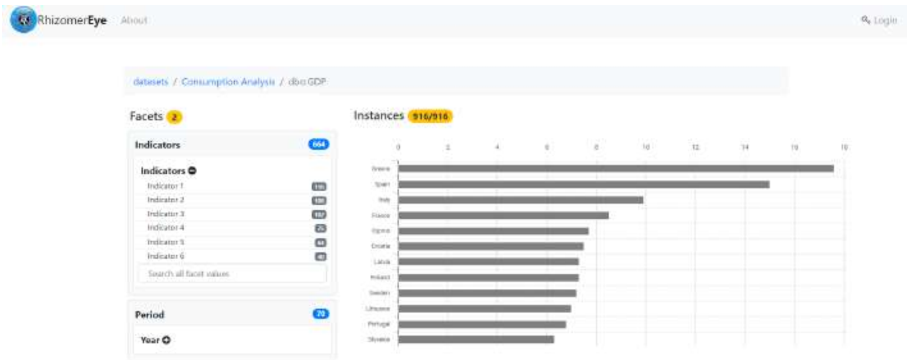

Figure 3.4: Design of the bar chart prototype

3.3 Iteration 3: Developing of the chart Visualization

This iteration is going to be the longest, the main goal is to develop a functional but not final chart visualizer with the design planned before. Usability concerns as well as the aesthetic of the feature won't be the goals of this iteration even though the work done will follow the line planned in the design and try to be done as correctly and usable as possible.

The feature that is going to be developed is part of a broader system, Rhizomer's front-end. This application, developed by Roberto García is divided into components and has its own idiosyncrasy when calling the API or when retrieving the data. Therefore, before starting developing the new features, it is necessary to comprehend the code already written and try to integrate the new functionalities without looking like patches.

3.3.1 Software Requeriments

All Angular Google Charts' charts work the same way. A typescript variable, or a group of them, contains the required information and by using the tag *google-chart* and referencing the right variables, the user can create the charts.

```
<google-chart
  [title]="chart.title"
  [type]="chart.type"
  [data]="chart.data"
  [columns]="chart.columnNames"
  [options]="chart.options"
>
</google-chart>
```

As it can be read in the previous code snippet, the chart is called from the *HTML* file but the dataframe represented is configured in the typescript. The necessary data for a complete chart is the following one:

- **Type:** The library offers an enumeration of chart types. By changing this attribute, a chart can change its form. ChartType.Table, ChartType.Bar, ChartType.Line and ChartType.PieChart will be the chart types used.
- **Data:** It is the most important attribute. It represents the data represented in the charts in a dataframe format. As Figure 3.5 illustrates, the required information that the instance needs in order to work properly. It is a two-dimension array with the rows' identifier and the values. Multiple columns are allowed in the vast majority of charts but not all of them so there has to be a precaution and resize the matrix when changing the chart type.
- **Columns:** As it can be noticed from the Data description and the Figure 3.5, column headings are not included in the dataframe but in a separate array. The size of this one dimension array must match the number of columns of the dataframe. Otherwise, an exception will be thrown.
- **Options and other optional attributes:** The three attributes mentioned before are mandatory but there are other optional attributes like the size, the colours, the titles,... However, these attributes are necessary in order to show the information in a better way.

Another aspect that has to be considered is how the data is retrieved from the API. Even though it is stored as RDF, the query result is in JSON format. When requesting all the

Column_name_1	Column_name_2	Column_name_3	Column_name_4

Row_Id_1	x	x	x
Row_Id_2	x	x	x
Row_Id_3	x	x	x
Row_Id_4	x	x	x
Row_Id_5	x	x	x
...			

Figure 3.5: Example of a column names' array (top) and the dataframe (bottom), where x represent values

instances of a class, the back-end will return an array with each instance represented as a JSON object. Let's remember that, in the case of the thesis, a result will be an observation with a country, a year and all the macroeconomic indexes of the country in the selected year.

An example of the response of the API is illustrated in the Figure 3.6.

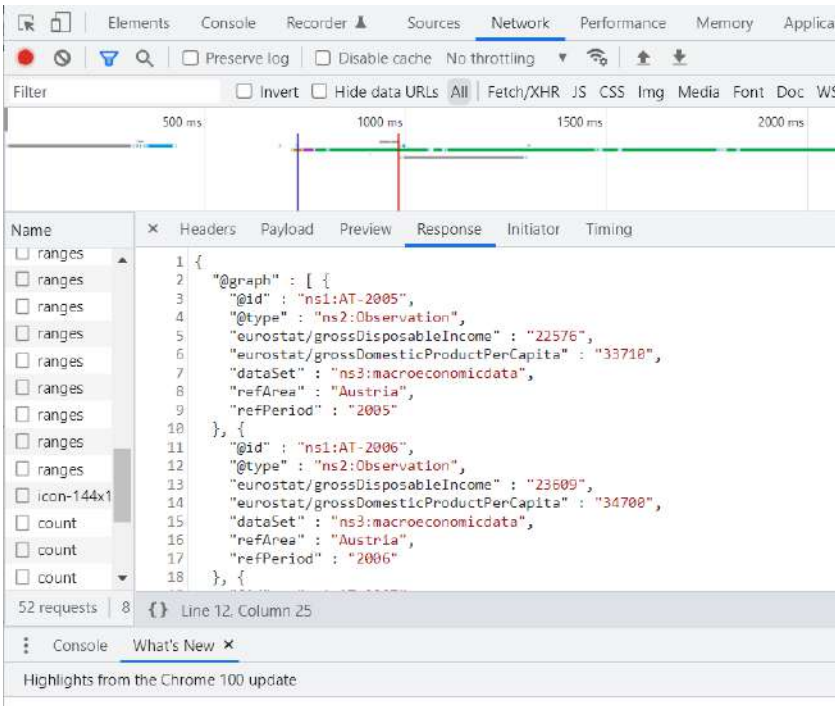

Figure 3.6: Example of the response of the API when retrieving data with all the attributes

Different queries can be done to the API in order to get, for example, the number of registers that a class has, the instances without the details or just the identifier and the label.

3.3.2 User Requirements

In order to develop an usable and intuitive application, it is necessary to state which are the actions that a user has to be able to do. It has to be remembered that the main objective of the chart visualization is giving the user another way of looking at numerical data as a whole other than reading the individual instances. The following bullet points are the actions that a user would be able to do using the new features added to Rhizomer:

- Be able to switch from the instances mode to the chart mode and back in few clicks.
- Let the user choose the two attributes that will be the references of the dataframe. Row_Id_X and Column_name_X from Figure 9.
- Once the attributes are created the user must be able to easily switch among the different types of charts available.
- The kind of values displayed in the charts should be also easy to change, among those available as different columns from the dataframe.
- User has to be able to filter the individual items displayed in the chart the same way the instances are filtered in the faceted view.

Moreover, it has to be considered that there are things that the system must restrict the user to do. The datasets uploaded in Rhizomer can work with large quantities of data and it would not be a good decision to let the user interact with all the set because it can led to create charts that doesn't make sense, cause uncontrolled exceptions and consume too many resources in the client-side.

- Not offering the chart mode visualization to classes that do not accomplish the necessary requirements: having numerical values to represent and having at least two attributes that can be categorized as references.
- When choosing the attributes that will be references, only offer the possibility of choosing between those facets that are part of every instance of the class. By doing this the system guarantees that the dataframe can be created correctly.

3.3.3 Implementation

In order to get the attributes necessary to create the tables used to generate the visualizations, it is necessary to analyze the instances of the class. As it has been explained in the requirements, a class is considered as graphically visualizable if there are at least two attributes that appear in all the instances and can work as the row index and column index of the table. The user selects this information by clicking on a button with a form collapsed where just the possible attributes are shown.

Once the user has selected the attributes, the system creates multiple tables. One table for each attribute that can be represented. After downloading all the instances, each value is placed in the right position according to the two attributes previously selected by the user.

After this process, the charts are displayed together with a group of buttons that can control them, as shown in Figure 3.7. The chart displayed by default is the table one, it contains all the data in a table format. The user can change the chart type by clicking on the different buttons. The underlying code will change the attribute Type of the Google Chart accordingly. Moreover, the user can change the displayed data by pressing the dropdown button and selecting the desired dataframe column.

Figure 3.7: Group of buttons displayed on top of the charts

However, each chart type has its particularities and cannot be configured the same way. For example, in order to make the information shown in a Line Chart more understandable, it is recommended that the attribute that describes the time period is in the horizontal axe. Nevertheless, there can be other situations where the opposite configuration is preferred. Because of this, there is a *Switch Axes* button. When the user clicks the button, the tables used to generate the visualisations are transposed. The difference is notable as it can be seen in the Figure 3.8.

