



GRADO EN INGENIERÍA EN TECNOLOGÍAS DE TELECOMUNICACIÓN

TRABAJO FIN DE GRADO
A censorship-resistant inflation feed

Autor: Juan Caminos Colmenarejo

Director: Pablo Zulaica Pérez

Madrid

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título

A censorship-resistant inflation feed

en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el

curso académico 2021/22 es de mi autoría, original e inédito y

no ha sido presentado con anterioridad a otros efectos.

El Proyecto no es plagio de otro, ni total ni parcialmente y la información que ha sido

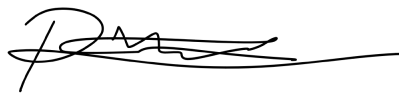
tomada de otros documentos está debidamente referenciada.

Fdo.: Juan Caminos Colmenarejo

Fecha: 05 / 06 / 2022

Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO



Fdo.: Pablo Zulaica Pérez

Fecha: 06 / 06 / 2022



GRADO EN INGENIERÍA EN TECNOLOGÍAS DE TELECOMUNICACIÓN

TRABAJO FIN DE GRADO
A censorship-resistant inflation feed

Autor: Juan Caminos Colmenarejo

Director: Pablo Zulaica Pérez

Madrid

Agradecimientos

A mis padres por darme la oportunidad de formarme profesionalmente, enseñarme sus valores y ayudarme en lo que han podido durante este trabajo de fin de grado y durante mi desarrollo académico.

A Pablo Zulaica, por guiarme durante el proyecto, ayudarme siempre que lo necesitaba y por todo lo que me ha enseñado a lo largo del año.

A CENSORSHIP-RESISTANT INFLATION FEED

Autor: Caminos Colmenarejo, Juan.

Director: Zulaica Pérez, Pablo.

Entidad Colaboradora: ICAI – Universidad Pontificia Comillas

RESUMEN DEL PROYECTO

Se ha estudiado la metodología actual, usada por el Gobierno, que se lleva a cabo para medir la inflación en España y se ha comparado con la métrica que usan otros países y las recomendaciones del FMI. Se identificaron cambios que no estaban justificados y aspectos de mejora. Mediante técnicas de web scraping, se monitorizaron precios online y se elaboró una métrica independiente con la que se obtuvieron unos resultados superiores a los publicados por el Gobierno. También se desarrolló una página web para mostrar los resultados.

Palabras clave: Economía, Inflación, Programación, Web scraping.

1. Introducción

La inflación se define como un aumento general de los precios de bienes y servicios en una economía durante un periodo determinado de tiempo. En España, la estadística más utilizada para medir la inflación es el índice de precios al consumidor (IPC) publicado mensualmente por el Instituto Nacional de Estadística (INE).

¿El índice publicado representa el aumento real de nuestro coste de vida? Verificaremos ese hecho mediante dos puntos de vista, uno metodológico y uno práctico.

2. Definición del proyecto

En el análisis metodológico, estudiaremos cómo se calcula el IPC en España utilizando los documentos oficiales que publica el Instituto Nacional de Estadística y podremos comprobar si la métrica actual es razonable o no. También estudiaremos el documento que emite cada año el Fondo Monetario Internacional (FMI) en el que define cómo debe ser una correcta métrica de la inflación. Por otro lado, analizaremos las metodologías que siguen otros países tanto de la Unión Europea, como Alemania o Francia, y de fuera de ella como Estados Unidos, Japón y China.

Por otro lado, en el análisis práctico, utilizaremos tecnologías de web scraping para monitorear los precios online todos los días e informar del coste real no ajustado y la

fluctuación del precio de los artículos en los que generalmente gastamos nuestros ingresos después de impuestos, obteniendo una métrica independiente. Estos resultados se mostrarán en una página web.

3. Descripción del sistema

La página web se ha desarrollado siguiendo la arquitectura cliente-servidor. Dispone de un front-end que establece una conexión directa a través de una red al back-end. Este back-end lo componen un servidor que contiene el programa que se conecta al servidor de la base de datos que contiene toda la información de los productos y sus precios.

Dentro de la arquitectura cliente-servidor, se ha usado la arquitectura de tres capas. Son el front-end, el back-end y el middle-end. En el front-end se encuentra la lógica de presentación que interactúa con el usuario, en el back-end se encuentra la base de datos y en el middle-end se encuentra la lógica de la métrica.

Usando técnicas de web scraping se obtienen los precios online y se almacenan en el servidor de la base de datos. Los datos se procesan en el servidor intermedio y se realiza el cálculo de la nueva métrica. Los resultados del servidor intermedio se muestran en el front-end a través de una página web.

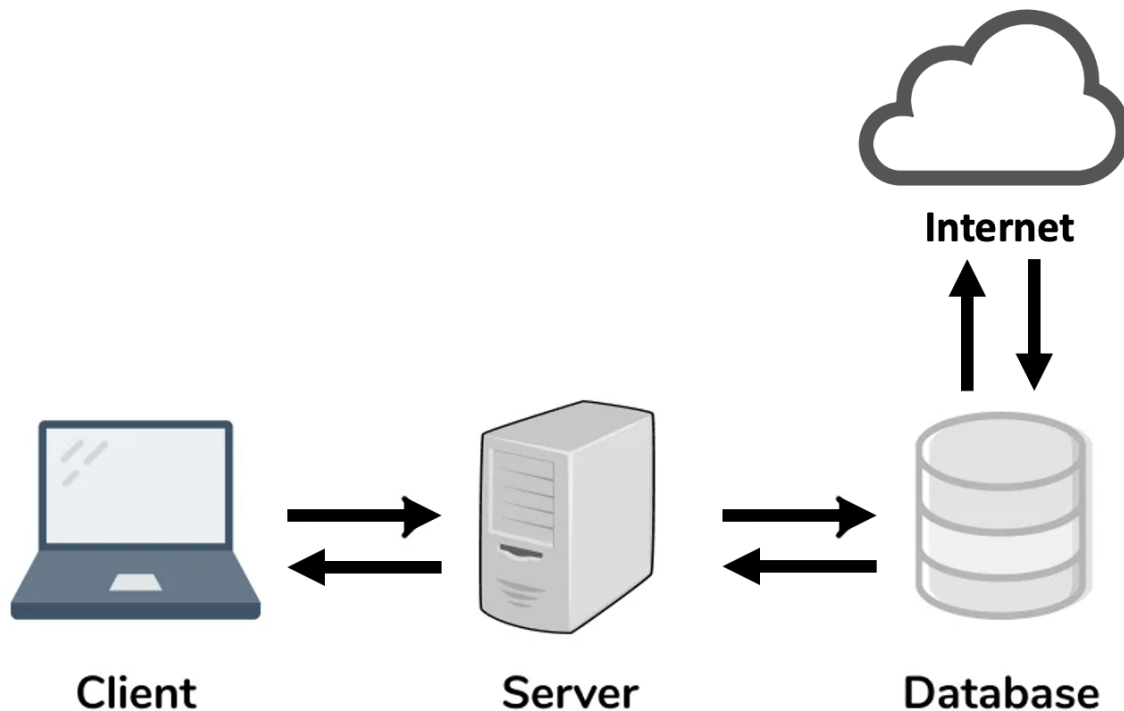


Image 1 - Esquema de la arquitectura de la aplicación web

4. Resultados

Los resultados obtenidos del cálculo independiente son superiores a los de la métrica actual que usa el Gobierno para medir la inflación. A continuación, se muestra una parte de la página web donde se comparan ambos resultados.

May 2022

Groups	Official Inflation	Real Inflation	Difference
General Index	4,1	5,2	1,1
Alimentos y bebidas no alcohólicas	7,5	8,1	0,6
Bebidas alcohólicas y tabaco	2,3	3,1	0,8
Vestido y calzado	-0,9	0,1	1,0
Vivienda	0,9	2,9	2,0
Menaje	4,4	4,4	0,0
Sanidad	0,6	0,6	0,0
Transporte	9,8	11,5	1,7
Comunicaciones	0,5	0,3	-0,2
Ocio y cultura	0,3	-0,9	-0,6
Enseñanza	0,0	0,0	0,0
Hoteles, cafés y restaurantes	4,2	6,4	2,2
Otros bienes y servicios	2,3	4,1	1,8

Source: [INE](#).

Spanish Inflation - May 2022

Image 2 - Comparación de resultados

Los resultados muestran la diferencia entre los índices de cada categoría y del índice general de la inflación. Esto confirma que el análisis de deficiencias realizado en el proyecto, del cual se dedujo que la métrica oficial mostraba resultados inferiores a los reales, es correcto o al menos una parte de él.

5. Conclusiones

A medida que ha ido evolucionando la metodología del Índice de Precios al Consumidor en España, muchos de los cambios que se han introducido en la métrica no han sido justificados y han provocado resultados inferiores al real. Especialmente desde la armonización de las metodologías de la Unión Europea.

Estos cambios suelen pasar desapercibidos y, a primera vista, no parecen producir cambios muy significativos en el resultado. Sin embargo, la suma de todos los cambios implantados a lo largo de los años se traduce en un resultado muy por debajo del real.

Las principales áreas de mejora que se han detectado en este proyecto han sido: la exclusión de gran parte del mercado de coches de segunda mano; la ponderación de los productos de la cesta no refleja los hábitos de consumo reales; la métrica no es representativa del coste de vida de todas las clases sociales, está sesgada; los cambios de calidad y sustitución de los bienes de la cesta; la unidad de medida de la métrica no es estable en el tiempo, por lo que se desvirtúan los resultados; y la falta de transparencia en diferentes áreas de la métrica.

Por otro lado, los resultados confirman que la métrica actual, usando las mismas ponderaciones, subestima la inflación. Casi todos los índices calculados fueron superiores a los oficiales, hasta 2.2 puntos, y el índice general de la inflación calculado era 1.1 puntos superior.

Sería más razonable tener una métrica personalizable para que cada persona pudiese tener su índice de inflación propio. Esto se debe a que cada persona gasta su dinero en lo que le gusta y la gente tiene distintos hábitos de consumo. Al menos, debería estar dividida por grupos sociales porque, como se ha demostrado en el capítulo Parte I4.3.2.4, la inflación afecta de forma diferente a las clases sociales y la métrica oficial trata a todo el mundo de la misma forma.

Disponer de una métrica independiente no controlada por el Gobierno es importante para evitar posibles censuras o manipulaciones del resultado que favorezcan sus predicciones, y porque puede haber ciertos hábitos de consumo que no se estén teniendo en cuenta ya sea porque son recientes o porque se eliminaron de la métrica por normativa de la Unión Europea.

A CENSORSHIP-RESISTANT INFLATION FEED

Author: Caminos Colmenarejo, Juan.

Supervisor: Zulaica Pérez, Pablo.

Collaborating Entity: ICAI – Universidad Pontificia Comillas

ABSTRACT

The Spanish methodology to measure inflation has been studied and compared with other countries' metric and with the IMF's recommendations. Unjustified changes have been identified as well as aspects that could be improved. Using web scraping techniques, online prices were tracked, and a new independent metric was created obtaining results higher than the published by the Government. A website was also developed to show the results.

Keywords: Economy, Inflation, Programming, Web scraping.

1. Introduction

Inflation is defined as a rise in the general level of prices of goods and services in an economy over a given period. In Spain, the most commonly used statistic for measuring inflation is the consumer price index ("CPI") published by the Instituto Nacional de Estadística (INE).

Does the published index represent the real increase in our cost of living? We'll check that fact from a theoretical and a practical standpoint.

2. Project definition

For the theoretical analysis, we will study how the CPI is calculated in Spain using the official documents published by the Instituto Nacional de Estadística and we will check if the metric is reasonable or not. We will also study the document published by the International Monetary Fund (IMF) in which it states guidelines which every inflation metric should follow to be correct. Moreover, we will analyze other countries' metric from the European Union, like Germany or France, and others like the United States of America, Japan, and China.

On the other side, for the practical analysis, we will use web scraping techniques to monitor online prices every day and report the unadjusted actual cost and price

fluctuation of the items on which we usually spend our after-tax income, obtaining an censorship-resistant inflation feed. These results will be published on a website.

3. System's description

The website has been developed following the client-server architecture. It has a front-end which establishes a connection to the back-end via a network. This back-end has a server which contains the program which connects to the database server which contains all the information of the products and their prices.

Inside the client-server architecture, we have decided to use the three-tier architecture. These three tiers are the front-end, the middle-end and the back-end. The front-end has the presentation logic and interacts with the user, the back-end stores the data and the middle-end contains the business logic of the application.

Using web scraping techniques online prices are obtained and them stored in the database server. The data is processed by the middle-end server and calculates the new metric. Then these results are shown in the front-end through a website

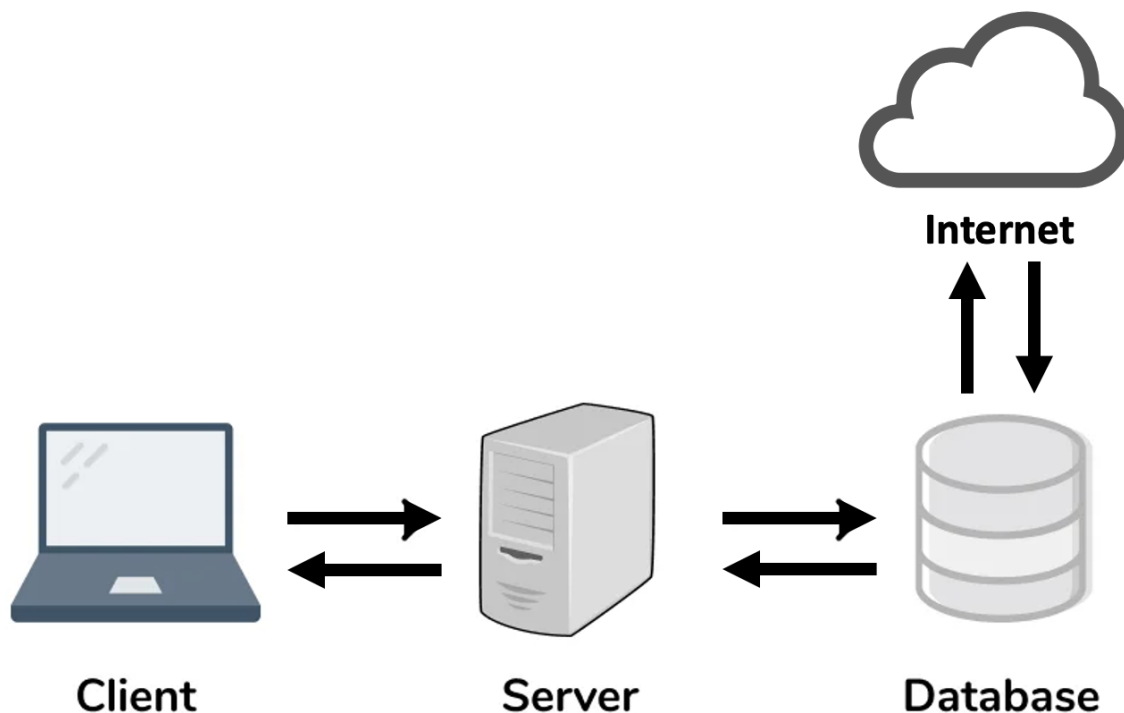


Image 3 - Architecture scheme of the web application

4. Results

The results obtained from the independent metric are higher than the ones published by the Government to measure inflation. Next, a part of the website where both results are compared is shown.

May 2022

Groups	Official Inflation	Real Inflation	Difference
General Index	4,1	5,2	1,1
Alimentos y bebidas no alcohólicas	7,5	8,1	0,6
Bebidas alcohólicas y tabaco	2,3	3,1	0,8
Vestido y calzado	-0,9	0,1	1,0
Vivienda	0,9	2,9	2,0
Menaje	4,4	4,4	0,0
Sanidad	0,6	0,6	0,0
Transporte	9,8	11,5	1,7
Comunicaciones	0,5	0,3	-0,2
Ocio y cultura	0,3	-0,9	-0,6
Enseñanza	0,0	0,0	0,0
Hoteles, cafés y restaurantes	4,2	6,4	2,2
Otros bienes y servicios	2,3	4,1	1,8

Source: [INE](#).

Spanish Inflation - May 2022

Image 4 - Results comparison

The results show a big difference in the indices and the general inflation rate. This confirms that the analysis of areas of improvement made in the project, from which the suspicion that the official metric understates inflation was originated, is entirely or at least partially correct.

5. Conclusions

As the methodology of the Consumer Price Index has evolved, in Spain, a lot of the metric changes have not been justified and, as a result, the value obtained is lower than the real one. Specially since the harmonization of the methodologies of the European Union.

These changes usually go unnoticed and, at first, they do not seem to produce any relevant change in the final result. However, the sum of all these changes implemented throughout the years translate into an outcome lower than the real one.

The main areas of improvement that have been detected in this project are: a great part of the used cars market is excluded, the weighting of the products of the basket does not show the real consumption habits, the metric is not representative of the cost of living of all social classes, it is biased, quality adjustments and substitution of products of the basket, the unit of measurement is not stable in time, so the results are biased, and the lack of transparency in different areas of the metric.

Moreover, the results confirm that the actual metric understates inflation even using the same weights which understate inflation. Almost every index calculated was higher than the official ones, up to 2.2 points, and the general inflation obtained was 1.1 points higher.

It would be better to have customizable metrics so anyone can have their own inflation rate. This is because everyone spends their money on what they like and people have different consumption habits. At least it should be divided by social groups because, as proven in chapter Parte I4.3.2.4, inflation affects social classes differently and the official inflation methodology treats everyone the same.

Having an independent, censorship-resistant methodology is important in order to avoid certain manipulations that favor Government's predictions. In addition, it is also important because some consumption habits may not be taking into account due to them being recent or because they were eliminated by regulations of the European Union.

Index

Chapter 1. Introduction	7
1.1 Project motivation	7
1.2 Project objectives	8
1.3 Work methodology	8
Chapter 2. Technology description	10
2.1 HTTP Protocols	10
2.2 REST API.....	12
2.3 Spring boot	14
2.4 Model View Controller	15
2.5 MySQL.....	16
2.6 Web scraping.....	18
2.7 Postman	19
Chapter 3. State of the art	21
3.1 Inflación verdadera Venezuela	21
3.2 Inflación verdadera Argentina.....	26
Chapter 4. Theoretical Analysis	30
4.1.1 Inflation Analysis.....	30
4.1.2 What is inflation?	30
4.1.3 What causes inflation?	31
4.1.4 What are the consequences of inflation?	33
4.1.5 How is inflation controlled?	34
4.2 Methodology Analysis	36
4.2.1 How is inflation measured in Spain?.....	36
4.2.2 How has the methodology evolved in recent years and what has caused the changes?	40
4.2.3 What are the differences between the Spanish and other countries' methodologies?....	58
4.2.4 Is the Spanish methodology aligned with the IMF's recommendations?	61
4.3 Metric Analysis	63
4.3.1 What are the main uses of the inflation metric in Spain?.....	63
4.3.2 Areas of improvement.....	64

4.3.3 Is the CPI a good metric to measure the cost of living in general terms?.....	73
Chapter 5. Practical Application.....	75
5.1 Independent methodology definition.....	75
5.2 Conceptual and technological design	76
5.3 Implementation.....	78
5.3.1 Web scraping module development	78
5.3.2 Back-end development.....	81
5.3.3 Middle-end development	85
5.3.4 Front-end development.....	87
5.4 How to make a censorship-resistant metric?.....	90
Chapter 6. Results Analysis.....	92
Chapter 7. Conclusions	94
Chapter 8. Future development.....	96
Chapter 9. Bibliography	98
ANEXO I: ALINEACIÓN DEL PROYECTO CON LOS ODS	103

Images Index

Image 1 - Esquema de la arquitectura de la aplicación web.....	9
Image 2 - Comparación de resultados	10
Image 3 - Architecture scheme of the web application	13
Image 4 - Results comparison	14
Image 5 - US M1 Money supply over the recent years (https://themarketear.com/)	7
Image 6 - HTTP Response Example (http://www.tcipguide.com/free/t_HTTPResponseMessageFormat.htm).....	12
Image 7 - REST API Model (https://appmaster.io/es/blog/que-es-la-api-rest-y-en-que-se-diferencia-de-otros-tipos)	13
Image 8 - Spring Boot Logo (https://cursosdedesarrollo.com/2022/03/spring-boot-con-live-reload-en-intellij/).....	15
Image 9 - Model View Controller Architecture (https://developer.mozilla.org/es/docs/Glossary/MVC).....	16
Image 10 - MySQL Logo (https://itsoftware.com.co/content/que-es-y-para-que-sirve-mysql/).....	17
Image 11 - Web Scraping Procedure (https://medium.com/@Emmitta/web-scraping-7f87930face4).....	18
Image 12 - Postman Logo (https://chicadev.medium.com/actualizando-postman-en-ubuntu-b5a58db0579a)	19
Image 13 - Postman Example.....	20
Image 14 - Inflacion Verdadera Venezuela Mobile App Display (http://www.inflacionverdadera.com/venezuela/ayudar/)	23
Image 15 - Inflacion Verdadera Venezuela Mobile App Add a Product (http://www.inflacionverdadera.com/venezuela/ayudar/)	23
Image 16 - Inflacion Verdadera Venezuela Data Validation (http://www.inflacionverdadera.com/venezuela/validacion-de-precios/)	24

Image 17 - Inflacion Verdadera Venezuela Food Category Monthly Variation (http://www.inflacionverdadera.com/venezuela/)	25
Image 18 - Inflacion Verdadera Venezuela CPI by Category (http://www.inflacionverdadera.com/venezuela/)	25
Image 19 - Inflacion Verdadera Argentina Online and Offline price indexes in four Latin American Countries (https://www.hbs.edu/ris/Publication%20Files/Cavallo_Alberto_Online%20and%20Official%20Price%20Indexes%20Measuring%20Argentinas%20Inflation_600c0e8f-cc57-4)....	27
Image 20 - Inflacion Verdadera Argentina Online vs Official series (https://www.hbs.edu/ris/Publication%20Files/Cavallo_Alberto_Online%20and%20Official%20Price%20Indexes%20Measuring%20Argentinas%20Inflation_600c0e8f-cc57-430d-b869-a650dda2b290.pdf).....	27
Image 21 - Inflacion Verdadera Argentina Price Index, Annual Inflation Rate and Monthly Inflation Rate (https://www.hbs.edu/ris/Publication%20Files/Cavallo_Alberto_Online%20and%20Official%20Price%20Indexes%20Measuring%20Argentinas%20Inflation_600c0e8f-cc57-430d)	28
Image 22 - Phillips Curve (https://www.researchgate.net/figure/The-modified-Phillips-curve_fig4_342734687)	32
Image 23 - M2 money supply of the US (https://fred.stlouisfed.org/series/M2)	33
Image 24 - Paul Volcker's years as president of the FED. Inflation vs Interest Rates vs Unemployment Rate (https://worldbuilding.stackexchange.com/questions/224117/why-would-my-central-bank-suddenly-increase-the-interest-rates-to-over-20)	35
Image 25 - CPI Group Distribution and amount of Articles that belong to them (https://www.ine.es/metodologia/t25/t2530138_16.pdf).....	37
Image 26 - Spanish CPI Demographic Breakdown for which data is published monthly (https://www.ine.es/metodologia/t25/t2530138_16.pdf).....	38
Image 27 - Time Period of each Spanish CPI system (https://www.ine.es/metodologia/t25/t2530138_16.pdf).....	42

