

# Análisis socioeconómico de la crisis Rusia-Ucrania y sus consecuencias a corto plazo mediante el empleo de herramientas de *Business Intelligence*

**Adrián Santos Menéndez**  
Grado en Ingeniería Informática

**Humberto Andrés Sanz**

15/01/2023



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

Copyright © 2023 Adrián Santos Menéndez. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no BackCover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

© (Adrián Santos Menéndez) Reservados todos los derechos. Está prohibido la reproducción total o parcial de esta obra por cualquier medio o procedimiento, comprendidos la impresión, la reprografía, el microfilme, el tratamiento informático o cualquier otro sistema, así como la distribución de ejemplares mediante alquiler y préstamo, sin la autorización escrita del autor o de los límites que autorice la Ley de Propiedad Intelectual.



## FICHA DEL TRABAJO FINAL

<b>Título del trabajo:</b>	Análisis socioeconómico de la crisis Rusia-Ucrania y sus consecuencias a corto plazo mediante el empleo de herramientas de <i>Business Intelligence</i>
<b>Nombre del autor:</b>	Adrián Santos Menéndez
<b>Nombre del consultor:</b>	Humberto Andrés Sanz
<b>Fecha de entrega (mm/aaaa):</b>	01/2023
<b>Área del Trabajo Final:</b>	Business Intelligence
<b>Titulación:</b>	<i>Grado en Ingeniería Informática</i>
<b>Resumen del Trabajo (máximo 250 palabras):</b>	
<p>Este trabajo pretende resaltar la creciente interconexión entre el mundo de las ciencias sociales y la política y el de las nuevas tecnologías. Concretamente, se busca plantear cuáles de estas tecnologías y herramientas podrían ser efectivas para analizar las consecuencias del actual conflicto Ucrania-Rusia, y el cómo podrían plantearse a tal fin.</p> <p>Más allá del marco estrictamente teórico, este trabajo se ha centrado en producir gráficos y reportes, esenciales dentro de la inteligencia de negocio y aplicados en este caso a la temática del conflicto bélico internacional.</p> <p>Al cierre del trabajo, se concluye que es mucho lo que se puede analizar con los datos y herramientas adecuadas, y que posiblemente el conflicto Ucrania-Rusia haya entrado ya en una etapa crónica en la que su impacto ya no es tan grande aunque siga entrañando riesgos potenciales.</p>	

**Abstract (in English, 250 words or less):**

This project aims to highlight the growing interconnection between the world of social sciences and politics, and the world of modern technology. Specifically, it aims to determine which technologies and tools could be effective to analyze the consequences of the present conflict Ukraine-Russia, and how they could be used to serve that purpose.

Beyond the purely theoretical framework, this work has also been focused on producing visuals and reports, vital for Business Intelligence, and applied in this case to the topic of international armed conflicts.

This work concludes that there is a lot that can be analyzed with the right tools and data, and that the Ukraine-Russia conflict had probably reached a phase where it is no longer as influential as it once was, but it still not without its risks.

<b>Palabras clave (entre 4 y 8):</b>
Business Intelligence, Análisis de Sentimientos, Power BI, Python, Ucrania, Rusia, Clusterización

## Índice

1. Introducción .....	1
1.1 Contexto y justificación del Trabajo .....	1
1.2 Objetivos del Trabajo .....	1
1.3 Enfoque y método seguido.....	2
1.4 Planificación del Trabajo .....	3
1.4.1. Análisis de riesgos .....	3
1.4.2. Actividades y tareas.....	5
1.4.3. Hitos .....	7
1.5 Breve resumen de productos obtenidos .....	7
1.6 ¿Qué es Business Intelligence? .....	8
1.7 Breve descripción de los otros capítulos de la memoria .....	9
2. Desarrollo del proyecto .....	10
2.1. Selección de las herramientas principales .....	10
2.1.1. ¿Tableau o Power BI? .....	10
2.1.2. ¿Python o R? .....	12
2.2. Análisis social .....	13
2.2.1. Opinión pública .....	13
2.2.1.1. ¿Qué es análisis de sentimientos? .....	13
2.2.1.2. Extracción de respuestas de Twitter .....	13
2.2.1.3. Preprocesado de texto: stop words, tokenización, lematización, otros .....	21
2.2.1.4. Carga y transformación de datos en Power BI .....	23
2.2.1.5. Visualización de datos y reporte .....	26
2.2.2. Crisis humanitaria (refugiados) .....	29
2.2.2.1. ¿Qué es análisis predictivo? .....	29
2.2.2.2. Carga y transformación de datos en Power BI .....	29
2.2.2.3. Visualización de datos y reporte .....	30
2.3. Análisis económico .....	33
2.3.1. Selección de indicadores .....	33
2.3.2. Inflación .....	37
2.3.2.1. Carga y transformación de datos .....	37
2.3.2.2. Visualización de datos y reporte .....	40
2.3.3. PIB.....	42
2.3.3.1. Carga y transformación de datos .....	42
2.3.3.2. Visualización de datos y reporte .....	42
2.3.4. Desempleo .....	43
2.3.3.1. Carga y transformación de datos .....	43
2.3.3.2. Visualización de datos y reporte .....	44
2.3.5. Clusterización .....	45
3. Conclusiones .....	47
4. Glosario .....	48
5. Bibliografía .....	49

## Lista de figuras

Ilustración 1: Ciclo de trabajo básico en analítica de datos	2
Ilustración 2: Análisis de riesgos	3
Ilustración 3: Acciones correctoras	4
Ilustración 4: Actividades y tareas	5
Ilustración 5: Diagrama de Gantt	6
Ilustración 6: Hitos	7
Ilustración 7: logo de Tableau	10
Ilustración 8: logo de Power BI	11
Ilustración 9: Tableau o Power BI – Tabla comparativa resumen	11
Ilustración 10: logo Python / R	12
Ilustración 11: Python o R – Tabla comparativa resumen	12
Ilustración 12: Cuenta de Twitter de Zelensky	14
Ilustración 13: Búsqueda avanzada Twitter (1)	15
Ilustración 14: Búsqueda avanzada Twitter (2)	15
Ilustración 15: Tweet Zelensky enero 2022	16
Ilustración 16: Tweet Zelensky febrero 2022	16
Ilustración 17: Tweet Zelensky marzo 2022	17
Ilustración 18: Tweet Zelensky abril 2022	18
Ilustración 19: extracción de comentarios a través de ExportComments	20
Ilustración 20: archivo con comentarios extraídos de Twitter	20
Ilustración 21: Muestreo representativo de gran volumen de datos de partida	21
Ilustración 22: vista de Power Query (comentarios Twitter)	23
Ilustración 23: Reemplazar valores en Power Query	24
Ilustración 24: Eliminación de enlaces en Power Query (1)	25
Ilustración 25: Eliminación de enlaces en Power Query (2)	25
Ilustración 26: Ejecución de script de Python para análisis de sentimientos en Power Query	26
Ilustración 27: Importación de objeto visual – Nube de palabras	27
Ilustración 28: Filtrado de palabras para el bando ucraniano	27
Ilustración 29: Reporte social – Opinión pública	28
Ilustración 30: carga de tabla con datos sobre población de países (Wikipedia)	30
Ilustración 31: Número de refugiados a fecha 16/05/2022. Gráfico predictivo con confianza del 95%.	30
Ilustración 32: Confianza 80% (1)	31
Ilustración 33: Confianza 95% (1)	31
Ilustración 34: Confianza 95% (2)	31
Ilustración 35: Confianza 80% (2)	31
Ilustración 36: Reporte social – Crisis humanitaria (refugiados)	32
Ilustración 37: Objetivo-Pregunta-Medida	33
Ilustración 38: Muestra de algunos datos disponibles en OECD.org	35
Ilustración 39: Carga de datos donde el delimitador decimal es un “.”	35
Ilustración 40: Importancia de mantener un formato sólido. Cambio de “.” por “,”	36



Ilustración 41: Uso de “agrupar por” para crear una nueva tabla con valores agrupados	38
Ilustración 42: Pivotar filas en Power Query	39
Ilustración 43: Variación porcentual	39
Ilustración 44: Unión de tablas en Power Query	40
Ilustración 45: Reporte económico – Inflación año a año 2021 y 2022	41
Ilustración 46: Reporte económico – PIB trimestre a trimestre 2021 y 2022	43
Ilustración 47: Reporte económico – Desempleo trimestre a trimestre 2021 y 2022	45
Ilustración 48: Reporte económico - Clusterización	46

# 1. Introducción

## 1.1 Contexto y justificación del Trabajo

Hoy día, la tecnología se abre paso en campos cada vez más diversos.

La estadística, la inteligencia artificial o las técnicas de Big Data juegan cada vez un papel más crucial en aspectos políticos y sociales, como por ejemplo en las campañas electorales. Por este motivo, parece importante incidir en esa unión creciente entre disciplinas y plantear nuevos escenarios de aplicación para los recursos tecnológicos ya disponibles.

Los conflictos armados involucran una gran variedad de aspectos no solo militares, sino de muy diversa índole: económicos, sociales, comunicativos, políticos, humanitarios, tecnológicos.

La actual crisis Rusia-Ucrania, con raíces en 2014 pero eclosión reciente pone a prueba los equilibrios geopolíticos, las estrategias y capacidades de los países intervinientes y la solidez y unión de las economías globales.

Además, dicha crisis cuenta con características muy interesantes y merecedoras de análisis. Se trata de un país largamente tensionado por el antagonismo entre los sentimientos pro-rusos y pro-europeos. Un país que ha experimentado la revolución del Euromaidán y también la guerra (aún no resuelta y oficial *cassus belli* para el Gobierno Ruso) en la región del Donbass. Y un país con un presidente marcadamente atípico tanto en su origen como en su estilo comunicativo y de liderazgo.

Además, es este un conflicto que involucra a un no menos interesante país como es Rusia, que fuerza a la Unión Europea a tomar decisiones rápidas e inéditas, que obliga a una activación y despliegue (si bien por ahora contenido) de la OTAN y que ha provocado una respuesta en forma de sanciones económicas nunca vistas de países de todo el globo.

Por todo lo anteriormente descrito y por los enormes retos que plantea, el estudio de la crisis Rusia-Ucrania podría proporcionarnos importante información sobre dónde nos encontramos históricamente y hacia dónde nos encaminamos.

## 1.2 Objetivos del Trabajo

1. Estudiar qué herramientas y técnicas asociadas al *Business Intelligence* se podrían utilizar para analizar el reciente conflicto Ucrania-Rusia desde un punto de vista social y económico, de cara a entender lo que ha ocurrido y cómo minimizar daños futuros.

2. Mostrar cómo se podrían aplicar de forma práctica dichas herramientas y técnicas, recopilando datos dispersos y generando gráficos e informes a modo de ejemplo.
3. Tratar de obtener algunas conclusiones a partir de los gráficos e informes realizados.

### 1.3 Enfoque y método seguido

El desarrollo del proyecto estará basado en los preceptos y recomendaciones recogidos en la Guía de los Fundamentos de la Gestión de Proyectos (PMBOK).

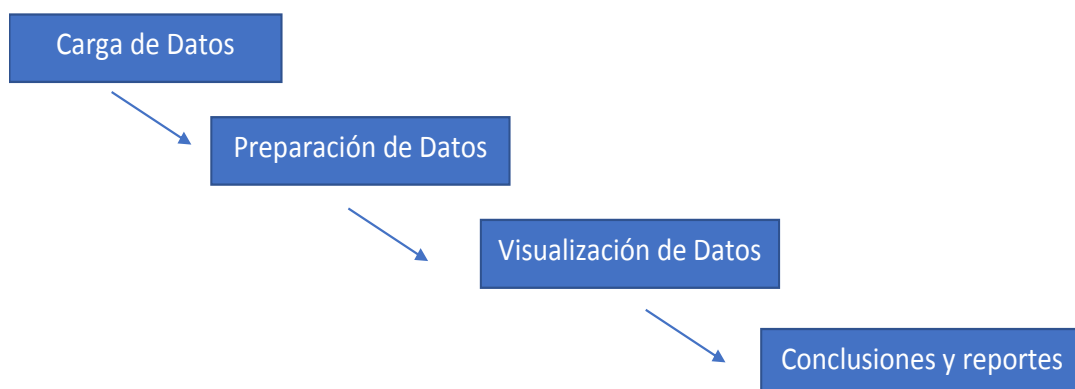
En cuanto a áreas de interés para un trabajo de BI, este proyecto explorará los siguientes ámbitos, entre otros posibles:

- Análisis de sentimientos en redes sociales y *social analytics*
- Análisis predictivo.
- Visualización de datos (previa obtención, limpieza, filtrado, etc).
- Elaboración de informes.

El proyecto tuvo dos enfoques fundamentales:

- **Social:** estudio de la efectividad de las campañas de comunicación de los distintos bandos involucrados mediante análisis de sentimientos en redes sociales (Twitter).
- **Económico:** estudio del impacto del conflicto en indicadores económicos clave. Se tratará de determinar hasta qué punto el conflicto ha influido en dichos indicadores, cómo han variado, si ya se ha alcanzado un equilibrio, etc.

El ciclo de trabajo a seguir con los datos que vamos a obtener será el clásico:



**Ilustración 1: Ciclo de trabajo básico en analítica de datos**

## 1.4 Planificación del Trabajo

Se estima que el proyecto requerirá una inversión de alrededor de 300h para ser completado.

Habrà un único recurso humano encargado tanto de la dirección del proyecto como del resto de aspectos del mismo (recopilación de información, limpieza y procesado de datos, etc). Dicho recurso humano cuenta con experiencia básica utilizando Python y R y realizando tareas sencillas de minería de datos. No cuenta con experiencia en el uso de herramientas comerciales de BI.

Se reserva una partida de 100,00€ en caso de que fuera necesario contratar prestaciones *premium* de alguna de las herramientas a emplear, no habiendo aparte de esto más gastos significativos previstos.

El equipo a utilizar cuenta con procesador i5-4210U con 3,90 GB de RAM utilizable, 2.40 GHz y sistema operativo Windows 8, de 64 bits. Espacio disponible: 400 GB.