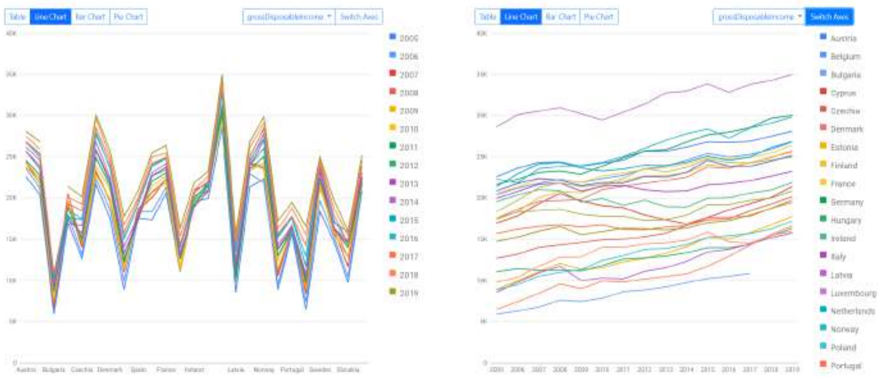

Figure 3.8: Effect of "Switch Axes" button in a line chart

3.3.4 Conclusion

At this point, an already functional version of a chart visualizer is already deployed, but there are still a lot of things to improve so the user can interact with the web app more easily. The following iteration will be focused on solving all these issues in order to make a more usable app.

3.4 Iteration 4: Improving usability and efficiency

3.4.1 Introduction

At this point, the web application worked perfectly and could be used for normal usage. For example, extracting conclusions from macroeconomic charts, which is the second purpose of the thesis.

However, there are some limitations that could cause the application to not work properly depending on the size of the data set uploaded, depending on the device the web app is displayed on and depending on the way the user interacts with it.

3.4.2 Development

Adding an endpoint structure

In the last iteration all the development related with the creation of the charts was done only using variables that changed the behaviour of the component and modified the view. No URL structure was used and this implied that when the user was on the chart visualizer and the page was refreshed the component restarted from scratch.

A URL is nothing more than the address of a given unique resource on the Web. However, it doesn't have to be perceived as a static route to a resource but as a dynamic one where parts of the URL represent choices made by the user. The following example helps clarify how the URL changes when the user is located in the *Observation* class from the *consump-semantic* dataset and displays a chart visualization choosing as axe variables the refArea and the refPeriod properties.

From:

`http://localhost/datasets/consump-semantic/qb:Observation`

To:

`http://localhost/datasets/consump-semantic/qb:Observation/charts:refArea&refPeriod`

The component detects that the last element of the route is not the class of the dataset but a string that starts with the key word *"charts:"*. After detecting this, it proceeds to initialize the chart visualizer with the two elements axes separated by an '&'.

The URL of Rhizomer saves as well the filters applied to the class as a query. Moreover, the chart representation wants to be able to be accessed from a natural way from the instances mode so the filters don't have to reset once we switch to the chart visualization mode.

Because of this, when navigating through the chart visualization site, the query in the URL has to remain.

***"Paginating"* the chart visualizer**

One of the things that has to be considered is that Rhizomer has to be able to manage big quantities of data. When working with a dataset that has a lot of data it would be risky to retrieve all the information at once. It could be a heavy task for the browser and it would be very demanding for the user's internet connection, specially on a mobile device.

The instance representation that was already developed when the thesis was started is paginated and because of this only 10 registers are retrieved at a time. However, this cannot be done with the chart visualizer because some times it requires to display the bigger picture so conclusions can be extracted.

With this premise, a system that allows the user to choose the size of the retrieved dataset had to be designed. Initially only 20 instances will be retrieved to ensure that anyone's browser is capable of managing this information. After this, the system will allow the user to choose the range of data to be retrieved. That will allow the user to display the whole dataset at once or to represent the dataset in different *"pages"* of the chosen size.

Adding a necessary feature after testing

After having finished the third iteration the software developed was tested by analyzing charts with private consumption data. This analysis is documented in the second part of the thesis.

The chart visualization was a very complete tool because it allowed the user to represent multiple two dimensional instances of the same type in the same chart. That meant that if the user was working with data of the Gross Domestic Product (GDP) from countries of the European Union, which is the case for the second part of the thesis, it was possible to compare the GDP of different countries. However, users cannot compare the GDP of a country with another variable.

Because of this, another feature was developed. It lets the user to choose as many variable as desired and then create a chart that combines all of them. This feature, extremely useful for finding correlations, was found very useful and was included in the final version of the application.

Making it responsive and improving the interface

Rhizomer is a responsive application and it can be used in any device. Because of this, the chart visualization had to be responsive as well. The problem is that an specific size has to be set as an attribute of the chart. So a *@HostListener* that monitored the resizing of the window has been used.

By monitoring the size of the window it can be calculated what is the best size for the charts no matter the device used. Moreover, it can be determined when the filter panel is collapsed so the charts can adapt to the available space.

Another improvement of the user experience is the use of icons. By using icons the buttons become more intuitive and having less text in the user interface makes it look cleaner. The Figure 3.9 shows the changes of the button group in charge of switching from the different types of charts.

Figure 3.9: Before (top) and after (bottom) of adding icons to buttons.

Another necessary change that had to be done is changing the different forms that were developed in order to create the charts or in order to create the correlation charts. During the third iteration, when those features where developed for the first time, a collapse button was used. A collapse button is a button that when pressed displays an inline form. The problem was that this form moved all the other elements that were next to it and broke the aesthetic of the interface.

In order to improve this, some research was made and it was decided to use a modal instead. A modal is a similar element as a collapse button but instead of displaying an inline form, it appears as a pop-up. By using a modal, the interface looks cleaner and there is more room for the form's elements. As it can be seen in the Figure 3.10 the difference is notable.

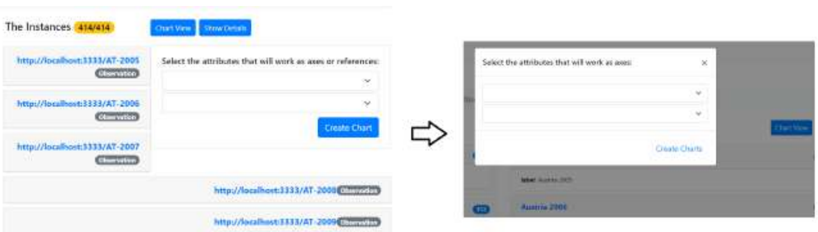

Figure 3.10: Before (left) and after (right) of adding modals instead of collapse's buttons

3.4.3 Conclusions

The evolution of the software from the third iteration to the fourth does not vary much from a visual point of view. However, those changes have allowed the system to be more usable. The final application has all the features required for analysing the target data through charts and it is flexible enough so it can be used with any dataset.

3.5 Future Work

As it has been said previously, the application works fine but there is still some work that could be done in order to improve its potential.

Determining the type of non-Classes facets

Each class recognised by Rhizomer in the input data has different facets, corresponding to properties of the instances of the class. Some of them are relationships and point to other classes, but others are attributes with literal values. These can have different types: strings, numbers, dates, hours... though by default the are all treated just as strings. By having the facets values labeled as number, strings, etc. it would be possible to improve the way the tools guides the users during visualisations configuration.

All the charts and figures analysed from this point are extracted from Rhizomer Chart Visualizer. Available in the following repository: `https://github.com/oriolaguilar/rhizomerEye` using the following SPARQL endpoint: `https://api.triplydb.com/datasets/OriolAguilarLarruy/1stDataSet/services/1stDataSet/sparql`

Range of values facets filters

Initially, that was one of the objectives planned for the thesis. The plan was making filters that were able to filter not only single value for a facet but a range of them. At the end, this improvement was dismissed because it was not requirement for the rest of the planned work and it required profound changes in both the frontend and the backend of the application.

The filter system of the application worked as it follows. When the user selects a value of one facet, it automatically selects all the registers that have the filtered value in that facet. This means that filtering different values of a facet is not allowed.

Moreover, if the type of the scalar facets can be determined. The numerical and date facets could be filtered with intervals using sliders for example. If using this option, the URL query structure has to be changed as well so other operators like *(<, >, <=, >=)* can be used.

4. Determinants of Household Consumption

4.1 About Household Consumption and Consumption Functions

Household consumption[1] consists of expenditure incurred by resident household on goods or services that are used for the satisfaction of needs or wants.[12]. It is constitutes around the 60% of the Gross Domestic Product of a country[13] as it will be explained later when the determinants and causes of consumption will be defined.