Image 28 - Evolution of number of items through the bases (https://www.ine.es/metodologia/t25/t2530138_16.pdf)	57
Image 29 - Wages growth vs Inflation (https://www.bbc.com/news/business-44857055). 63	
Image 30 - Chained CPI vs CPI	64
Image 31 - Housing Inflation and CPI compared (1987-2003). Source:.....	67
Image 32 - Housing Inflation and CPI compared (1998-2003). Source:.....	68
Image 33 - Housing Prices compared to Housing category CPI (1993-2000). Source:	69
Image 34 - Expexted Inflation in One Year by Income Bracket. Source: University of Michigan.....	70
Image 35 - Architecture of the web application. Own elaboration	77
Image 36 - MySQL database table with raw data. Own elaboration.....	81
Image 37 - Test of the Web application API. Own elaboration	85
Image 38 - Web Application Home Page.....	87
Image 39 - Web Application Comparison Page	88
Image 40 - Web Application Comparison Page	89
Image 41 - Results comparison and difference between them	92

Tables Index

Table 1 - HTTP Methods and Operations	11
Table 2 - Spanish CPI 1936 Base. Categories and Weights. Own elaboration	41
Table 3 - Spanish CPI 1936 Base. Weights distribution. Pie Chart. Own elaboration.....	41
Table 4 - Spanish CPI 1958 Base. Categories and Weights. Own elaboration	43
Table 5 - Spanish CPI 1958 Base. Weights distribution. Pie Chart. Own elaboration.....	43
Table 6 - CPI 1968 Base. Categories and Weights. Own elaboration.....	44
Table 7 Spanish CPI 1968 Base. Weights distribution. Pie Chart. Own elaboration.....	44
Table 8 - CPI 1976 Base. Categories and Weights. Own elaboration.....	45
Table 9 - Spanish CPI 1976 Base. Weights distribution. Pie Chart. Own elaboration.....	45
Table 10 - CPI 1983 Base. Categories and Weights. Own elaboration.....	46
Table 11 - Spanish CPI 1983 Base. Weights distribution. Pie Chart. Own elaboration.....	46
Table 12 - CPI 1992 Base. Categories and Weights. Own elaboration.....	47
Table 13 - Spanish CPI 1992 Base. Weights distribution. Pie Chart. Own elaboration.....	48
Table 14 - CPI 2001 Base. Categories and Weights. Own elaboration.....	49
Table 15 - Spanish CPI 2001 Base. Weights distribution. Pie Chart. Own elaboration.....	50
Table 16 - CPI 2006 Base. Categories and Weights. Own elaboration.....	51
Table 17 - Spanish CPI 2006 Base. Weights distribution. Pie Chart. Own elaboration.....	51
Table 18 - CPI 2011 Base. Categories and Weights. Own elaboration.....	52
Table 19 - Spanish CPI 2011 Base. Weights distribution. Pie Chart. Own elaboration.....	53
Table 20 - CPI 2016 Base. Categories and Weights. Own elaboration.....	54
Table 21 - Spanish CPI 2016 Base. Weights distribution. Pie Chart. Own elaboration.....	55
Table 22 - CPI 2016 Base. Categories and Weights. Own elaboration.....	56
Table 23 - Spanish CPI 2021 Base. Weights distribution. Pie Chart. Own elaboration.....	56
Table 24 - CPI comparison of general characteristics. Own elaboration	60

Chapter 1. INTRODUCTION

1.1 PROJECT MOTIVATION

During the past years, inflation has been a concept that hasn't stopped causing interest in the world population. However, after the covid-19 pandemic, the concern of high inflation rates started to appear due to the monetary policies of central banks. They began to increase the money supply to a point where, in December 2021, 80% of all the US dollars in the economy were printed in the previous 22 months going from \$4 trillion to \$20 trillion (see Image 5 below). This decision was taken in order to save the economy and avoid bankruptcy of a lot of companies. People started to fear that all that money printed would have a big impact on the inflation rates in the near future and they were right.

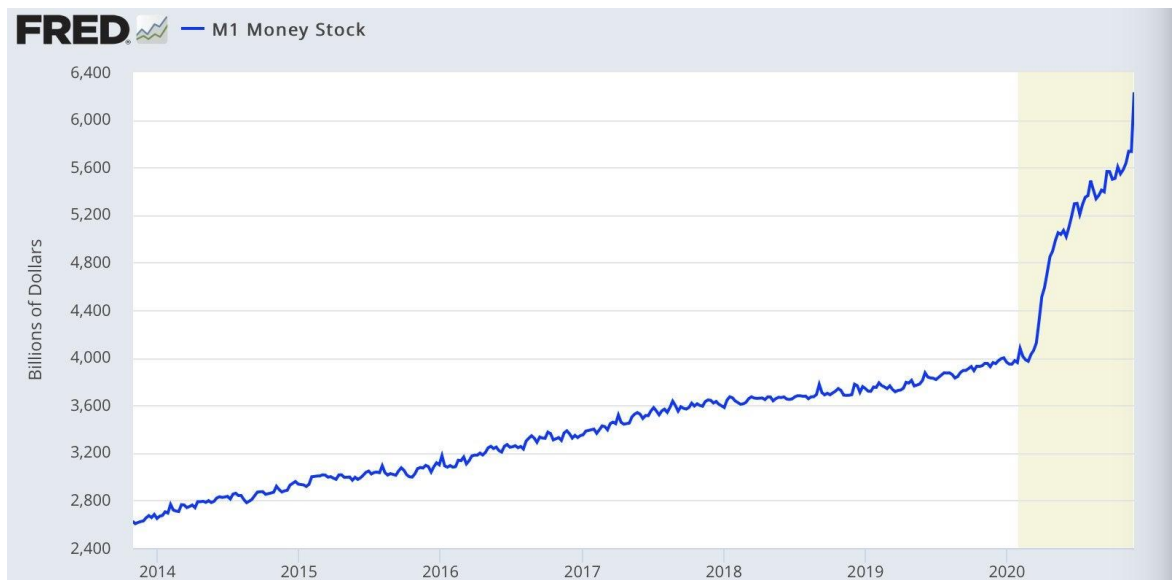


Image 5 - US M1 Money supply over the recent years (<https://themarketear.com/>)

Today inflation has risen up to 10% over a year in Spain. This means that Spain could be entering a galloping inflation (explained in chapter 4.1.2) or a stagflation because it hasn't recovered from the pandemic crisis yet and it is not growing economically as they expected.

It is for this reason and the Billion Prices Project from the MIT (explained in Chapter 3. that this project was born. Developing an independent, censorship-resistant inflation metric that calculates the inflation rate more accurately and changes some criteria that the Government is using to understate inflation.

1.2 PROJECT OBJECTIVES

The objectives of this project are:

- Understand the concept of inflation, what it means, which are the causes of it, how is it controlled, and what are the consequences of it.
- Study and understand the Spanish methodology and metric to measure inflation.
- Compare the Spanish methodology with the one used by other countries.
- Check if the Spanish methodology is in compliance with the IMF's recommendations.
- Know what the uses of the inflation metric are.
- Research areas of improvement.
- Develop an independent methodology and metric of inflation.
- Develop a web scraping module to monitor online prices and store them in a database.
- Develop a web application to show the results.
- Compare the results with the ones published by the Government.

1.3 WORK METHODOLOGY

The work dynamic will consist of setting fortnightly objectives and every 15 days attending a meeting in which progress is discussed.

Regarding the development of the website, the Agile methodology will be used with sprints of 2 to 4 weeks to increase productivity and obtain higher quality results. At the end of a sprint, there will be an increase in product that will add value to the final result.

Chapter 2. TECHNOLOGY DESCRIPTION

For the practical application, several programming languages, concepts, and tools were used. Next, those concepts and technologies used in the project will be described.

2.1 HTTP PROTOCOLS

In order to understand the main technologies used in this project, it is important to describe what HTTP Protocols are and how they work.

Hypertext Transfer Protocol (HTTP) is the protocol which lets users request data and resources. The answers to each HTTP request are in HTML programming language. HTTP is the base of every internet exchange of information. It has a client-server architecture which means that a user makes a request of specific information to one or more servers through a network. The user will receive the answer from the servers usually in a website format.

An HTTP request has several fields: method, path, version of the protocol, headers, and body.

- **Methods.** They define the type of operation that the client wants to perform. These operations are summarized in the acronym CRUD: create, read, update and delete. These are the actions that can be used to handle stored information. CRUD operations are mostly used in relational databases as they perform the same actions as some SQL functions. The next table [Table 1] shows each method along with its CRUD operation.

HTTP METHOD	CRUD OPERATION
GET	READ
POST	CREATE
PATCH	UPDATE
DELETE	DELETE
PUT	UPDATE

Table 1 - HTTP Methods and Operations

- Path. Is the address of the server where the information is stored, the URL. It has two main parts: the protocol and the domain. If a domain is located in a specific TCP port of the server, this can be added to the path. In this project, the database server is runned locally so the TCP port will be added to the path. More specifically on port 80.
- Version of the protocol. The standard one is HTTP 1.1.
- Headers. These let the client and the server exchange extra information in both the requests and the answers. The information exchanged vary from operating system used to version of the browser, date, specified connection options, warnings, content location, content type, content length, content encoding, content language, page requested, when it was last modified, etc.
- Body. In the request, the information shared in the body depends on the method that the client is using. For example, in a POST method, the body will have the information that the client wants to add to the database server. On

the other hand, the body of a response is the information retrieved by the server. This is usually a website in HTML format.

HTTP/1.1 200 OK	Status Line	HTTP Response
Date: Thu, 20 May 2004 21:12:58 GMT	General Headers	
Connection: close		
Server: Apache/1.3.27	Response Headers	
Accept-Ranges: bytes		
Content-Type: text/html	Entity Headers	
Content-Length: 170		
Last-Modified: Tue, 18 May 2004 10:14:49 GMT		
<html>	Message Body	
<head>		
<title>Welcome to the Amazing Site!</title>		
</head>		
<body>		
<p>This site is under construction. Please come back later. Sorry!</p>		
</body>		
</html>		

Image 6 - HTTP Response Example (http://www.tcpipguide.com/free/t_HTTPResponseMessageFormat.htm)

2.2 REST API

An API (Application Programming Interface) is a group of protocols that are used to design and implement applications software. It is considered as a contract between a the client and the information provider. A client makes a request and the provider gives him the information in an answer. APIs let computers interact with servers to receive information or use functions. They let programmers use functions that are already created and obtain specific information from a company making it easier for them.

On the other hand, REST is not a protocol but a group of architecture limits. REST can be implemented in different ways. For an API to be considered RESTful, it has to comply with certain criteria:

- Client-server architecture. The application must have clients and servers in order to handle requests through HTTP. Clients make requests using the

front-end and the requests are handled by the back-end which prepare the answer with the data requested.

- Serverless communication between client and server. This means that all the requests are independent one from another and they do not interfere between them.
- The data must be storable in the cache to optimize communications between client and server.
- The resources stored in the server are unique and are different from the representations sent to the client.
- The information sent to the client must contain information about how to process it.
- As a result of habing a layered architecture, servers are organized hierarchically. There are servers in charge of security, load balance, etc.

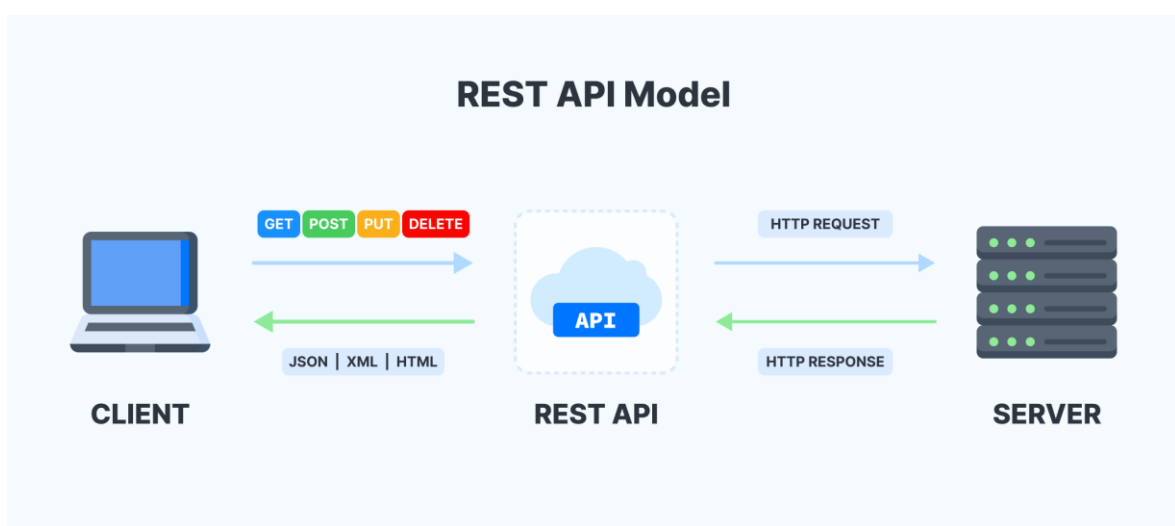


Image 7 - REST API Model (<https://appmaster.io/es/blog/que-es-la-api-rest-y-en-que-se-diferencia-de-otros-tipos>)

2.3 SPRING BOOT

Due to the rise in web systems' complexity caused by functionality, graphics, and security, the programming community developed a framework to not repeat code, and to reduce time and space for applications development.

It is the most popular Java enterprise framework to create high performance code, simple and reusable. The main goal is to standardize, speed up, manage, and solve problems that may appear during the project development.

An important advantage of using Spring's framework is that the development teams can focus on the business logic for the application instead of building the structure and framework of the application. This makes programming more efficient, shorter, and faster avoiding repetitive tasks.

The main characteristics of a Spring framework are:

- Dependency Injection. When building a complex Java application, the classes must be as independent as possible from each other so they can be reused and tested independently. Injecting dependencies help unifying classes and maintaining them.
- Security frameworks.
- Model View Controller (explained later).
- Testing frameworks.
- Implementation.
- Transaction management.
- Databases access and frameworks.
- Aspect Oriented Programming (AOP).
- Remote administration modules.

Moreover, Spring Boot contains an infrastructure which eliminates most of the work of building Spring applications. The goal of Spring Boot is to provide tools to build Spring applications faster and easy to configure.

It provides an embedded sapplication server, a dependency manager, automated configurations and many more.



Image 8 - Spring Boot Logo (<https://cursosdedesarrollo.com/2022/03/spring-boot-con-live-reload-en-intellij/>)

2.4 MODEL VIEW CONTROLLER

MVC is an achitecture pattern that has 3 components (model, view, controller) and separate the application logic from the presentation logic. Almost every modern framework use MVC. This architecture lets the programmer change parts of the code of a specific section without affecting to other sections.

The controller has the application logic, when he detects a request from a client, he answers it and asks the model the information the client is requesting. The controller receives requests from the client and it is in charge of requesting the data to the model and send it back to the view. The model is in charge of the application data, it queries the database and answers the controller with the data he asked from.

Once the controller has the data, he sends it to the view which applies the styles, organizes the information and builds the website that appears on the computer.

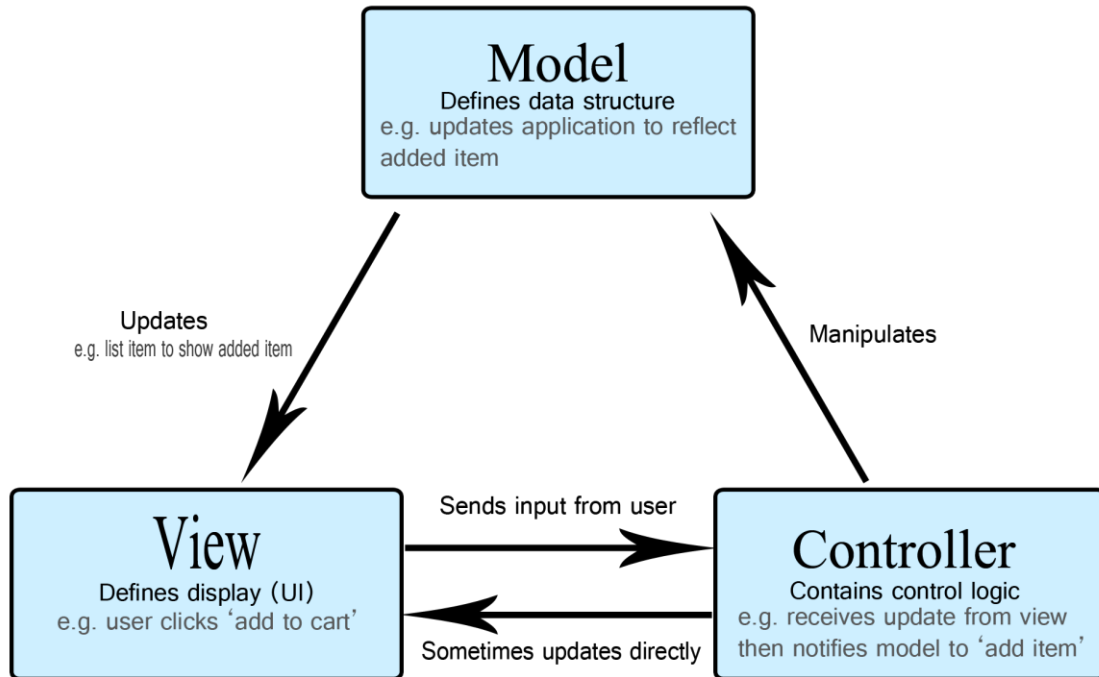


Image 9 - Model View Controller Architecture (<https://developer.mozilla.org/es/docs/Glossary/MVC>)

2.5 MYSQL

MySQL is a relational database manager (Relational Database Manager System, RDBMS) with a client-server architecture. RDBMS is a software as a service used to create and manage relational databases.

The databases are coded in SQL (Structure Query Language) language and organized in tables. SQL lets the user perform the next operations:

- Query specific data.
- Add data.
- Delete data.
- Update data.
- Sort data.
- Stablish data types.
- Stablish the relationships between tables via foreign keys.
- Access control to data.

MySQL provides a server for relational databases. This gives the users a number of benefits like:

- Create databases to store and handle data.
- Clients can query in SQL format to the MySQL server.
- The server application will answer the client with the information requested.



Image 10 – MySQL Logo (<https://itsoftware.com.co/content/que-es-y-para-que-sirve-mysql/>)

2.6 WEB SCRAPING

Web Scraping is the process in which text from websites is read to extract data and store it. This automates a copy and paste procedure. For images is called image scraping.

First, the program reads the HTML file of the website and parses it so it can be interpreted by the software. Then it searches for a specific label in it and get the data in that label.

It is commonly used by companies to analyze competitor prices and compare them with their own prices. Another example is Google which uses web scraping to show information the weather forecast or prices comparisons of flights and hotels. A lot of websites that compare prices use these techniques to show the information of other websites and retailers.

In this project, web scraping has been implemented in Python using the libraries requests and bs4. Requests allows the user to send HTTP requests using the protocol version HTTP/1.1. The user can also choose the method the want (GET, POST, DELETE). Requests also lets the user parse the HTTP answer to a JSON format. Moreover, bs4 library (Beautiful Soup) is used to parse the HTTP answer and look for specific labels and attributes of labels.

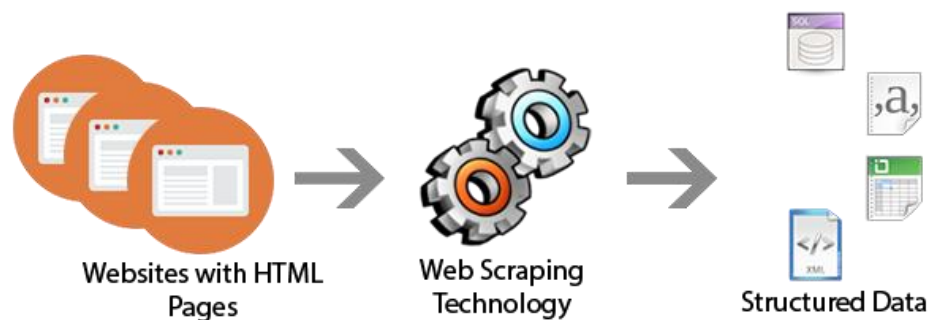


Image 11 - Web Scraping Procedure (<https://medium.com/@Emmitta/web-scraping-7f87930face4>)

2.7 POSTMAN

Postman is a desktop application which helps programmers develop their APIs. It lets users perform tasks in the API REST constraint. From creating APIs locally to performing tests to validate API's behavior and creating workspaces.



Image 12 - Postman Logo (<https://chicadev.medium.com/actualizando-postman-en-ubuntu-b5a58db0579a>)

In this project, Postman has been used to validate the behavior of the API and to check if it retrieved the data correctly. It lets the user get the response from the API in different formats. Users can create a workspace and store all the different API tests to its functions. If the method performed is POST then it lets you enter the parameters you want to add to the database and check the result of the operation.

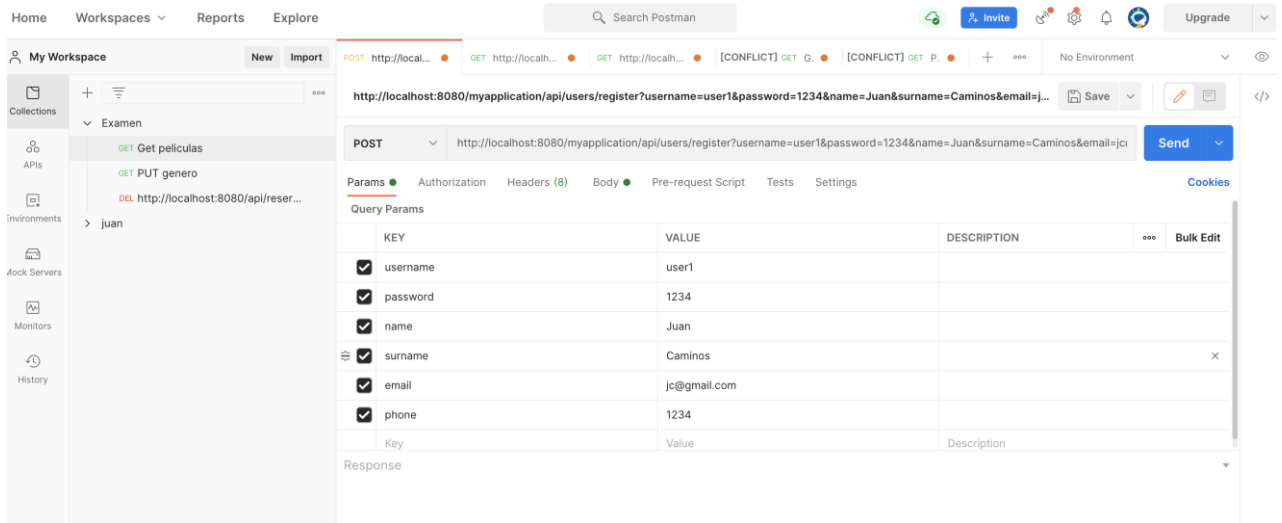


Image 13 - Postman Example

Chapter 3. STATE OF THE ART

In Spain, the consumer price index is calculated by the Instituto Nacional de Estadística. The CPI is calculated using the Sistema Base 2021 which states how the different categories are calculated. The weightings are calculated using the data from a survey called Encuesta de Presupuestos Familiares which is filled out by 24,000 families approximately from the entire national territory. These families provide basic information about their expenses on goods and services.