### 1.4.1. Análisis de riesgos

RIESGOS				
ID	DENOMINACIÓN	DESCRIPCIÓN	PROBABILIDAD	IMPACTO
R01	Falta de información	Ausencia de datos suficientes para desarrollar alguno de los objetivos planteados	MEDIA	MEDIO
R02	Problemas de hardware	Daños y averías en el equipo de trabajo	BAJA	ALTO
R03	Problemas de software	Fallos imprevistos en la configuración, problemas de compatibilidad	MEDIA	MEDIO
R04	Imprevistos de índole personal	Problemas laborales o familiares	BAJA	MEDIO
R05	Desviaciones	Fallo en la estimación de coste temporal	BAJA	MEDIO

Ilustración 2: Análisis de riesgos

ACCIONES CORRECTORAS			
ID	RIESGO	ACCIÓN	COSTE
A01	R01	Redimensionar el proyecto	MEDIO
A02	R02	Sustitución del hardware	ALTO
A03	R03	Buscar ayuda en la red / un tercero	MEDIO
A04	R04	Jornadas más intensivas	BAJO
A05	R05	Jornadas más intensivas	BAJO

**Ilustración 3: Acciones correctoras**

#### 1.4.2. Actividades y tareas.

	Denominación	Fecha de finalización	Coste (horas)
0	Gestión del TFG	15/01/2023	80
1	Propuesta		
2	Análisis inicial	03/10/2022	30
3	Estimación de costes	05/10/2022	5
4	Definición de objetivos	07/10/2022	5
5	Definición de hitos	09/10/2022	5
6	Análisis de riesgos	11/10/2022	5
7	Documentación inicial	16/10/2022	10
8	Entrega PEC 1		
9	Análisis de herramientas	26/10/2022	20
10	Instalación y toma de contacto	31/10/2022	10
12	Tratamiento de datos	12/11/2022	30
13	Revisión y conclusiones	16/11/2022	10
14	Entrega PEC 2		
15	Estudio de datos e indicadores	29/11/2022	20
16	Estudio de funcionalidades	02/12/2022	10
17	Tratamiento de datos	14/12/2022	30
18	Revisión y conclusiones	16/11/2022	10
19	Entrega PEC 3		
20	Revisión	31/12/2022	10
21	Integración y pruebas complementarias	15/01/2023	10
	Entrega final		

Ilustración 4: Actividades y tareas

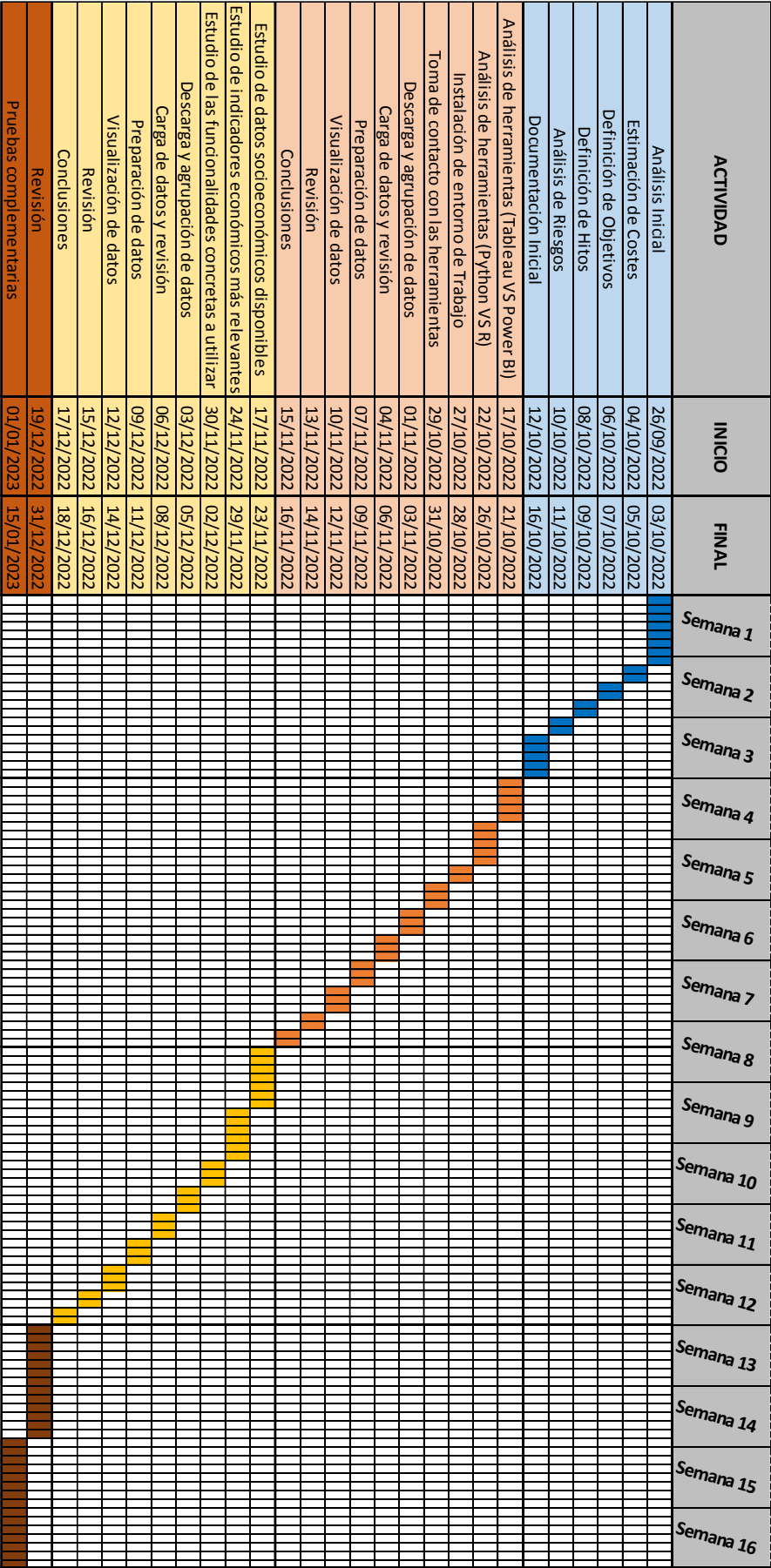


Ilustración 5: Diagrama de Gantt

### 1.4.3. Hitos

NOMBRE	FECHA	TIPO
Propuesta inicial aceptada	02/10/2022	EJECUCIÓN
Análisis inicial cerrado	03/10/2022	EJECUCIÓN
Planificación cerrada (PEC 1)	16/10/2022	SEGUIMIENTO
Análisis secundario - selección de herramientas y datos - cerrado	26/10/2022	EJECUCIÓN
Desarrollo del primer enfoque – Análisis de sentimientos – cerrado (PEC 2)	18/11/2022	SEGUIMIENTO
Análisis complementario al segundo enfoque – Indicadores económicos – cerrado	29/11/2022	EJECUCIÓN
Desarrollo del segundo enfoque – Indicadores económicos – cerrado (PEC 3)	18/12/2022	SEGUIMIENTO
Revisión y pruebas complementarias cerradas	15/01/2023	EJECUCIÓN
Entrega final	15/01/2023	EJECUCIÓN

Ilustración 6: Hitos

### 1.5 Breve resumen de productos obtenidos

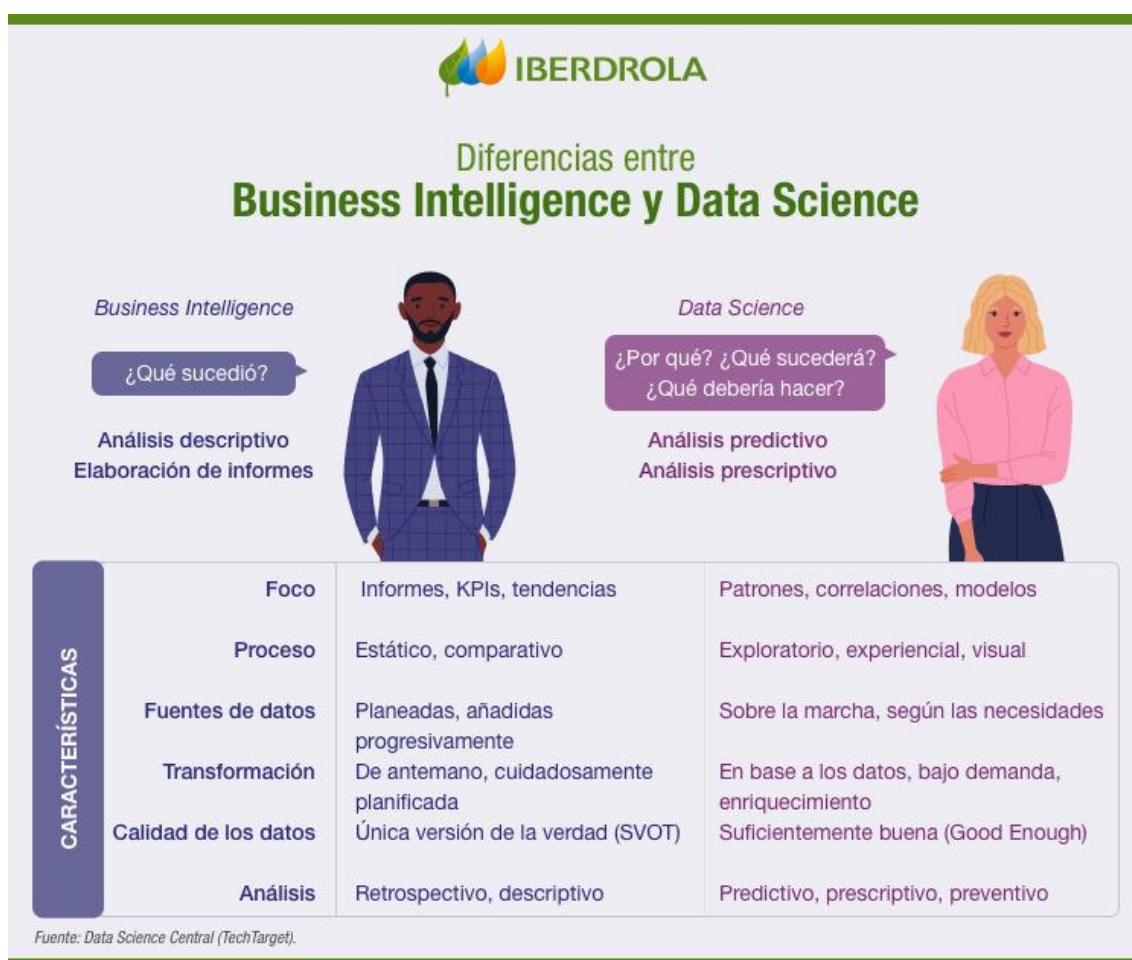
- Memoria del proyecto
  - Introducción, desarrollo, conclusiones.
- Presentación del proyecto
  - Resumen detallado y comprensible
- Otros productos tangibles
  - Datos extraídos para el proyecto final (.csv, .xlsx, etc)
  - Reportes en formato Power BI



## 1.6 ¿Qué es Business Intelligence?

El *Business Intelligence (BI)* o Inteligencia de Negocio comprende un conjunto de técnicas y herramientas enfocadas a analizar datos con el objetivo de mejorar el rendimiento y proceso de toma de decisiones dentro de un entorno empresarial.

El BI bebe de otras disciplinas como por ejemplo la Minería de Datos o el Aprendizaje Computacional, y es muchas veces confundido con otras como el *Big Data* o el *Data Science*. Sin embargo, lo que caracteriza al BI es el enfoque práctico basado en el negocio, se trata de una disciplina aplicada y, en muchas ocasiones, más enfocada a analizar lo que ya existe para adelantarse a los acontecimientos y ver en qué punto está situado un proyecto.



## **1.7 Breve descripción de los otros capítulos de la memòria**

A lo largo del capítulo 2 se expone el proceso seguido para el desarrollo del proyecto.

En cada apartado se va haciendo una exposición ordenada en base a los objetivos descritos:

1º Herramientas / técnicas que se plantean para el proyecto en sí y para cada escenario concreto de análisis.

2º Resumen de los pasos seguidos para poder elaborar un producto que aporte respuestas (reporte con visualizaciones de datos)

3º Pequeña conclusión referente al tema planteado.

## 2. Desarrollo del proyecto

### 2.1. Selección de las herramientas principales

Para llevar a buen término la parte práctica del trabajo, se seleccionaron 2 herramientas principales que fueran capaces de cubrir todas las necesidades de obtención de datos, transformación de los mismos y creación de gráficos. Es decir, que nos permitieran cumplir con todo el ciclo anteriormente descrito de analítica de datos.

#### 2.1.1. ¿Tableau o Power BI?

En primer lugar se planteó la elección de la que sería la herramienta que actuara como base para todo el apartado práctico del proyecto, enfocada específicamente al trabajo amplio con datos para propósitos de BI, con capacidad para abarcar gran cantidad de tareas y, especialmente, la creación de gráficos interactivos.

Son numerosas las herramientas de ese estilo presentes hoy día en el mercado pero, para simplificar la elección, se analizaron únicamente aquellas dos que actualmente se encuentran presentes en un mayor número de ofertas de trabajo, empresas y proyectos como este que estamos realizando.

Dichas herramientas son Tableau y Power BI.



Ilustración 7: logo de Tableau

[Fuente de la imagen](#)

Tableau es una herramienta muy potente, con más de 20 años de recorrido y enfocada al BI. La calidad de los gráficos es elevada y se pueden obtener resultados visualmente muy atractivos. Sin embargo, no es la herramienta más sencilla ni intuitiva de manejar y no es un proyecto abierto al que se puedan ir añadiendo algunas funcionalidades como, por ejemplo, más tipos de gráficos.

A pesar de que puede soportar grandes cantidades de datos y las visualizaciones que posee de partida son buenas, no tiene demasiadas prestaciones a la hora de transformar los datos.

Por otro lado, su comunidad, si bien grande, resulta pequeña en comparación con la de otras herramientas como Power BI. Por último, las versiones de pago tienen un coste algo elevado.



**Ilustración 8: logo de Power BI**

[Fuente de la imagen](#)

Power BI es la herramienta específica para BI de Microsoft. Cuenta con una enorme comunidad de millones de usuarios y está ampliamente extendida, hasta el punto de que la mayoría de *dashboard* e infografías sobre la pandemia del COVID-19 se realizaron con esta herramienta por parte de las administraciones públicas.

Power BI cuenta con una gran capacidad para transformar los datos que cargamos en ella, admite la adición de gráficos personalizados y cuenta con versiones de pago asequibles.

	TABLEAU	POWER BI
<b>Usuarios (2016-2017)</b>	+350.000	+5.000.000
<b>Potencia y prestaciones</b>	Altas	Medias
<b>Dificultad de aprendizaje</b>	Media-Alta	Media-Baja
<b>Interfaz</b>	Menos amigable	Más amigable (similar a otros productos conocidos como Excel)
<b>Enfocada a</b>	Grandes proyectos y empresas	Proyectos y empresas de todo tipo
<b>Flexibilidad</b>	Media	Alta
<b>Versión gratuita</b>	Limitada	Potente
<b>Versiones de pago</b>	Costosas	Baratas

**Ilustración 9: Tableau o Power BI – Tabla comparativa resumen**

La elección final para el proyecto fue Power BI, al entender que presentaba menos inconvenientes y se ajustaba mejor a las necesidades y dimensiones del trabajo a realizar.

### 2.1.2. ¿Python o R?

Una vez seleccionada Power BI como herramienta base, queda aún determinar otra herramienta auxiliar que nos permita realizar operaciones para las que Power BI no sea adecuada o eficiente. Lo ideal es un lenguaje de programación que nos permita desarrollar tareas de análisis y preprocesado de datos.

Los dos lenguajes compatibles con Power BI son Python y R, pudiendo conectar gráficos y *scripts* con ella.



Ilustración 10: logo Python / R

[Fuente de la imagen](#)

Python es un lenguaje de programación muy sencillo y versátil que actualmente es empleado para todo tipo de tareas, y muy especialmente por ingenieros y científicos que necesitan un lenguaje de programación eficaz pero no engorroso.

Por su parte, R es un lenguaje más especializado en estadística, aunque muy útil en tareas relacionadas con el análisis de datos. Es menos sencillo y versátil que Python y suele ser la herramienta preferida de aquellos con formación extensa en matemáticas y estadística.

	PYTHON	R
Versatilidad	Muy alta	Baja. Especializado
Gráficos	Aceptables	Numerosos y de calidad
Dificultad de aprendizaje	Media-Baja	Media-Alta
Enfocada a	Todo tipo de proyectos	Estadística y gráficos
Flexibilidad	Alta	Media

Ilustración 11: Python o R – Tabla comparativa resumen

La elección final para el proyecto fue Python, al entender que presentaba menos inconvenientes y se ajustaba mejor a las necesidades y dimensiones del trabajo a realizar.

## **2.2. Análisis social**

El primer componente que se planteó para el proyecto es el análisis social de la situación. Se trataba de preguntarse qué aspectos sociales, susceptibles de ser analizados mediante BI, podrían ser interesantes a la hora de estudiar el conflicto Ucrania-Rusia.

### **2.2.1. Opinión pública**

El primer aspecto a medir que se decidió fue el de la opinión pública. Para cualquier negocio es imprescindible conocer qué quieren y no quieren sus clientes. Esto es extrapolable a cualquier campaña de *marketing* y a escenarios políticos como el que trabajamos en este proyecto.

Una de las técnicas más empleada en la actualidad para medir estados de opinión y tendencias de un grupo es el análisis de sentimientos.

#### **2.2.1.1. ¿Qué es análisis de sentimientos?**

Se trata del conjunto de técnicas de minería de datos y análisis del lenguaje natural que busca extraer información subjetiva de las reproducciones, normalmente escritas, de los seres humanos.

Es una herramienta muy importante tanto en el mundo de los negocios como de la política, ya que puede predecir el éxito de una campaña de marketing o ayudar a entender los gustos de los consumidores.

#### **2.2.1.2. Extracción de respuestas de Twitter**

En este trabajo se decidió desarrollar esta técnica a través de un análisis de comentarios en redes sociales, concretamente Twitter, y más específicamente las respuestas a tweets del presidente ucraniano Zelensky.

Twitter es una plataforma ampliamente utilizada para obtener información sobre la opinión de usuarios, con lo que hay una importante comunidad de desarrolladores y herramientas especializadas en trabajar con datos de Twitter.

Por otro lado, la cuenta de Zelensky ha sido uno de los lugares donde se ha congregado un mayor número de usuarios para expresar su opinión sobre el conflicto, sobre todo durante los primeros meses.



Ilustración 12: Cuenta de Twitter de Zelensky

[Fuente de la imagen](#)

Se decidió hacer una selección de Tweets, teniendo en cuenta que tuviesen bastantes respuestas y que estuviesen repartidos a lo largo de varios meses (enero-abril 2022).