Households are the center of the economy, so it can be said that private consumption is the angular piece of the economy. Without consumption the world we live in would not be the same. Even the other important components of economics, some of them more abstract, such us government expenditures, business investments, interest rates or monetary policies turn around consumption. In fact, crisis and economic boom periods are determined by measuring and comparing the levels of consumption with the predecessor levels.[14]

Moreover, consumption can be conceived as a measurement of growth, for example the rise of consumption of electric energy is positively correlated with economical growth as electric energy is one of the most important inputs of the economy because energy is needed to produce goods and to provide services to consumers.[8].

In microeconomics, consumption is the theory behind the behaviour of households when deciding what to buy. Gender, age, wealth and the human nature among others are the determinants studied and analysed when analysing consumption from a microeconomic approach. This field of study, named behavioural economics, has been widely explored by

[1] During the whole thesis the concept "private consumption" is going to be used as synonym even though the latter includes other sources of consumption

scholars and different opinions have been held. The crucial point of all theories is trying to determine the relation between consumption and income.

4.1.1 Keynesian Consumption Function

John Maynard Keynes, described for the first time in 1936 a formula that connected consumption and income. Even though it is very simple it is considered as the first approach to link these two components. Income and consumption are connected with a variable called marginal propensity to consume (MPC) that acts as a regulator of consumption depending on the context. The MPC is the slope of the function and is calculated as $\Delta C/\Delta Y$.

$$C = C_o + cY \tag{4.1}$$

where: C = Consumption
C_o = Autonomous consumption[2]
c = Marginal propensity to consume. Decimal value between 0 and 1[3]
Y = Disposable Income

However, Keynes noticed of the tendency for the propensity to consume to decrease when income increased.[4] The decline in marginal propensity to consume as the income increases implies that the proportion of income that is saved increases with the increase in income. The reason behind that is that those consumers who have higher incomes will already have most of their needs satisfied.

Another variable that sometimes is used in Keynesian consumption functions in order to connect the income with consumption is the Average Propensity to Consume. The concept is similar to MPC but instead of being calculated with the variation of income and consumption, is calculated dividing the expected income between the consumption.

Even though Keynesian consumption is function very important for being the first one, is characterized of being too simplistic, it groups all the possible situations consumers can face into a single formula and does not take into account other determinants. During the following decades various authors published their theories based on Keynes' formula but adding the required complexity that consumption needs.

[2]Autonomous consumption is the minimum level of consumption that exists for basic necessities.

[3]MPC is calculated by dividing the the change in consumption is between the change in income.

[4]Keynes, John M. (1936). The General Theory of Employment, Interest and Money. New York: Harcourt Brace Jovanovich. The marginal propensity to consume is not constant for all levels of employment, and it is probable that there will be, as a rule, a tendency for it to diminish as employment increases; when real income increases, that is to say, the community will wish to consume a gradually diminishing proportion of it."

4.1.2 Franco Modigliani's life-cycle hypothesis

The life-cycle hypothesis suggests that individuals plan their consumption and savings behaviour over their life-cycle. Modigliani, added the life expectancy and the retirement age factors in his formula in order to describe a situation where consumers accumulate wealth when they earn and dis-save when they are retired. Another difference between the Keynesian function is that the life-cycle hypothesis incorporates the factor of wealth. Establishing that the consumption does not only depend on income but on accumulated wealth as well. Another key assumption is that all individuals choose to maintain stable lifestyles so the consumption will not vary much from one period to another.

The following formulas describe the life-cycle hypothesis. The first one (4.2) represents the consumption *(C)* function based on the three parameters that Modigliani includes: wealth *(W)*, income *(Y)* and age; years until retirement *(R)* and years of life *(T)*. The following formulas (4.3) and (4.4) are the consumption function (4.2) but transformed in order to obtain the marginal propensity to consume from wealth (α), the marginal propensity to consume from income (β) and the marginal propensity to consume (C/Y).

$$C = \frac{W + RY}{T} \tag{4.2}$$

$$C = \left(\frac{1}{T}\right) W + \left(\frac{R}{T}\right) Y = \alpha W + \beta Y \tag{4.3}$$

$$\frac{C}{Y} = \alpha \left(\frac{W}{Y}\right) + \beta \tag{4.4}$$

With the different equations, a model can be done in order to visualize the evolution of the different metrics. The module is initialized with an individual of age 30 that will receive a 2000 m.u income until 65 years old and will live until 80 years old. The wealth is calculated by adding the prior wealth to the saved money from income.

Age	Income	Wealth W	Consumption	α	β	C/Y
30	2000	0	1400	.0200	.7	.7
31	2000	600	1400	.0204	.694	.7
32	2000	1200	1400	.0208	.687	.7
...	...	...	...	...	...	...
77	0	4200	1400	0.33	0	.7
78	0	2800	1400	0.5	0	.7
79	0	1400	1400	1	0	.7
80	0	0	0	0	0	0

Table 4.1: Franco Modigliani's life-cycle hypothesis Model

The charts in the Figure 4.1 describe the model and the theory explained before. The chart describes the accumulation of wealth/savings during the work time until retirement

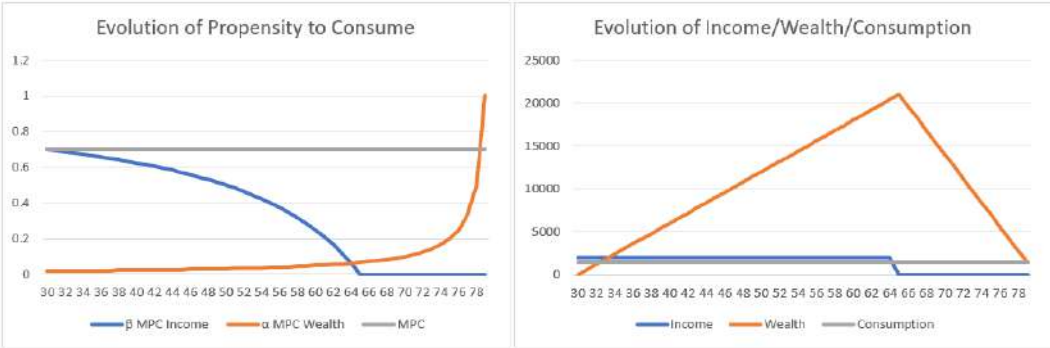

Figure 4.1: Chart Representation of the model of Table 4.1

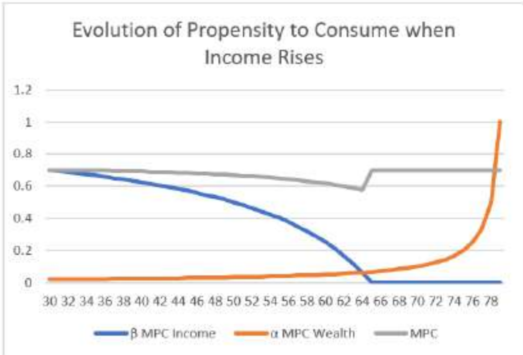

Figure 4.2: Chart Representation of the prior model when income raises

and the expenditure of all of it during the retirement years. When the consumer has an income the main weight of consumption comes from there, however when the consumer approaches and then surpasses the retirement age the weight of consumption transfers to wealth. Moreover, Modigliani's life-cycle hypothesis represents the behaviour of consumer that Keynes described when theorizing that marginal propensity to consume decreased when income increased. In the Figure 4.2 the chart represents the same model as the Figure 4.1 but with an annual increase of 2% of the income. Even though that consumption incremented as well, the marginal propensity to consume keep decreasing until the retirement age, situation that Keynes did not include in his theories.

4.1.3 Milton Friedman's permanent income hypothesis

Friedman's theory complements Keynes and life-cycle theories by adding a nuance. While neither Modigliani nor Keynes differentiated between different types of income and didn't put into the equations crisis periods, Friedman differentiates between permanent income, the part of the income that consumers expect to continue receiving in a future, in other words, their expected long-term average income and transitory income, the part that don't expect to continue receiving.[11]

Friedman agrees with the prior authors by establishing that there is marginal propensity to consume affecting but the main difference in his hypothesis is that while in the other conjectures, the whole current income was fully included in the formula whereas in the permanent income hypothesis only the permanent income is included. In other words, Milton Friedman believed that consumer spending is a result of estimated future income as opposed to consumption that is based on current income.

4.2 Macroeconomic determinants of Consumption

In order to determine the causes and the consequences of household consumption a group of different indexes have been selected. When analyzing something that has such an important weight in global economics like consumption where everything is connected to a certain point is important to look at the broader picture so not to miss anything. The goal of selecting different metrics is to be able to detect correlations and possible causalities between them so conclusions can be extracted.

In the following subsections, each chosen metric will be analysed theoretically. This analysis is going to focus in how different metrics cause a rise or reduction of household consumption based on the literature of different authors. The explanations will be followed by figures, charts and formulas in order to explain much better the situations. The indexes chosen will be the indexes represented graphically in the Rhizomer tool and analysed in the next chapter, because of this, if the index chosen has a special condition, per capita form, adjusted data to a certain year,... it will explained as well. The goal of defining the determinants before the analysis is to understand the theory behind them so the analysis can be more understandable.