This project is inspired by The Billion Prices Project. An academic initiative at MIT and Harvard that collected prices from hundreds of online retailers around the world every day to later carry out research in different economic realms. It was founded in 2008 by Alberto Cavallo and Roberto Rigobon. There are two main projects that derived from The Billion Prices Project which are the main inspiration for this one: *Inflación Verdadera Venezuela* and *Inflación Verdadera Argentina*.

Moreover, they have another big project called PriceStats in which data is collected via web scraping and identify the best retailers they can use to measure inflation and purchasing power parity (PPP). They also conduct additional research on several topics and publish papers frequently.

3.1 INFLACIÓN VERDADERA VENEZUELA

The main objective of this project was to have an independent metric of Venezuela's inflation. More specifically, they calculated an independent CPI and then they compared the results with the one published by the Instituto Nacional de Estadística which is the Venezuelan public body in charge of measuring the country's inflation.

Since May 2017, they started to collect prices from different Venezuelan cities. The way they collected data was using a mobile app for Android devices. All the people involved were separated into small groups and each group would go every week to the same local

store and look for the prices of the same ten products. They would open the InflationVerdadera app and introduce the bar code, select the category, type the price, and then take a picture of the product to verify the data. After the data is uploaded, an Excel file is generated automatically and sent to their servers where the prices are processed and added to the database. Then they use that database to measure inflation and different metrics.

The data is uploaded by a lot of people and processed automatically. However, they have a software that detects anomalies automatically and a group of volunteers that verify that the prices uploaded correspond to the images attached.

After the data is collected, they create a price index for every category using a geometric average of the daily changes of the prices. Then they calculate an aggregated price index using the weighted arithmetic average with the official weights of each category published by the Government. They mention that those weights probably understate the categories that have higher price changes and overestimate the importance of the ones that have lower price changes.

The source code of the mobile app, the database and software used to calculate the different statistics are published in their website. Today, they are still looking for volunteers to help them in three different ways: collecting price data from Venezuela, validating price data from anywhere, and sharing the project on social media in order to make the project more popular.

For the people that want to help from Venezuela, they would have to download the mobile application and follow the instructions published on their website or follow a tutorial in which they make a detailed explanation about how to upload costs. The app has a simple and intuitive design as you can see in the next image.



Image 14 - Inflacion Verdadera Venezuela Mobile App Display
(<http://www.inflacionverdadera.com/venezuela/ayudar/>)



Image 15 - Inflacion Verdadera Venezuela Mobile App Add a Product
(<http://www.inflacionverdadera.com/venezuela/ayudar/>)

On the other hand, for the people that want to help from outside Venezuela, they can validate the data uploaded by the previous volunteers. The process is very simple, in their website, there is a section where there is a picture from the database and the user has to type the price (taxes included). Here is an example.

Cual es el precio que se ve en la imagen? El total a pagar, incluyendo el IVA (PMVP + IVA)

- El precio es:
- No se ve el precio/No se ve la imagen

Siguiente

Image 16 - Inflacion Verdadera Venezuela Data Validation
(<http://www.inflacionverdadera.com/venezuela/validacion-de-precios/>)

The results are published on their website. They focus more on the food category which they represent in different formats (yearly, quarterly, monthly [Image 17]) and they also publish their calculated consumer price index by category [Image 18]. They also have links to the specific data that has been used to elaborate each graph and table, and a brief definition of the metrics they calculate. Moreover, they explain that usually last month's data changes due to data delays.

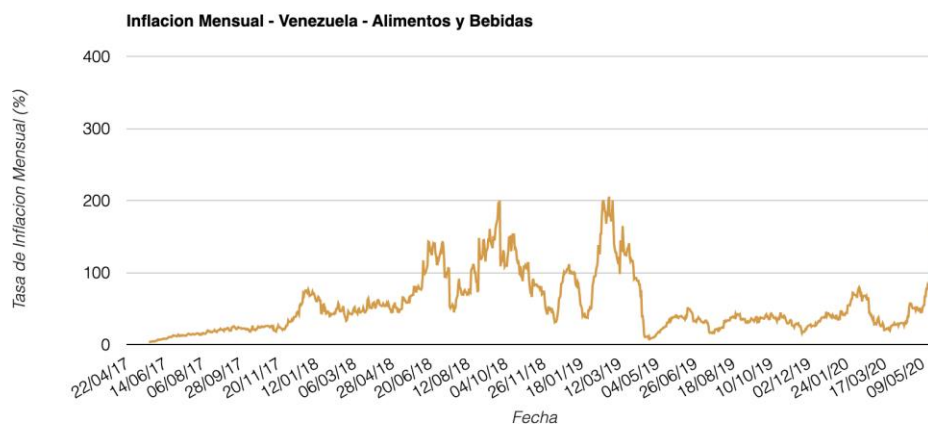


Image 17 - Inflacion Verdadera Venezuela Food Category Monthly Variation

(<http://www.inflacionverdadera.com/venezuela/>)

CATEGORIA	MENSUAL	INICIO	FIN	OBS	PRODUCTOS	CIUDADES	TIENDAS	PONDERACION
Alimentos y Bebidas	305,65	22apr2017	04jun2020	18.830	7.175	10	237	32,19
Bebidas Alcoholicas y Tabaco	55,48	24jun2017	31may2020	4.154	868	5	52	2,96
Vestido y Calzado	31,98	07sep2017	31may2020	439	140	3	12	7,16
Alquiler de Vivienda	0	05may2017	22may2020	16	8	6	10	9,83
Servicios de la Vivienda Excepto Telefono	0	06may2017	07apr2020	242	74	7	50	2,33
Equipamiento del Hogar	361,26	22may2017	30may2020	3.158	464	4	32	5,6
Salud	26,56	07may2017	31may2020	6.956	376	6	128	4,31
Transporte	42,86	04may2017	07apr2020	204	47	6	57	10,84
Comunicaciones (Telefonos y Servicios Asociados)	100	06may2017	06may2020	108	28	8	11	3,81
Electronicos y Entretenimientos	22,37	22may2017	17mar2020	1.613	272	5	19	3,61
Servicios de Educacion	0	06may2017	11may2020	31	10	6	15	2,7
Restaurantes y Hoteles	29,54	21may2017	31may2020	12.924	850	6	129	8,83
Bienes y Servicios Diversos	339,96	06may2017	28may2020	300	212	6	28	5,82

Image 18 - Inflacion Verdadera Venezuela CPI by Category (<http://www.inflacionverdadera.com/venezuela/>)

The most shocking part of the project is that the results they obtained were different from the ones that the Government was publishing. They proved that the metric that the Government was using, was not accurate or it had been manipulated in order to get lower results.

3.2 INFLACIÓN VERDADERA ARGENTINA

This project started in 2007 under the suspicion that the Government was manipulating the official inflation rate. The objective was to provide alternative inflation metrics to the official ones provided by the Government. The project lasted until 2012 and it would later develop into the Billion Prices Project. The results of this project were published in a paper at the Journal of Monetary Economics.

On the 14 pages long paper, Alberto Cavallo, explained how he proved that the inflation measure was being manipulated by the Government. Like in the Venezuelan project, he used data prices from top supermarkets of Argentina and South America in general. He used we scraping techniques to monitor online prices and stored them to later calculate the inflation rate.

He wanted to prove that official prices and online prices were not equal, and the official prices were manipulated. The reason why he chose top South America supermarkets was so that he could check if this was something that happened all over South America or only in Argentina. More specifically, he compared Argentina with Brazil, Chile, Colombia and Venezuela. In the next image [Image 19], he checked that in the neighbor countries, online and official prices were very similar. Then he checked the same metric in Argentina and the online prices were much higher than the official ones [Image 20].

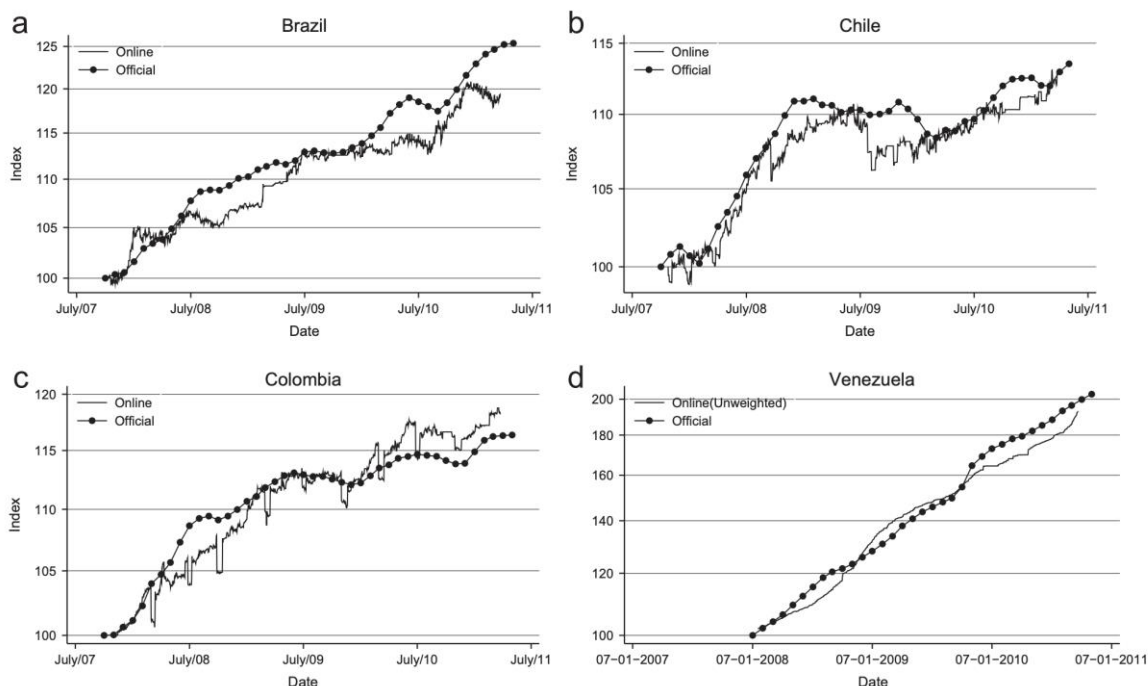


Image 19 - Inflacion Verdadera Argentina Online and Offline price indexes in four Latin American Countries (https://www.hbs.edu/ris/Publication%20Files/Cavallo_Alberto_Online%20and%20Official%20Price%20Indexes%20Measuring%20Argentinas%20Inflation_600c0e8f-cc57-4)

Online vs. official series.

	Argentina	Brazil	Chile	Colombia	Venezuela
<i>Mean annual inflation (%)</i>					
Official CPI index	8.53	5.28	2.44	3.79	27.37
Official supermarket index	8.38	5.91	3.19	3.73	29.38
Online supermarket index	20.14	4.72	3	4.88	27.43
<i>Correlations between online and official supermarket series</i>					
Price index	0.98	0.96	0.97	0.95	0.92
Annual inflation	0.84	0.09	0.97	0.89	-0.08
Monthly inflation	0.6	0.5	0.38	0.43	0.18
<i>Regression - official supermarket on online monthly inflation rates (12 lags)</i>					
Constant	0.84	-0.54	0.14	0.03	-1.96
Constant p-value	0.000	0.19	0.17	0.86	0.23
R2	0.9	0.55	0.66	0.59	0.66
<i>Monthly inflation rate volatility (standard deviation)</i>					
Official supermarket index	0.57	0.51	0.58	0.48	0.98
Online supermarket index	1.11	0.73	0.62	0.76	0.86

Image 20 - Inflacion Verdadera Argentina Online vs Official series

(https://www.hbs.edu/ris/Publication%20Files/Cavallo_Alberto_Online%20and%20Official%20Price%20Indexes%20Measuring%20Argentinas%20Inflation_600c0e8f-cc57-430d-b869-a650dda2b290.pdf)

He then proceeded to calculate the price index, annual inflation rate and monthly inflation rate using online data because, in the neighbor countries, online data matched official inflation rates but not in Argentina [Image 21].

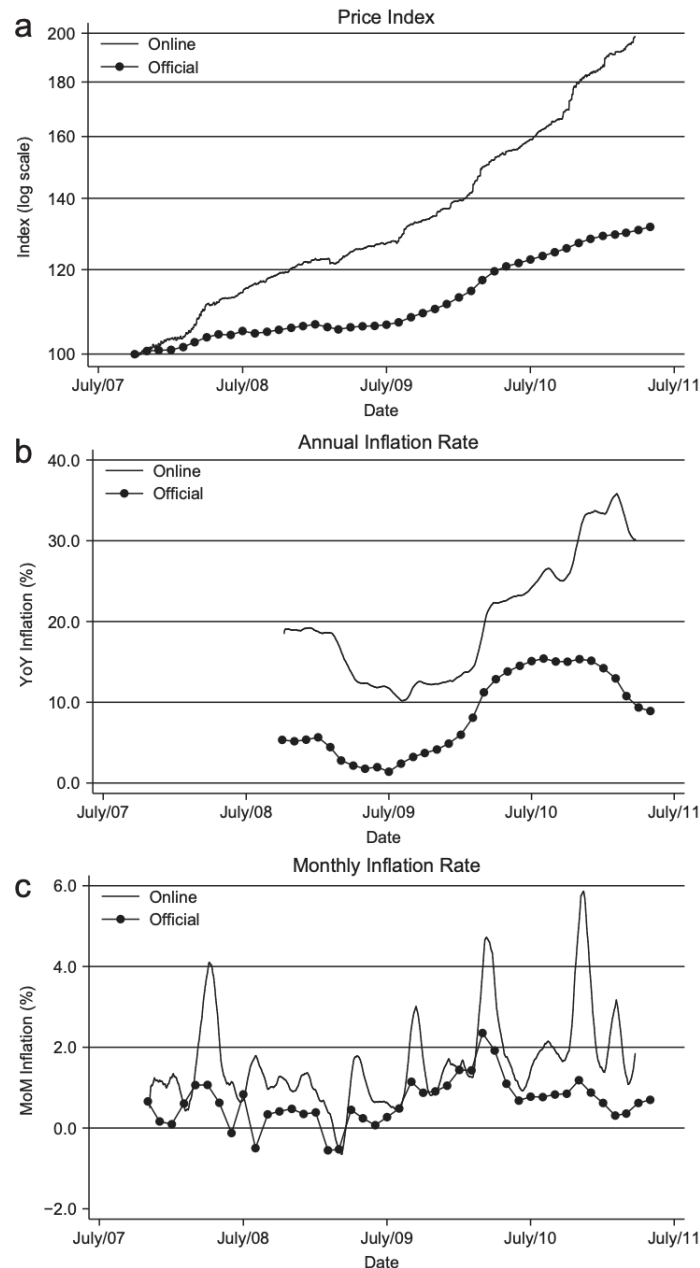


Image 21 - Inflacion Verdadera Argentina Price Index, Annual Inflation Rate and Monthly Inflation Rate
[https://www.hbs.edu/ris/Publication%20Files/Cavallo Alberto Online%20and%20Official%20Price%20Indexes%20Measuring%20Argentinas%20Inflation 600c0e8f-cc57-430d](https://www.hbs.edu/ris/Publication%20Files/Cavallo%20Alberto%20Online%20and%20Official%20Price%20Indexes%20Measuring%20Argentinas%20Inflation%20600c0e8f-cc57-430d)

The results showed huge discrepancies in the metrics. He also tried alternative methods to calculate inflation and added robustness to his calculation by adding more supermarkets, calculating the metrics independently for every supermarket, and changing the basket of goods. He then concluded that the difference in the online vs. the official metrics was so big that it confirmed the suspicion that the Government was manipulating the official inflation measure. He stated that the real inflation had been consistently two and three times higher than the official one. The annual inflation rate was, on average, 2.5 times higher and the Government could have many reasons to manipulate it for their benefit.

Chapter 4. THEORETICAL ANALYSIS

This chapter is about the theoretical analysis of inflation and its different aspects to understand it. It focuses on the Spanish way of measuring inflation and the methodology followed to calculate the CPI. The chapter is divided into a study of inflation in general as a concept, a study about the Spanish methodology comparing it with other countries' methodology, and a study about the metric and its uses.

4.1.1 INFLATION ANALYSIS

This part focuses on the general aspects of inflation to understand what it is, why it happens, what consequences it has, and how it is controlled.

4.1.2 WHAT IS INFLATION?

Inflation is defined as a rise in the general level of prices of goods and services in an economy over a given period. This translates into being able to purchase less goods and services with the same amount of money. The metric used to measure it is the Consumer Price Index (CPI), which calculates the percentage of change in price of a basket of goods and services.

Underlying inflation is the CPI calculation which does not include categories like energy and food. This metric does not include the most volatile products and, as a result, the rate of inflation obtained is more stable. This metric is more appropriate to predict the inflation trend in the medium term.

The main types of inflation are:

- Walking/Moderate inflation. Happens when inflation rates are between 3% and 10% annually.
- Galloping inflation. Happens when the causes of inflation are such that the rates go up dual or triple digits annually.
- Hyperinflation. Happens when inflation rates are higher than 50% monthly.

- Deflation. Happens when there are negative inflation rates. This is a general decline in prices of goods and services. As opposed to inflation, deflation makes the value of a currency rise over time.
- Stagflation. It is a combination of both inflation and an economic stagnation in a country. Stagflation happens when the economy of a country does not grow and the cost of living increases simultaneously. The main consequence of stagflation is the impoverishment of the population.

4.1.3 WHAT CAUSES INFLATION?

There are multiple causes of inflation. The main causes of inflation are:

- Demand-pull Inflation. Happens when demand rises and the supply is not able to fulfill it. As a result, inflation increases due lower unemployment and higher GDP as the economy moves along the Phillips Curve. The Phillips curve shows that if unemployment rises, inflation will be lower due to consumers having less purchase power and, as a result, demand will be lowered [Image 22].

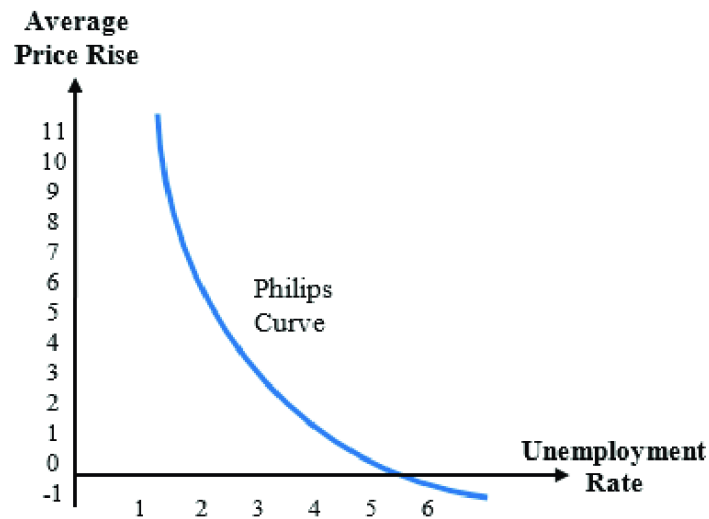


Image 22 - Phillips Curve (https://www.researchgate.net/figure/The-modified-Phillips-curve_fig4_342734687)

- Cost-Push Inflation. Happens when the supply of goods and services decreases due to an increase in production costs, an increase in the prices of commodities, an increase in workforce costs, taxes, and the rest of the factors of production. As a result, vendors increase their prices to make up for these costs increases and try to maintain their profits.
- Built-in Inflation. Happens when suppliers expect a rise in prices in the future and they anticipate it by increasing the prices. They also take into account that if prices increase, the wages they would have to pay their employees would be higher.
- Monetary Policy. The supply of currency in a market also affects inflation. If the supply of money is very high, the value of it will decrease due to supply and demand laws. This results in having higher inflation. This specific cause of inflation is one of the main reasons why inflation is very high nowadays. During the covid-19 pandemic, countries started to print money to have it in circulation and so that businesses would not go bankrupt. In the short term, this policy helps the economy and saves a lot of businesses. However, in the

medium/longer term results in higher inflation rates. In the next chart, you can see the M2 money supply of the US [Image 23], which is the amount of money supply in circulation in an economy and considers the cash, checking deposits, saving goods deposits, and money market securities.

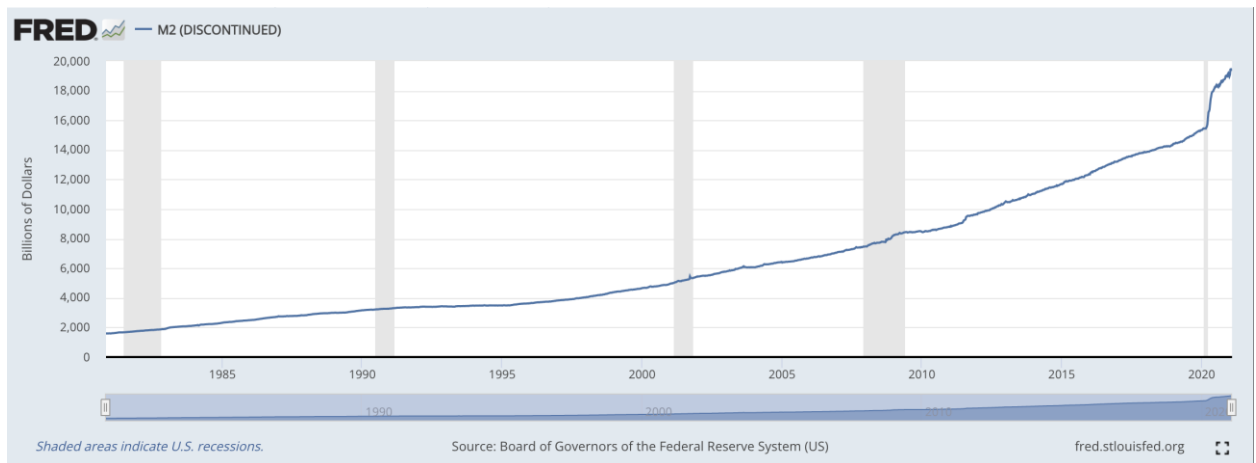


Image 23 - M2 money supply of the US (<https://fred.stlouisfed.org/series/M2>)

4.1.4 WHAT ARE THE CONSEQUENCES OF INFLATION?

The main consequences of inflation are the following:

- Decrease of the value of the currency. When prices of goods and services increase, you cannot purchase the same amount of them with the same resources. Therefore, inflation causes a decrease in purchasing power.
- Discourages medium and long term investments. Investors look for stable and safe situations when investing their money. An economy with high inflation generates uncertainty. Moreover, as I will explain in the next part, one measure to control inflation is to increase interest rates, which leads to a decrease in investments due to having to pay back more money.
- Benefits the debtors and harms the creditors. When giving a credit, if both parts agree to return the same amount of money and there is inflation, after

some time, that same amount of money will have less value than in the present time. This is due to the first cause of inflation I explained, the decrease of the value of the currency. However, most of the times a credit is given, the amount that the debtor must give back to the creditor is adjusted to the expected inflation.