Para acceder cómodamente a dichos tweets se empleó el mecanismo de búsqueda avanzada de Twitter.

X
Búsqueda avanzada
Buscar

Idioma  
inglés

Cuentas

De estas cuentas

@ZelenskyUa

Ejemplo: @Twitter · enviado desde @Twitter

A estas cuentas
Ejemplo: @Twitter · enviado como respuesta a @Twitter

Mención de estas cuentas
Ejemplo: @SFBART @Caltrain · menciona a @SFBART o menciona a @Caltrain

Filtros

Ilustración 13: Búsqueda avanzada Twitter (1)

X
Búsqueda avanzada
Buscar

Ejemplo: 280 · Tweets con un mínimo de 280 respuestas

Mínimo de Me gusta
600

Ejemplo: 280 · Tweets con un mínimo de 280 Me gusta

Mínimo de Retweets
600

Ejemplo: 280 · Tweets con un mínimo de 280 Retweets

Fechas

Desde

Mes  
enero

Día  
1

Año  
2022

Hasta

Mes  
abril

Día  
30

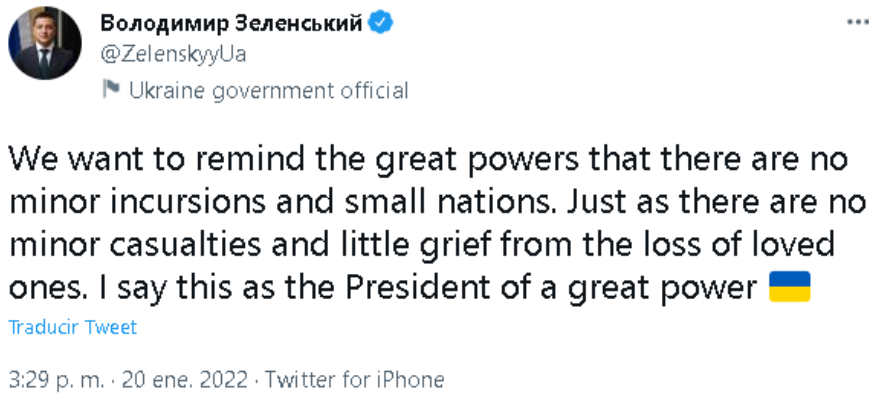
Año  
2022

Ilustración 14: Búsqueda avanzada Twitter (2)

15



Los tweets escogidos fueron los siguientes:



8.673 Retweets 1.641 Tweets citados 45 mil Me gusta

**Ilustración 15: Tweet Zelensky enero 2022**

<https://twitter.com/ZelenskyyUa/status/1484171183264129025>



16,5 mil Retweets 1.176 Tweets citados 75,4 mil Me gusta

**Ilustración 16: Tweet Zelensky febrero 2022**

<https://twitter.com/ZelenskyyUa/status/1496787304811315202>



**Болондмнр Зеленський** ✓  
@ZelenskyyUa  
Ukraine government official



Breaking!

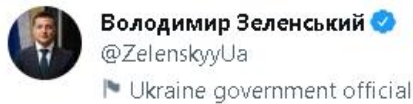


3:03 p. m. · 6 mar. 2022 · Twitter Web App

57,7 mil Retweets · 6.090 Tweets citados · 241 mil Me gusta

**Ilustración 17: Tweet Zelensky marzo 2022**

<https://twitter.com/ZelenskyyUa/status/1500472014452273157>



Without additional weaponry, this war will become an endless bloodbath, spreading misery, suffering, and destruction. Mariupol, Bucha, Kramatorsk – the list will be continued. Nobody will stop Russia except Ukraine with Heavy Weapons. #ArmUkraineNow

Traducir Tweet



Ilustración 18: Tweet Zelensky abril 2022

<https://twitter.com/ZelenskyyUa/status/1514242356949704709>

Para poder extraer las respuestas a un determinado tweet hay varias opciones, que se pueden resumir en 2:

1. Hacer uso de la API de Twitter a través de una pequeña cantidad de código. Python proporciona al efecto bibliotecas como tweepy y permite exportar los comentarios a archivos .csv.
2. Hacer uso de una herramienta comercial específica, como <https://exportcomments.com/>

En el primer caso es necesario crear una cuenta de desarrollador y solicitar permisos elevados. En los últimos tiempos ha habido numerosos cambios sobre los permisos con lo que es posible que para cierto volumen de datos o ciertas solicitudes no sea suficiente con permisos elevados y por tanto haya que pagar acceso premium.

Se muestra a continuación un ejemplo de código funcional para realizar la descarga de comentarios por esta vía, una vez obtenidas las credenciales:

```

import csv
import tweepy
import ssl

ssl._create_default_https_context = ssl._create_unverified_context

#Credenciales necesarias, obtenidas tras recibir permisos de
desarrollador
consumer_key = "Ed3*****UufVJ"
consumer_secret = "F1RD*****soKg"
access_token = "123088*****1TiufDF"
access_token_secret = "WpKy*****1Gb"

#Comprobación de identidad
auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)
api = tweepy.API(auth)

#Identificar cuenta y tweet concreto cuyos comentarios se desean extraer
name = 'ZelenskyUA'
tweet_id = '1613573375972052994'

replies=[]
for tweet in tweepy.Cursor(api.search_tweets,q='to:'+name,
result_type='recent', timeout=999999).items(1000):
    if hasattr(tweet, 'in_reply_to_status_id_str'):
        if (tweet.in_reply_to_status_id_str==tweet_id):
            replies.append(tweet)

with open('tweetZelensky00.csv', 'w') as f:
    csv_writer = csv.DictWriter(f, fieldnames=('user', 'text'))
    csv_writer.writeheader()
    for tweet in replies:
        row = {'user': tweet.user.screen_name, 'text':
tweet.text.replace('\n', ' ')}
        csv_writer.writerow(row)

```

Código adaptado al caso y actualizado a las nuevas versiones de la API a partir del ejemplo del siguiente enlace: <https://hackernoon.com/scrapping-tweet-replies-with-python-and-tweepy-twitter-api-a-step-by-step-guide-z11x3yr8>

La segunda opción es mucho más sencilla ya que no requiere código ni solicitudes, pero para volúmenes de comentarios elevados también puede ser necesario abonar una suscripción premium.

The screenshot shows the ExportComments website. At the top, there's a navigation bar with links for HOME, MY PAYMENTS, MY EXPORTS, and LOGOUT. Below this, a status bar indicates 'PERSONAL TILL 2022-06-12'. The main content area shows 'Total 9 exports found'. There's a search bar and a 'FILTER' button. Below, a table lists exports with columns: Export File, Url, Paid, Nested, Total, Replies, Comments, Date Exported, Notified, Days to delete, Status, and Actions. The table shows four rows of exports, each with a link to the export file and a status indicator (Progress or Done).

Ilustración 19: extracción de comentarios a través de ExportComments

Ambas opciones (API + Python o herramientas comerciales específicas) proporcionan resultados similares, aunque no son infalibles y en algunos momentos no han sido capaces de descargar toda la información de algún determinado tweet.

Al final del proceso de extracción contamos con varios archivos .csv o .xlsx que albergan el texto de los comentarios a cada Tweet, entre otros datos.

The screenshot shows a Microsoft Excel spreadsheet with a table of extracted tweet comments. The table has columns for ID, User, Comment, and Date. The data is organized into rows, with each row representing a tweet and its associated comments. The spreadsheet is titled 'Sheet1' and shows a list of tweets with their respective user avatars, names, and the text of the comments. The dates are listed in the rightmost column.

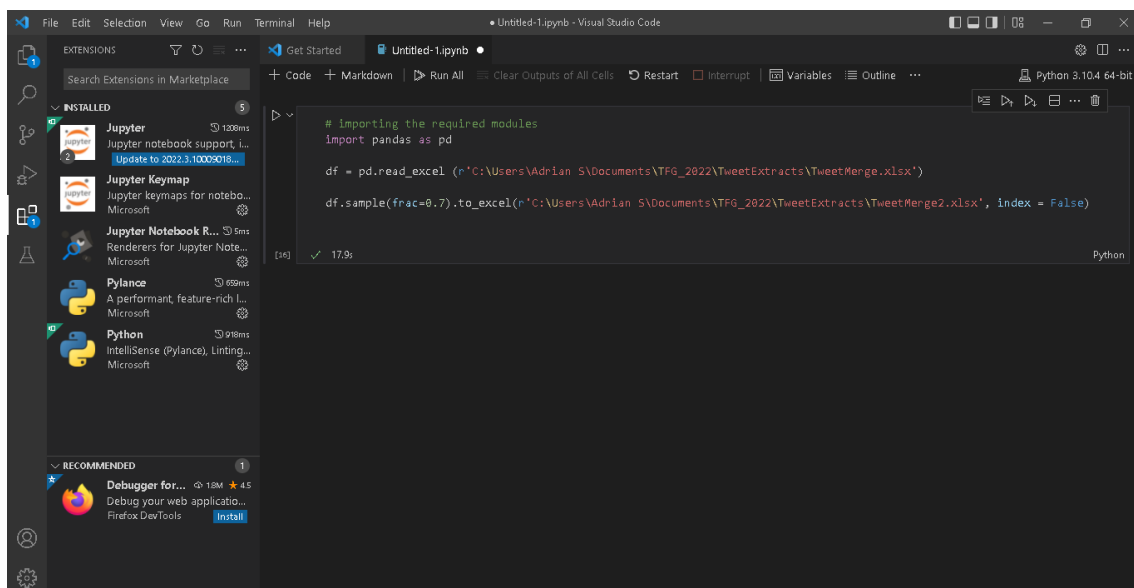
Ilustración 20: archivo con comentarios extraídos de Twitter

A partir de ese punto disponemos de unos 20.000 comentarios asociados a los Tweets indicados. Podemos emplear Python(funciones dataframe.append() y pandas.concat()) o Excel para unir todos los datos en un único documento.

### 2.2.1.3. Preprocesado de texto: stop words, tokenización, lematización, otros

Una posible opción para facilitar el procesamiento de datos, especialmente cuando estamos experimentando problemas de rendimiento, es seleccionar un subconjunto de los de partida y trabajar sobre él.

Python nos permite tomar una muestra aleatoria de nuestro conjunto, que mientras no sea muy inferior al original debería ser razonablemente representativa. Por ejemplo, durante las pruebas se tomó un subconjunto equivalente al 70% de los datos de partida para aliviar la carga del programa.



**Ilustración 21: Muestreo representativo de gran volumen de datos de partida**

Otras opciones que tenemos para procesar con Python antes de cargar los comentarios en Power BI son:

Eliminación de stop words: se trata de eliminar aquellas palabras que, por sí solas, no aportan información sobre el sentimiento asociado al comentario (artículos, determinantes, etc).

Tokenización: se trata de dividir el texto en fragmentos (tokens), uno por cada palabra, para simplificarlo.

Lematización: se basa en, dada una palabra, obtener su palabra canónica. Por ejemplo, comí → comer.

Eliminación de otros elementos redundantes (almohadillas, arrobas...).

A continuación se muestra un ejemplo de código de estos procesos de limpieza previos:

```
pandas as pd
import nltk
nltk.download('all')
from nltk.corpus import stopwords
from nltk.tokenize import TweetTokenizer
from nltk import pos_tag
from nltk.stem import WordNetLemmatizer

stopWords = nltk.download('stopwords')
stopWords = set(stopwords.words('english')) #Listado de stopwords
inglesas
lemmatizer=WordNetLemmatizer()

df = pd.read_excel (r'C:\Users\Adrian
S\Documents\TFG_2022\EntregaFinal\Twitter\ComentariosZelenskyTwitter.xlsx
')

df['Text'] = df['Text'].str.replace('@[^\s]+','',case=False) #Eliminar
alias
df['Text'] = df['Text'].str.replace('\d+','',case=False) #Eliminar
numeros
df['Text'] = df['Text'].str.replace('[\.\!;\?]', '',case=False) #Eliminar
signos de puntuacion
df['Text'] = df['Text'].str.lower().str.strip() #Eliminar mayusculas y
espacios al inicio y final
df['Text'] = df['Text'].str.replace('http\S+|www.\S+', '') #Eliminar
enlaces

#Eliminar stopWords establecidas
df['Text']=df['Text'].apply(lambda x: ' '.join([word for word in
x.split() if word not in (stopWords)]))

#pd.isnull(df['Text'])

#Tokenizar
tweetTokenizer = TweetTokenizer()
df['Text']=df['Text'].apply(tweetTokenizer.tokenize)
print(df['Text'])

#Lematizar
df['Text']=df['Text'].apply(lambda lst:[lemmatizer.lemmatize(word) for
word in lst])
df['Text']=df['Text'].apply(lambda x: '
'.join([lemmatizer.lemmatize(word) for word in x.split()])))
```

```
df.sample(frac=0.8).to_excel(r'C:\Users\Adrian
S\Documents\TFG_2022\EntregaFinal\Twitter\ComentariosZelenskyTwitterLEMA.
xlsx', index = False)
```

Sin embargo, estos pasos eliminan información a cambio de simplificar el texto a analizar. La idoneidad de aplicar este preprocesado depende del algoritmo que vayamos a usar posteriormente para llevar a cabo el análisis de sentimientos.

En este caso el algoritmo empleado es VADER el cual funciona bien sin necesidad de pasos previos, y puede fallar si el texto original está demasiado adulterado.

Por ese motivo, se decidió para este proyecto no emplear lematización ni tokenización. En cuanto a la eliminación de elementos redundantes, es posible hacerlo a través de Power BI, incluso a nivel de gráfico, reduciendo las modificaciones del texto original al mínimo indispensable.

#### 2.2.1.4. Carga y transformación de datos en Power BI

Cargar datos en Power BI es tan sencillo como ir al icono de “Obtener datos” seleccionar el tipo de fuente (.csv, web, excel...).

Las tareas de transformación de datos las realizamos a través de Power Query (pestaña de transformación/edición de datos)..

	Tweet Text	Author Followers	Author Friends	Author Favorites	Author Statuses	Author Bio	Author Image
16	@ZelenskyUa @AndzejDuda https://t.co/dU0KTSvQr	0	34	0	449		http://abs.twimg.com/sticky/default_profile_image
14	@ZelenskyUa You are such an inspiration... wish all leader	0	25	0	8		http://abs.twimg.com/sticky/default_profile_image
17	@ZelenskyUa Apparently? Hunter didn't make any money	0	2	0	249		http://abs.twimg.com/sticky/default_profile_image
13	@ZelenskyUa @AndzejDuda fight on and defeat the russi	0	3	0	9		http://abs.twimg.com/sticky/default_profile_image
14	@ZelenskyUa Уважаемый Президент, а бразилец и проку	0	3	0	13		http://abs.twimg.com/sticky/default_profile_image
13	@Lonestarmoncom @ZelenskyUa @SproAgnewGhost @D	0	10	0	3		http://abs.twimg.com/sticky/default_profile_image
8	@ZelenskyUa Didn't you say that it's going ukraine's way s	0	3	0	237		http://abs.twimg.com/sticky/default_profile_image
14	@ZelenskyUa 🇺🇦💙	0	28	0	8		http://abs.twimg.com/sticky/default_profile_image
5	@ZelenskyUa Totally agree we should give Ukraine the fig	0	6	0	5		http://abs.twimg.com/sticky/default_profile_image
12	@Akhandbharat_23 @ZelenskyUa Maybe you should read	0	13	0	47		http://abs.twimg.com/sticky/default_profile_image
15	@ZelenskyUa @OlafScholz I am German, I am ashamed th	0	3	0	2		http://abs.twimg.com/sticky/default_profile_image
11	@ZelenskyUa @OlafScholz finally some real help and not	0	10	0	2		http://abs.twimg.com/sticky/default_profile_image
9	@ZelenskyUa Faith without works is dead. More provisio	0	47	0	51		http://abs.twimg.com/sticky/default_profile_image
13	@ZelenskyUa He die when documenting this war. Many di	0	2	0	51		http://abs.twimg.com/sticky/default_profile_image
18	@ZelenskyUa @OlafScholz America @POTUS , uk and Franc	0	20	0	3		http://abs.twimg.com/sticky/default_profile_image
17	@ZelenskyUa It is the whole world's aspect to help a coun	0	19	0	92		http://abs.twimg.com/sticky/default_profile_image
7	@ZelenskyUa once a comedian always a comedian	0	0	0	16		http://abs.twimg.com/sticky/default_profile_image
13	@alices_headband @Bobby_Don_Velch @ZelenskyUa Ge	0	8	0	9		http://abs.twimg.com/sticky/default_profile_image
7	@ZelenskyUa @OlafScholz If he promised u must keep pu	0	18	0	50		http://abs.twimg.com/sticky/default_profile_image
9	@Cassizzi @ZelenskyUa Weaken Biden? How? He is bare	0	1	0	41		http://abs.twimg.com/sticky/default_profile_image
9	@ZelenskyUa @AndzejDuda Handle your own troubles yo	0	4	0	150		http://abs.twimg.com/sticky/default_profile_image
5	@ZelenskyUa @OlafScholz German people are behind you	0	0	0	25		http://abs.twimg.com/sticky/default_profile_image