4.2.1 Gross Domestic Product per Capita

Gross Domestic Product (GDP) per capita shows a country's GDP divided by its total population. The GDP is the market value of all the finished goods and services produced within a country's borders in a specific time period, in this case it will be based in an annual basis [2]. The GDP is the synonym of the economy's size of a country. This metric will be used in the per capita form instead of the nominal because by doing this the population of the country won't affect the result nor the conclusions.

$$GDP = Consumption + Investments + GovernmentSpending + NetExports$$

Household consumption is about 60 percent of GDP making it the largest component of GDP besides investment, government spending and net exports.[13] There are, however, large differences across countries that can range from about 45 percent of GDP to over 70 percent of GDP as it can be seen in the Figure 4.4. For example, if the country has an economy

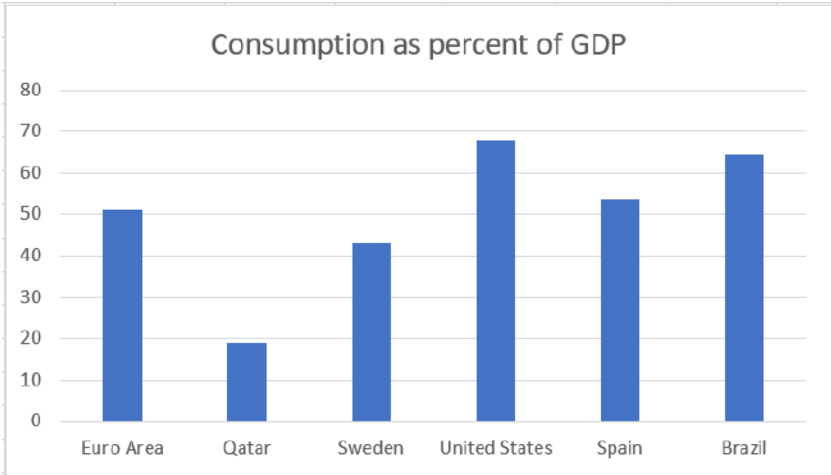

Figure 4.3: Percentage of consumption in GDP

based on exportation like Qatar or if the country is considered a fiscal paradise, household consumption won't be the most important component. The weight of the government and the state will determine as well whether the consumption has more weight or less. GDP cannot be considered a determinant of consumption but a metric to understand the weight of expenditure in a economy.

4.2.2 Employment Rate

The Employment Rate is the ratio between the number of employed population and the number of people in age to work (15 - 64 years old) no matter if the last ones are not in a situation of looking for a job [3]. The employment rate is a sign of the state of the labour force of a location and as consequence the health of incomes in families.

Usually,the higher the employment rate is, the higher is the gross disposable income of the families and as it will be explained in the following section, there is a correlation between disposable income and consumption.

4.2.3 Gross Disposable Income (GDI)

Gross Disposable Income is the amount of money that the individuals in the household sector have available for spending or saving after taxes, social contributions, etc. have taken effect.[7]

Even though the Employment Rate and Gross Disposable Income can seem similar indexes it is necessary to select both of them because they can offer different points of view on the same matter. For example, the first one doesn't take into account different measures that are taken during scarcity and crisis periods like reductions of working hours or the reduction of supplements and incentives that affect consumption and expenditure.

At the macro level, disposable personal income is closely monitored as one of the key economic indicators used to gauge the overall state of the economy.

Even though that seems logical that the higher the income is the higher the expenditure will be a higher Gross Domestic Income doesn't always guarantee that it would transfer into consumption. Personal savings rates, marginal propensity to consume (MPC), and marginal propensity to save (MPS) are other indicators that must be taken into account when wanting to forecast the consumption of an individual or a household.[16] As it has been commented before when explaining the different functions of consumption the fear of a lower income in a future can affect consumption in a period where income has not suffered any dramatic downfall.

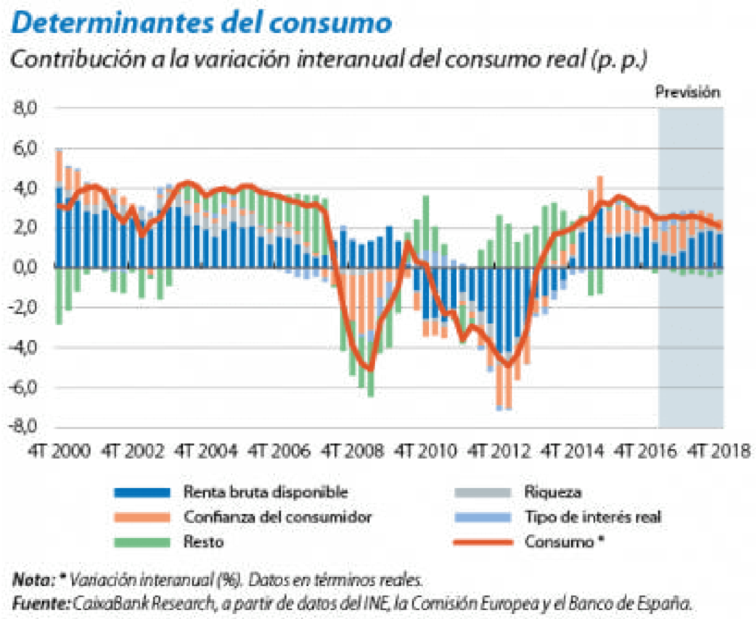

Figure 4.4: Evolution of consumption based on Disposable Income and Customers' Confidence

The Figure 4.4 illustrates the prior affirmation by showing the causes of the fall of consumption during the Financial Crisis of 2008 in Spain and the almost perfect correlation between disposable income and consumption. The first recession of consumption that started in late 2007 and was not caused by the reduction of the gross disposable income but due to the distrust of consumers against the markets, specially the housing market, that started collapsing. After the first recession and a small recuperation a second recession took place at 2010 when the effects of the crisis affected the average consumer lowering their wages and destroying jobs. It wasn't until 2013 when, thanks to the job creation and wage improvements that consumption recovered its levels pre-crisis.

4.2.4 Interest Rate

The interest rate is the amount charged on top of the principal[5] by a lender to a borrower for the use of assets.

The interest rates are issued by the central banks in order to favour the economy. Historically the interest rates were raised when there was a risk of a bubble[6] and were lowered when there was a necessity to heighten the economy.

The effect of low interest rates, the most common ones since the end of the 2008 Financial crisis, have had a direct effect on consumers' pockets. Low rates imply a reduction in the rate at which the European Central Bank lends money to banks, so they should offer lower interest rates to households and companies that want to finance themselves. Morever, those families who have their mortgages referenced to Euribor, which is the majority of families [9], will experience lower payments. With lower payments, households are able to save more and have more money to consume.

Low interest rates, however, are not beneficial for everyone. Low interest rates mean that banks won't give the customers returns for their savings. According to data offered by the Bank of Spain the average remuneration of demand deposits in Spain is 0.03%, compared to 0.04% for term deposits up to 1 year and 0.06% for term deposits of more than 2 years. These returns that do not outpace inflation[7] that stood at 1.5% on average back then. [9]

Another drawback that must be taken into account, even though is not very frequent, of low interest rates is the possibility of falling into a liquidity trap. A liquidity trap is the situation where consumers have no confidence that they can earn a higher rate of return by investing and they believe deflation in on the horizon. People and investors don't spend or invest. They believe goods and services will be cheaper tomorrow, so they wait to consume. Japan experienced a period of liquidity trap and stagnation from 1991 to 2001 labeled as Japan's lost decade. [17]

4.2.5 Harmonised Index of Consumer Prices (HICP)

Is an indicator offered by the Central Bank in order to provide a common measurement of inflation It is obtained as a result of homogenising the most significant methodological aspects of each one of the Consumer Price Indices from each one of the member states of the European Union to make them comparable. [5]

The Consumer Price Index (CPI) is calculated by taking the average weighted cost of a basket of goods in a given period and dividing it by the same basket from the previous period [1]. The HICP is divided in different selections of similar items that go from Food

[5]The principal term is most commonly used to refer to the original sum of money borrowed in a loan or put into an investment.

[6]An economic bubble is a situation in which asset prices are much higher than the underlying fundamentals can reasonably justify

[7]In economics, inflation refers to a general increase in prices of goods and services in an economy.

and beverages, Transport, Housing and electricity, etc. Depending on the country and its economy the weight of the selection varies in order to harmonize the indicator.[5]

In the case of the Eurostat data, the data source of this thesis, the comparisons are made with the cost of the basket of goods of 2015, in other words, the HICP data in 2015 for all the countries is always 100% and the other years are compared with that value.