- Increases the costs of production. As a result of the rise of prices, raw materials get more expensive and the wages are increased as a result of higher cost of living.
- Favors assets and their owners. Inflation makes assets appreciate over time. Therefore, investors see how their assets appreciate as time goes by.

4.1.5 HOW IS INFLATION CONTROLLED?

Inflation is generally controlled by Central Banks and Governments. To do so, there are different methods to achieve it.

- The primary strategy to prevent inflation is to increase interest rates. When interest rates are increased, the demand in the economy is decreased. This makes the economy grow slower and therefore, decrease inflation. This also increases the bank deposits' yield and therefore, it encourages saving money. Today, as a result of such high inflation rates, both the FED (Federal Reserve of the US) and Europe's Central bank have increased their interest rates to try to lower inflation.

Back in the 1970s, due to the losing the Vietnam war, the support given to Israel which ended up with the Arabic petroleum boycott, and other factors, the US was immersed in a large deficit. Then the president at that time, Nixon, announced that the gold standard would be finished. This resulted in the rise of inflation rates up to 14% at one point. In 1979, Paul Volcker was elected president of the FED with the mission of lowering the inflation. His solution to inflation was to increase interest rates at all costs even if the country was not going to be able to pay its debt. He increased interest rates up to 20%. The

aftermath of this measure was that the US entered a recession and the unemployment rate skyrocketed . However, he managed to lower inflation rates from 14% to lower than 3% in only 3 years.

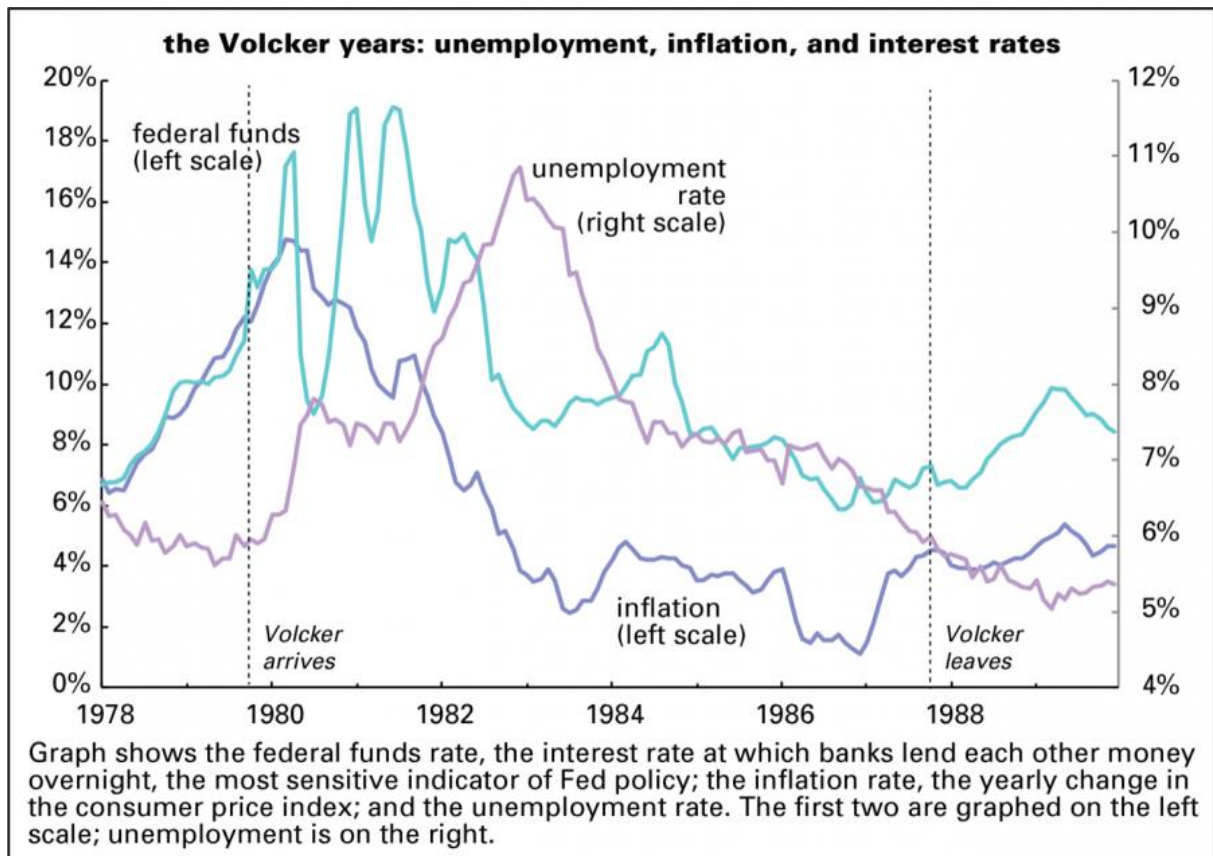


Image 24 - Paul Volcker's years as president of the FED. Inflation vs Interest Rates vs Unemployment Rate
(<https://worldbuilding.stackexchange.com/questions/224117/why-would-my-central-bank-suddenly-increase-the-interest-rates-to-over-20>)

- Control money supply. As stated previously, too much money supply leads to higher inflation. Also, the lower the money supply the less demand will be in an economy.
- Increasing income taxes. This results in less purchase power on the citizens and therefore, less demand.

4.2 METHODOLOGY ANALYSIS

This section focuses on the Spanish methodology to measure inflation, its evolution, a comparison with other countries methodology, and a study of IMF's recommendations to check if it's in compliance.

4.2.1 HOW IS INFLATION MEASURED IN SPAIN?

In Spain, inflation is measured using the Consumer Price Index (CPI) and the Harmonized Index of Consumer Prices (HICP) and it is published by the Instituto Nacional de Estadística (INE).

The consumer price index (CPI) is an indicator that measures the variation in the prices of a basket of goods and services in a specific place during a certain period of time.

The harmonized consumer price index (HICP) is calculated following the standards agreed upon in the European Union in order to have indicators that can be compared between member countries.

The current CPI is calculated using the 2021 Base System that sets how the different metric indices are calculated. The calculation of the weights is made with data from the Encuesta de Presupuestos Familiares carried out by approximately 24,000 households throughout the national territory, which provides basic information on family spending on consumer goods and services.

In the survey, each household remains in the sample for two consecutive years, with a collaboration each year consisting of a period of 14 days during which the households write down its expenses and provide information through various interviews. In the 2021 Base System, a technology called scanner data was introduced to add more sources of information. Scanner data consists in the access to supermarkets' databases to extract data directly from them. They also announced that they are working on implementing web scraping techniques to add more sources of information.

Due to the new international classification of consumption from the EU, the ECOICOP, each consumption category is broken down into subclasses. Before, a category was represented by one or more items and, the price change of one, supposed the change of all the elements of the category.

The items in the basket of goods are added in subclasses, these in classes, then in subgroups, and finally, the subgroups in groups. It consists of 12 groups, 43 subgroups, 101 classes, and 219 subclasses as of January 2017.

The 12 groups are distributed like this:

Número de artículos IPC Base 2016		
Grupos		Número de artículos
01	Alimentos y bebidas no alcohólicas	170
02	Bebidas alcohólicas y tabaco	12
03	Vestido y calzado	66
04	Vivienda	18
05	Menaje	57
06	Medicina	11
07	Transporte	31
08	Comunicaciones	5
09	Ocio y cultura	43
10	Enseñanza	7
11	Hoteles, cafés y restaurantes	22
12	Otros bienes y servicios	37
Total		479

Image 25 - CPI Group Distribution and amount of Articles that belong to them

(https://www.ine.es/metodologia/t25/t2530138_16.pdf)

The demographic breakdown for which data is published monthly is as follows:

ÍNDICE	Nacional	Comunidad Autónoma	Provincia
General	X	X	X
Grupos	X	X	X
Subgrupos	X	X	X
Clases	X		
Subclases	X		
Rúbricas	X	X	
Grupos Especiales	X	X	

Image 26 - Spanish CPI Demographic Breakdown for which data is published monthly
(https://www.ine.es/metodologia/t25/t2530138_16.pdf)

For the selection of municipalities, it is important that they are distributed throughout the province to avoid concentration in big cities. Emphasis is placed on the representativeness of small municipalities, and all municipalities must contain articles from all groups, for which prepares a reduced basket in which basic consumption items are included. However, later we will see that the information about the selection of municipalities is not public at all.

To select the representative articles of the survey's spending categories, different organizations, business associations, manufacturers, merchants and establishments are consulted, which provide information on those articles that best represent the different categories, in compliance with the following selection criteria : the evolution of the prices of the selected articles must be similar to that of the rest of the articles of the category they represent, the articles must be those consumed commonly by the population, they must have easily observable prices, and they must offer reasonable guarantees of permanence in the market.

The basket consists of 462 items from traditional sourcing and 493 from scanner data. The number of prices processed monthly is approximately 210,000.

The formula used to calculate the CPI indices, base 2021, is the chained Laspeyres formula which understates inflation as I will explain later. The general index corresponding to month m of year t is mathematically expressed as follows:

$${}_0 I_{LE}^t = \prod_{k=1}^t \frac{\sum_i p_i^k q_i^{k-1}}{\sum_i p_i^{k-1} q_i^{k-1}}$$

An elementary aggregate is the consumption component with the lowest level of aggregation for which indices are obtained and, in whose calculation, weights are not involved; the indices of these aggregates are called elementary indices. In the Spanish CPI, an elementary index is calculated for each item in the shopping basket in each of the provinces, so the elementary aggregate is the item-province.

The index of an elementary aggregate i is obtained as the quotient of the average price of said elementary aggregate in the current period and the average price in the price reference period, that is, December of the previous year.

The weights that intervene in the calculation of the aggregate indices come from the Encuesta de Presupuestos Familiares. This survey provides estimates of spending on consumer products by households living in Spain.

The weights of each article represent the relationship between the expenditure made in the categories represented by said article and the total expenditure made in all the categories covered by the index:

$$W_i = \frac{\text{expenditure in a category represented by the article } i}{\text{Total expenditure}}$$

The weights are different in each of the geographical aggregations (provinces, autonomous communities, and the national group) and from them the weights of the different

functional aggregations are obtained. Thus, the weighting of the functional aggregate A is obtained as the sum of the weightings of the items that make up said aggregation:

$$W_A = \sum_{i \in A} W_i$$

Depending on the periodicity and frequency in collecting prices, a way of classifying the articles is established. There are two types, monthly and quarterly.

The monthly ones are observed every month in all the establishments of the sample. Seasonal items for example.

The quarterly ones are those whose prices are more stable (furniture, restoration, repair, etc). The establishments are divided into 3 sub-samples and one sub-sample is visited per month.

4.2.2 HOW HAS THE METHODOLOGY EVOLVED IN RECENT YEARS AND WHAT HAS CAUSED THE CHANGES?

The INE introduced the first Cost of Living Index System in 1939. Before 1936, they published simple indexes and average prices that served to create the first System. In 1938, the Servicio Nacional de Estadística began to compile cost-of-living indices for some provincial capitals. In 1939, with the expansion to all provincial capitals, the Cost of Living Index System was implemented, based on July 1936.

Next table shows the categories and weights of the base of 1936:

Base July 1936	
Category	Weights
Alimentación	601.0
Vestido y calzado	94.0
Vivienda (alquiler)	146.0
Gastos de casa	85.0
Gastos diversos	74.0
Total	1000.0

Table 2 - Spanish CPI 1936 Base. Categories and Weights. Own elaboration

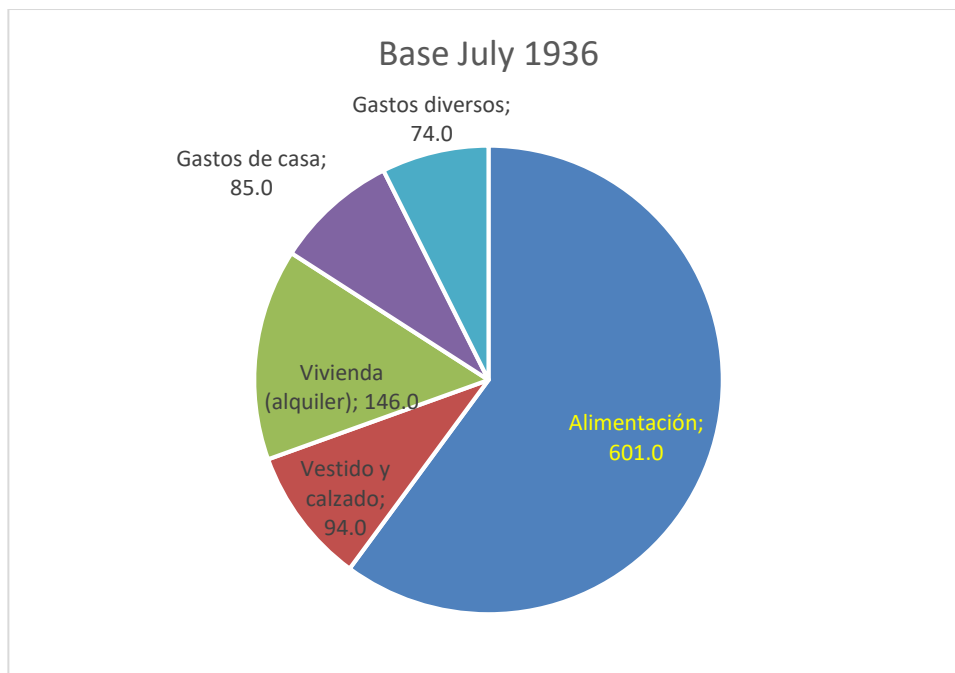


Table 3 - Spanish CPI 1936 Base. Weights distribution. Pie Chart. Own elaboration

Since then, there have been nine different systems of Consumer Price Indices. The base periods have been: July 1936, 1958, 1968, 1976, 1983, 1992, 2001, 2006, 2011 and 2016.

Period of each system:

Sistemas de Índices

Base	Tiempo de vigencia
1936 (julio)	julio de 1939 – diciembre de 1960
1958	enero de 1961 – diciembre de 1968
1968	enero de 1969 – diciembre de 1976
1976	enero de 1977 – julio de 1985
1983	agosto de 1985 – diciembre de 1992
1992	enero de 1993 – diciembre de 2001
2001	enero de 2002 – diciembre de 2006
2006	enero de 2007 – diciembre de 2011
2011	enero de 2012 – diciembre de 2016
2016	enero de 2017

Image 27 - Time Period of each Spanish CPI system (https://www.ine.es/metodologia/t25/t2530138_16.pdf)

- 1958 Base. It is the first CPI that was established in Spain. It included the entire Spanish territory, but population coverage was limited to households whose income was less than 80,000 pesetas (480€).

The Encuesta de Cuentas Familiares was used, which interviewed more than 4,000 families.

It offered information both for the national group and for the provincial capital and non-capital municipalities.

The basket size was 181 items.

Base 1958	
Category	Weights
Alimentación	553.0
Vestido y calzado	136.0
Vivienda (alquiler)	49.6
Gastos de casa	82.1
Gastos diversos	179.3
Total	1000.0

Table 4 - Spanish CPI 1958 Base. Categories and Weights. Own elaboration

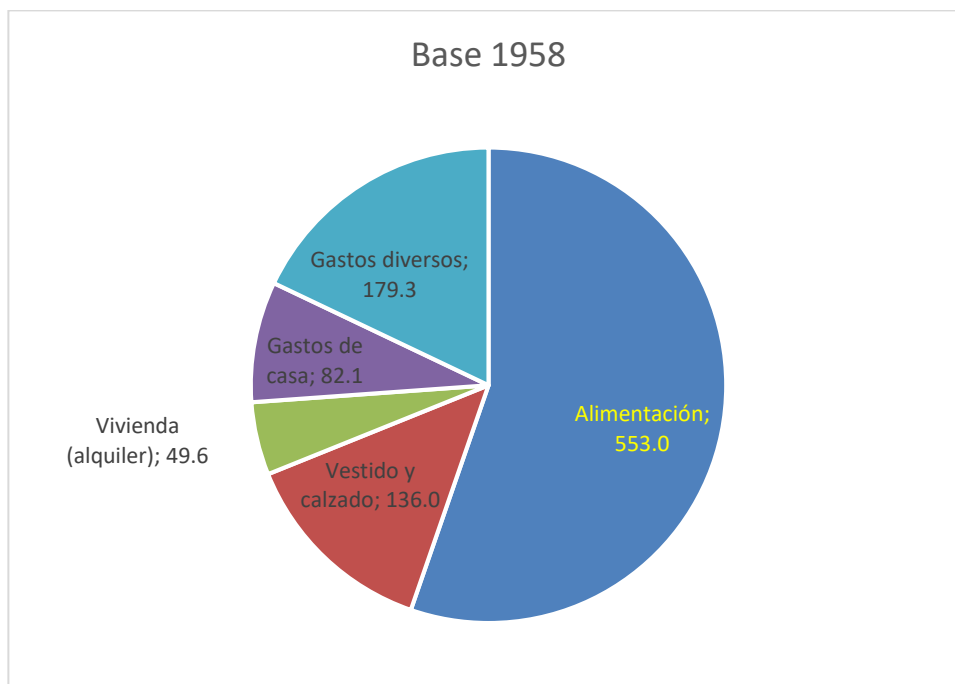


Table 5 - Spanish CPI 1958 Base. Weights distribution. Pie Chart. Own elaboration

- 1968, 1976 and 1983 Base. In the following three CPI bases, population coverage was restricted by annual income within a threshold.

The 1968 base used the Encuesta de Presupuestos Familiares to establish the weighting structure with sample sizes of about 20,000 households.

The results included the national set, provinces and distinction between urban and non-urban. The base of 1983 incorporated the autonomous communities.

The basket size increased from 225 items in the 1968 base to 428 items in the 1983 base.

Base 1968	
Category	Weights
Alimentación	493.9
Vestido y calzado	139.4
Vivienda (alquiler)	79.5
Gastos de casa	98.9
Gastos diversos	188.3
Total	1000.0

Table 6 - CPI 1968 Base. Categories and Weights. Own elaboration

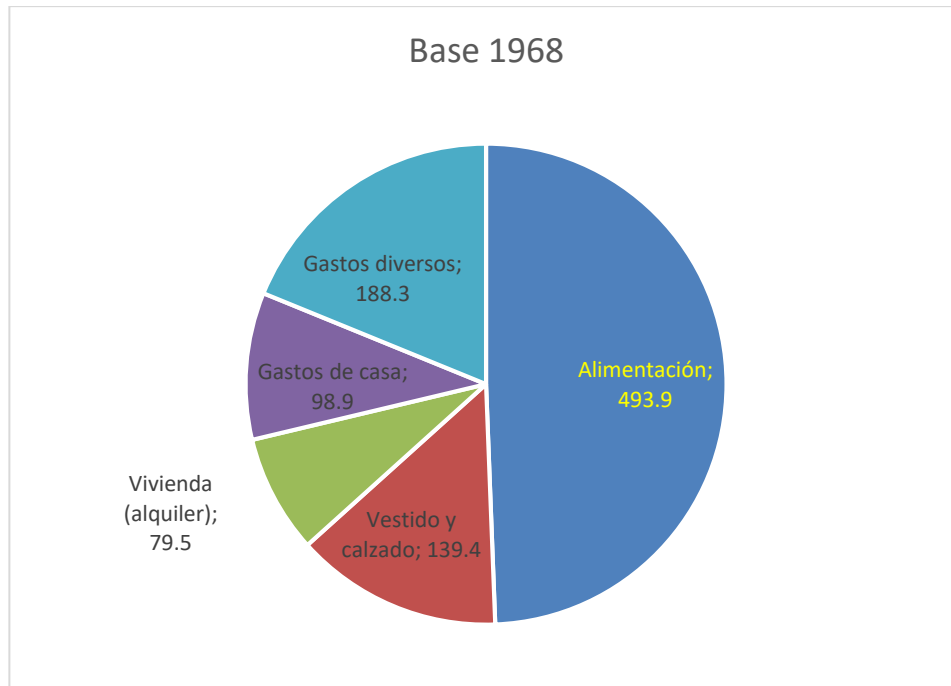


Table 7 Spanish CPI 1968 Base. Weights distribution. Pie Chart. Own elaboration

Base 1976	
Category	Weights
Alimentación	405.20
Vestido	81.71
Vivienda	140.01
Menaje	77.51
Medicina	33.74
Transporte	97.44
Cultura	69.44
Otros	94.95
Total	1000.00

Table 8 - CPI 1976 Base. Categories and Weights. Own elaboration

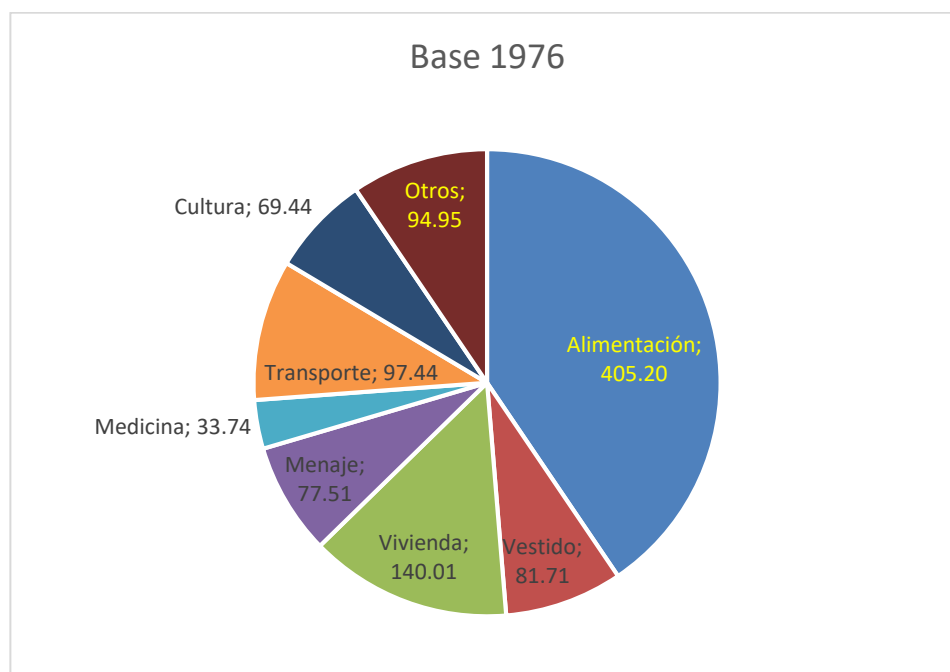


Table 9 - Spanish CPI 1976 Base. Weights distribution. Pie Chart. Own elaboration

Base 1983	
Grupo	Ponderación
Alimentación	330.27
Vestido	87.39
Vivienda	185.65
Menaje	74.15
Medicina	23.93
Transporte	143.81
Cultura	69.60
Otros	85.20
Total	1000.00

Table 10 - CPI 1983 Base. Categories and Weights. Own elaboration

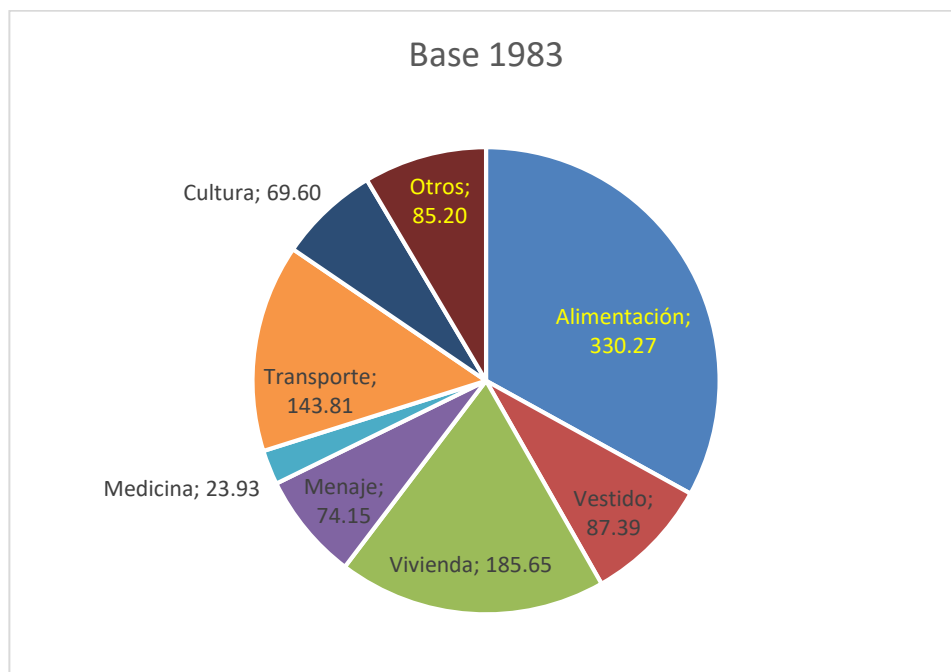


Table 11 - Spanish CPI 1983 Base. Weights distribution. Pie Chart. Own elaboration

- 1992 Base. This base completely changed the concepts and methods that had been used previously. It coincides with the first steps of harmonization of the CPIs of the European Union. It is considered the first CPI comparable to those of surrounding countries.