Ilustración 22: vista de Power Query (comentarios Twitter)

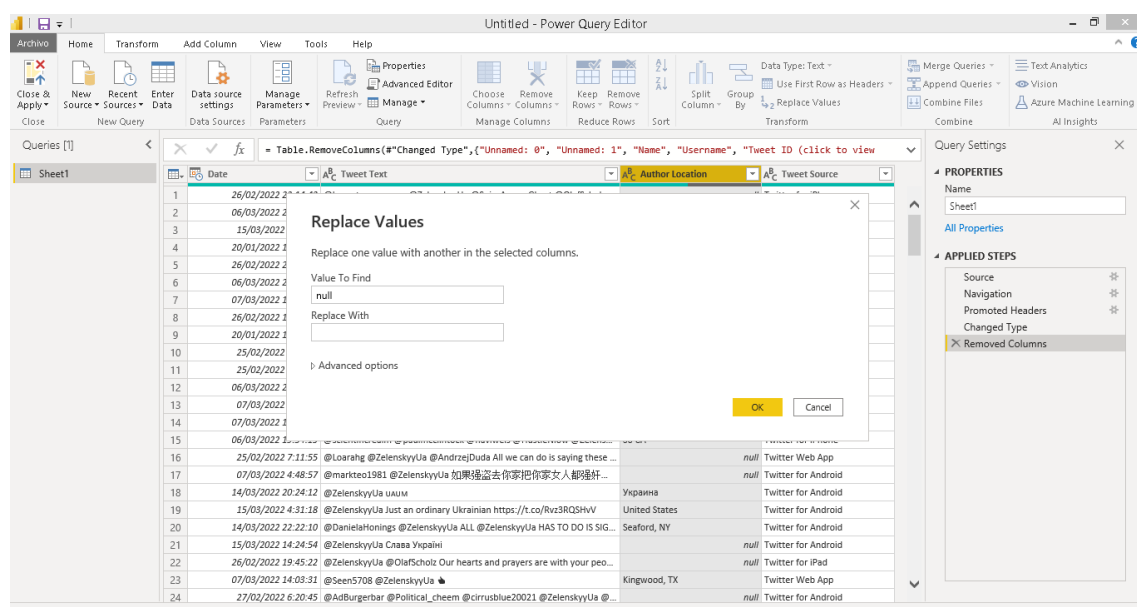


Eliminamos la mayoría de columnas, que no nos resultan de interés analítico. Lo principal es el texto de los comentarios dejados por los usuarios de Twitter. También conservaremos fechas y localizaciones.

Limpieza de datos:

Análisis estadístico y distribución de valores: nos centraremos en valores textuales con lo que en principio no será necesario.

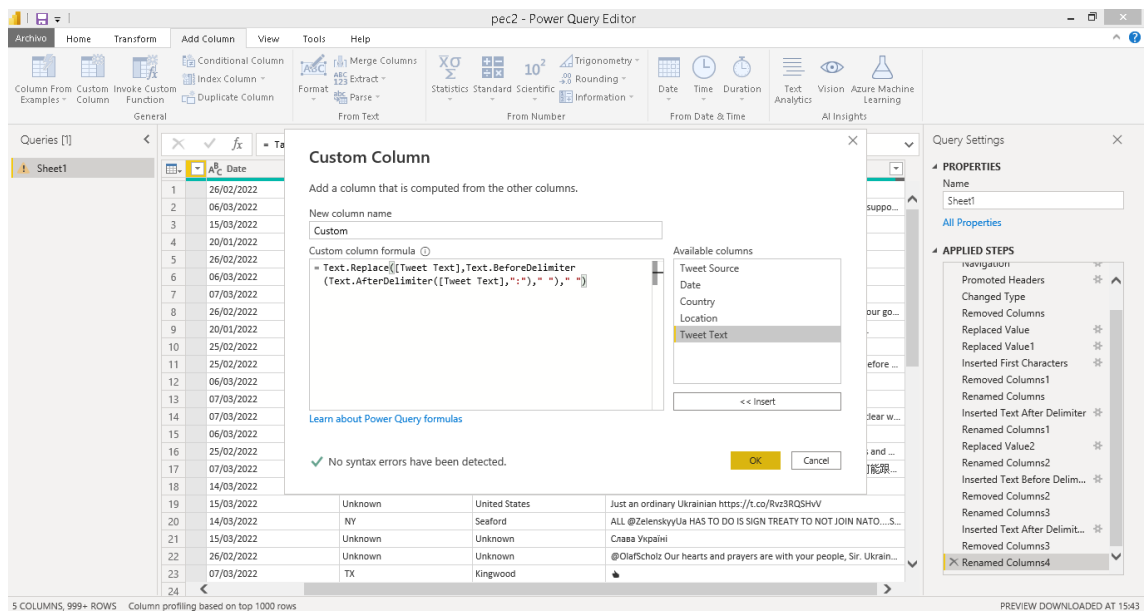
Valores nulos: la mayoría forman parte de las columnas que no usaremos. En caso de que se mantuviese alguno, deberemos decidir, según su naturaleza, si sustituirlos por un 0 o por un espacio vacío o similar (son las alternativas más lógicas en este conjunto de datos concreto).



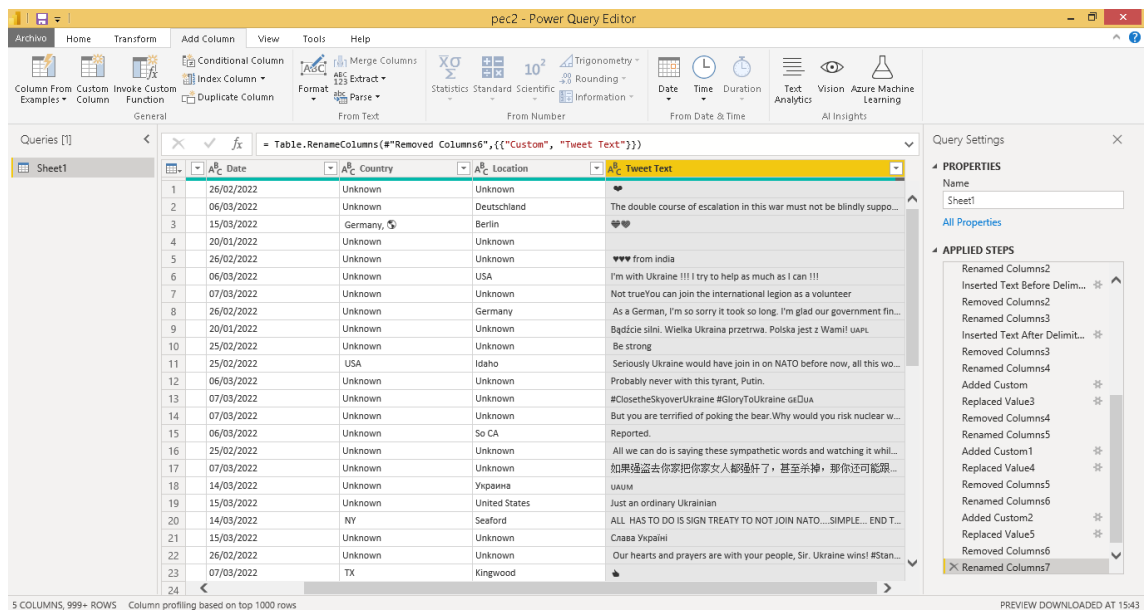
**Ilustración 23: Reemplazar valores en Power Query**

De cara a realizar un análisis de sentimientos y visualización de datos textuales, es conveniente eliminar aquellos elementos que no nos aportan información, si es que no los eliminamos previamente con Python.

Por ejemplo, eliminamos las menciones (@alias) y los enlaces (https:...). En la siguiente imagen se muestra una posible forma de eliminar los enlaces (a partir de “.”).



**Ilustración 24: Eliminación de enlaces en Power Query (1)**

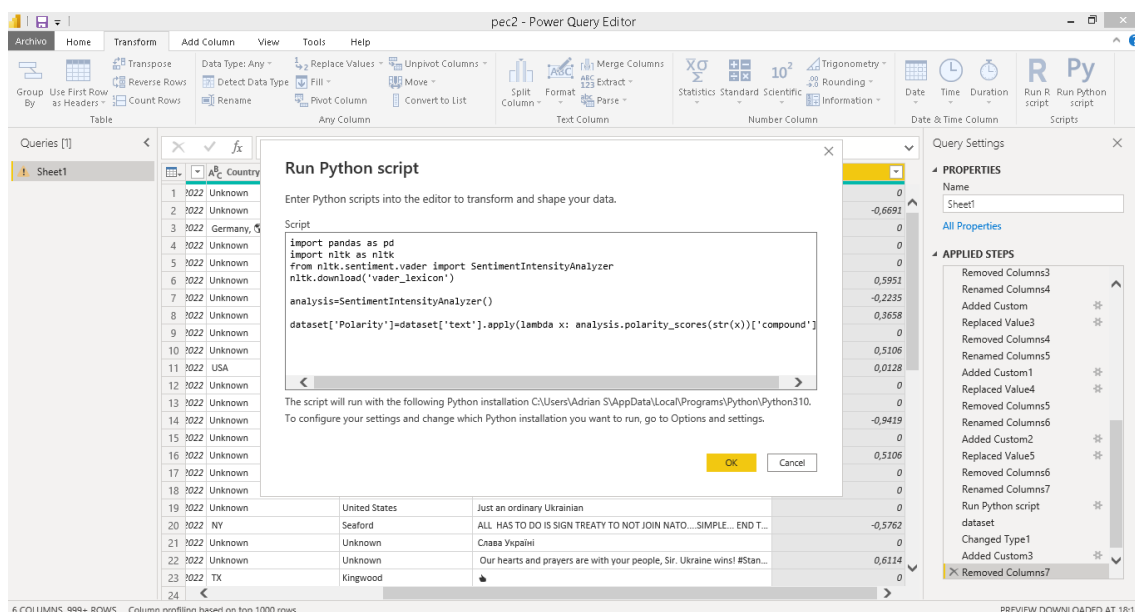


**Ilustración 25: Eliminación de enlaces en Power Query (2)**

Para poder realizar sobre el texto depurado un análisis de sentimientos en Power BI tenemos dos opciones.

1. La primera sería mediante el servicio de Microsoft Cognitive Services, para lo que necesitaríamos una suscripción de pago.
2. La segunda forma es emplear scripts externos (en nuestro caso, de Python).

Emplearemos esta segunda opción, ejecutando el siguiente script sobre nuestros datos textuales. Previamente deberemos haber instalado tanto Python como las librerías pandas y nltk (cmd -> pip install).



**Ilustración 26: Ejecución de script de Python para análisis de sentimientos en Power Query**

El script anterior accede a la librería de procesamiento de lenguaje natural, nltk, y genera una nueva columna de polaridad que refleja el tipo de sentimiento asociado a cada comentario, mediante una puntuación numérica. Las puntuaciones positivas denotan un sentimiento positivo (apoyo, satisfacción, etc). Las puntuaciones negativas sugieren enfado, rechazo, ira. En algunos casos el tipo de sentimiento no queda claro y se simboliza con una polaridad neutra (0).

Esta opción de análisis de sentimientos no funciona demasiado bien con idiomas que no sean el inglés, pero hemos escogido específicamente aquellos mensajes de Zelensky escritos en dicho idioma y donde la mayoría de interacciones se realizan también en esa lengua.

### 2.2.1.5. Visualización de datos y reporte

En esta fase en la que ya tenemos los datos textuales preparados y con una polaridad asignada, nos planteamos qué objetos visuales podrían ser útiles para analizar la opinión de los usuarios de Twitter sobre el conflicto.

Se decidió dividir el problema en 3 líneas de estudio, empleando la funcionalidad de duplicar tablas y de añadir columnas condicionales en Power Query:

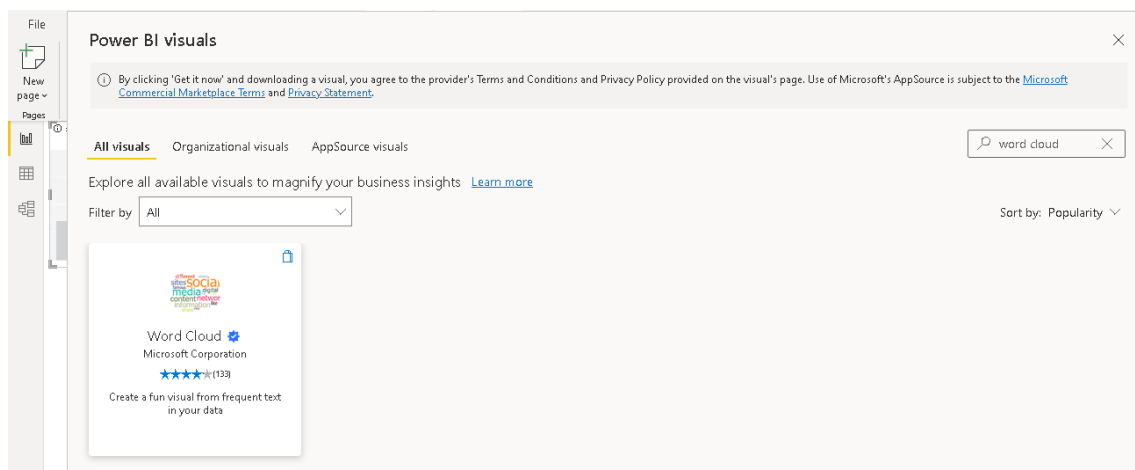
- **Bando ucraniano**, con el texto original filtrado a través de Power Query, de forma que se eliminaran todos aquellos comentarios que hicieran referencia a Putin, Rusia, los rusos, la OTAN, la UE o el presidente Biden.

- **Bando ruso**, con aquellos comentarios que hicieran referencia directa a Putin, Rusia o los rusos, pero no al resto de actores.
- **Bando occidental (UE-OTAN)**, con aquellos comentarios que mencionaran a estas potencias o al presidente Biden, pero no al resto.

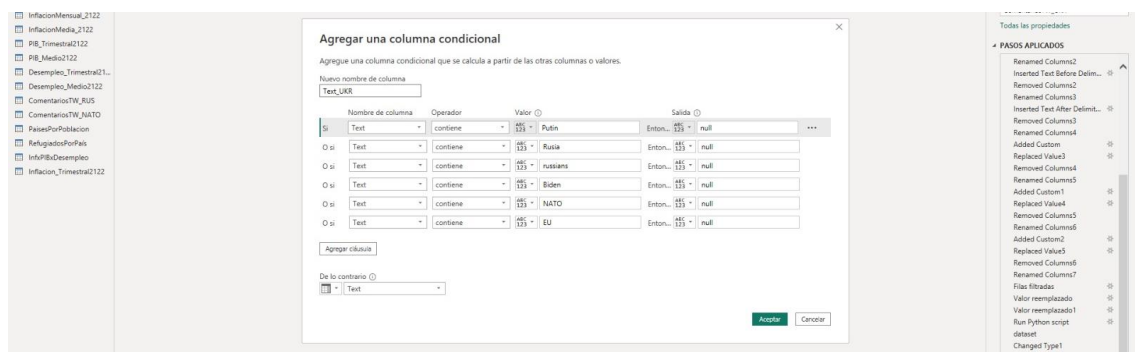
Para visualizar los comentarios de cada una de esas líneas de estudio se planteó un reporte con:

- Gráfico de líneas por polaridad media en función del tiempo.
- Medidor de polaridad media del momento seleccionado.
- Tabla que muestre el texto de los comentarios, pudiendo ordenarlo por polaridad.
- Nube de palabras con aquellas más frecuentes entre los comentarios, eliminando las stop words y estableciendo un máximo de 200.