A controlled inflation around 1.5-2% is considered as stability[5] but an inflation much higher than that can cause an erosion of purchasing power because goods become more expensive but at the same time can contribute to more investment and spending because money's value decreases.[4]

On opposite, there is deflation. It is a descend in the overall level of prices in an economy and an increase in the purchasing power of the currency. It can be caused by an increase on productivity due to technological improvements, a decrease of the aggregate demand[8], or by a decrease in the supply of money and credit.[?] Even though a deflation can have a good impact on consumers due to the increase of purchasing power, a rapid deflation is often linked with a contraction of economy.[?]

4.2.6 Purchasing Power

As it has been previously when explaining the consequences of a high inflation, the terms purchasing power and Consumer Price Index are very closed. The definition of purchasing power is the amount of goods or services that a unit of currency can buy at a given point in time.[6] It doesn't have be believed that CPI and Purchasing Power are exactly correlated and draw the same function but that one monitors the other. Moreover, Eurostat offered both metrics so both of them have been selected.

Purchasing power affects every aspect of economics, from consumers buying goods to a country's economic prosperity. When a currency's purchasing power decreases due to excessive inflation, serious negative economic consequences arise, including rising costs of goods and services contributing to a high cost of living.[6]

[8]Aggregate demand is a measurement of the total amount of demand for all finished goods and services produced in an economy

5. Analysis of the Consumption in EU Countries

5.1 Conclusions obtained from Rhizomer charts

In contrast to the prior chapter where the analysis was made according to the existing literature and theory about consumption and macroeconomics, this chapter will focus in the specific analysis of the same determinants analysed before in European Union countries during the last decade. The analysis will be done based on the chart visualizer developed in Rhizomer's software. The main goal of the analysis is to demonstrate with empirical cases the theories explained earlier.

The data used for the representation and following analysis is selected from Eurostat's database and from the Organization for Economic Co-operation and Development (OECD) database. Eurostat offers data related to all kind of topics in a user friendly way. The tables are classified in topics that go from Economy and Finance to Transport are stored as TSV files[1] and include information of a metric classified per country and per year, usually the period of time available is a range from 10 to 15 years prior the publication of the table. Due to the dates that this thesis is being made, first semester of 2022, the latest data available when regarding the indexes that are posted in an annual basis is 2020. In order to compare Household Consumption data with GDP per capita and Gross Disposable Income per capita, that are the indexes that Eurostat offers, it has been used the Individual Household Consumption data that OCDE offers.

Even though the Rhizomer application offers multiple charts to visualize in order to have different views of the same dataset, the charts chosen in order to analyze the data is the line chart and the table chart. By choosing the line chart, the evolution through the years and

[1] A tab-separated values (TSV) file is a simple text format for storing data in a tabular structure, e.g., a database table or spreadsheet data

comparison between different countries can be analysed easily and the table chart offers the possibility to sort the different indexes in order to see the best countries or years.

5.1.1 Correlations and Patterns

The general characteristics that can be obtained from looking to all the charts is that most of the countries follow the same pattern. Economy is interconnected and the tendencies are shared between EU members. For example consumption itself has a similar behaviour for almost all the countries, obviously there are some exceptions like Cyprus that is going to be commented later, but the line chart of the countries are almost parallel. The reasons of this homogeneous behaviour can be explained by referencing the different functions of consumption explained before when various authors like Modigliani exposed that consumption remains stable through years.

Even though that it has been said before that consumption charts show stability, all the countries experience a growth. This is because the value of money decreases in time when there is a positive inflation. Which is the case of all the countries.

Another index that transmits a feeling of homogeneity is purchasing power, the charts show a common behaviour and only in certain points this homogeneity is broken, for example in 2016 when United Kingdom suffers a downfall of the purchasing power probably due to the panic after the Brexit referendum's result.

On contrast, one index that varies from country to country are the employment rates (Figure 5.1). Employment depends on the power of a country to generate labour-force and in contrary to consumption it depends on the specific scenarios of the countries. What the charts illustrate is the differences between countries that historically have had low unemployment like Nordic Countries, Germany and Czechia among others and the those with higher unemployment rates such the Southern Countries, Spain, Portugal and Italy among others.

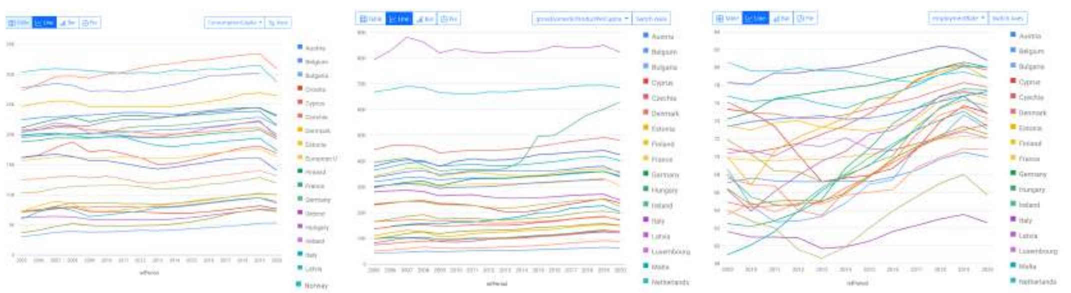

Figure 5.1: From left to right: Household Consumption, GDP per capita and Employment rate

Correlation between Consumption and Gross Disposable Income

One of the most noticeable correlations that can be easily detected is the correlation between Gross Disposable Income and Household Consumption. However, every country has its own correlation. In other words, the variation of income does not translate equally to consumption in all the countries nor all the years. The same rise in income in two different countries can suppose different rises in consumption.

The following figure (Figure 5.2) represent the situation expressed above and justify the consumption function theories explained in the prior chapter. Consumption relies mainly on income but there are as well other parameters related with it such us the propensity to consume and the propensity to save. In the Figure 5.2, it can be observed that even though the correlation is not perfect and depending on the countries it is stronger or weaker, changes in income affect consumption.

Income, however, cannot forecast consumption with a 100% accuracy. A good example is the 2020 data of consumption. Due to COVID-19 and the different measures for preventing the propagation of the virus, lock-downs, quarantines, curfews... consumption suffered a recession in most of the countries even though Gross Disposable Income did not. Later on this thesis, there is going to be a dedicated section to consumption during COVID pandemic and its relation with Disposable Income.

Figure 5.2: Correlation of Gross Disposable Income and Individual Household Consumption from 2005 to 2020 in (from left to right and top to bottom) Portugal, Italy, Cyprus and the European Union

Correlation between Consumption and GDP per capita

As it has been said before, household consumption is a very important part of the GDP of a country, the more household consumption a country has, the higher GDP per capita will have.

The next table, Table 5.1, shows 11 countries with the highest GDP per capita and the 11 countries with the highest Household Consumption of the EU in 2020. All the countries that appear in one list appear in the other one and all the countries except Ireland, whose case it is going to be explained later, occupy similar positions.

Rank.	GDP per Capita	Indiv. Household Consumption
#1	Luxembourg	Luxembourg
#2	Norway	Norway
#3	Ireland*	Denmark
#4	Denmark	Netherlands
#5	Sweden	Germany
#6	Netherlands	Austria
#7	Finland	Belgium
#8	Austria	Sweden
#9	Germany	Finland
#10	Belgium	France

Table 5.1: Top 10 countries with highest GDP per capita and Household Consumption
Source: Own elaboration with Eurostat data

The Figure 5.3 depicts the correlation between the two indexes. As it can be seen the correlation is even stronger in this case. That is because unlike Consumption and Income, where the relation is causal, the relation between Individual Household Consumption and GDP per capita is different. As it has been said before, Private Consumption is the largest component of GDP so changes on Consumption will affect on GDP.

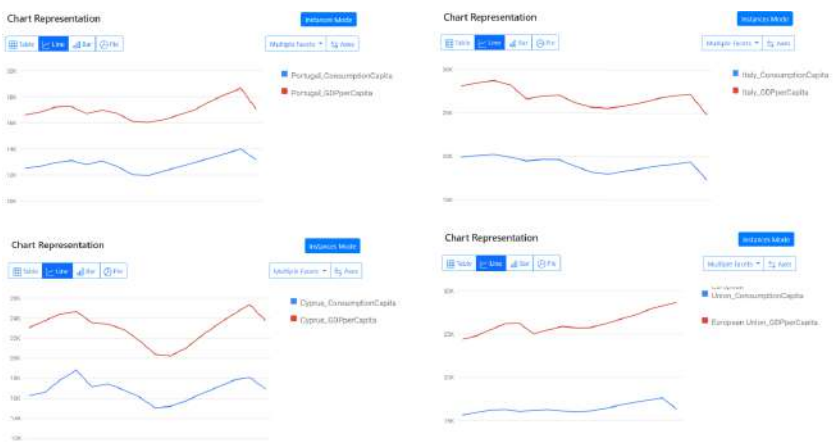

Figure 5.3: Correlation of GDP per capita and Individual Household Consumption from 2005 to 2020 in Portugal, Italy, Cyprus and the European Union

Correlation coefficient between Consumption and its determinants

The two determinants explained are the only ones that their correlation with Consumption can be depicted in a chart because the scales are the same. However, the other determinants, Harmonized Index of Consumer Prices (HICP), employment rate and real interest rate cannot be represented in the same chart with consumption because while consumption is in a monetary scale of thousands of euros, the other determinants are represented as percentages.