The most relevant changes were: the general calculation formula (the fixed-base Laspeyres formula is introduced), significant expansion of the sample (municipalities, shopping areas, establishments, products and number of prices collected), specific treatments for certain groups of items (seasonal products, housing), new procedures for purging and validating prices, and a publication calendar is implemented (introduces the dissemination of the CPI no later than two weeks after the end of the reference month).

The basket size increased to 471 items.

Subsequent IPC Systems introduced changes from the conceptual structure of this base.

Base 1992	
Category	Weights
Alimentación	293.61
Vestido	114.79
Vivienda	102.80
Menaje	66.84
Medicina	31.26
Transporte	165.42
Cultura	72.67
Otros	152.61
Total	1000.00

Table 12 - CPI 1992 Base. Categories and Weights. Own elaboration

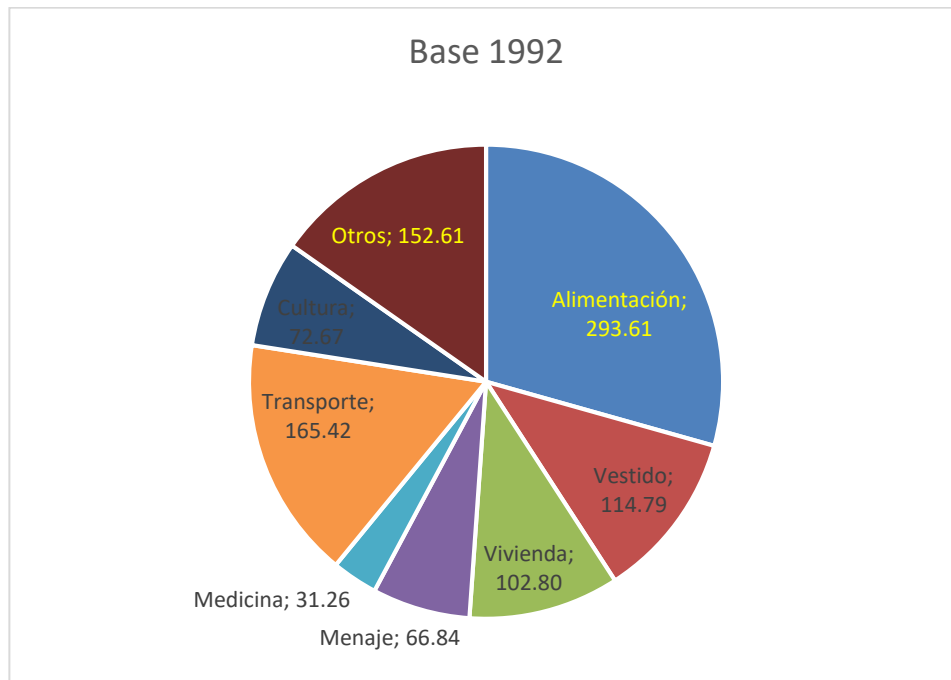


Table 13 - Spanish CPI 1992 Base. Weights distribution. Pie Chart. Own elaboration

- 2001 Base. Starting from the 1992 base, the changes applied to the systems are focused on the harmonization of the EU CPIs and reviewing the calculation methods and the elements that comprise it. The two most important are the composition of the shopping basket and the weight structure. Both fields are built with the results of the Encuesta de Presupuestos Familiares.

In 1997, the INE prepared a survey that provided information on a quarterly basis. Thanks to this, an indicator was designed that allowed its components to be reviewed more frequently. In 2001, the CPI started to be reviewed every year.

The main change was that the CPI was calculated using the chained Laspeyres index, instead of using the fixed-base one.

Reduced prices and offers by decision of the EU were also incorporated. This meant that the current indices could not be compared with those of the

previous year. This was solved by recalculating the indices from the previous year with the new prices and weights of the new base.

The basket size was 484 items.

Base 2001	Year 2002
Category	Weights
Alimentos y bebidas no alcohólicas	218.630
Bebidas alcohólicas y tabaco	32.170
Vestido y calzado	99.280
Vivienda	110.260
Menaje	63.571
Medicina	28.062
Transporte	155.760
Comunicaciones	25.729
Ocio y cultura	67.263
Enseñanza	17.444
Hoteles, cafés y restaurantes	112.708
Otros bienes y servicios	69.124
Total	1000.001

Table 14 - CPI 2001 Base. Categories and Weights. Own elaboration

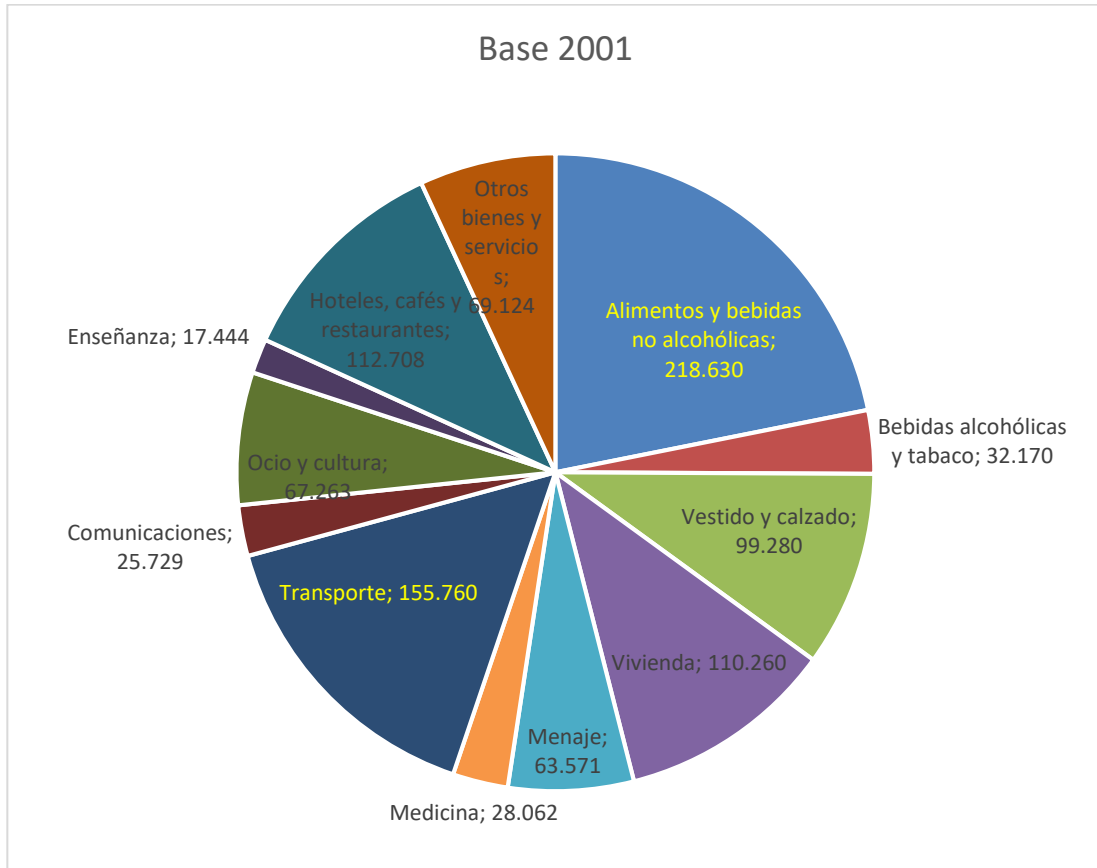


Table 15 - Spanish CPI 2001 Base. Weights distribution. Pie Chart. Own elaboration

- 2006 Base. Indices are prepared for the whole country, the 17 autonomous communities, the 50 provinces, Ceuta, and Melilla.

Base 2006	Year 2007
Category	Weights
Alimentos y bebidas no alcohólicas	220.556
Bebidas alcohólicas y tabaco	28.229
Vestido y calzado	90.280
Vivienda	103.607
Menaje	61.520
Medicina	28.259
Transporte	148.879
Comunicaciones	35.845
Ocio y cultura	71.089
Enseñanza	16.027
Hoteles, cafés y restaurantes	115.477
Otros bienes y servicios	80.230
Total	999.998

Table 16 - CPI 2006 Base. Categories and Weights. Own elaboration

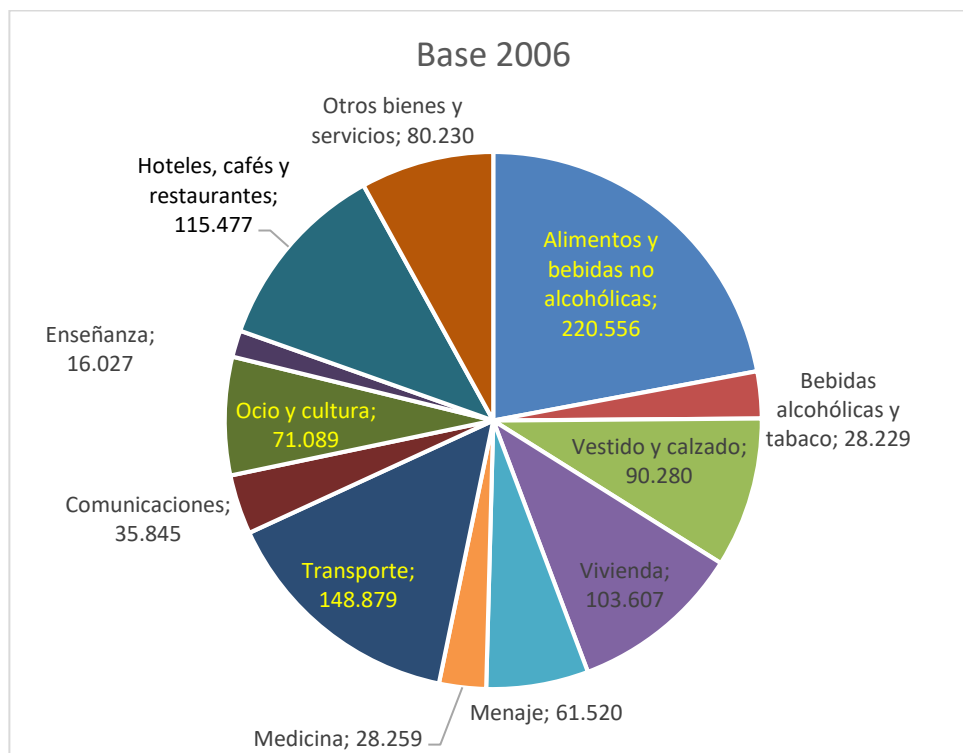


Table 17 - Spanish CPI 2006 Base. Weights distribution. Pie Chart. Own elaboration

- 2011 Base. There wasn't any relevant change. Just some articles were substituted and weights updated.

Base 2011	Year 2012
Category	Weights
Alimentos y bebidas no alcohólicas	182.642
Bebidas alcohólicas y tabaco	28.872
Vestido y calzado	83.437
Vivienda	120.006
Menaje	66.750
Medicina	31.398
Transporte	151.630
Comunicaciones	38.498
Ocio y cultura	75.420
Enseñanza	14.175
Hoteles, cafés y restaurantes	114.608
Otros bienes y servicios	92.563
Total	999.999

Table 18 - CPI 2011 Base. Categories and Weights. Own elaboration

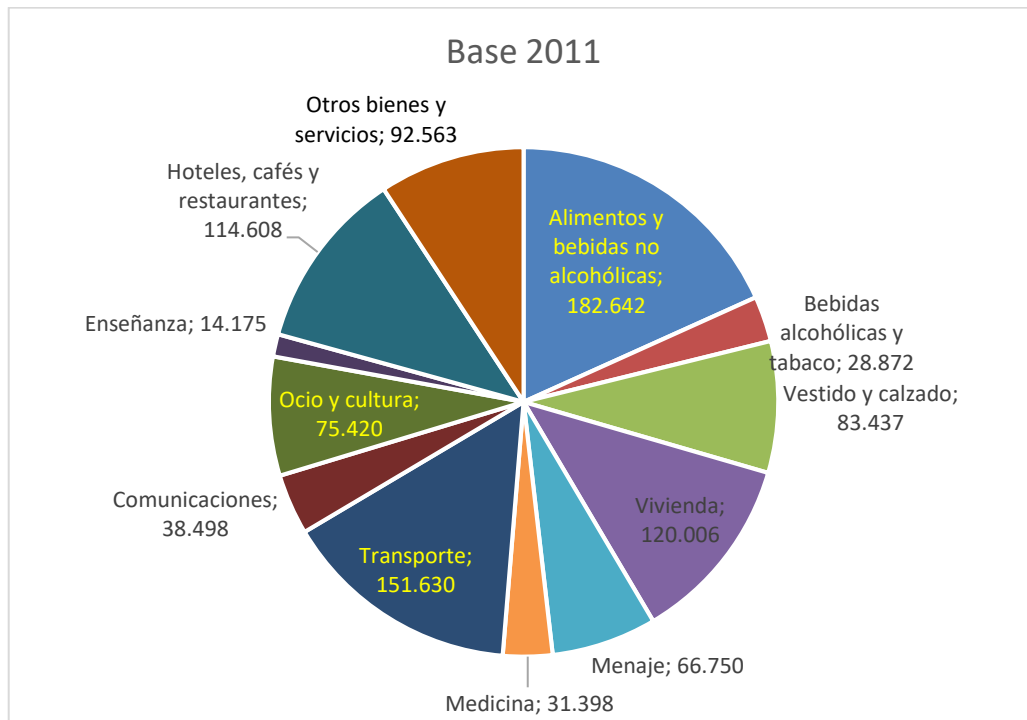


Table 19 - Spanish CPI 2011 Base. Weights distribution. Pie Chart. Own elaboration

- 2016 Base. A new definition of consumption field is introduced, based on the concept of household final consumption expenditure of the European System of Accounts (ESA). It involved a change in: gross spending vs. net spending (expenditure estimates must be made taking into account the monetary considerations received by households, e.g. Insurance takes into account the compensation received), transactions between the household sector and companies (there are no to take into account transactions between households, e.g. purchase of second-hand cars), and the incorporation of gambling.

Change in the field of consumption. Before, a batch was represented by one or more items and, the price change of one, supposed the change of all the elements of the batch. The change of the EU implies the breakdown of subclasses into two or more batches.

Base 2016	Year 2017
Category	Weights
Alimentos y bebidas no alcohólicas	197.711
Bebidas alcohólicas y tabaco	30.190
Vestido y calzado	67.342
Vivienda	133.033
Menaje	58.765
Medicina	39.612
Transporte	146.693
Comunicaciones	35.960
Ocio y cultura	85.209
Enseñanza	16.793
Hoteles, cafés y restaurantes	121.193
Otros bienes y servicios	67.499
Total	1000.000

Table 20 - CPI 2016 Base. Categories and Weights. Own elaboration

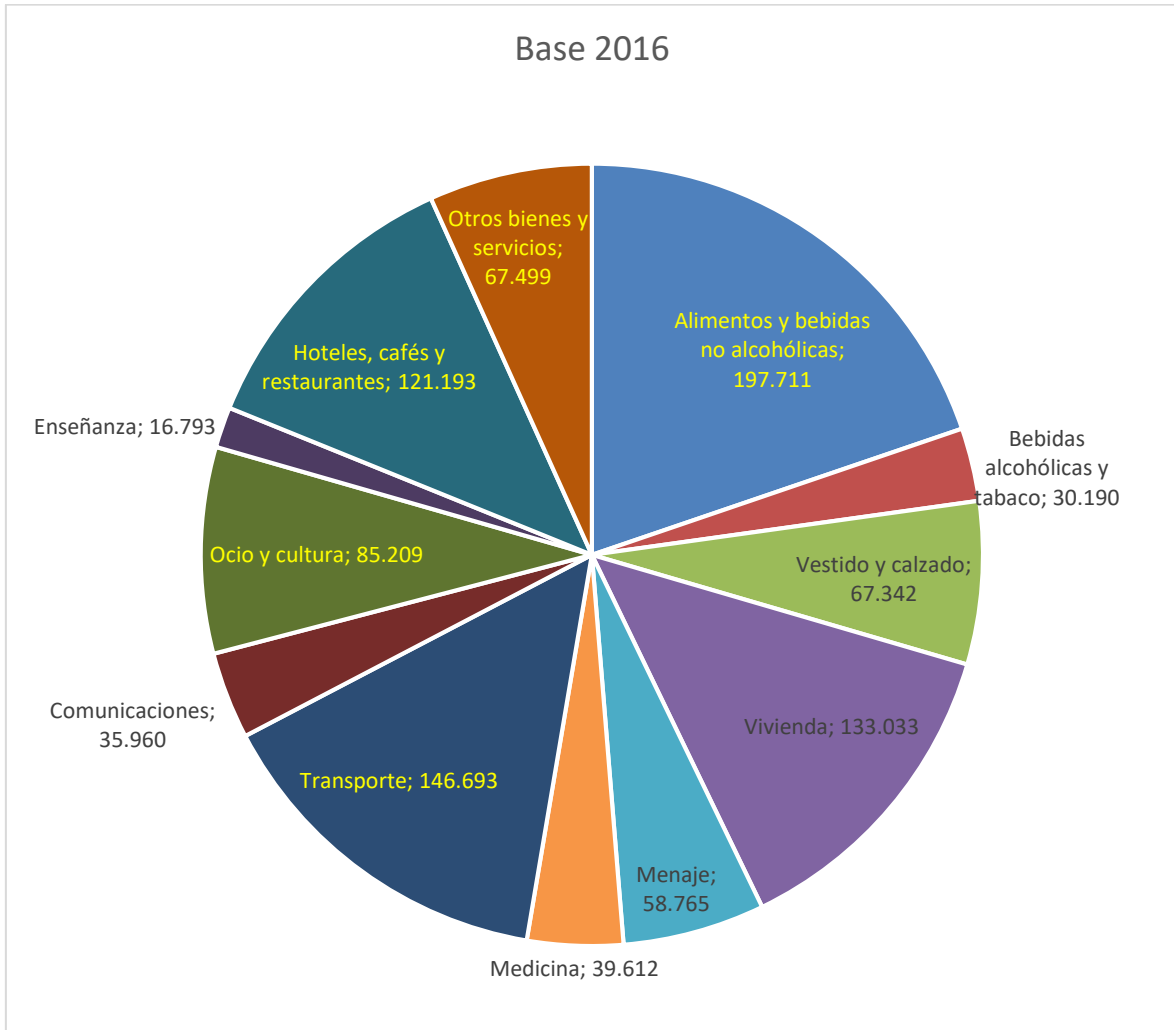


Table 21 - Spanish CPI 2016 Base. Weights distribution. Pie Chart. Own elaboration

- 2021 Base. Added and substituted items to adapt the metric to the covid-19 consumption habits. Introduced the idea of adding web scraping techniques as another source of information. Focused on data quality checks. Adjusted weightings and changed criteria on deals.