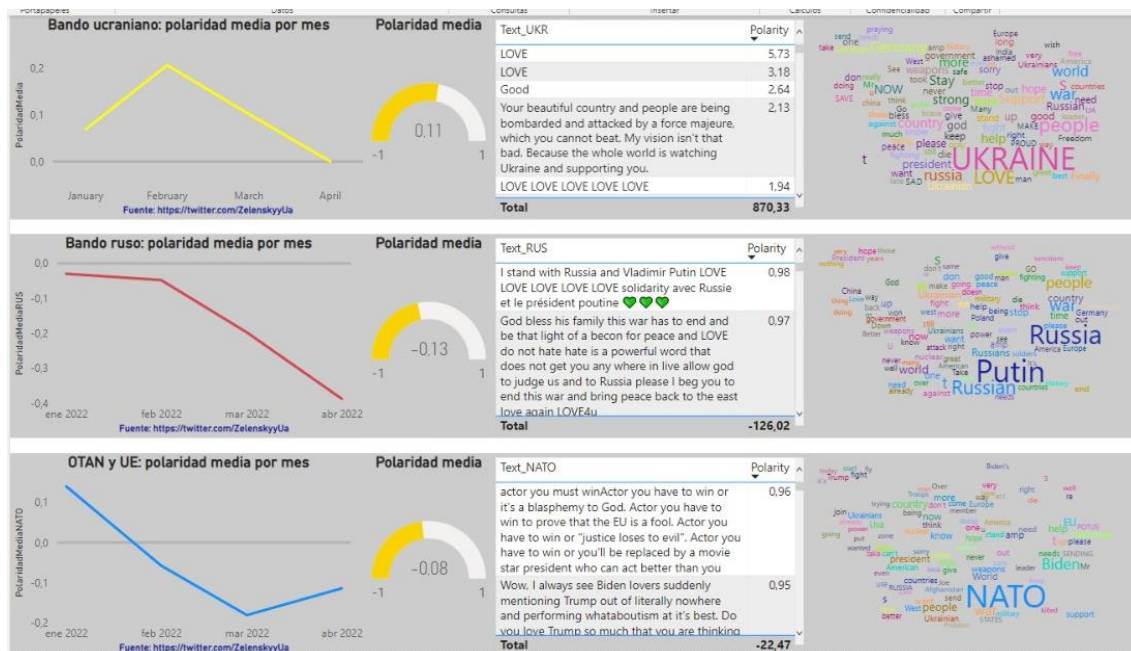
Para crear nubes de palabras, deberemos importar el gráfico, que no está disponible por defecto:



**Ilustración 27: Importación de objeto visual – Nube de palabras**



**Ilustración 28: Filtrado de palabras para el bando ucraniano mediante columnas condicionales**



**Ilustración 29: Reporte social – Opinión pública**

Un primer vistazo al reporte parece indicar que ningún bando obtiene unas puntuaciones claramente positivas o negativas, sino que son bastante neutras.

Hay que tener presente que ni el algoritmo de análisis de sentimientos ni los filtros que aplicamos al texto son perfectos, y, además, los comentarios muy positivos pueden compensarse con los muy negativos y viceversa.

Si comparamos un bando con otro vemos que el bando ucraniano es el único que logra mantenerse en valores positivos. Por su parte, el bando ruso es el que peor parado sale en cuanto a valoración, y la OTAN y UE están en un punto intermedio, aunque más cercano al resultado ruso.

También parece haber una tendencia negativa, ya que la polaridad de todos los bandos tiende a caer (salvo por el último mes estudiado en el caso OTAN-UE). Quizá podría significar que la gente empieza a cansarse del conflicto y los mensajes de apoyo se reducen, mientras aumentan el rechazo y la indiferencia.

Cada línea de estudio puede filtrarse por fecha (desde el gráfico de líneas), comentario (desde la tabla), o palabra de la nube.

### **2.2.2. Crisis humanitaria (refugiados)**

El segundo y último aspecto social que se decidió incluir sobre el proyecto fue el tema de los refugiados. A la hora de analizar el conflicto, es importante saber cuántas personas se vieron desplazadas, dónde están siendo acogidas, y qué países se encuentran más sobrecargados.

Por otro lado, puede ser interesante hacer un análisis predictivo de cuántas personas desplazadas puede haber en una fecha futura partiendo de los datos ya disponibles.

#### **2.2.2.1. ¿Qué es análisis predictivo?**

Se trata de la aplicación de técnicas y herramientas cuyo objetivo es tratar de ofrecer datos futuros basándose en aquellos de los que ya se dispone y en las tendencias mostradas.

El análisis predictivo no es necesariamente parte del BI, ya que como se indicó al principio, este último suele centrarse en lo que ya ha ocurrido y en los datos que ya están disponibles.

No obstante, el análisis predictivo es muy útil para los negocios como manera de adelantarse a los acontecimientos y poder visualizar escenarios realistas en el corto y medio plazo.

Un ejemplo muy básico de técnica de análisis predictivo es la regresión lineal, pero podemos encontrarnos otros mucho más complejos como las redes neuronales.

#### **2.2.2.2. Carga y transformación de datos en Power BI**

Los datos se extrajeron del portal de datos del Alto Comisionado de las Naciones Unidas para los Refugiados. (<https://data.unhcr.org/en/situations/ukraine>). Sin embargo, el número de refugiados huidos de Ucrania a lo largo del tiempo se dejó de actualizar en formato descargable durante el verano. Para recuperar los datos se ha usado Wayback machine. Esto permitió acceder al documento donde se actualizaban dichas cifras, que ya no estaba disponible en la web desde al menos julio, pero que seguía existiendo (<https://web.archive.org/web/20220713231017/https://data.unhcr.org/es/situations/ukraine>).

Los datos a los que se ha tenido acceso son aquellos que recogen las cifras desde el inicio del conflicto hasta mediados del mes de octubre de este año. Los archivos se proporcionan en formato JSON, que en caso necesario se podría transformar a formato Excel.

Por otro lado, se copiaron también los datos de las tablas interactivas disponibles recientemente, que recogen el número de refugiados por país.

Un dato del que no disponemos y que nos será útil para posteriores visualizaciones es la población total de cada país. Una forma rápida de obtener esta información es acudir a la Wikipedia e importar, mediante “Obtener datos” desde Power BI, la tabla correspondiente que incluye nombres de países y su población.

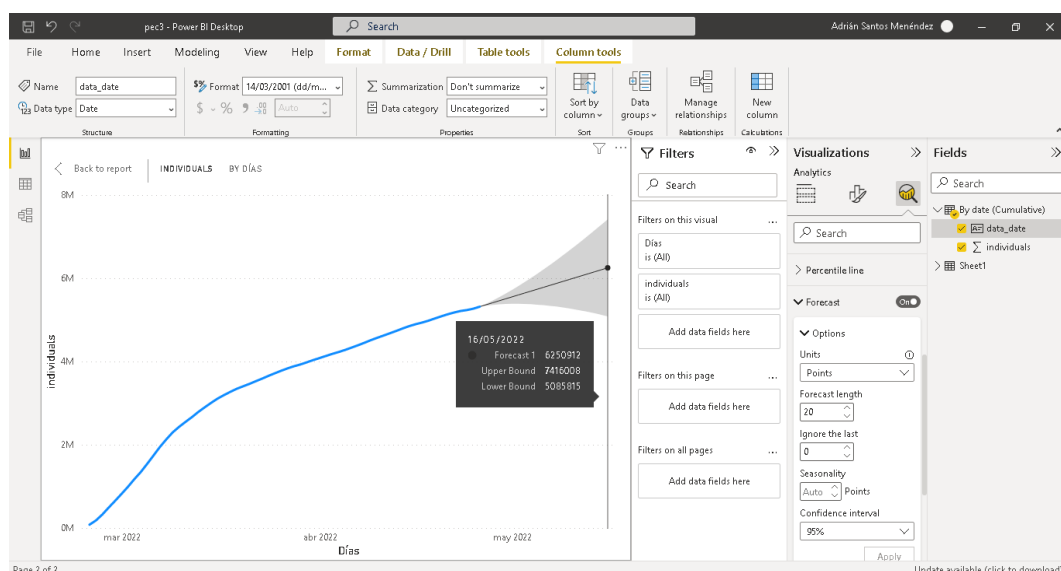


**Ilustración 30: carga de tabla con datos sobre población de países (Wikipedia)**

Posteriormente, podemos hacer una combinación de consultas desde Power Query y así obtener una única tabla que tenga ambos datos: número de refugiados en el país, por un lado, y su población autóctona, en otra columna.

### 2.2.2.3. Visualización de datos y reporte

Prueba de análisis predictivo en Power BI: estimación a 20 días, para la fecha 16/05/2022, partiendo de un conjunto de datos recopilados desde el 24/02/2022 hasta el 26/04/2022



**Ilustración 31: Número de refugiados a fecha 16/05/2022. Gráfico predictivo con confianza del 95%.**



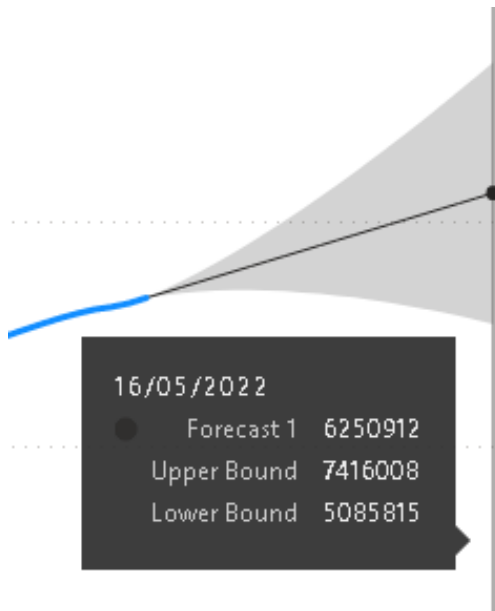


Ilustración 33: Confianza 95% (1)

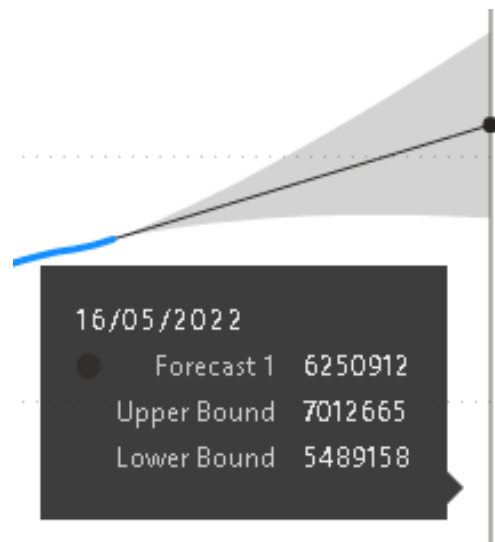


Ilustración 32: Confianza 80% (1)

La cifra real, según datos oficiales del *Operational Data Portal* del Alto Comisionado de las Naciones Unidas para los Refugiados (UNHCR), fue de 6.266.843. Dicha cifra está dentro de ambos intervalos y es muy ajustada a lo predicho.

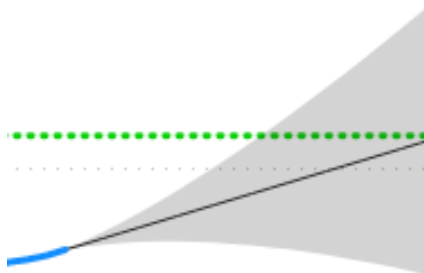


Ilustración 34: Confianza 95% (2)

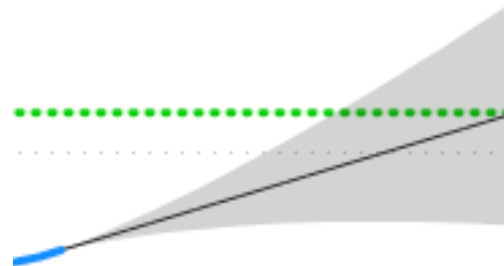


Ilustración 35: Confianza 80% (2)

**En verde, el valor realmente alcanzado en dicha fecha**

$$\text{Error porcentual: } \left| \frac{\text{Valor predicho} - \text{Valor real}}{\text{Valor real}} \right| \times 100 = 0.25\%$$

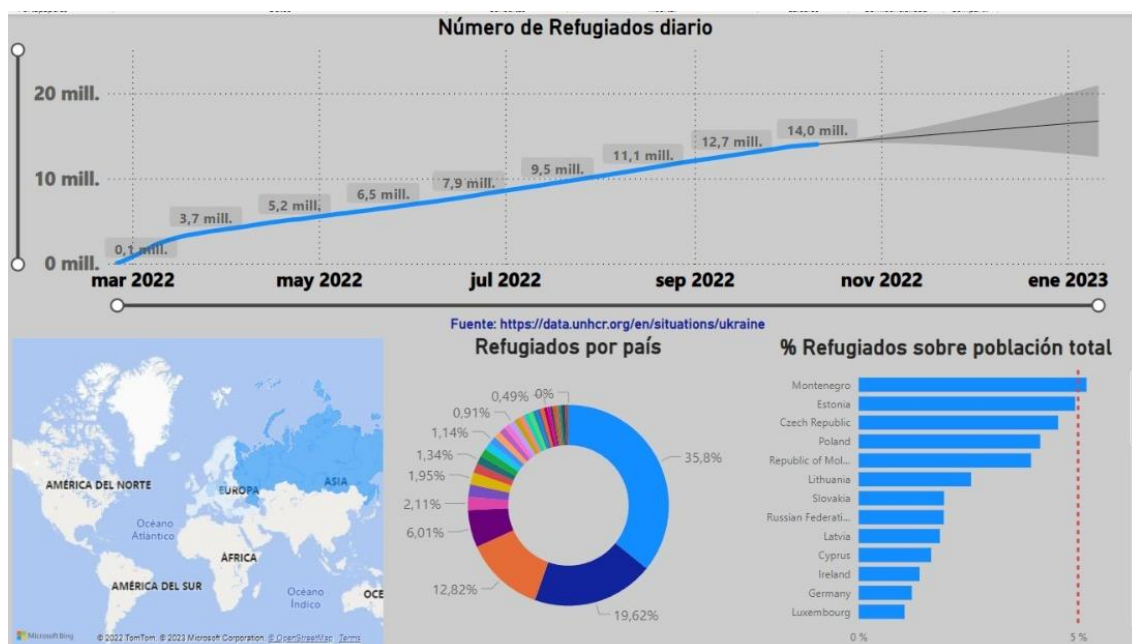
La precisión tan elevada a veces puede ser sinónimo de *overfitting*, pero en este caso es debida a que claramente se sigue una tendencia lineal. Power BI está preparado para escoger los métodos de predicción más adecuados, en función del tipo de datos que tengamos.

En algunos casos se emplean algoritmos capaces de funcionar en series con cierta periodicidad, y en otros aquellos que se fijan más en la tendencia general, mediante regresiones lineales y no lineales que se ajusten al modelo.



Finalmente, para este apartado, se ha considerado útil contar con los siguientes gráficos:

- Un gráfico de líneas que muestre la evolución del número de refugiados a lo largo del tiempo. Además, este gráfico admite análisis predictivo muy sencillo desde Power BI mediante la opción de *forecasting*, y así podemos estimar cuántos refugiados habrá en el futuro próximo.
- Un mapa de color que muestre visualmente la cantidad de refugiados que hay en cada país.
- Un gráfico de donut que indique en qué proporción (%) están distribuidos los refugiados por cada país.
- Un gráfico de barras que muestre la proporción (%) entre los refugiados acogidos por cada país y su población. Dicho gráfico permitiría tener una medida objetiva de hasta qué punto un país se encuentra sobrecargado. Se añade además como umbral máximo de proporción un 5%, de forma que se vea de manera más inmediata cuáles tienen demasiados refugiados con respecto a su población total.



**Ilustración 36: Reporte social – Crisis humanitaria (refugiados)**

Se observa que el número de refugiados que abandonan suelo ucraniano no dejó de crecer en 2022, y la predicción es que siga creciendo a buen ritmo.

Por otro lado, Rusia y Polonia son quienes más refugiados reciben, pero quienes acogen a un mayor número con respecto a la población que tienen son Montenegro y Estonia. Esta información podría ser útil a la hora de tomar decisiones sobre dónde enviar a nuevos refugiados en un futuro.

## 2.3. Análisis económico

El segundo componente que se planteó para este trabajo se refiere a la parte económica del conflicto. Se trata de preguntarse qué indicadores económicos podría ser interesante monitorizar de cara a comprobar los efectos de la guerra en las economías mundiales.

### 2.3.1. Selección de indicadores

Para analizar la componente económica del proyecto, debemos escoger unos *Key Performance Indicator* razonables. Estas variables son las que, dentro del *Business Intelligence*, representan aquellos valores que mejor representan la evolución favorable o desfavorable de una compañía de cara a cumplir sus objetivos.