Thus, Pearson Correlation Coefficient (see following formula) has been used in order to determine the correlations between the determinants (x and y). When r is close to 1 means that both determinants are directly proportional and when it is close to -1 they are inversely proportional. When it is 0 it means that there is no correlation.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Relation	**r**
GDP per capita and Individual Household Consumption	0.97906
Gross Disposable Income and Individual Household Consumption	0.90885
Employment Rate and Individual Household Consumption	0.91227
Purchasing Power and Individual Household Consumption	0.78519
Interest rate (Cost of borrowing money) and Ind. Hous. Consumption	-0.6870
HICP and Individual Household Consumption	0.68364
HICP and Purchasing Power	0.90053
Gross Disposable Income and Employment rate	0.92802

Table 5.2: Pearson's Correlation Coefficient of studied determinants

As it could be seen in the prior charts, both GDP per capita and Gross Disposable income are very correlated with Consumption. Moreover, employment rate has a very high correlation as well. That's because employment rate have a clear direction with income, the highest is the number of people working the highest of people earning a salary will be.

However, there seems to a high correlation between HICP and consumption. And this makes completely sense. As it has been said in the past chapter, a controlled inflation is a good sign for economy but a too high consumption can be fatal. The available data of HICP used is from 2012 to 2020 so it has been a period of stability and controlled inflation.

Another sign of stability and controlled inflation is the fact that the correlation between HICP and Purchasing Power is close to 1. According to the definition offered before Purchasing Power is the amount of goods or services that a unit of currency can buy at a given point in time. So if there was a savage inflation purchasing power would decrease and the correlation (r) would be negative. Which is not the case.

Interests rates regarding the cost of borrowing money for households and Consumption have a negative correlation. That means that consumption have increased the most when interests have decreased. That is because after 2008 crisis most of central banks, including European Central Bank entered a low interest rate environment and that means more spending money in consumers' pockets.

5.1.2 Analysis of macroeconomic atypical situations

Even though the performance of the different EU countries in the different indexes seems similar because all the countries follow the same patterns and tendencies of the the global economy, there are countries' indexes that attract the attention of the user when analysing them due to breaking the named above global tendency. In this part of the thesis two special moments will be explained in order to analyse how consumption responded to the important rise of Ireland's GDP in 2015 and to the financial crisis of Cyprus during 2012 and the bank freeze.

Looking at the Figure 5.4, it can be seen that the two countries that have a special behaviour are Ireland and Cyprus. Ireland has a steep GDP rise in 2015 when the rest of the countries remain steady and Cyprus is the country that suffered the worst consumption recession during the period listed.

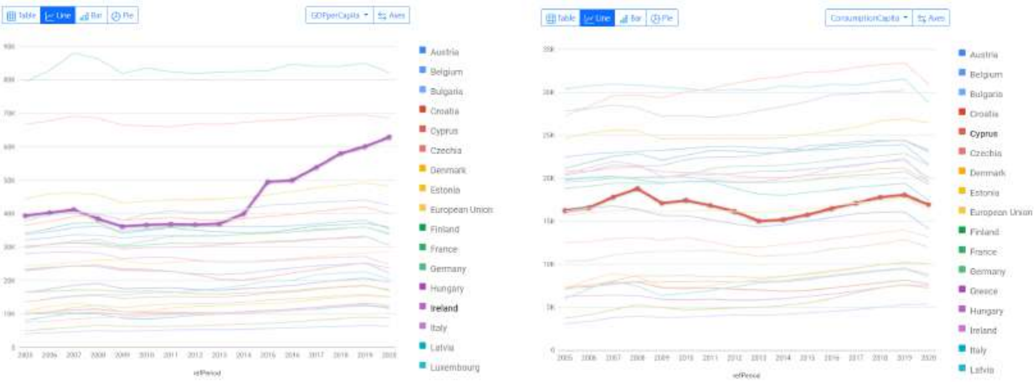

Figure 5.4: GDP of EU countries with Ireland highlighted (left) and consumption of EU countries with Cyprus highlighted (right)

How affected Consumption Ireland's GDP rise of 26% in 2015

One of the most noticeable aspects when looking the charts is the sudden rise of Ireland's GDP per capita from 2015 to the current times. In 2014 the country's GDP per capita was at 40.010€ and the following year, 2015, it raised up to 49.610€, a relative increase of 24%. Even though that the year before GDP had an increase of 8%, the prior 5 years, from 2008 to 2013, the chart line was almost flat, increasing only 2.2%. After 2015 there was not any year with increases as steep as the one in 2015 but the growth was constant with an ascent of 57.5% during the following six years. The most noticeable point is that meanwhile Ireland experienced all this growth the chart line of the other countries remained almost flat as it can bee seen in Figure 5.4.

As it has been said before, GDP is considered as the main indicator to determine the economy of a country. However, this situation raised the interest of many economist that questioned the reliability and the conceptual basis of the measurement of GDP. Did the population of Ireland, the individuals and the households, benefit from this rise?

By analyzing other metrics it can be determined what have been the consequences of this alleged economic improvement. Since 2009, when there was a consumption decline of 5.6%, the consumption chart experienced a slow but steady growth until 2015. Since that point the slope of the chart became steeper with a total increase in 5 years of 17.8%. However, even though that this evolution can be considered as very positive, it does not stand out in front of the other EU members because the other countries had similar growths. The same thing happens with the other indexes like purchasing power. The evolution is positive but it does not correspond to a rise of 26% of the Gross Domestic Product. As it can be seen in Figure 5.5.

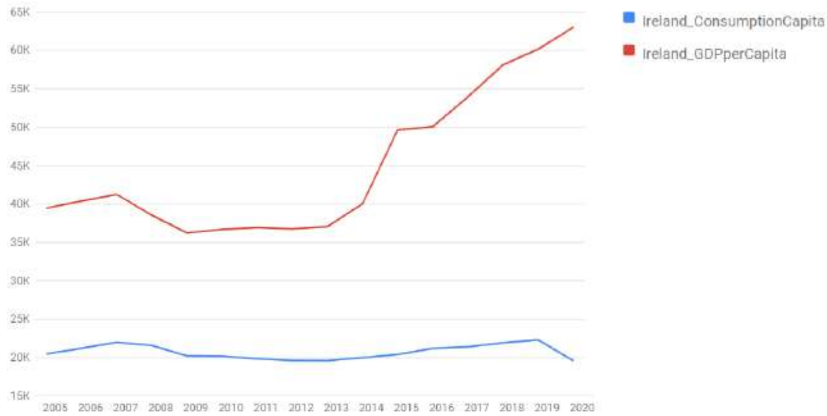

Figure 5.5: Ireland's GDP per capita and Individual Household Consumption

Not only the evolution of the consumption determinants falls short in comparison to the growth of the GDP but the ranking[2] of Ireland in indexes like the Gross Disposable Income, it occupies the 13th position, differs a lot of the position of Ireland in the GDP per capita, where it places 3rd overall, only after Luxembourg and Norway. On the other side, the rest of countries that occupy the top positions in the GDP index remain at the top in the Income charts as well.

The main reason of the 26% rise of GDP, not affected proportionally neither by a rise in consumption nor an increase in productivity, is attributed to an increase of the number of

[2]It is considered as ranking the position of a country in a list sorted from best to worst

large multinational corporations that have relocated their economic activities in Ireland due to the low corporation taxes. It is known, for example, that aircraft leasing company AerCap moved the domicile of its fleet of aircraft into Ireland, which has a total value of over €35 billion, during those years. Moreover, there have also been reports that Apple moved some of its valuable intellectual property assets to Ireland.[18]

Many economists have talked and written about this situation, criticising whether GDP adequately reflects the economic activity of a country. Paul Krugman, an American economist and Nobel prize winner, incubated the term *Leprechaun Economics* to describe that artificial GDP ascent.

How affected consumption 2012–2013 Cyprus' financial crisis

All the countries represented in the different charts show stability and uniformity but Cyprus breaks the rule. Meanwhile the rest of European countries were recovering from the recession caused by 2008 financial crisis, Cyprus' economy keep worsen. Prior to 2012 due to the consequences of the crisis in tourism and shipment employment rate decreased from 75% in 2010 to 67.3% in 2013.