Base 2021	Year 2021
Category	Weights (%)
Alimentos y bebidas no alcohólicas	23.620
Bebidas alcohólicas y tabaco	3.200
Vestido y calzado	6.370
Vivienda	13.580
Menaje	5.940
Medicina	3.930
Transporte	12.450
Comunicaciones	3.730
Ocio y cultura	6.790
Enseñanza	1.660
Hoteles, cafés y restaurantes	11.640
Otros bienes y servicios	7.100
Total	100.010

Table 22 - CPI 2016 Base. Categories and Weights. Own elaboration

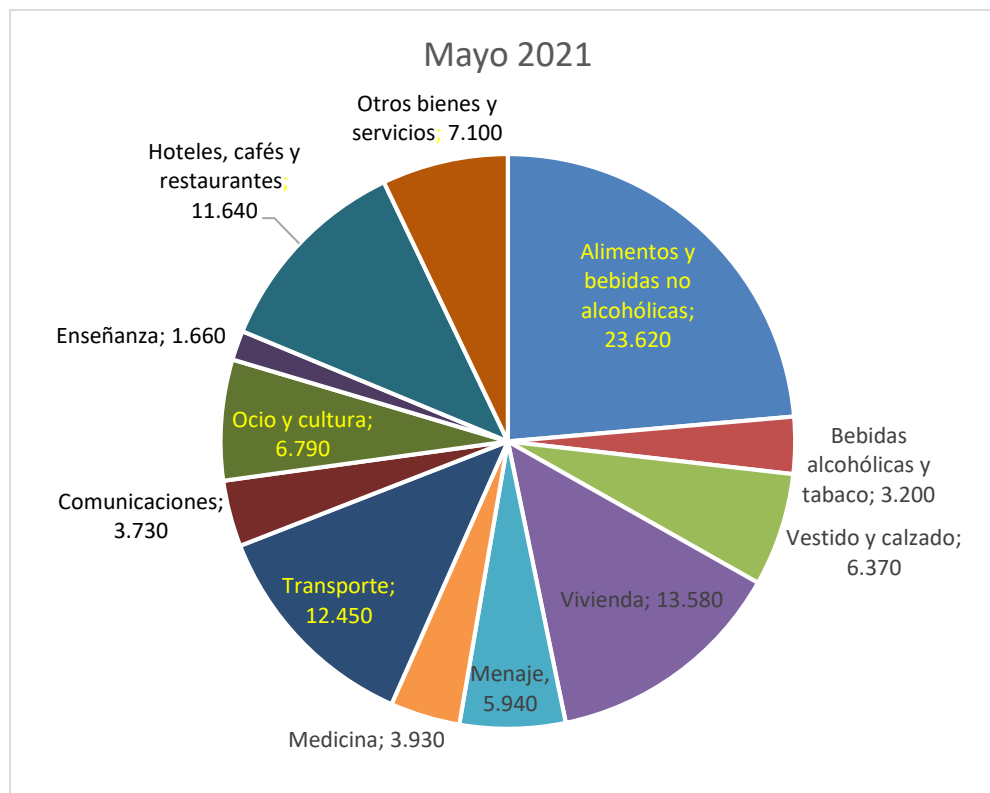


Table 23 - Spanish CPI 2021 Base. Weights distribution. Pie Chart. Own elaboration

The next chart shows the evolution of the number of items taking into consideration in the CPI calculation through the bases:

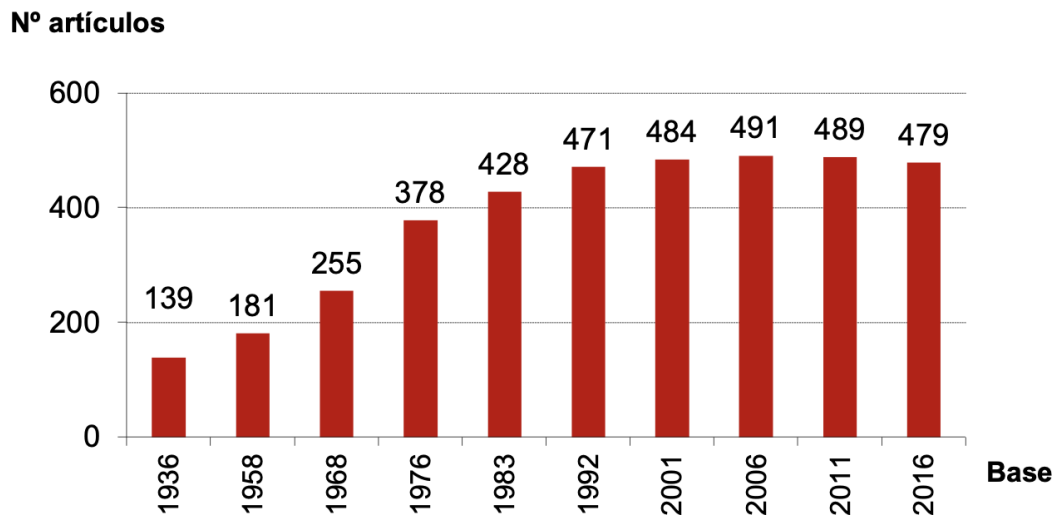


Image 28 - Evolution of number of items through the bases
(https://www.ine.es/metodologia/t25/t2530138_16.pdf)

Currently, the options to improve the methodology and data processing are limited, since any change will have little effect on the result. However, the use of new technologies in information collection processes can make the process more efficient and accurate. Thanks to the computerized collection of information (greater speed, optimization of the use of information, quality control) and the use of the databases of the reporting units (classifications, product monitoring, detection of atypical values, data integration).

Each municipality is made up of geographical zones called census sections. There are a total of 36,000 census tracts. The census sections are grouped according to sociodemographic characteristics. In a first stage, a sample of census sections is drawn, and, in a second stage, a sample of dwellings is drawn within each section. This sample, which will have been duly selected (by means of sampling design techniques) will fulfill the

function of representing the entire population, so that with the sample results it is possible to obtain (estimate), by means of statistical inference, the population results.

Depending on the survey, information is requested from all the residents of the dwelling or from a sample of them. The number of dwellings in the sample depends on the degree of detail of the estimates provided by the sample.

This sample, which will have been duly selected (by means of sampling design techniques) will fulfill the function of representing the entire population, so that with the sample results it is possible to obtain (estimate), by means of statistical inference, the population results.

The items are chosen in such a way that they offer guarantees of staying in the market and that they are regularly consumed by the population.

Thanks to the interviews, a representative sample of what is consumed is obtained and those that aren't, are eliminated. In addition, what they have acquired representation is incorporated. Greater precision is gained.

There may be bias when substituting products since using a fixed base basket defined in the base period does not take into account the substitutions of goods that consumers make in response to price changes.

4.2.3 WHAT ARE THE DIFFERENCES BETWEEN THE SPANISH AND OTHER COUNTRIES' METHODOLOGIES?

In general, the basis of the methodologies are very similar among countries. Calculations of the indices are performed using the chained Laspeyres formula. To obtain the indices for each group, interviews are conducted with a certain number of households over a period of time. The selected households are distributed throughout the territory and the intention is that they be as representative as possible. Items are grouped into groups and

subgroups. Each country has a base year that is set when the chained Laspeyres formula begins to be used. All countries publish results each month of the monthly variation and the variation so far that year. They also publish results from different municipalities and cities in the territory.

The next table shows the comparison of general aspects of the methodologies of Spain, United States, Japan, China, Germany, and France [Table 24]. The cells that have a “-” on them is because the data was not found when looking for it.

We can see that the methodologies are very similar. The differences are the number of prices processed, the amount of items that they consider, the base year for using the chained Laspeyres formula and the number of families that participate in the surveys. The biggest difference is the number of prices processed by each country. China is at the top with 1.2 million prices processed annually, which is 6 times the Spanish one, followed by the United States with almost 1 million prices processed.

The other feature where there is a big difference is the number of families that participate in the surveys. Again, China is at the top with 110,000 participants, which is 5 times the Spanish quantity, followed by the United States with 87,000.

	España	United States	Japan	China	Germany	France
Categories	12	8	10	8	12	12
Items	479	299	582	262	-	1600
Families Participants	24,000	87,000	26,300	110,000	-	-
Processed items	220,000	80,000 (monthly)	-	1.2 million	200,000	160,000
Formula	Chained Laspeyres	Chained Laspeyres	Chained Laspeyres	Chained Laspeyres	Chained Laspeyres	Chained Laspeyres
Base Year	2015	1984	2010	2005	2015	1998

Table 24 - CPI comparison of general characteristics. Own elaboration

4.2.4 IS THE SPANISH METHODOLOGY ALIGNED WITH THE IMF'S RECOMMENDATIONS?

In general, Spain's current methodology complies with the recommendations of the International Monetary Fund.

As for the prerequisites. The legal environment (responsibility for collecting, processing, and disseminating statistics, data sharing with producing agents, confidentiality of individual informant data, and ensuring statistical reporting), the resources employed (staff, facilities, media, funding, and ensuring efficient use), the supervision of user requirements and quality management (quality policy, quality monitoring and quality planning) are aligned with the IMF's recommendations.

As for integrity. Professionalism (impartiality of statistics; selection of sources, methodology, and modes of dissemination; feedback on misinterpretation and misuse of statistics), transparency (disclosure of terms and conditions for collection, processing, and dissemination of statistics; internal government access to statistics prior to publication; advance notice of major changes in methodology, source data, and statistical techniques) and ethical standards are also in line with the IMF.

Regarding the methodology. The concepts and definitions, the scope of the data, the classification/sectorization and the basis for registration (assessment, basis for registration and extrapolation procedures) follow the recommendations.

In terms of accuracy and reliability, data sources (source data collection schedules; definitions, scope, classifications, valuation, and timing of source data recording; timeliness of source data), statistical techniques (of source data and other statistical procedures) and data validation are also aligned.

Regarding the ability to offer the service. Periodicity and timeliness, coherence (internal consistency and temporal consistency) and review (schedule for review and identification of preliminary and revised data) are also aligned.

Finally, regarding accessibility. The data (statistical presentation, media and dissemination formats, publication schedule, simultaneous release) and metadata also follow the IMF's recommendations.

4.3 METRIC ANALYSIS

This chapter focuses on the metric and the main uses of it in Spain, the shortcomings it has, whether it is a good metric to measure inflation or not, and whether it shows the reality of all social groups or not.

4.3.1 WHAT ARE THE MAIN USES OF THE INFLATION METRIC IN SPAIN?

In theory, the index is mainly used to readjust rents, credits, salaries and wages, and different public and private contracts. It is also used to readjust various rates regulated by the authority such as basic services, for example, electricity and drinking water, collective locomotion, among others.

However, the reality is very different. The main example is salaries, they do not follow the inflation rate. They are usually adjusted months later and not as much as the percentage of inflation. Moreover, in Spain, the same thing happens with the retirement pension. They don't follow inflation.

The next image shows a graph that compares wages growth vs inflation rate:

Comparing inflation and wage growth

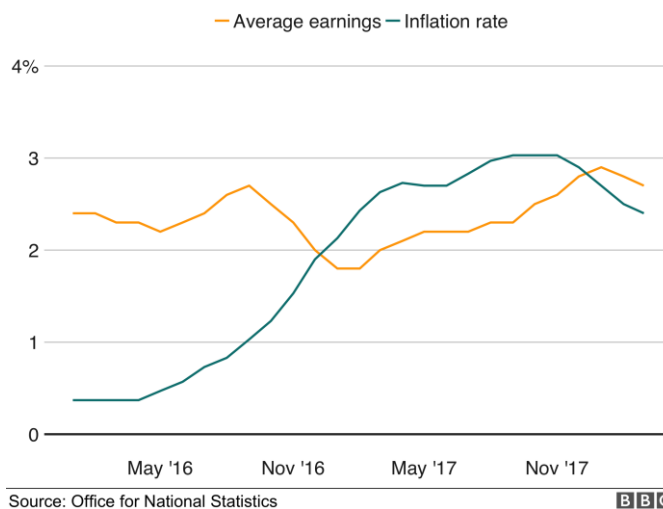


Image 29 - Wages growth vs Inflation (<https://www.bbc.com/news/business-44857055>)

As you can see in the image [Image 29], when inflation rose back in 2017, the wages maintained its growth. As a result, wages grew slower than inflation.

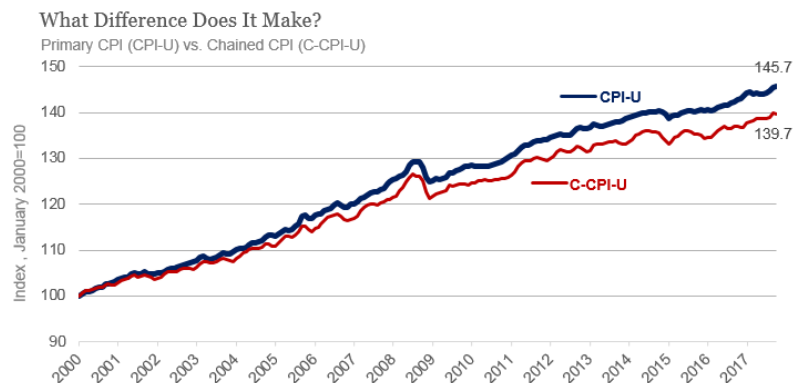
What follows inflation is the prices of food, electricity, and gas, among others. The generic rule is that people’s income is not as correlated to inflation as the things people spend their income on.

4.3.2 AREAS OF IMPROVEMENT

This section will focus on aspects that could be improved and understate inflation. In general, most of the changes applied to the metric are not justified or the justification hides something behind it.

4.3.2.1 *Chained Laspeyres formula*

The chained consumer price index was introduced in the 2001 base and the difference with the previous formula, fixed Laspeyres, is that they added expenditure weights. This change at first, doesn’t seem to harm the outcome of the metric. However, it understates inflation as you can see in the next image where inflation has been calculated using both formulas, the red line is the fixed CPI and the blue line is the Chained CPI.



Source: FRED, BLS.

Image 30 - Chained CPI vs CPI

This proves that the formula used to measure inflation, understates it.

4.3.2.2 The products of the basket don't show the real consumption habits

Great part of the used cars market is excluded.

In the 2016 Base, the definition of the field of consumption was changed based on the concept of final consumption expenditure of households in the European System of Accounts (ESA). Among other changes, the concept of consumption has to include only transactions between households and the business sector, not those between households. This affects the estimation of spending and weighting for used cars, since only transactions between households and businesses are taken into account.

Moreover, the prices of used cars have skyrocketed in the last two years as you can see in the next chart.

Used car prices are up nearly 40% during the pandemic

Change in prices since March 1, 2020

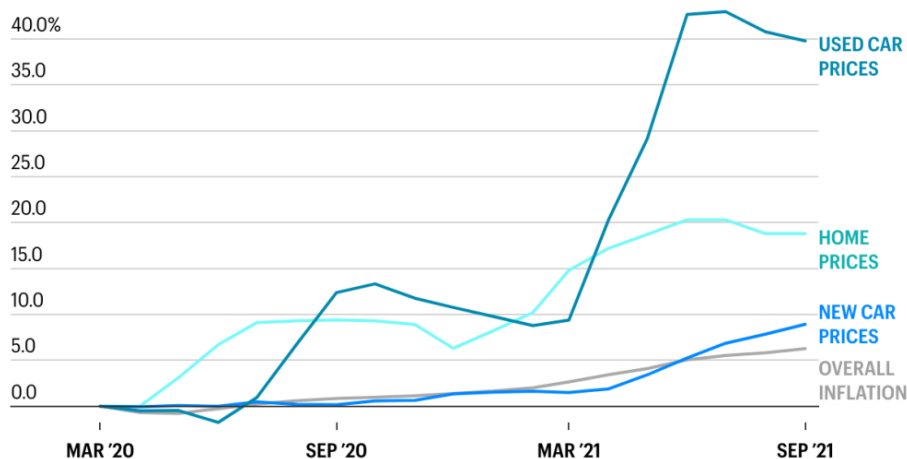


CHART: LANCE LAMBERT • SOURCE: U.S. BUREAU OF LABOR STATISTICS, REALTOR.COM

FORTUNE

Ilustración 1 - Used cars price rates (2020-2021). (<https://fortune.com/2021/11/01/used-car-prices-high-carmax-2021/>)

The solution would be to include used car purchase prices from both companies and individuals in the metric. To obtain both averages, both for the prices of second-hand cars from individuals and companies, the prices on the website coche.com will be consulted. Specifically, for used cars of a particular: <https://www.coches.com/coches-segunda-mano/coches-segunda-mano-particulares.htm>. And for enterprise cars: <https://www.coches.com/coches-segunda-mano/coches-ocasion.htm>

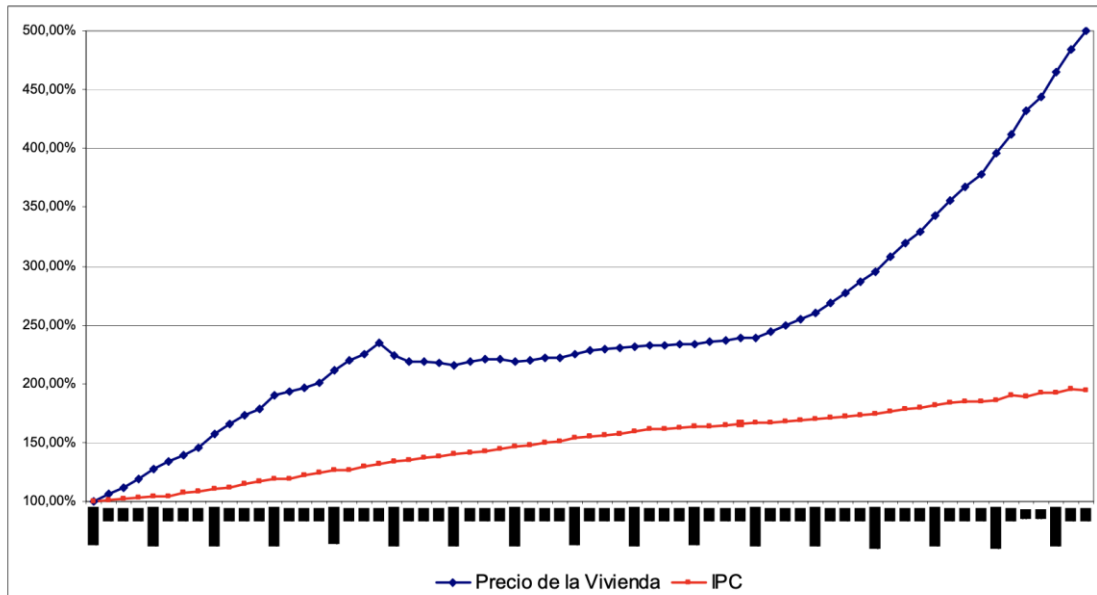
4.3.2.3 The weights do not reflect the real consumption habits

Housing only weights 11% of the entire expenditure.

From 1993 to 2001, both the cost of home ownership (imputed rent) and rental housing were considered. The first with a contribution to the index (1992 base) equivalent to 5.3051% and the second with 1.3947%, representing the set of costs related to housing at 10.2803%. Subsequently, in 2002, the changes in the Consumer Price Index eliminated the consideration of home ownership, compensating for said disappearance with the increase in the contribution of residential rent to the CPI, until it stood at 2.2073%. Housing represented 11.0259% of the CPI. However, this apparent relative revaluation should be qualified by the fact that residential prices went from weighing 6.6998% (rent and property) to represent only 2.2073% (rent only). Housing, therefore, has a certainly reduced weight in the calculation of the CPI. Certainly, a disproportionate situation given the importance of this concept of spending in family economies.

In the following image you can see that from 1987 to 2003, real estate prices increased by 400% while the CPI increased by 95%.

Figura nº 1: Evolución de los precios de la vivienda y del IPC (1987-2003)

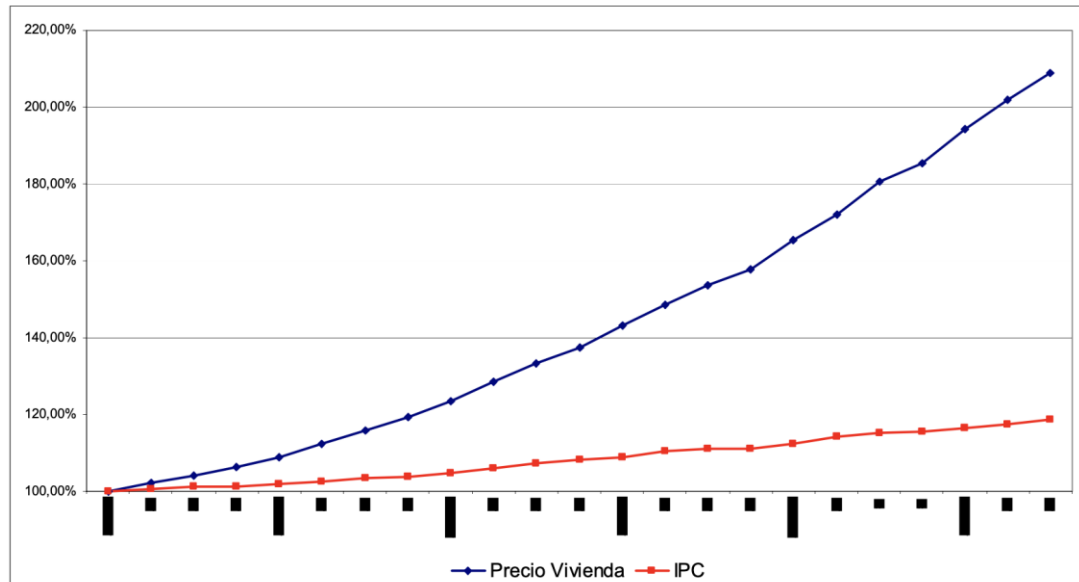


Fuente: Ministerio de Fomento (precios vivienda) e INE (IPC). Elaboración propia

Image 31 - Housing Inflation and CPI compared (1987-2003). Source:

This difference became apparent between 1998 and 2003. House prices increased by 109%, yet the CPI increased by only 18.7% [Image 32].

Figura nº 2: Evolución de los precios y de la vivienda (1998-2003)

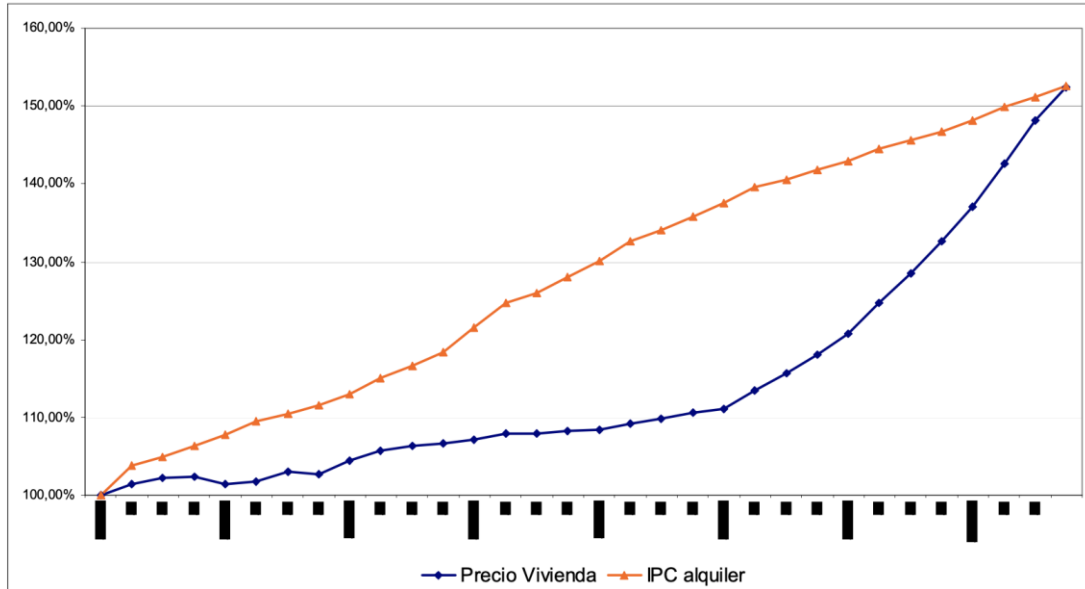


Fuente: Ministerio de Fomento (precios vivienda) e INE (IPC). Elaboración propia

Image 32 - Housing Inflation and CPI compared (1998-2003). Source:

If we analyze the relationship between house prices and the house price index, we see that they have a negative correlation. Between 1993 and 1998, the CPI for housing rose more sharply when house prices stagnated. On the other hand, between 1998 and 2000, the housing CPI smooths out when house prices rise more sharply.

Figura nº 3: Precio de la vivienda e índice de alquiler (1993 – 2000)



Fuente: Ministerio de Fomento (precios vivienda) e INE (IPC alquiler). Elaboración propia

Image 33 - Housing Prices compared to Housing category CPI (1993-2000). Source:

The solution is to take into account both home ownership and rental housing. Living in a house implies living in it, and living in it implies a cost, regardless of whether the (main) home is rented or owned.