Para analizar posibles indicadores podemos seguir el sistema Meta-Pregunta-Medida (Goal-Question-Metric).

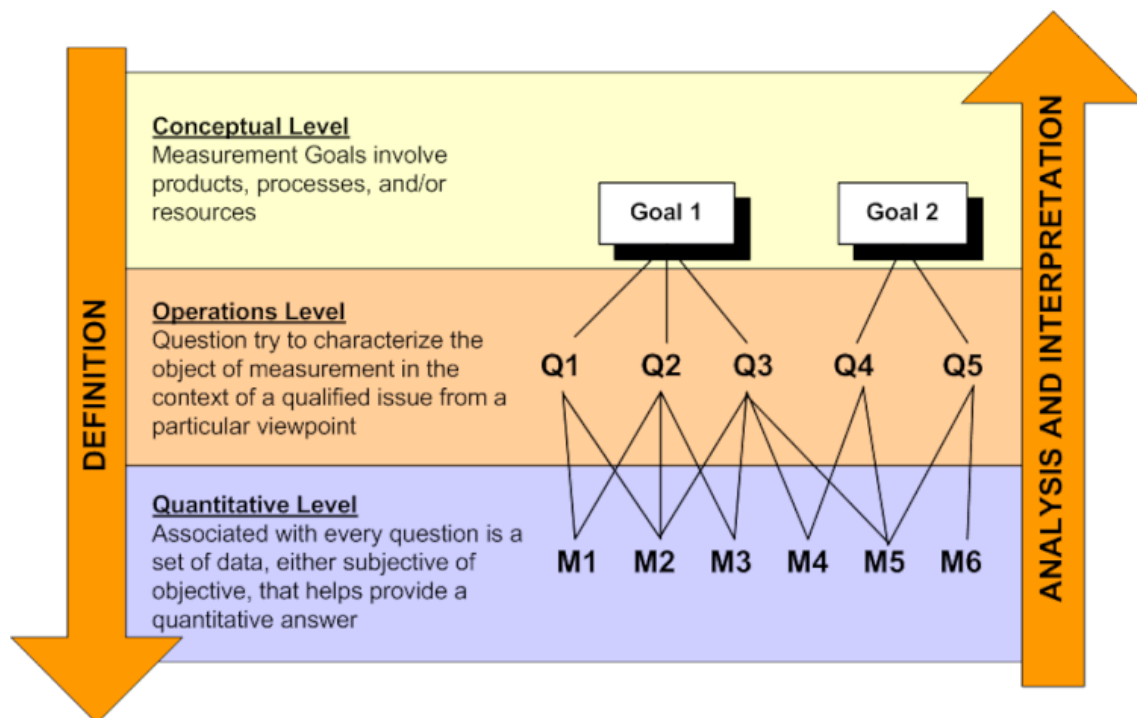


Ilustración 37: Objetivo-Pregunta-Medida

[Fuente de la imagen](#)

Si en el caso de este proyecto la compañía son los países que pueden verse afectados por el conflicto, podemos pensar:

Meta 1: mantener una evolución de precios en los mercados que sea razonable

Pregunta: ¿cómo están cambiando los precios?

Medida: INFLACIÓN

Meta 2: mantener un crecimiento razonable del valor de la producción

Pregunta: ¿cómo está evolucionando el valor de lo que se produce?

Medida: PRODUCTO INTERIOR BRUTO

Meta 3: mantener unos niveles razonables de renta dentro de la población

Pregunta: ¿cómo ha evolucionado la renta de los habitantes de cada país?

Medida: RENTA NACIONAL BRUTA

Meta 4: mantener unos niveles de desempleo reducidos

Pregunta: ¿cómo ha evolucionado el número de trabajadores?

Medida: TASA DE DESEMPLEO

De este modo llegamos a los 4 indicadores económicos básicos de cualquier economía.

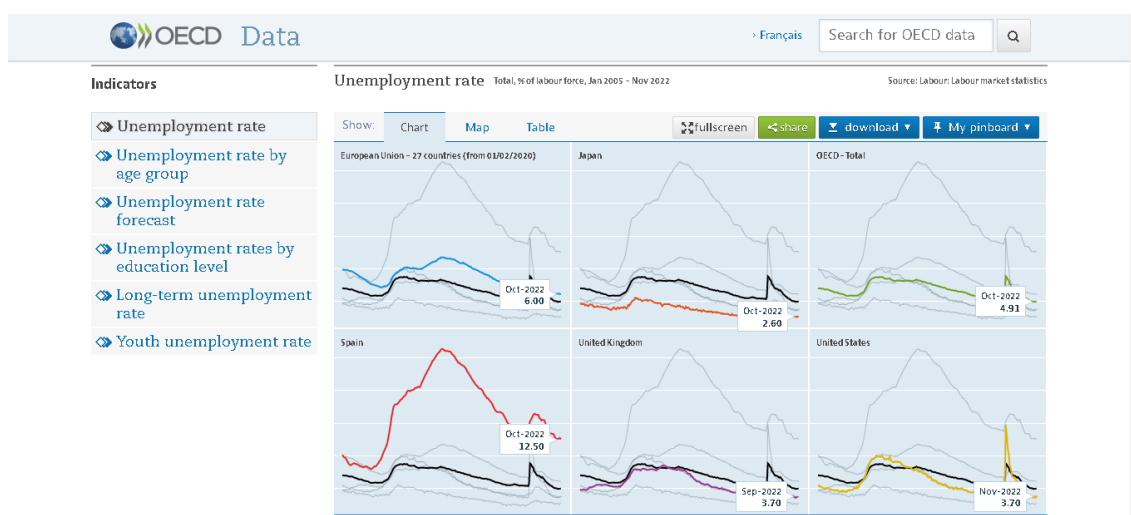
**Inflación:** medida de la subida del precio de los productos, que a su vez reduce el valor del dinero disponible.

**Producto Interior Bruto:** medida del valor de todos los bienes producidos en un país.

**Renta Nacional Bruta:** medida de las retribuciones generadas a raíz de la producción realizada por un país.

**Tasa de desempleo:** medida del número de personas que forman parte de la población activa pero no tienen un puesto de trabajo.

Podemos encontrar datos fiables sobre todos los mentados indicadores en <https://data.oecd.org/>

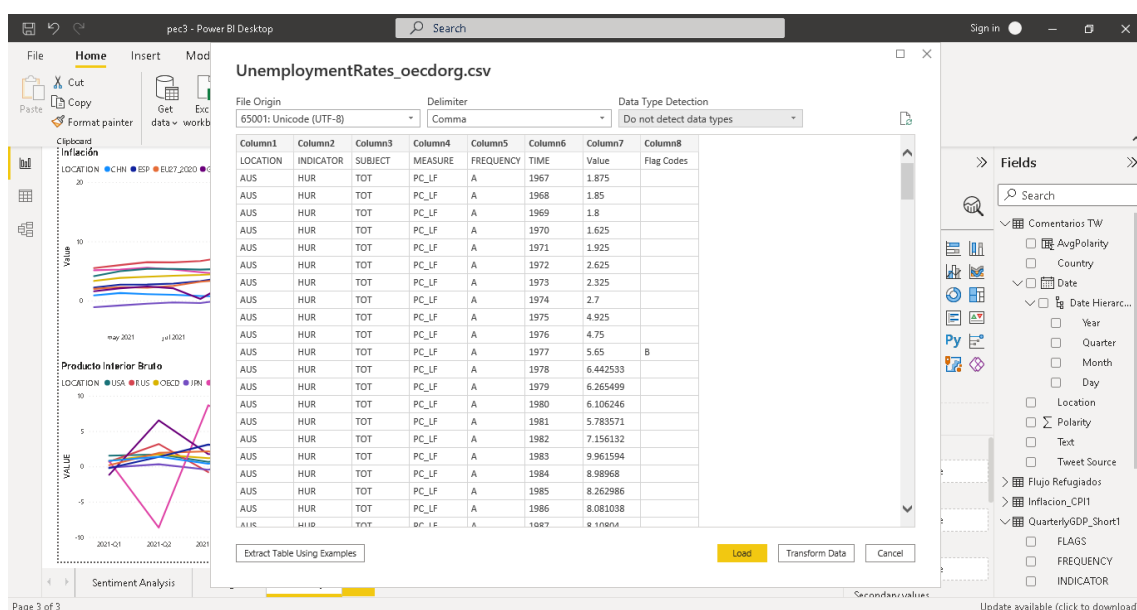


**Ilustración 38: Muestra de algunos datos disponibles en OECD.org**

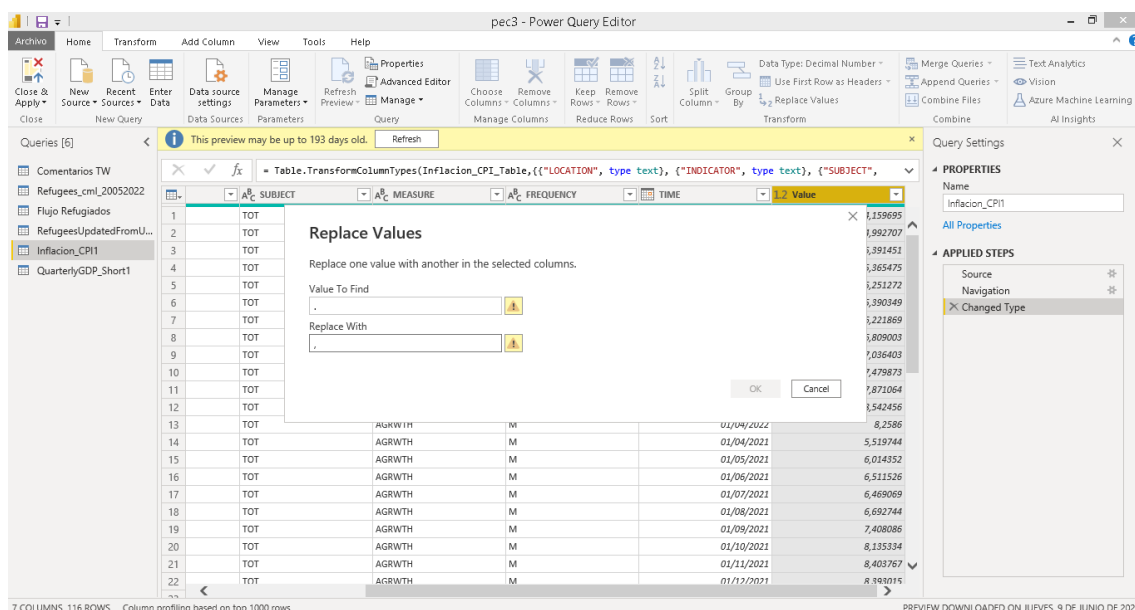
Los datos se pueden descargar ya parcialmente filtrados (eliminando años o países que no se requieran para el análisis).

En este punto cabe destacar la importancia de los formatos de los datos que se emplean. Por ejemplo, en este caso los datos están disponibles en un formato .csv en el que el separador de los valores decimales es un “.”.

Esto generó algunos problemas en Power BI a la hora de procesar correctamente los valores decimales. Para subsanarlo, es necesario cargar los datos con la opción “No detectar tipo de datos”. En caso de que Power BI no separe correctamente las columnas del documento, se deberán mantener los valores numéricos como texto para posteriormente sustituir los puntos por una coma a través de Power Query.



**Ilustración 39: Carga de datos donde el delimitador decimal es un “.”**



**Ilustración 40: Importancia de mantener un formato sólido. Cambio de “.” por “,”**

Además de los problemas técnicos de formato, encontramos otras dificultades con los datos de nuestros indicadores económicos:

- **Todos los indicadores:** no incluyen a Ucrania. Sin embargo, podemos prescindir de esos datos ya que el trabajo se centra más en saber qué ha ocurrido con el resto de países.

- **Inflación:** la tasa de inflación de Rusia está desactualizada (solo cubre hasta marzo de 2022). Solución elegida: ya que no hay datos mensuales o trimestrales de organismos fiables, comprobar cuál es la inflación media de Rusia prevista por los organismos internacionales y a partir de ahí completar de forma aproximada los valores faltantes. Esto permite evitar excluir del análisis a un país tan importante para la temática del trabajo como es Rusia. La tasa de inflación media prevista para Rusia en 2022 ronda 12.5% según el Fondo Monetario Internacional:

<https://www.imf.org/external/datamapper/PCPIEPCH@WEO/RUS?zom=RUS&highlight=RUS>

- **Producto Interior Bruto:** también está desactualizado para Rusia. No hay registros desde finales del tercer cuatrimestre de 2021. Solución: análoga a la seleccionada para la inflación. Por otro lado, falta un dato de Argentina, el tercer trimestre de 2022. Solución: lo obtenemos de las previsiones del propio gobierno argentino: <https://www.indec.gob.ar/indec/web/Nivel4-Tema3-9-47#:~:text=La%20estimaci%C3%B3n%20preliminar%20del%20producto,mismo%20per%C3%A9odo%20del%20a%C3%B1o%20anterior>

- **Renta Nacional Bruta:** los valores de 2022 no están disponibles. Solución: se decide prescindir de este indicador ya que los datos

disponibles no abarcan en absoluto el período a estudiar en este proyecto.

- **Tasa de desempleo:** no hay registros recientes para varios países como Rusia o China. Solución: se prescinde de dichos países en lo que respecta al apartado del desempleo al no haber encontrado datos recientes y fiables.

Otros países que no tenían información sobre datos varios, como Costa Rica, han sido eliminados del análisis.

### 2.3.2. Inflación

La inflación es uno de los grandes indicadores económicos que interesaría medir para determinar hasta qué punto la guerra en Ucrania ha afectado a la economía mundial. Es, además, uno de los indicadores de los que más se habla, el que marca cuánto y a qué ritmo suben o bajan los precios.

La inflación considerada en este trabajo es el dato de variación año a año, uno de los más comunes.

#### 2.3.2.1. Carga y transformación de datos

Obtenemos datos fiables en formato .csv de <https://data.oecd.org/price/inflation-cpi.htm>.