Moreover its banking system was about to crash. Cyprus, that had a small economy, had amassed €22 billion of Greek private sector debt and was disproportionately affected by the measures taken by creditors, in connection with the Greek sovereign debt crisis. This situation not only increased the difficulty to credit access but caused that the European Union had to intervene with an economic rescue in 2012. On March 16, 2013, there was a freeze on the current accounts of Cypriots, with the aim of raising €5.8 billion through a write-off of bank deposits to pay the European bailout.

Even though these measures are used in order to save the banking system and avoid the bankrupt of the country it is inevitable that consumption, in a short term, is adversely affected. In the Cyprus case, even though the measures took place in the first semester of 2013, household consumption keep decreasing until 2014.

From 2015 onwards, the situation in the island improved, the employment increased and so the gross disposable income and the consumption. Thanks to the economical aid from the European Union and the recover of international tourism sector the situation improved and it got up to higher levels of pre-2008 crisis.

5.2 Other conclusions obtained from other indexes and charts

Eurostat's available data has a main drawback. It only offers data from 12 to 15 antiquity periods, some indexes just offer the last 12 months or 12 trimesters, so it is impossible have a picture that shows both the long-term evolution of an index and at the same time the evolution during a certain year. For the general analysis, data sources with annual periodicity have been chosen in order to have a broader picture to analyze.

However, due to the unique period of time when this thesis was written, first semester of 2022. Different sources of data have been chosen in order to explain the last two years more precisely.

5.2.1 Analysis of household consumption during COVID pandemic (March 2020 - June 2021)

During the whole thesis, it has been said that most of the time there is a correlation between disposable income and consumption. In other cases, it was an exceptional external factor that affected consumption without affecting the other determinants of consumption. In this section the goal will be similar, detect the relation between income and consumption prior to 2020 and how affected COVID.

This analysis could be very complete and be a separate thesis itself because its a complex matter with different situations that varied depending on the country and the stage of the pandemic. However, it is going to be more focused in the consequences of COVID in consumption and the causes of the decrease of from a general point of view.

As it can bee seen in the Figure 5.6 the situation before the pandemic started, described an absolute correlation between disposable income and consumption and saving rates did not vary much. Usually when saving rates do not vary means that is a period of stability because the consumers don't feel that they have to save for future uncertain times and they don't have to user their savings either in order to manage their lifestyle.

Figure 5.6: Year over-year change of household saving rate consumption and income
Source: Eurostat data

However, when COVID pandemic started in the first quarter of 2020 situation decontrolled and the correlation between income and consumption disappeared. The panic originated by the uncertainty of the moment and an additional difficulty for consumption, due to the lock-downs imposed in order to stop the virus spread did not allow consumers to consume as usual. Even though that income increased all the quarters of 2020 minus the second one, the hardest in terms of the pandemic, consumption always suffered a decrease.

Another thing it is worthy to notice is that the decrease of consumption was not the same. As it can be seen in the Figure 5.7, the main consumption purposes that suffered a decrease were the ones related with leisure activities and such us restaurants and recreational activities, which were forbidden in most of the countries during long periods of time, everything related with mobility and transport and retail businesses whose products are more likely to be bought in the physical shop rather than being bought online, like clothing and footwear. Countries and regions whose economy is more focused on this purposes got more harmed.

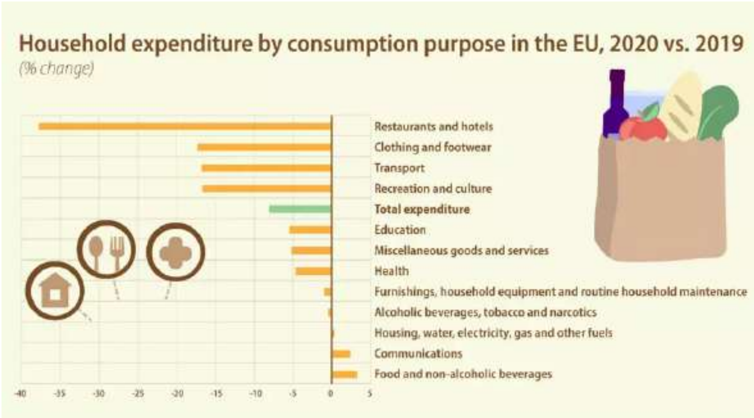

Figure 5.7: Household expenditure per purposes
Source: Eurostat data

5.2.2 Analysis and causes of post-COVID inflation's period (April 2021 - present)

If there is a word that summarizes all the economic headlines of the last 12 months is inflation. Since the COVID restrictions started to cease in most of the countries inflation started to increase.

It was originally caused by an excess of monetary mass and an excessive indebtedness due to the COVID-19 measures taken by governments and companies. Another cause is the drop in production capacity. Bottlenecks are occurring in the production of certain key intermediate goods and some raw materials. The Russian invasion to Ukraine and the later consequences have been another important cause for the constant inflation.

However, inflation has not been the same in all the products. In fact it has been some items that have dragged the overall inflation to this levels. As it has been said before when defining the HICP, it is divided in different items. European Central Bank (ECB) offers monthly data of all these items for all the countries so by analysing the charts displayed in their portal it would be possible to determine the principal causes. As it can be seen in the Figure 5.8, the inflation rate was lead by two components: Housing, electricity, gas and Transport which includes fuels, vehicles purchase and transport services among others.

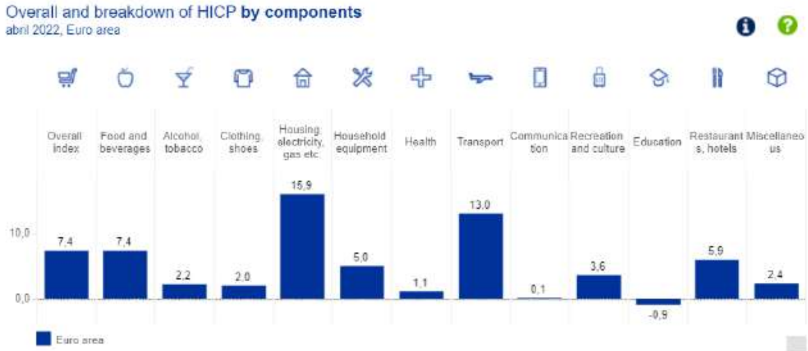

Figure 5.8: HICP per components
Source: ECB Data

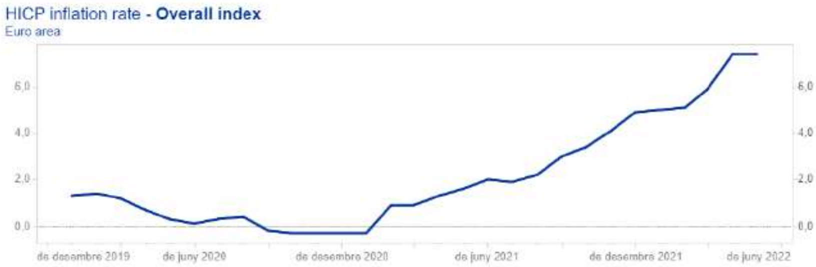

Figure 5.9: Evolution of HICP in the EU from 2019 to 2022
Source: ECB Data

Gas shortages across Europe, caused by a prolonged cold winter between 2020 and 2021, the high demand for liquefied natural gas from Asia, the delays and complications to the Nord Stream 2 pipeline, other bottleneck logistic problems due to COVID and the invasion of Ukraine by Russia, the biggest importer of gas and oil to Europe, have been the causes of the energy prices increase.[22] The increase of gas prices has a direct correlation with the rise of oil prices as well, the supply of which has not risen at the same level as its demand.

The Figure 5.10 depicts the real growth of consumption, income and employment from the first quarter of 2018 to the fourth quarter of 2021. The growth is calculated from the difference between the data from one quarter and the data from the quarter before. Moreover, inflation is included into the equation because both the income and the consumption growth are adjusted to inflation or deflation.

At first glance, what it can be seen is the major impact that COVID had. Before 2020, the fluctuations between quarters were minimal and regular. However, since the first quarter 2020 changes between one quarter to another one have been huge and it can be described as a chaotic situation.

Employment followed a regular pattern, growth during summer quarters and decreases during winter suffered the worst decrease of the listed years during the second quarter of 2020 and the first quarter of 2021. After this point, the situation improved quarter after quarter.

Consumption experienced two consecutive decreases in the first quarters 2020 diminishing more than 13%. The second quarter of 2020 it was the worst clearly of all the indicators (income, consumption and employment) decreased. Nevertheless, when there were strong drops, they were often followed by solid increases of consumption. That is because a lot of times the lack of consumption was not caused by the financial contraction but by the temporal lock-downs and restrictions.