4.3.2.4 The metric does not represent the cost of living of all social groups

Below is a chart from research conducted by the University of Michigan that represents the impact of inflation based on income:

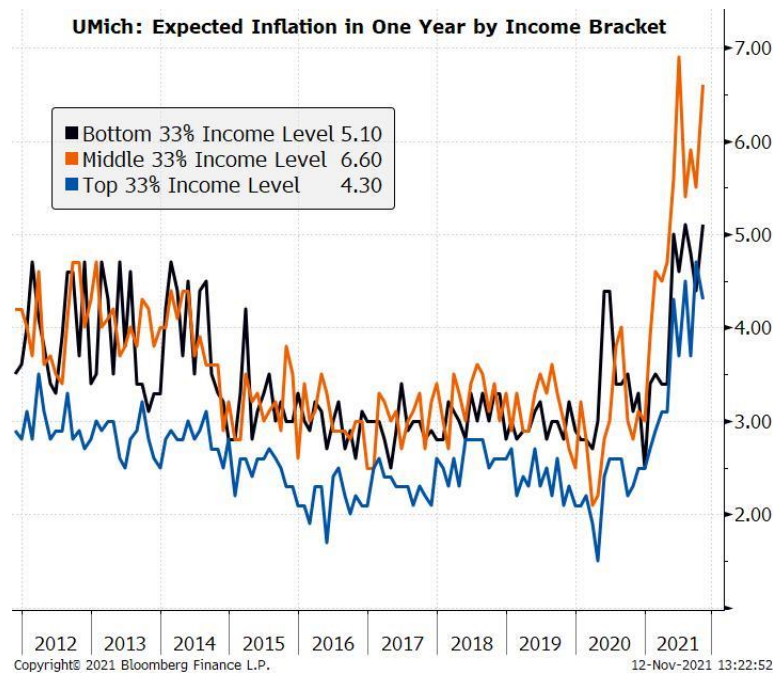


Image 34 - Expected Inflation in One Year by Income Bracket. Source: University of Michigan

As seen on the chart, over the last 10 years, inflation has affected the lower and middle class more, with the middle class being the most affected in the last two years. The upper class is the least affected by inflation.

However, the CPI shows a fixed increase in prices without taking into account social groups.

The cause is that in the CPI, for the calculation of the indices, the products that are most consumed at the national level are used. It does not take into account the consumption habits of each social group separately.

The solution would be to carry out independent CPI measurements taking municipalities of the corresponding social class as a sample and taking into account their consumption habits and the quality of the products they consume. Differentiating between upper, middle, and lower class.

4.3.2.5 Changes in quality and substitution of goods of the basket

Substitutions can be applied in different scenarios:

- The article ceases to be representative and another more representative appears on the market.
- The article disappears from the market.
- The store where the price of the item is collected ceases to be representative, closes or changes its economic activity.

Starting with the 2011 CPI base, the most currently used quality adjustment methods were implemented. Those are the following:

- Total quality adjustment. It starts from the assumption that the difference between the price of the substituted article and the substitute article is entirely motivated by the difference in quality between the two or that the articles are so different that they cannot be compared.

It is considered, then, that the price difference between the two items is due solely to their different quality, so the index will not reflect price variation. With this adjustment it is assumed that if the replaced item had remained on sale, its price would not have changed.

- Identical quality adjustment. It is based on the idea that the substitute item has the same quality as the replaced item, that is, that the price difference between the two is due to a real price variation. With this adjustment it is assumed that if the replaced item had remained on sale, its price would have been the same as that of the substitute item.
- Other options. Other options prices, production costs, imputation prices, expert feedback, overlap prices and hedonic regression methods.

The most widely used methods to solve discontinuities in price evolution are based on the information provided by the experts (who they are or the criteria they use to choose them is not mentioned), overlapping prices and option prices.

The cause is that the precision with which the CPI measures the evolution of prices depends, among other factors, on the stability of the characteristics that define the selected products. The objective is to monitor the prices of a basket whose products remain unchanged over time, so that the measurement of the variation in prices does not depend on changes in the characteristics of those products.

The solution is to avoid quality adjustments unless the product is completely different after the quality change. In this way, the real increase in the prices of the items that are consumed is seen. Although the product improves in some aspect, the effort that the citizen has to make to be able to buy it is greater.

As for the substitution of an item, it would only be taken into account if it disappears from the market or ceases to be consumed. On the contrary, if, for example, the price of beef goes up and you begin to buy meat of lower quality for the same price as beef, and the item is replaced, reflecting that there has been no increase in beef price; certain price increases would not be taken into account. This causes the CPI result to be lower than reality. It also does not take into account the change in the standard of living of citizens who will now eat lower quality products.

4.3.2.6 The unit of measurement of the metric is not stable over time, so the results are distorted.

Price changes of the metric are measured with the value of the euro.

When the value of the euro falls, people have less purchasing power and cannot buy the same products as before. Therefore, they are replaced by other lower quality products that they can afford.

Prices are the result of purchasing decisions which, in turn, depend on prices.

This is one of the hardest problems to solve. The approach to solve the problem would be to use another calculation basis such as gold.

4.3.2.7 Lack of transparency and government influence

The Instituto Nacional de Estadística is a public organization and, as such, they are under the influence of the Government. In this case, they are influenced directly from the Government. It is controlled by a person who was an affiliate to the current Government.

In addition, there is lack of transparency in several aspects of the methodology. The criteria with which they choose the families that will participate in the Encuesta de Presupuestos Familiares is not defined. They mention that the families they choose are representative of all the population. They do not state how they choose them. They also don't share any information about the families chosen to protect their privacy.

Moreover, they mention that the basket of goods is formed by the most consumed products, but they do not state how those goods are selected and who selects them.

Lastly, for the quality adjustments, one of the most used methods is the information provided by the experts. They don't mention who these experts are and the criteria they use to judge whether a change in quality happens or not.

4.3.3 IS THE CPI A GOOD METRIC TO MEASURE THE COST OF LIVING IN GENERAL TERMS?

As stated in the Areas of Improvement chapter, there are multiple features that could be improved and others that could be implemented. There are multiple areas that understate inflation and all of them added together understate inflation to a greater degree. The main problem is that it is a very generic index, and it probably does not show the exact rise of the cost of living of anyone. It is an approach.

Given the importance of the metric, it should focus on changing the methodology to make it more customizable for different groups of population like social groups, different regions with their local products. There should be multiple metrics depending on these groups. Each metric should focus on the group they are targeting and be independent from the other metrics.

Chapter 5. PRACTICAL APPLICATION

In this chapter, an independent methodology will be defined and tested to later compare the results with the ones published by the Government. First, there will be a theoretical solution, the methodology, and then a practical one, a website where the results are published.

5.1 INDEPENDENT METHODOLOGY DEFINITION

The aim of this project is to create a censorship-resistant inflation feed more accurate than the one that the Government publishes. In order to do so, the following procedure will be carried out.

First, web scraping techniques will be used to monitor online prices and store them in a public database. Then, those prices will be used to calculate the indices of the different categories. Finally, the results along with the all the information about the methodology will be published on a website.

The main attributes of the methodology are:

- Independent. It is not supervised or regulated by the government or other public bodies.
- Open source. Available for everyone with complete transparency. The information about the methodology, the databases, and the sources.
- Censorship-resistant. Decentralized and unalterable.

Moreover, in this metric, there will be some items and aspects that will be considered.

Used cars sold by individuals, not only used cars sold by companies.

Home ownership, not only rent. This is because living in a house implies living in it, and living in it implies a cost, regardless of whether the (main) home is rented or owned.

There will not be substitution of products unless they disappear from the market. As stated before, if the price of beef goes up and you begin to buy meat of lower quality for the same price as beef, and the item is replaced, reflecting that there has been no increase in beef price; certain price increases would not be considered. This results in a lower inflation rate and it would not reflect that the standard of living will be worse due to eating less quality food.

There will not be quality adjustments unless the product is completely different after the quality changes. In this way, the real increase in the prices of the items that are consumed is seen. Although the product improves in some aspect, the effort that the citizen has to make to be able to buy it is greater.

Like in the Billion Prices Project, to calculate the indices, a price index will be created for every category using a geometric average of the changes of the prices. Then aggregated price index will be calculated using the weighted arithmetic average with the official weights of each category published by the Government. Even though we have seen in the previous chapters that these weightings understate inflation, the lack of data to compute the weightings makes it impossible to calculate them accurately. It is preferred to use the Government ones right now to compare both metrics than having very inaccurate weightings. This will be explained in Chapter 8.

5.2 CONCEPTUAL AND TECHNOLOGICAL DESIGN

For the web application development, the Agile methodology will be used with sprints of 2 to 4 weeks to increase productivity and obtain higher quality results. At the end of a sprint, there will be an increase in product that will add value to the final result.

The webs application has been developed following the client-server architecture. It has a front-end which establishes a connection to the back-end via a network. This back-

end has a server which contains the program which connects to the database server which contains all the information of the products and their prices.

Inside the client-server architecture, we have decided to use the three-tier architecture. These three tiers are the front-end, the middle-end and the back-end. The front-end has the presentation logic and interacts with the user, the back-end stores the data and the middle-end contains the business logic of the application.

The MVC architecture was also implemented to separate the business logic part of the application from the visual part. This means that changes in one part will not affect the other. MVC is explained in chapter 2.4.

Using web scraping techniques online prices are obtained and then stored in the database server. The data is processed by the middle-end server and calculates the new metric. Then these results are shown in the front-end through a website.

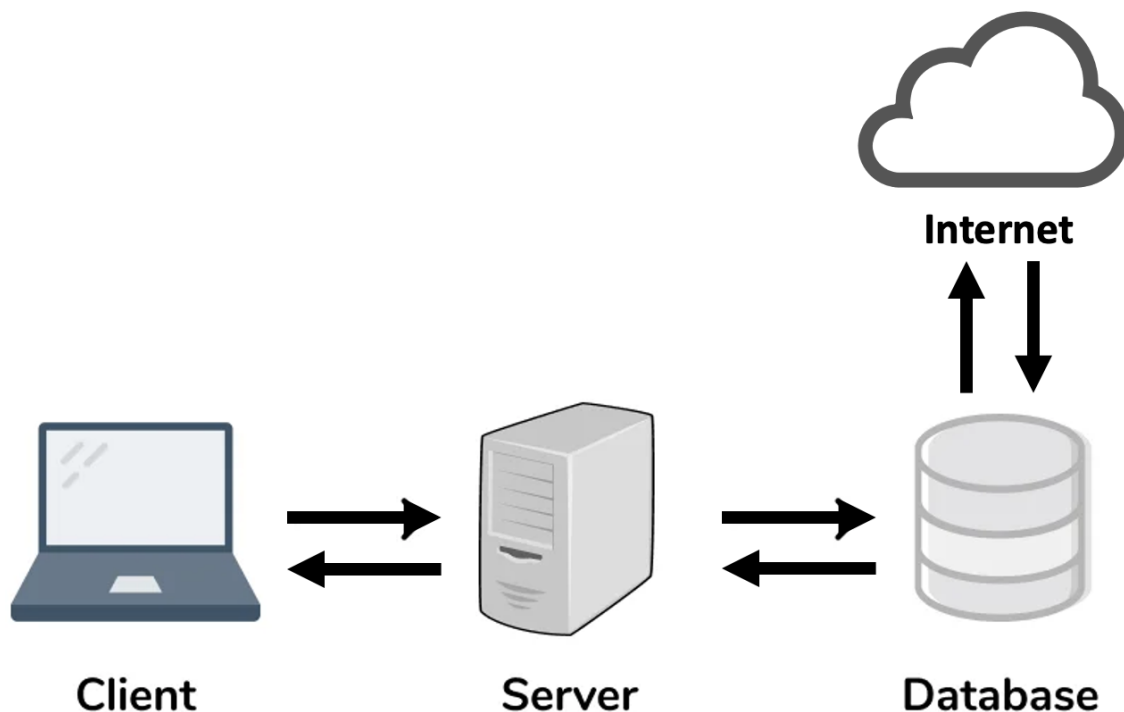


Image 35 - Architecture of the web application. Own elaboration

For the development of the web application a Spring framework was used (explained previously in chapter 2.3) to focus on the application logic instead of spending a lot of time building the framework.

For the web scraping module, Python programming language was chosen for simplicity and velocity.

For the database server and management, MySQL was chosen. MySQL has a client-server architecture like the web application and manages relational databases. Lets the user query using SQL format which was implemented in the Spring framework. MySQL is explained in chapter 2.5.

5.3 IMPLEMENTATION

In this section, there will be a deep-dive explanation of the different parts of the architecture and how they were programmed.

5.3.1 WEB SCRAPING MODULE DEVELOPMENT

The script for web scraping was programmed in Python programming language. The python modules used in the program are Pandas, BeautifulSoup (bs4), Request, Pymysql and Sqlalchemy, and datetime.

```
import requests
import bs4
import pandas as pd
import pymysql
from sqlalchemy import create_engine
from datetime import date
```

As explained in the chapter 2.6, the process is the following:

- Gets the table that already exists in the database

```
sqlEngine = create_engine('mysql+pymysql://root:12345@localhost/tfg')
dbConnection = sqlEngine.connect()
df2 = pd.read_sql("select * from Precios", dbConnection);
dbConnection.close()
```

- The HTML file of the website is obtained through an HTTP GET method.

```
URL = "https://soysuper.com/c/lacteos-y-huevos/leche?page=1#products"
response = requests.get(URL)
```

- The file is parsed for it to be examined and easy to use.

```
soup = bs4.BeautifulSoup(response.text, "lxml")
```

- Searches for the label that the user wants and gets the information in that label.

```
cases = soup.find_all("span", class_="unitprice")
cases2 = soup.find_all("span", class_="productname")

for i in cases:
    prices.append((i.text).strip())

for j in cases2:
    products.append((j.text).strip())
    if len(products) == len(prices):
        break
```

- Stores the data in a dataframe with its corresponding category and the date it was obtained.

```
today = date.today()
```

```
df = pd.DataFrame({"Price": prices, "Name of the product": products,  
"Group": "Alimentos", "Date": today})
```

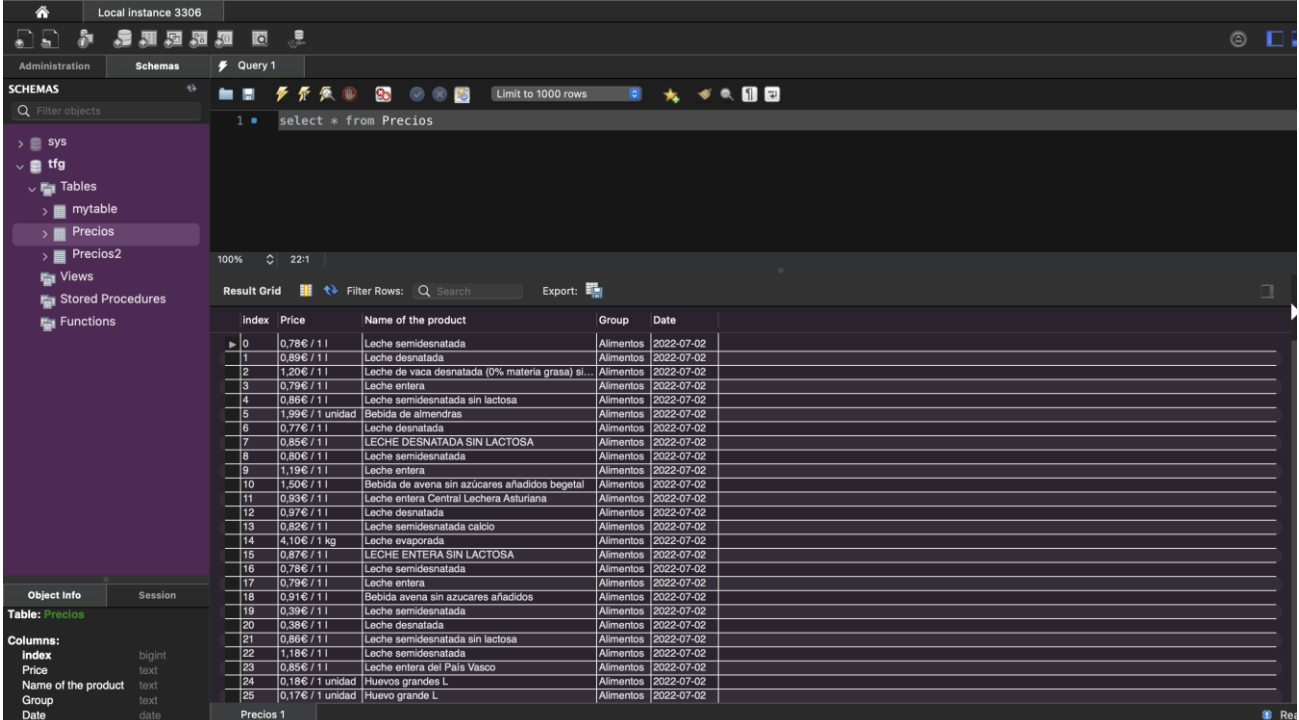
- Appends the old dataframe with the new one.

```
df3 = df2.append(df, ignore_index=True)
```

- The dataframe is stored directly into MySQL database

```
tableName = "Precios"  
  
sqlEngine = create_engine('mysql+pymysql://root:12345@localhost/tfg')  
  
dbConnection = sqlEngine.connect()  
  
try:  
    frame = df3.to_sql(tableName, dbConnection, if_exists='fail');  
except ValueError as vx:  
    print(vx)  
except Exception as ex:  
    print(ex)  
else:  
    print("Table %s created successfully."%tableName);  
finally:  
    dbConnection.close()
```

The database in MySQL looks like the following:



index	Price	Name of the product	Group	Date
0	0.78€ / 1 l	Leche semidesnatada	Alimentos	2022-07-02
1	0.89€ / 1 l	Leche desnatada	Alimentos	2022-07-02
2	1.20€ / 1 l	Leche de vaca desnatada (0% materia grasa) si...	Alimentos	2022-07-02
3	0.79€ / 1 l	Leche entera	Alimentos	2022-07-02
4	0.86€ / 1 l	Leche semidesnatada sin lactosa	Alimentos	2022-07-02
5	1.99€ / 1 unidad	Bebida de almendras	Alimentos	2022-07-02
6	0.77€ / 1 l	Leche desnatada	Alimentos	2022-07-02
7	0.85€ / 1 l	LECHE DESNATADA SIN LACTOSA	Alimentos	2022-07-02
8	0.80€ / 1 l	Leche semidesnatada	Alimentos	2022-07-02
9	1.19€ / 1 l	Leche entera	Alimentos	2022-07-02
10	1.50€ / 1 l	Bebida de avena sin azúcares añadidos vegetal	Alimentos	2022-07-02
11	0.93€ / 1 l	Leche entera Central Lechera Asturiana	Alimentos	2022-07-02
12	0.97€ / 1 l	Leche desnatada	Alimentos	2022-07-02
13	0.82€ / 1 l	Leche semidesnatada calcio	Alimentos	2022-07-02
14	4.10€ / 1 kg	Leche evaporada	Alimentos	2022-07-02
15	0.87€ / 1 l	LECHE ENTERA SIN LACTOSA	Alimentos	2022-07-02
16	0.78€ / 1 l	Leche semidesnatada	Alimentos	2022-07-02
17	0.79€ / 1 l	Leche entera	Alimentos	2022-07-02
18	0.91€ / 1 l	Bebida avena sin azúcares añadidos	Alimentos	2022-07-02
19	0.39€ / 1 l	Leche semidesnatada	Alimentos	2022-07-02
20	0.38€ / 1 l	Leche desnatada	Alimentos	2022-07-02
21	0.86€ / 1 l	Leche semidesnatada sin lactosa	Alimentos	2022-07-02
22	1.18€ / 1 l	Leche semidesnatada	Alimentos	2022-07-02
23	0.85€ / 1 l	Leche entera del País Vasco	Alimentos	2022-07-02
24	0.18€ / 1 unidad	Huevos grandes L	Alimentos	2022-07-02
25	0.17€ / 1 unidad	Huevo grande L	Alimentos	2022-07-02

Image 36 - MySQL database table with raw data. Own elaboration

5.3.2 BACK-END DEVELOPMENT

The database is designed in SQL following a relational model to avoid data redundancy. There are 2 tables, one has all the raw data and the other one has the categories with their rates calculated.

The table with the raw data has 4 columns: price, product name, group, and date. This table is later used to calculate the inflation rate and the categories rates. This table is the result of the web scraping script described previously.

The other table is the result of calculating the inflation rate and the categories rates. It has 2 columns: category name and rate. This is the one that the front end reads and shows. The calculus of the metrics are done independently with python. It is done this way in order to have a faster website without bugs.

To connect the database to Spring Boot, the following steps were taken:

- Add the MySQL dependencies to the file pom.xml

```
<dependency>
  <groupId>mysql</groupId>
  <artifactId>mysql-connector-java</artifactId>
</dependency>
```

- Add the Spring Boot MySQL connection configuration

```
spring.datasource.url=jdbc:mysql://localhost:3306/tfg
spring.datasource.username=root
spring.datasource.password=12345
```

- Create a Repository class for Spring Boot with the query

```
package com.practica.TFG.repository;

import com.practica.TFG.model.Category;
import org.springframework.data.repository.CrudRepository;
import org.springframework.data.jdbc.repository.query.Query;
import org.springframework.data.repository.query.Param;

import java.util.List;

public interface CategoryRepository extends CrudRepository<Category,
Integer>{

}
```

- Add a Java class for each category with two attributes, group name and rate.

```
package com.practica.TFG.model;

import lombok.Data;
import org.springframework.data.annotation.Id;
import org.springframework.data.relational.core.mapping.Column;

import javax.persistence.Entity;
import javax.persistence.GeneratedValue;
import javax.persistence.GenerationType;
import javax.persistence.Table;
```

```
@Data
public class Category {
    @Id
    private String group_name;
    private float rate;
}
```

- Create the service for the categories and the interface. This part has the business logic of the API.

```
package com.practica.TFG.service.impl;

import com.practica.TFG.model.Category;
import com.practica.TFG.service.CategoryService;
import com.practica.TFG.repository.CategoryRepository;
import org.springframework.beans.factory.annotation.Autowired;
import org.springframework.stereotype.Service;

import java.util.ArrayList;
import java.util.Iterator;
import java.util.List;

@Service
public class CategoryServiceImpl implements CategoryService{

    @Autowired
    CategoryRepository categoryRepository;

    @Override
    public List<Category> getCategories() {
        Iterable<Category> categories = categoryRepository.findAll();

        Iterator<Category> it= categories.iterator();

        ArrayList<Category> list_categories= new ArrayList<Category>();

        while(it.hasNext()) {
            list_categories.add(it.next());
        }

        return list_categories;
    }
}
```

```
package com.practica.TFG.service;

import com.practica.TFG.model.Category;
```

```
import java.util.List;

public interface CategoryService {

    List<Category> getCategories();

}
```

- Create the Controller to handle the requests and add the API direction.

```
package com.practica.TFG.controller;

import com.practica.TFG.model.Category;
import com.practica.TFG.repository.CategoryRepository;
//import com.practica.TFG.service.CategoryService;
import com.practica.TFG.service.CategoryService;
import org.springframework.beans.factory.annotation.Autowired;
import org.springframework.http.HttpStatus;
import org.springframework.http.ResponseEntity;
import org.springframework.stereotype.Controller;
import org.springframework.web.bind.annotation.*;

import java.util.List;

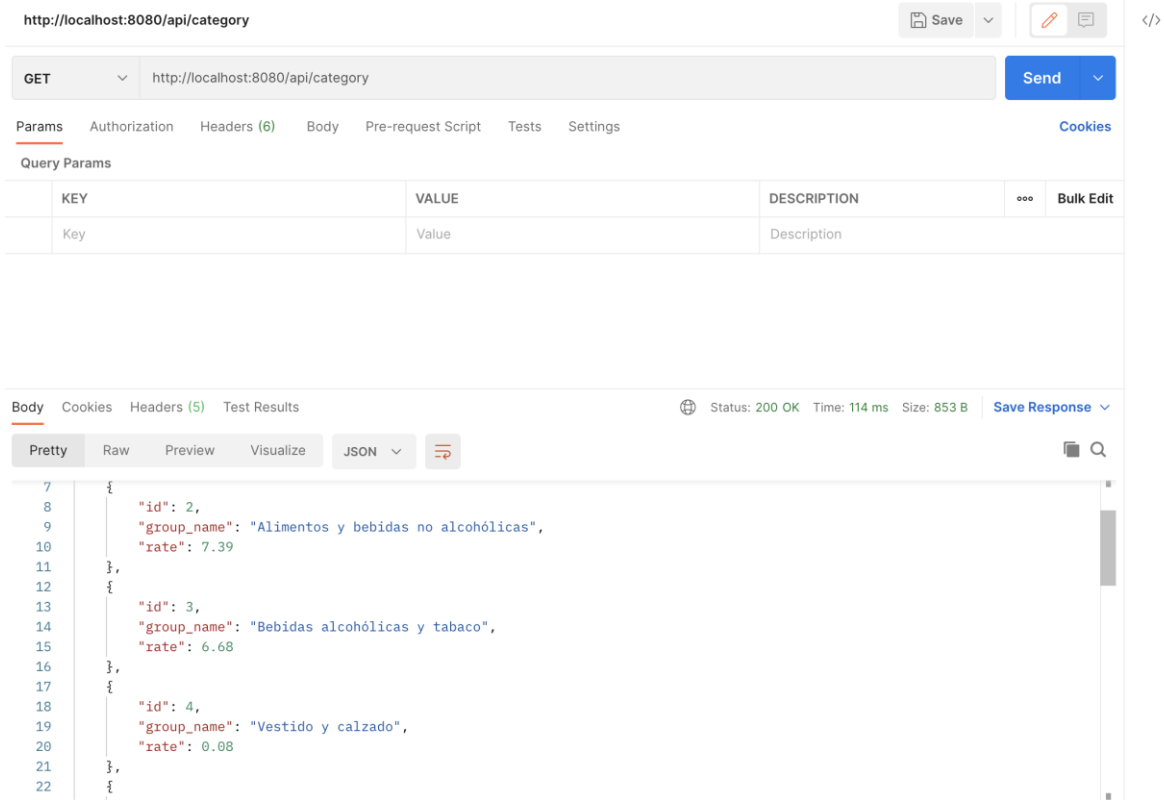
@Controller
@RequestMapping("/api/category")
public class CategoryController {

    @Autowired
    CategoryService categoryService;

    @GetMapping()
    public @ResponseBody Iterable<Category> categories() {
        return categoryService.getCategories();
    }

}
```

- Test with Postman if the connections are working.