Como se indicó anteriormente, hay datos faltantes de Rusia. Podemos emplear Python para completarlos a partir de los datos existentes y el hecho, ya señalado, de que la inflación media prevista para Rusia es de 12.5.

```
import csv

expectedY2022 = 12.5 #Inflacion esperada para Rusia para 2022

q1_2022 = 11.55 #Inflacion de Rusia para el primer trimestre de 2022

q_2022 = ((expectedY2022*4)-q1_2022)/3 #La inflacion de los trimestres
que faltan, para cumplir con la estimacion anual

data = [
["RUS", "CPI", "TOT", "AGRWTH", "Q", "2022-Q2", q_2022, None], #Fila segundo
trimestre 2022
["RUS", "CPI", "TOT", "AGRWTH", "Q", "2022-Q3", q_2022, None] #Fila tercer
trimestre 2022
]
```

```

with open(r'C:\Users\Adrian
S\Documents\TFG_2023\PEC3\OECD_inflacion202122_Q.csv', 'a',newline='') as
file:

    writer = csv.writer(file)

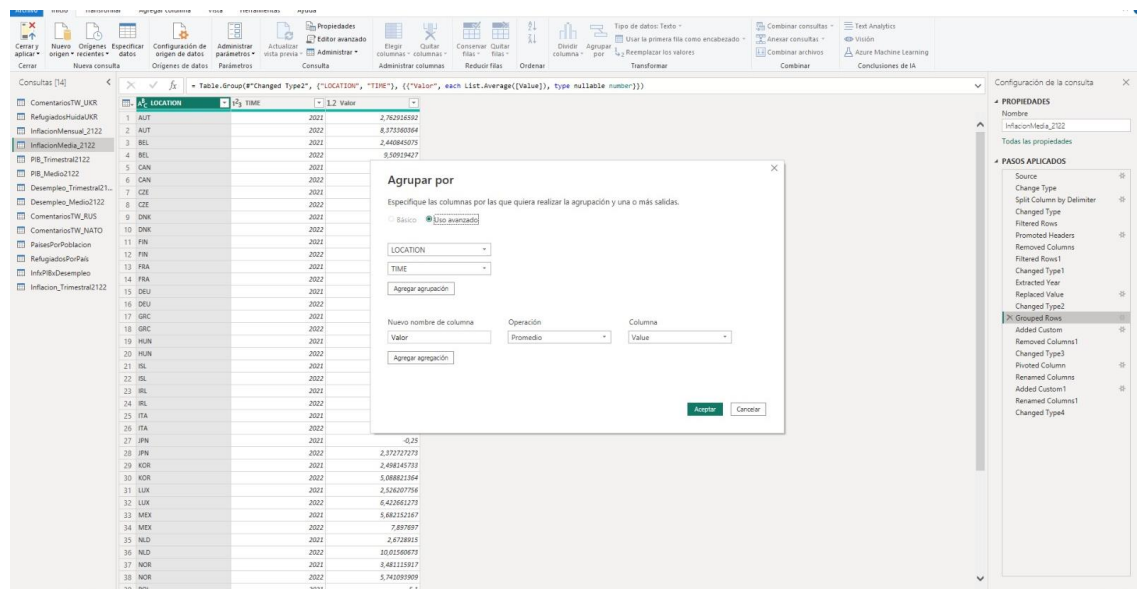
    writer.writerows(data) #Añadir filas del segundo y tercer
cuatrimestre

    print((q_2022*3+q1_2022)/4) #Comprobar que se obtiene la inflacion
anual prevista (12.5)

    file.close()

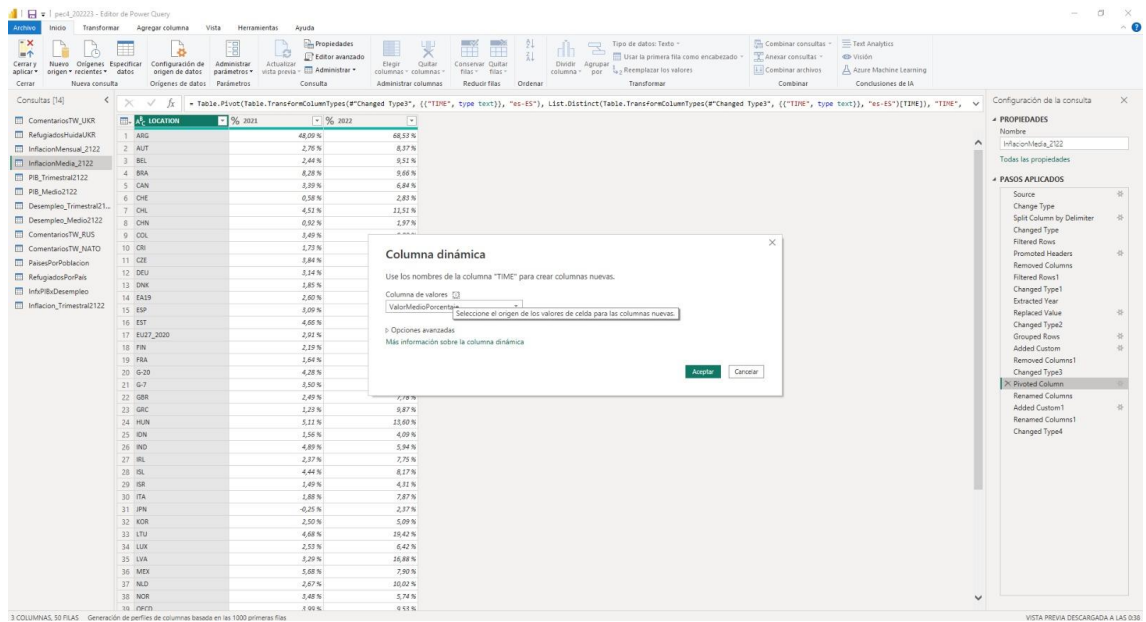
```

Mediante Power Query realizamos las transformaciones necesarias. Aparte de las operaciones básicas que ya vimos, o convertir los valores en porcentaje (modificando su tipo y dividiendo las columnas por 100), podemos emplear otras opciones para construir nuevas tablas a partir de las que ya tenemos, realizar cálculos y combinar consultas.

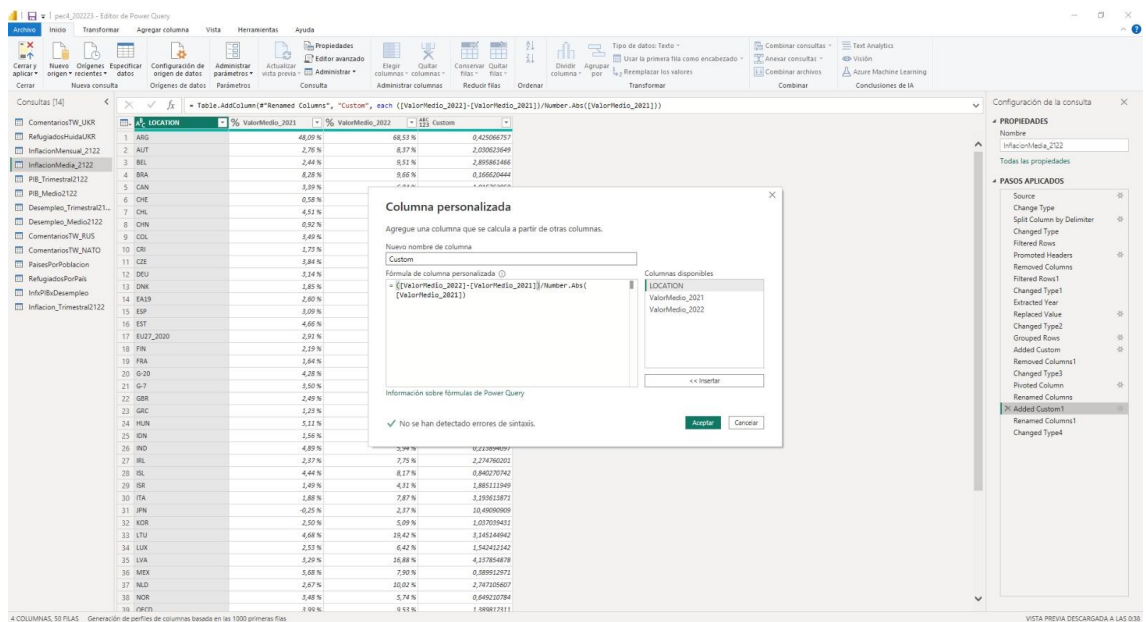


**Ilustración 41: Uso de “agrupar por” para crear una nueva tabla con valores agrupados**



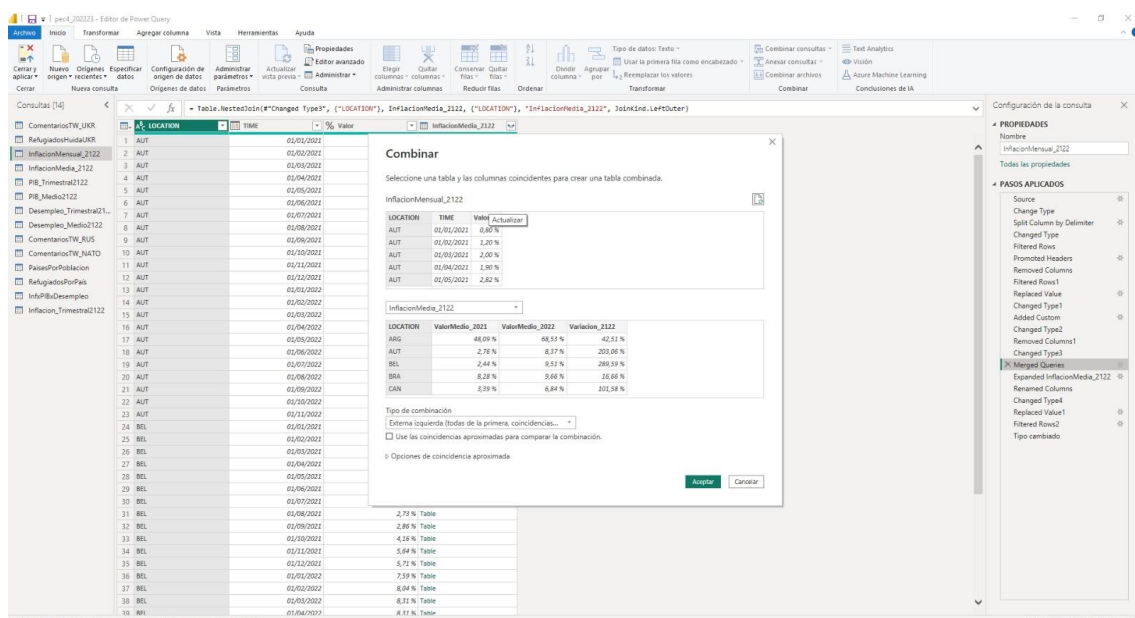


**Ilustración 42: Pivotar filas en Power Query**



**Ilustración 43: Variación porcentual**





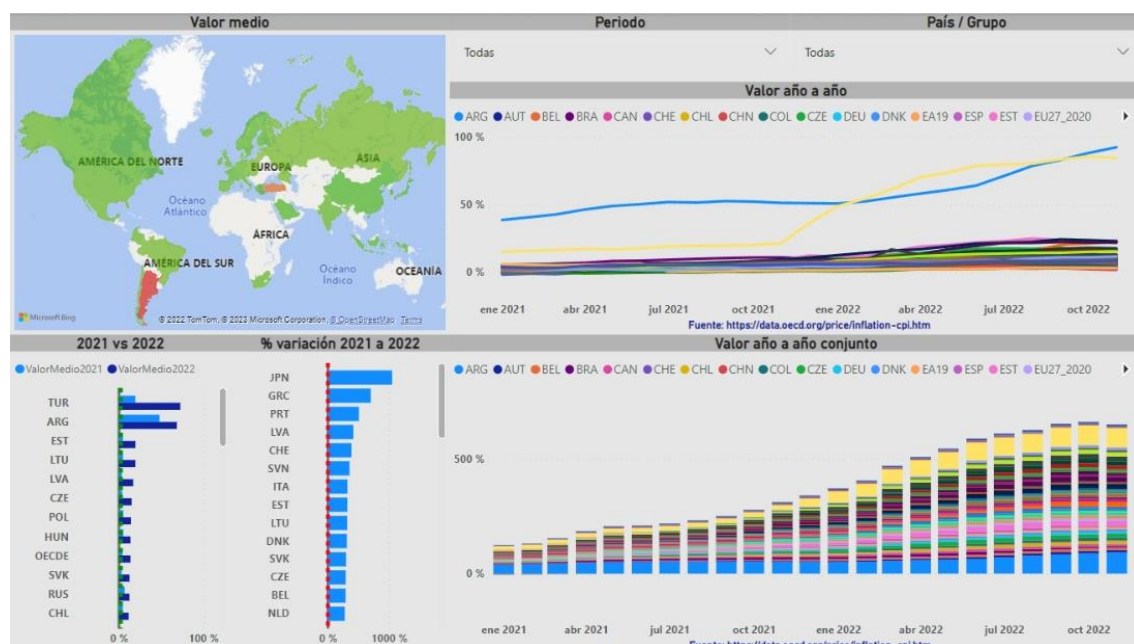
**Ilustración 44: Unión de tablas en Power Query**

### 2.3.2.2. Visualización de datos y reporte

Para este apartado, se ha considerado útil contar con los siguientes gráficos:

- Un mapa de color que muestre visualmente el valor medio para cada país en el período seleccionado.
- Un gráfico de líneas que muestre la evolución del valor para cada país en función del momento.
- Un gráfico de barras apiladas de los países que estemos visualizando, en el período escogido. Esto nos da visualmente la medida de la evolución de los valores agregados. Es decir, si el valor del conjunto visualizado sube o baja.
- Un gráfico de barras horizontales donde se muestre, para cada país seleccionado, el valor medio en 2021 y en 2022. Además, se añade una franja de valores ideales con dos líneas verticales en el 1% y 3%. Esto es debido a que los economistas generalmente consideran una inflación de en torno al 2% como ideal.
- Otro gráfico de barras horizontales donde se muestre, para cada país seleccionado, el porcentaje de variación entre el valor medio que había en 2021 y el que hay en 2022. Esto nos puede dar una medida de si el crecimiento o bajada de la inflación está acelerando o decelerando. Establecemos dos umbrales límite en el 10% y -10% ya que no es deseable que la inflación varíe mucho de un año para otro.

En este reporte se han añadido *slicers* por localización y período para poder filtrar los valores que deseamos visualizar en cada momento.



**Ilustración 45: Reporte económico – Inflación año a año 2021 y 2022**

La inflación es uno de los indicadores que más se ha disparado, con consecuencias que pueden ser variadas. El reporte nos indica que, en general, todos los países han sufrido un pico de inflación coincidiendo con el estallido de la guerra. No obstante, también nos dice que ese pico se moderó rápidamente y ahora el ritmo de crecimiento de la inflación se está ralentizando, como vemos en el gráfico de columnas apiladas.

Hay casos curiosos como el de Japón, que tiene más de un 1.000% más de inflación que en 2021, pero en cambio ahora sus niveles son mucho más saludables. Esto es debido a que partía de una inflación negativa (deflación).

Por otro lado, la mayoría de países que en 2021 estaban cerca de los valores ideales de inflación, ahora están muy lejos.

Podríamos concluir que la guerra, efectivamente, ha tenido un efecto significativo en la inflación, pero también que esta está empezando a controlarse.

### 2.3.3. PIB

El crecimiento del PIB es el que mide si un país avanza o retrocede económicamente. A raíz de la guerra y las malas perspectivas en el resto de aspectos se especuló con la posibilidad de que la mayoría de países acaben entrando en recesión, es decir, en crecimiento negativo o decrecimiento.

Nos podemos preguntar hasta qué punto la guerra ha podido influenciar en esto y qué países están sufriendo más o están más cerca de esa temida recesión.

#### 2.3.3.1. Carga y transformación de datos

Obtenemos datos fiables en formato .csv de <https://data.oecd.org/gdp/quarterly-gdp.htm#indicator-chart>

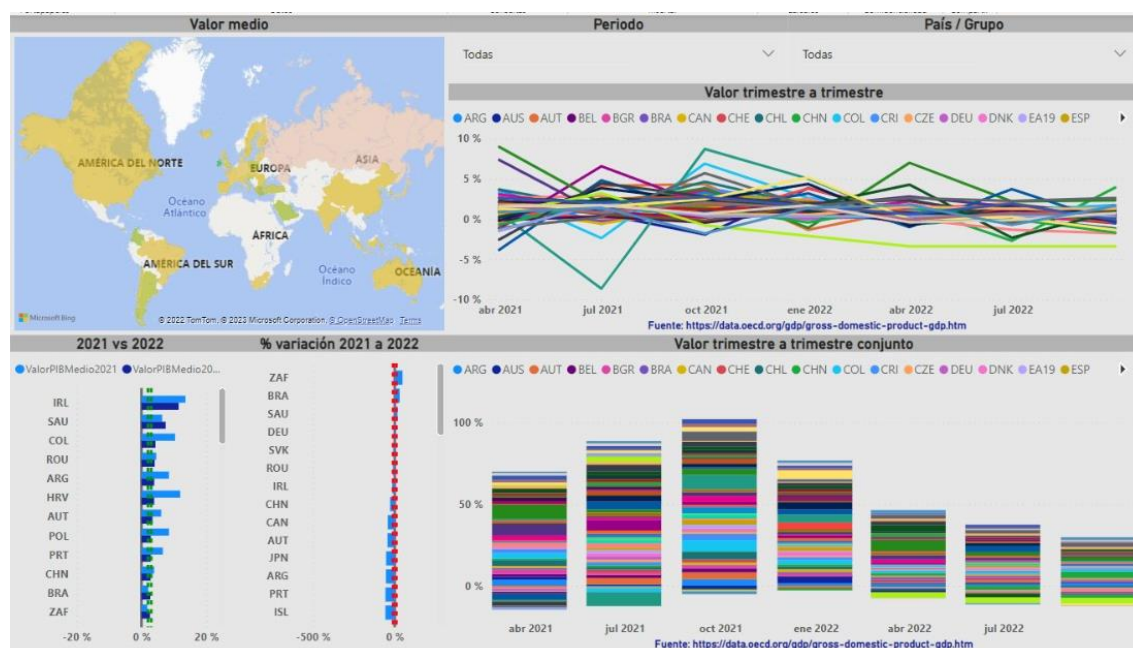
Procedemos de la misma manera que hicimos con la inflación, tanto para valores faltantes como para la carga de datos, ya que es el mismo formato de tabla.