Moreover, what is worth noticing is that in the last quarters of 2021 income decreased even though the the employment rate was increasing. That is because of the high inflation. As it has been said before, the growths are adjusted to inflation, so if the income growth is lower than the inflation the real income growth will be negative which is the case. It can be seen in the Figure 5.9 that the inflation was increasing steeply during this period.

Figure 5.10: Growth of Income, Consumption and Employment per quarters

6. Conclusions

After having seen the theory that is behind Household Consumption like the different consumption functions; that are basically theories made from a microeconomic approach because they describe decisions of an individual about their allocation of income, and having studied consumption from a macroeconomic approach, analyzing the behaviour of consumption in different countries based on different macroeconomic determinants such as GDP; Inflation and Gross Disposable Income among others, it can be determined that not only Household Consumption is the most important indicator in order to determine the economic health of a region but it is the best way, always comparing with other indicators, to forecast the near future.

For being able to analyze and extract all these conclusions it is necessary to have a powerful tool to represent all the data in charts that allows comparisons in an easy way. There will be cases like the increase of Ireland's GDP that requires external information such us the investments of multinational companies to that country but in most of the cases by selecting the right information and displaying the right charts, clear conclusions can be made easily.

Rhizomer's Chart Visualizer was not developed with the main goal of representing macroeconomic data but of representing the maximum number of datasets in an usable way. However, its Chart Representation option has features that neither Eurostat or European Central Bank data browsers have available. Representing multiple indexes in the same chart, filtering and switching the index represented in just two clicks. Moreover, Rhizomer's main goal is the exploration of semantic and linked data so a wide range of possibilities open from the side as well. The results speak for themselves and it has been demonstrated that some conclusions can be extracted. One of the thesis goals was to demonstrate with empirical data the theory of consumption and it has been achieved in a certain way.

The visualization of the charts has helped to demonstrate the theory of explained in the fourth chapter as it has been demonstrated during the analysis. By having the possibility to compare the data of each country, extracting correlations and having all the indexes in the same dashboard conclusions can be extracted easily and the causes of the variations in the trend can be determined. However, charts cannot give to the user all the inputs required to analyze the whole picture. Like the case of Ireland or the last years with the COVID pandemic a general knowledge of current affairs is necessary.

A dilemma could open right now asking whether if it is better to use Data Visualization applications like PowerBI, Tableau or Looker which offer the possibility to create dashboards with a high level of personality but with the limits of the software and with a User Interface that allows the user to create applications without being an expert or developing a less personalized chart visualizer in a data semantic faceted browser with unlimited opportunities that requires both a lot of time and technical skills. The answer is, obviously, "depends".

It depends on what are the user's goal and requirements. If the user only wants to analyze certain numerical data and what prioritizes is having a large library of charts with a high level of personalizing, then using a Data Visualization/Business Intelligence tool is the right option. If the user prefers visualizing a much more complete dataset with information that goes beyond numerical values for example; a dataset that describes the economic state of the world and not only includes numerical information but lists and links data from economic events, scientific papers or rating agencies, then using Rhizomer would be the right choice.

Nevertheless, whichever option is more suitable for the user, what it is clear is that when working with big quantities of data the best way of guaranteeing the user a good experience is developing a system that follows the Schneiderman's mantra: Overview first, allow zooming and filtering second and finally offer the details on demand third.

Bibliography

[1] David Lavie and Benjamin Curry, 3rd January 2022 *How to calculate inflation*. Forbes (Accessed on April 2022), <https://www.forbes.com/advisor/investing/how-to-calculate-inflation/>

[2] Fernando J., 2022 *Gross Domestic Product (GDP)*. Investopedia (Accessed on March 2022), <https://www.investopedia.com/terms/g/gdp.asp>

[3] Anonymous, 2022 *Employment rate*. OECD iLibrary (Accessed on March 2022), <https://www.oecd-ilibrary.org/employment/employment-rate/indicator/english_1de68a9b-en>

[4] David Floyd, 17th May 2021 *9 common effects on inlation*. Investopedia (Accessed on April 2022), <https://www.investopedia.com/articles/insights/122016/9-common-effects-inflation.asp>

[5] Anonymous, 21st April 2022 *Measuring inflation – the Harmonised Index of Consumer Prices (HICP)*. European Central Bank (Accessed on April 2022), <https://www.ecb.europa.eu/stats/macroeconomic_and_sectoral/hicp/html/index.en.html>

[6] Adam Hayes, 14th March 2021 *Purchasing Power*. Investopedia (Accessed on April 2022), <https://www.investopedia.com/terms/p/purchasingpower.asp>

[7] Anonymus, *Gross disposable household income*. Office for National Statistics (Accessed on April 2022), <https://www.ons.gov.uk/economy/regionalaccounts/grossdisposablehouseholdincome>

[8] Esen, Ömer; Bayrak, Metin, 12 June 2017 *Does more energy consumption support economic growth in net energy-importing countries?*. Journal of Economics, Finance and Administrative Science (Accessed on April 2022), 22 (42): 75–98. doi:10.1108/JEFAS-01-2017-0015. S2CID 55359390.

[9] Anonymus, 8 February 2019 *El efecto de los tipos de interés en tu bolsillo*. Cinco Dias - El Pais (Accessed on April 2022),
<https://cincodias.elpais.com/cincodias/2019/02/08/abante_asesores/1549621570_204596.html>

[10] Barry Nielsen, 14 January 2022 *The Lost Decade: Lessons From Japan's Real Estate Crisis*. Investopedia (Accessed on April 2022),
<https://www.investopedia.com/articles/economics/08/japan-1990s-credit-crunch-liquidity-trap.asp>

[11] Julia Kagan, 27 November 2020 *Permanent Income Hypothesis*. Investopedia (Accessed on April 2022),
<https://www.investopedia.com/terms/p/permanent-income-hypothesis.asp>

[12] Anonymus, Undated *Personal Consumption Expenditures*. HigherRock Education (Accessed on April 2022),
<https://www.higherrockeducation.org/glossary-of-terms/personal-consumption-expenditures-pce->

[13] Anonymus, Undated *Consumption as percent of GDP by country: the latest data*. The Global Economy (Accessed on April 2022),
<https://www.theglobaleconomy.com/rankings/consumption_GDP/>

[14] Ekrem Eddy Güzeldere, 27 July 2021 *Consumption and crisis*. Duvar English (Accessed on April 2022),
<https://www.duvarenglish.com/consumption-and-crisis-article-58310>

[15] The Investopedia Team, 22 January 2022, *Marginal Propensity to Consume (MPC)*. Investopedia (Accessed on May 2022),
<https://www.investopedia.com/terms/m/marginalpropensitytoconsume.asp>

[16] J. Singh, Undated *Mathematical Relationship between MPC and MPS*. Economics Discussion (Accessed on May 2022),
<https://www.economicsdiscussion.net/income/mathematical-relationship-between-mpc-and-mps-micro-economics/673>

[17] Barry Nielsen, 14 January 2022 *Mathematical Relationship between MPC and MPS*. Investopedia (Accessed on April 2022),
<https://www.investopedia.com/articles/economics/08/japan-1990s-credit-crunch-liquidity-trap.asp>

[18] Cliff Taylor, 15 July 2016 *Ireland's GDP figures: Why 26% economic growth is a problem*. The Irish Times (Accessed on May 2022),
<https://www.irishtimes.com/business/economy/ireland-s-gdp-figures-why-26-economic-growth-is-a-problem-1.2722170>

[19] Berners-Lee, Tim; Fischetti, Mark (1999). *Weaving the Web* HarperSanFrancisco chapter 12. ISBN 978-0-06-251587-2.

[20] Anonymus, Undated *The WWW and Web 2.0.* Penn State University (Accessed on March 2022),
<https://psu.pb.unizin.org/ist110/chapter/1-6-the-www-and-web-2-0/>

[21] Ora Lassila, 05 January 1999 *Resource Description Framework Model and Syntax Specification.* W3C (Accessed on March 2022),
<https://www.w3.org/TR/PR-rdf-syntax/Overview.html>

[22] Les Roberts, 27 April 2022, *Why are energy prices rising?* Bionic, (Accessed on May 2022),
<https://bionic.co.uk/blog/noticed-energy-prices-have-fallen-over-past-few-months/>

[23] Brooke Fitzgerald, 26 February 2016, *Schneiderman's Mantra* HAMPDATAVISUALIZATION, (Accessed on April 2022),
<https://hampdatavisualization.wordpress.com/2016/02/26/schneidermans-mantra/>

[24] Uriel Fernández, 22 February 2015, *MVC (Model, View, Controller) Explicado* Codigo Facilito, (Accessed on April 2022),
<https://codigofacilito.com/articulos/mvc-model-view-controller-explicado>