The screenshot shows a REST client interface. The URL bar contains `http://localhost:8080/api/category`. The request method is `GET`. The response status is `200 OK` with a time of `114 ms` and a size of `853 B`. The response body is displayed in JSON format:

```

7   {
8     "id": 2,
9     "group_name": "Alimentos y bebidas no alcohólicas",
10    "rate": 7.39
11  },
12  {
13    "id": 3,
14    "group_name": "Bebidas alcohólicas y tabaco",
15    "rate": 6.68
16  },
17  {
18    "id": 4,
19    "group_name": "Vestido y calzado",
20    "rate": 0.08
21  },
22  }

```

Image 37 - Test of the Web application API. Own elaboration

5.3.3 MIDDLE-END DEVELOPMENT

To calculate the indices, a python script was created that calculates the mean of prices of each category of two different timestamps introduced by the user. The user writes in the corresponding place the dates from which he wants to know the price difference. The program calculates the percentage of difference in prices for each category between the dates introduced and then stores the results in the MySQL database. The script is the following.

```

from shutil import register_unpack_format
import pandas as pd
import pymysql
from sqlalchemy import create_engine

sqlEngine = create_engine('mysql+pymysql://root:12345@localhost/tfg')
dbConnection = sqlEngine.connect()

```

```
df = pd.read_sql("select * from Precios", dbConnection);
dbConnection.close()

df2 = df[df["Date"]==XX] #date format: yyyy-mm-dd
grouped_df2 = df2.groupby("Group")
mean_df2 = grouped_df2.mean()

#print(mean_df2)

df3 = df[df["Date"]==XX] #date format: yyyy-mm-dd
grouped_df3 = df3.groupby("Group")
mean_df3 = grouped_df3.mean()

#print(mean_df3)

result_df = mean_df3.subtract(mean_df2).div(mean_df2).mul(100)

#print(result_df)

tableName = "Results"

sqlEngine = create_engine('mysql+pymysql://root:12345@localhost/tfg')

dbConnection = sqlEngine.connect()

try:
    frame = result_df.to_sql(tableName, dbConnection, if_exists='fail');
except ValueError as vx:
    print(vx)
except Exception as ex:
    print(ex)
else:
    print("Table %s created successfully."%tableName);
finally:
    dbConnection.close()
```

The database stored in the server is the one that the front-end requests and shows.

Moreover, the Service created in the previous section for the database connection, also has part of the business logic. It defines the function of the API to obtain the data of the different categories.

5.3.4 FRONT-END DEVELOPMENT

The front-end consists of 3 parts: home page with the results, a methodology description, and a comparison with the inflation rates published by the Government.

The home page shows each category with their respective inflation rate and the general inflation rate.

Real Inflation

[Home](#)

[Methodology](#)

[Comparison](#)

General index: 5.2
Alimentos y bebidas no alcohólicas: 8.1
Bebidas alcohólicas y tabaco: 3.1
Vestido y calzado: 0.1
Vivienda: 2.9
Menaje: 4.4
Sanidad: 0.6
Transporte: 11.5
Comunicaciones: 0.3
Ocio y cultura: -0.9
Enseñanza: 0
Hoteles, cafés y restaurantes: 6.4
Otros bienes y servicios: 4.1

Image 38 - Web Application Home Page

The methodology page contains the details of the metric, the general characteristics of it and the data sources.

Real Inflation

[Home](#) [Methodology](#) [Comparison](#)

General Characteristics

Independent

It is not supervised or regulated by the government or other public bodies.

Open source

Available for everyone with complete transparency. The information about the methodology, the databases, and the sources.

Censorship-resistant

Decentralized and unalterable

The aim of this project is to create a censorship-resistant inflation feed more accurate than the one that the Government publishes. In order to do so, the following procedure will be carried out.

First, web scraping techniques will be used to monitor online prices and store them in a public database. Then, those prices will be used to calculate the indices of the different categories. Finally, the results along with the all the information about the methodology will be published on a website.

The main attributes of the methodology are:

- Independent. It is not supervised or regulated by the government or other public bodies.
- Open source. Available for everyone with complete transparency. The information about the methodology, the databases, and the sources.
- Censorship-resistant. Decentralized and unalterable.

Moreover, in this metric, there will be some items and aspects that will be considered.

Used cars sold by individuals, not only used cars sold by companies.

Image 39 - Web Application Comparison Page

The comparison page shows a table with the comparison of the metric calculated and the official inflation rate published by the government.

May 2022

Groups	Official Inflation	Real Inflation	Difference
General Index	4,1	5,2	1,1
Alimentos y bebidas no alcohólicas	7,5	8,1	0,6
Bebidas alcohólicas y tabaco	2,3	3,1	0,8
Vestido y calzado	-0,9	0,1	1,0
Vivienda	0,9	2,9	2,0
Menaje	4,4	4,4	0,0
Sanidad	0,6	0,6	0,0
Transporte	9,8	11,5	1,7
Comunicaciones	0,5	0,3	-0,2
Ocio y cultura	0,3	-0,9	-0,6
Enseñanza	0,0	0,0	0,0
Hoteles, cafés y restaurantes	4,2	6,4	2,2
Otros bienes y servicios	2,3	4,1	1,8

Source: [INE](#).

Spanish Inflation - May 2022

Image 40 - Web Application Comparison Page

The front-end makes the request to the API created in the previous section through a JavaScript program. It makes a request using the HTTP method GET. Calls the API function, gets the information and passes it to the HTML. The script is the following.

```
document.addEventListener("DOMContentLoaded", function(event) {

    return categories();
});

function categories() {

    try{

        const direction = '/myapplication/api/category';
        fetch(direction, {
            method: 'GET',
            headers: {
                'Content-Type': 'application/json'
            }
        })
```



```

.then(response => {
  if(response.status==200){
    return response.json();
  }else {
    alert("Categories not found");
  }
})
.then(r => {
  console.log(r);
  //alert("Categories loaded");
  let html = '';
  r.forEach(function(category,i){
    html += '<div class="row-lg-6">'
      + '<div class="col-lg-6">'
      + '<div class="card" style="width:
40rem;">'
      + '<div class="card-body">'
      + `<h5 class="card-title"
id="category"> <b>${category.group_name}</b> <b></b>
<b>${category.rate}</b></h5>`
      + '</div>'
      + '</div>'
      + '</div>'
    })
    document.getElementById("list-output").innerHTML =
html;
  })

}catch (e) {
  console.error(e.message);
}

return false;
}

```

5.4 HOW TO MAKE A CENSORSHIP-RESISTANT METRIC?

There are different approaches when it comes to making the metric censorship-resistant. The aspect that makes a metric censorship-resistant is transparency and making all the information auditable. This means that every change made in the formulas or in the databases are tracked and need to be explained.

Having a public database that everyone can check, and track is also necessary because the moment data is hidden any manipulation can happen. For example, if a product has risen its price by a lot, with a private database, one could exclude that item from the metric in order to get lower inflation rates.

Moreover, stating all the sources where the data in the metric comes from. This is important so that anyone can check if those prices are correct and keep track of them. Publishing the data sources also makes the metric more reliable because one can show that he has nothing to hide.

Another of gaining transparency is to store the data in a decentralized database using blockchain technology. This will make it auditable and the data stored in it could not be removed by a particular due to the proof of stake, which makes it impossible for an individual to change the ledger or information stored in a blockchain. In the past years, decentralized apps (DApps) have gained popularity and people are starting to build their applications on a blockchain to obtain its benefits like security. Those blockchains work as a decentralized database in which people store their programs.

Chapter 6. RESULTS ANALYSIS

The next table shows the comparison of the results obtained and the official inflation rate published by the government, and the difference between them.

May 2022

Groups	Official Inflation	Real Inflation	Difference
General Index	4,1	5,2	1,1
Alimentos y bebidas no alcohólicas	7,5	8,1	0,6
Bebidas alcohólicas y tabaco	2,3	3,1	0,8
Vestido y calzado	-0,9	0,1	1,0
Vivienda	0,9	2,9	2,0
Menaje	4,4	4,4	0,0
Sanidad	0,6	0,6	0,0
Transporte	9,8	11,5	1,7
Comunicaciones	0,5	0,3	-0,2
Ocio y cultura	0,3	-0,9	-0,6
Enseñanza	0,0	0,0	0,0
Hoteles, cafés y restaurantes	4,2	6,4	2,2
Otros bienes y servicios	2,3	4,1	1,8

Source: [INE](#).

Spanish Inflation - May 2022

Image 41 - Results comparison and difference between them

The results obtained are higher than the official ones except for communications and leisure and culture. The main difference is the General Index, which refers to the general inflation rate. There is a 1.1 point difference between them and both metrics have used the same weights to calculate them. So, as it was previously stated, the real inflation would be higher as the weights understate inflation.

Moreover, there is not a big difference in the food category rate and it is the category where more data was used to calculate it. However, housing and restaurants had the biggest difference of all the categories. The housing difference could be explained by the addition

of the cost of living in a house, as stated in the chapter 4.3.2.3 and the restaurant difference may be due to the sample of restaurants used in the metric. Another big difference is in the category other goods and services which the official metric does not state what goods and services they have selected to calculate it.

Chapter 7. CONCLUSIONS

Throughout the years, as the methodology of the Consumer Price Index has evolved, in Spain, a lot of the metric changes have not been justified and, as a result, the value obtained is lower than the real one. Specially since the harmonization of the methodologies of the European Union.

These changes usually go unnoticed and, at first, they do not seem to produce any relevant change in the final result. However, the sum of all these changes implemented throughout the years translate into an outcome lower than the real one.

The main areas of improvement that have been detected in this project are: a great part of the used cars market is excluded, the weighting of the products of the basket does not show the real consumption habits, the metric is not representative of the cost of living of all social classes, it is biased, quality adjustments and substitution of products of the basket, the unit of measurement is not stable in time, so the results are biased, and the lack of transparency in different areas of the metric.

Moreover, the results confirm that the actual metric understates inflation even using the same weights which understate inflation. Almost every index calculated was higher than the official ones, up to 2.2 points, and the general inflation obtained was 1.1 points higher.

It would be better to have customizable metrics so anyone can have their own inflation rate. This is because everyone spends their money on what they like and people have different consumption habits. At least it should be divided by social groups because, as proven in chapter 4.3.2.4, inflation affects social classes differently and the official inflation methodology treats everyone the same.

Having an independent, censorship-resistant methodology is important in order to avoid certain manipulations that favor Government's predictions. In addition, it is also important

because some consumption habits may not be taking into account due to them being recent or because they were eliminated by regulations of the European Union.

Chapter 8. FUTURE DEVELOPMENT

This project can be upgraded in order for the metric to be more accurate and have better results. There are four main areas of improvement for this project, the amount of data obtained from the internet, the quality of the data processed, the weighting system, and customization of the metric.

In terms of the amount of data, the improvements could come from scanning more prices of the same product. This could be done by adding more supermarkets and data sources to the methodology. Moreover, more items that are not being considered right now could be added if they are consumed frequently by the population. To spot these items it would be necessary to know exactly where the families spend their income. This will be discussed later.

In terms of the quality of the data, it is important to check from time to time if the data obtained from the web scraping module from the products already implemented and the new ones that may be added. This process is manual but could be partly automatized by checking the labels that are close or related to the price.

In terms of the weighting system, right now this project uses the weights provided by the government which understate inflation as it has been proven in chapter 4.3.2.1 and chapter 4.3.2.2. One way to obtain the weights could be to elaborate a new survey, public to everyone, in which families would introduce everything they spend their income on. There would be several samples of families that would participate would have to be representative of all the different social groups and regions. This solution is very manual, and families could enter the wrong products by accident. Another way of obtaining these products would be to check where they spend the money on their bank accounts. Of course, families would have to agree to share this data and banks would have to agree to share that data to make the process more automatic.

Finally, in terms of customizing the metric, the user that enters the website would be able to customize the metric and add the products in which he spends their money on and exclude the ones that he does not buy. By having this functionality, the metric would be fully adapted to the users spending. They could also add their income and what percentage of their income they spend on each item or category. The web application would calculate the inflation rate based on the items selected. This would solve problems like the representation of the different social groups and small towns of the Spanish territory.

Chapter 9. BIBLIOGRAPHY

- [1] INE (2016). Descripción oficial metodología IPC España. Extraído el 1 de septiembre de 2021 desde fuente: <https://www.ine.es/>
- [2] IMF. Alineación de la metodología española con las recomendaciones del Fondo Monetario Internacional. Extraído el 3 de noviembre de 2021 desde fuente: <https://dsbb.imf.org/sdds-plus/dqaf-base/country/ESP/category/CPI00#notes>
- [3] Economipedia. Definiciones generales sobre conceptos económicos. Extraído el 10 de septiembre de 2021 desde fuente: <https://economipedia.com/>
- [4] Investopedia. Definiciones generales sobre conceptos económicos. Extraído el 11 de septiembre de 2021 desde fuente: <https://www.investopedia.com/>
- [5] Anibes (2021). Consumo actual de alimentos y bebidas entre la población del estudio científico ANIBES en España. Extraído el 20 de marzo de 2022 desde fuente: https://www.fen.org.es/anibes/archivos/documentos/ANIBES_numero_32.pdf
- [6] IAB (2021). Presentación estudio e-commerce 2021. Extraído el 25 de marzo de 2022 desde fuente: <https://iabspain.es/presentacion-estudio-e-commerce-2021/>

- [7] Coches. Coches de segunda mano y ocasión de particulares en Madrid. Extraído el 10 de mayo de 2022 desde fuente: <https://www.coches.net/segunda-mano/madrid/particulares/>
- [8] Coches. Coches de segunda mano y ocasión en Madrid. Extraído el 10 de mayo de 2022 desde fuente: <https://www.coches.net/segunda-mano/?arrProvince=28&st=1>
- [9] Financial Food (2020). Qué alimentos y bebidas son los más consumidos habitualmente por los españoles. Extraído el 15 de mayo de 2022 desde fuente: <https://financialfood.es/cuales-son-los-alimentos-y-bebidas-mas-consumidos-habitualmente-por-los-espanoles/>
- [10] Citibox (2022). Los 10 productos más vendidos online en España. Extraído el 12 de mayo de 2022 desde fuente: <https://citibox.com/es/blog/10-productos-mas-vendidos-online-espana>
- [11] Santander (2022). España: llegando al consumidor. Extraído el 15 de mayo de 2022 desde fuente: <https://santandertrade.com/en/portal/analyse-markets/spain/reaching-the-consumers>
- [12] OECD (2020). Final consumption expenditure of households. Extraído el 20 de mayo de 2022 desde fuente: https://stats.oecd.org/Index.aspx?DataSetCode=SNA_TABLE5

- [13] Eurostat. Key European Statistics Database. Extraído el 20 de mayo de 2022 desde fuente: <https://ec.europa.eu/eurostat/data/database>
- [14] Joan es mi pastor (2022). Comparación mensual de precios de alimentos. Extraído el 30 de mayo de 2022 desde fuente: <https://joan-es-mi-pastor-client.herokuapp.com/>
- [15] Datamarket (2022). Base de datos de precios de supermercados. Extraído el 10 de junio de 2022 desde fuente: <https://datamarket.es/#productos-de-supermercados-dataset>
- [16] Idealista (2022). Evolución del precio de la vivienda en ventas en España. Extraído el 11 de junio de 2022 desde fuente: <https://www.idealista.com/sala-de-prensa/informes-precio-vivienda/>
- [17] Farmacia Eva Contreras (2022). Precios de medicamentos analgésicos y antiinflamatorios. Extraído el 15 de junio de 2022 desde fuente: <https://www.farmaciaevacontreras.com/comprar/medicamentos/antiinflamatorios-y-analgescicos/>
- [18] Developer Mozilla. Generalidades del protocolo HTTP. Extraído el 20 de mayo de 2022 desde fuente: <https://developer.mozilla.org/es/docs/Web/HTTP/Overview>

- [19] Orodcs. Available HTTP Methods. Extraído el 20 de mayo de 2022 desde fuente: <https://doc.oroinc.com/api/http-methods/>
- [20] Developer Mozilla. CRUD. Extraído el 20 de mayo de 2022 desde fuente: <https://developer.mozilla.org/es/docs/Glossary/CRUD>
- [21] Red Hat. Qué es una API REST. Extraído el 21 de mayo de 2022 desde fuente: <https://www.redhat.com/es/topics/api/what-is-a-rest-api>
- [22] Spring. Spring Boot overview. Extraído el 10 de abril de 2022 desde fuente: <https://spring.io/projects/spring-boot>
- [23] Codigofacilito. MVC (Model, View, Controller) explicado. Extraído el 22 de mayo de 2022 desde fuente: <https://codigofacilito.com/articulos/mvc-model-view-controller-explicado>
- [24] The Billion Prices Project. An academic initiative to improve macroeconomics measurement. Extraído el 10 de octubre de 2021 desde fuente: <http://www.thebillionpricesproject.com/>
- [25] Inflacion Verdadera Argentina. Explicación proyecto inflación verdadera Argentina. Extraído el 10 de octubre de 2021 desde fuente: <http://www.inflacionverdadera.com/argentina/>

- [26] Inflacion Verdadera Venezuela. Explicación proyecto inflación verdadera Venezuela. Extraído el 10 de octubre de 2021 desde fuente: <http://www.inflacionverdadera.com/venezuela/>
- [27] PriceStats. Explicación proyecto PriceStats. Extraído el 11 de octubre de 2021 desde fuente: <https://www.pricestats.com/>
- [28] The Market Ear. US money supply chart over the recent years. Extraído el 20 de noviembre de 2021 desde fuente: <https://themarketear.com/>
- [29] The TCP/IP Guide. HTTP response message format. Extraído el 5 de marzo de 2022 desde fuente: http://www.tcpipguide.com/free/t_HTTPResponseMessageFormat.htm
- [30] ITSoftware. Qué es y para qué sirve MySQL Database. Extraído el 6 de marzo de 2022 desde fuente: <https://itsoftware.com.co/content/que-es-y-para-que-sirve-mysql/>

ANEXO I: ALINEACIÓN DEL PROYECTO CON LOS ODS

Los Objetivos de Desarrollo Sostenible son 17 objetivos globales diseñados para ser un plan para lograr un futuro mejor y más sostenible para todos. Este proyecto está relacionado con los siguientes objetivos:

Paz, justicia e instituciones sólidas. Al ser una métrica transparente y open source, ayudaría al Instituto Nacional de Estadística a mejorar los aspectos que se han investigado en el proyecto. Los cambios y artículos añadidos en ciertas categorías, la base de cálculo, el sistema de ponderaciones, y la falta de transparencia en ciertos aspectos de la métrica. Uno de los temas principales que ayudaría mucho a que los resultados publicados por el INE fuesen más precisos, es la personalización de la métrica para cada ciudadano. Es un objetivo ambicioso, pero añadir esta funcionalidad al cálculo de la inflación marcaría un antes y un después en la precisión de esta. Por lo menos una separación por grupos sociales diferenciados por rangos de ingresos. Estos resultados se ajustarían más a la realidad. También podrían tomar como inspiración el script del módulo de web scraping dado que, en la información publicada de la base más reciente, la de 2021, mencionan que están trabajando para añadir técnicas de web scraping como una tercera fuente de información. El script elaborado en este proyecto ayudaría a que se avanzase más rápido en esa parte de la métrica.

Trabajo decente y crecimiento económico. Disponer de métricas precisas, a nivel macroeconómico, ayuda a un mayor crecimiento económico del país. Cuanto mayor sea la presión de estas métricas, mejores decisiones se tomarán respecto a ellas. Esto ayudaría a prever posibles crisis y actuar antes para evitarlas o para reducir su efecto. Por otro lado, los ciudadanos estarían mejor informados sobre la situación económica y tomarían mejores decisiones sobre su gasto. Como se ha comentado durante el proyecto, la métrica también se usa para ajustar salarios y pensiones, aunque no sea del todo correcto, por lo que los ciudadanos pueden estar perdiendo poder adquisitivo al no ajustarse a la inflación real. No pasaría al revés porque, como se ha demostrado en

el proyecto, la métrica actual hace que los resultados sean inferiores a los reales. También se beneficiarían las empresas a la hora de negociar los intereses de préstamos para financiar proyectos y pactar la cantidad real ajustada a la inflación verdadera.