#### 2.3.3.2. Visualización de datos y reporte

Para este apartado, se ha considerado útil contar con los siguientes gráficos:

- Un mapa de color que muestre visualmente el valor medio para cada país en el período seleccionado.
- Un gráfico de líneas que muestre la evolución del valor para cada país en función del momento.
- Un gráfico de barras apiladas de los países que estemos visualizando, en el período escogido. Esto nos da visualmente la medida de la evolución de los valores agregados. Es decir, si el valor del conjunto visualizado sube o baja.
- Un gráfico de barras horizontales donde se muestre, para cada país seleccionado, el valor medio en 2021 y en 2022. Además, se añade una franja de valores ideales con dos líneas verticales en el 2% y 3%. Esto es debido a que los economistas generalmente cifran el ritmo ideal de crecimiento del PIB en ese intervalo.
- Otro gráfico de barras horizontales donde se muestre, para cada país seleccionado, el porcentaje de variación entre el valor medio que había en 2021 y el que hay en 2022. Esto nos puede dar una medida de si el crecimiento o bajada de la inflación está acelerando o decelerando. Establecemos dos umbrales límite en el 10% y -10% ya que no es deseable que el PIB varíe mucho de un año para otro.

En este reporte se han añadido *slicers* por localización y período para poder filtrar los valores que deseamos visualizar en cada momento.



**Ilustración 46: Reporte económico – PIB trimestre a trimestre 2021 y 2022**

Podemos observar que, por desgracia, el crecimiento se ha reducido mucho y tiene una tendencia a la baja.

Algunos países muy importantes, como Estados Unidos, están en riesgo de entrar en recesión técnica, al estar muy cerca del crecimiento negativo en algunos momentos.

No obstante, la caída del crecimiento parece haber comenzado antes de la guerra, hacia finales de 2021, con lo que puede haber otros motivos subyacentes que escapan a este trabajo.

## 2.3.4. Desempleo

El desempleo es uno de los problemas que más preocupan en todos los países. Una pregunta que nos podemos hacer es si la guerra ha contribuido a destruir empleos de una forma significativa.

### 2.3.3.1. Carga y transformación de datos

Obtenemos datos fiables en formato .csv de <https://data.oecd.org/unemp/unemployment-rate.htm>

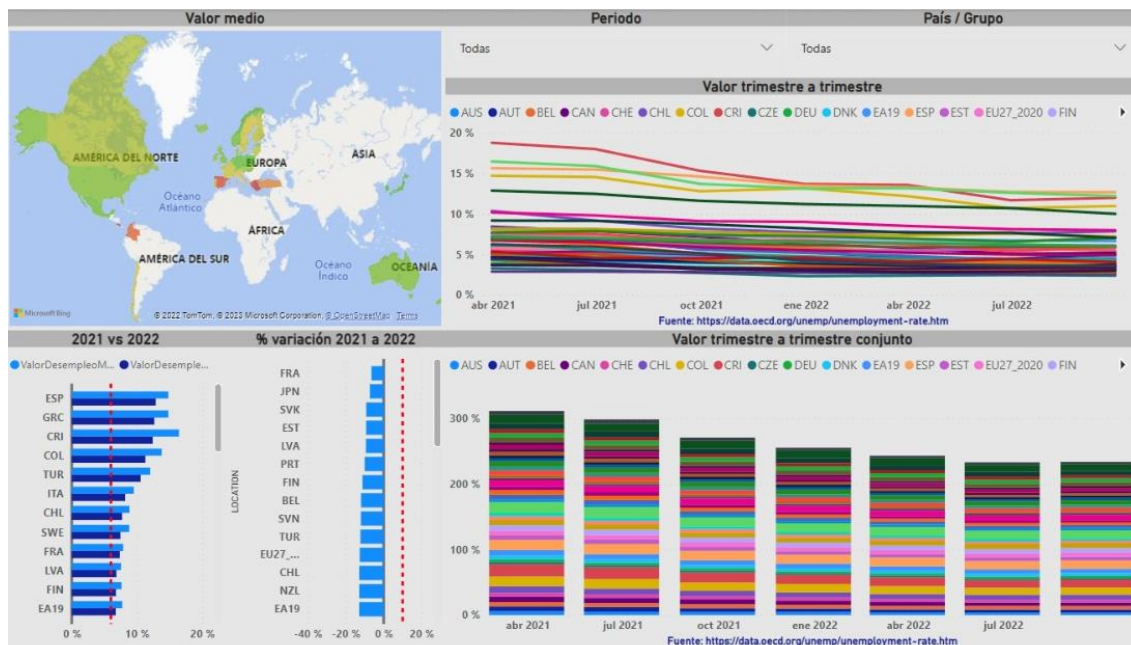
Procedemos de la misma manera que hicimos con la inflación y el PIB para la carga de datos, ya que es el mismo formato de tabla. No reconstruimos valores en este caso.

### 2.3.3.2. Visualización de datos y reporte

Para este apartado, se ha considerado útil contar con los siguientes gráficos:

- Un mapa de color que muestre visualmente el valor medio para cada país en el período seleccionado.
- Un gráfico de líneas que muestre la evolución del valor para cada país en función del momento.
- Un gráfico de barras apiladas de los países que estemos visualizando, en el período escogido. Esto nos da visualmente la medida de la evolución de los valores agregados. Es decir, si el valor del conjunto visualizado sube o baja.
- Un gráfico de barras horizontales donde se muestre, para cada país seleccionado, el valor medio en 2021 y en 2022. Además, se añade un umbral máximo del 6%. Esto es debido a que los economistas generalmente identifican valores de desempleo de en torno al 5% o menos como pleno empleo realista.
- Otro gráfico de barras horizontales donde se muestre, para cada país seleccionado, el porcentaje de variación entre el valor medio que había en 2021 y el que hay en 2022. Esto nos puede dar una medida de si el crecimiento o bajada del desempleo está acelerando o decelerando. Establecemos un umbral límite en el 10% ya que no es deseable que la tasa de desempleo acelere su crecimiento exageradamente.

En este reporte se han añadido *slicers* por localización y período para poder filtrar los valores que deseamos visualizar en cada momento.



**Ilustración 47: Reporte económico – Desempleo trimestre a trimestre 2021 y 2022**

Afortunadamente, en este caso parece que el indicador mejora de forma constante. No parece que la guerra haya destruido una cantidad significativa de puestos de trabajo, o quizá necesitemos datos y conocimientos más amplios para poder valorarlo.

### 2.3.5. Clusterización

El aprendizaje no supervisado se refiere al aprendizaje automático en el que no existe conocimiento a priori. Es decir, no se conoce el resultado final que debe producir el algoritmo empleado.

Uno de los ejemplos más conocidos es el del agrupamiento o clustering, habitualmente realizado mediante el algoritmo K-Medias. En *Business Intelligence* es posible beneficiarse de esta técnica para segmentar productos o clientes y ser así conscientes de sus diferencias y similitudes. En el caso que nos ocupa, podemos realizar agrupamiento de países a partir de los indicadores económicos que escogimos para ver cómo se distribuyen. Esto nos puede dar una idea de qué le podría ocurrir a un país o qué efecto pueden tomar determinadas medidas, fijándose en otro cercano de su grupo.

Tanto Python como Power BI permiten implementar agrupamiento. Escogemos Power BI por su simplicidad y de cara a elaborar el producto final con las visualizaciones.

Para realizar la clusterización bidimensional solo necesitamos crear un gráfico de puntos, abrir el menú de tres puntos y seleccionar la opción de encontrar clústeres automáticamente.

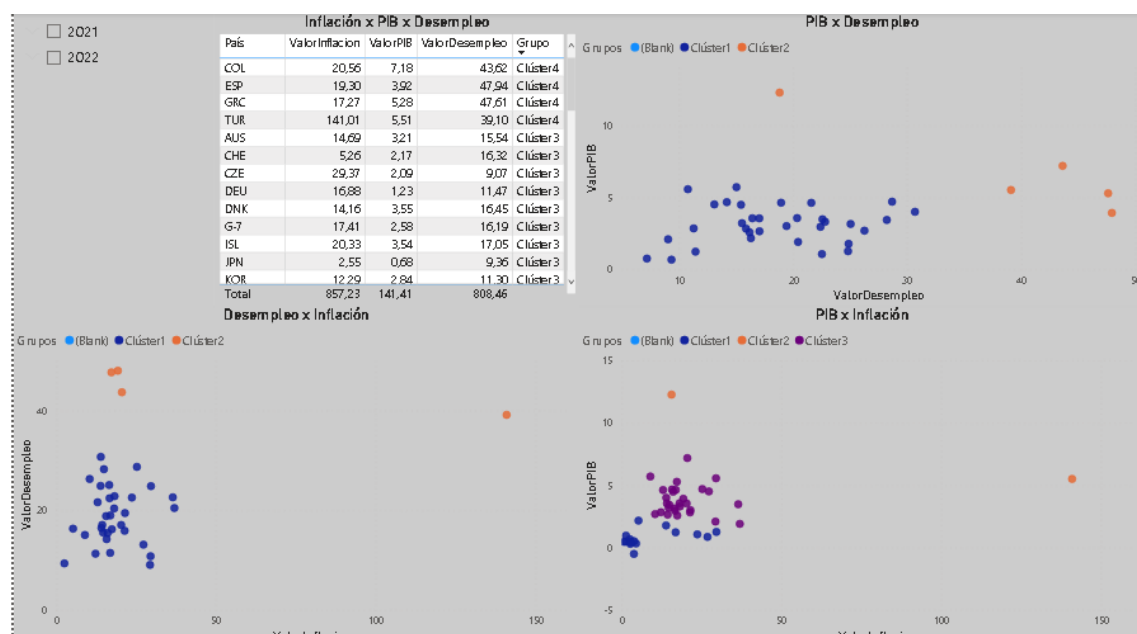


Podemos indicar nosotros cuántos grupos queremos que haga, o dejar que los halle de forma automática.

Si queremos hacer clusterización tridimensional, deberá ser con una tabla y no con un gráfico, ya que Power BI no lo permite.

Los datos que necesitamos para los gráficos ya fueron descargados y/o calculados en pasos anteriores. Básicamente, los valores por trimestre de cada una de nuestras 3 variables económicas. Podemos unir las todas en una única tabla como ya se hizo en otros pasos y construir los gráficos.

Añadiremos además un slicer por fecha para poder seleccionar el momento o período a analizar.



**Ilustración 48: Reporte económico - Clusterización**

Del reporte vemos que, por ejemplo, España se parece mucho a Grecia y a Colombia.

También vemos que, a la hora de agrupar los países por PIB e inflación, se encuentran todos muchos más juntos unos de otros, con menos diferencias o puntos solitarios.

Este tipo de información, como se comentaba anteriormente, puede servir para el proceso de toma de decisiones económicas y políticas.

### 3. Conclusiones

A través de este trabajo se ha podido investigar sobre diversas herramientas y técnicas de BI y su aplicación fuera del mundo empresarial.

La dificultad de acceso a fuentes de datos actualizadas y fiables ha supuesto un problema que ha habido que sortear, pero que no ha afectado significativamente al producto final.

Durante el desarrollo del proyecto se materializaron, además, varios de los riesgos previstos en el planteamiento inicial. Principalmente, problemas de rendimiento del equipo en el que se trabajó e imprevistos de índole personal, que ralentizaron el trabajo.

En cuanto a los objetivos inicialmente planteados, se considera que se han cumplido de manera aceptable. Sería no obstante posible, con más tiempo y recursos, ampliar y mejorar el apartado gráfico y los reportes. Por ejemplo, la clusterización podría mejorarse trabajando directamente sobre Python, los gráficos del reporte de análisis de sentimientos podrían repensarse para hacerlos más interconectados, y se podría buscar alguna forma más eficiente de lograr las mismas transformaciones de datos (mediante fórmulas avanzadas, aprovechando más las relaciones entre tablas, etc).

Por último, acerca del conflicto Ucrania-Rusia, se podría aventurar, en base a los datos observados, que se está cronificando. Ya no tiene el mismo impacto que al principio, ni en la opinión pública ni en los principales indicadores económicos, pero sigue ahí y, con ella, una influencia y riesgos latentes que habría que monitorizar.

Vemos, pues, que el *Business Intelligence* emplea un buen número de herramientas y técnicas que nos permiten dar forma y significado a los datos y marcar la diferencia en ámbitos más allá del estrictamente empresarial.



## 4. Glosario

**BI:** Business Intelligence (Inteligencia de Negocio)

**VADER:** *Valence Aware Dictionary for sEntiment Reasoning*. Modelo que permite asignar una polaridad a un texto, determinando si se asocia a emociones positivas o negativas.

**Overfitting:** en modelos estadísticos, cuando una función pretende ajustarse exageradamente a los datos disponibles para ser más exacta, y en el proceso deja de ser útil para otros datos

**Forecasting:** predicción/pronóstico

**Clusterización:** agrupación de datos por similitud

**Slicer:** filtro editable que permite segmentar los datos de un reporte en función de un determinado parámetro (fecha, lugar...)

**K-Means:** algoritmo que permite, de forma iterativa, dividir valores en grupos diferenciados.

## 5. Bibliografía

<https://www.iberdrola.com/innovacion/que-es-business-intelligence>

[Consultado: 15/09/2022]

<https://absentdata.com/advantages-and-disadvantages-of-tableau/>

[Consultado: 15/09/2022]

<https://datascientest.com/es/python-o-r-que-elegir>

[Consultado: 15/09/2022]

[https://circabc.europa.eu/sd/a/2c8378ab-c686-449d-9dd1-65371ab30889/Oilseeds-dashboard\\_en.pdf](https://circabc.europa.eu/sd/a/2c8378ab-c686-449d-9dd1-65371ab30889/Oilseeds-dashboard_en.pdf)

[Consultado: 15/09/2022]

<https://covid19.who.int/>

[Consultado: 15/09/2022]

[https://twitter.com/ZelenskyyUa?ref\\_src=twsrc%5Egoogle%7Ctwcamp%5Eserp%7Ctwqr%5Eauthor](https://twitter.com/ZelenskyyUa?ref_src=twsrc%5Egoogle%7Ctwcamp%5Eserp%7Ctwqr%5Eauthor)

[Consultado: 15/09/2022]

<https://openaccess.uoc.edu/bitstream/10609/96846/7/avillaTFG0619memoria.pdf>

[Consultado: 15/09/2022]

<https://openaccess.uoc.edu/bitstream/10609/64985/6/ggirosTFG0617memoria.pdf>

[Consultado: 15/09/2022]

<https://data.oecd.org/price/inflation-cpi.htm>

[Consultado: 10/10/2022]

<https://data.unhcr.org/en/situations/ukraine>

[Consultado: 10/10/2022]

<https://www.gartner.com/en/information-technology/glossary/social-analytics#:~:text=Social%20analytics%20is%20monitoring%2C%20analyzing,workplace%20and%20external%2Dfacing%20communities.>

[Consultado: 10/10/2022]

<https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17>

[Consultado: 10/10/2022]

<https://hackernoon.com/scraping-tweet-replies-with-python-and-tweepy-twitter-api-a-step-by-step-guide-z11x3yr8>

[Consultado: 10/10/2022]

<https://analitza.com/blog/business-intelligence-bi-vs-predictive-analytics>

[Consultado: 10/10/2022]

<https://www.geeksforgeeks.org/how-to-merge-multiple-excel-files-into-a-single-files-with-python/>

[Consultado: 10/10/2022]

<https://datascientest.com/es/power-bi-vs-tableau-es>

[Consultado: 10/10/2022]

<https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17>

[Consultado: 10/10/2022]

<https://learn.microsoft.com/en-us/power-bi/transform-model/desktop-ai-insights>

[Consultado: 10/10/2022]

<https://www.simplilearn.com/tutorials/power-bi-tutorial/power-bi-vs-tableau>

[Consultado: 10/10/2022]

<https://geekflare.com/es/tableau-vs-power-bi/>

[Consultado: 10/10/2022]

<https://www.interviewbit.com/blog/python-vs-r/>

[Consultado: 10/10/2022]

<https://towardsdatascience.com/python-vs-r-for-data-science-did-you-make-the-right-choice-7d090dca4786>

[Consultado: 10/10/2022]

<https://exportcomments.com/resources/twitter-advanced-search/>

[Consultado: 10/10/2022]

<https://data.unhcr.org/en/situations/ukraine>

[Consultado: 10/11/2022]

<https://web.archive.org/web/20220713231017/https://data.unhcr.org/es/situations/ukraine>

[Consultado: 10/11/2022]

<https://towardsdatascience.com/are-you-scared-vader-understanding-how-nlp-pre-processing-impacts-vader-scoring-4f4edadbc91d>

[Consultado: 10/11/2022]

<https://www.analyticsvidhya.com/blog/2021/06/vader-for-sentiment-analysis/>

[Consultado: 10/11/2022]

<https://datascience.stackexchange.com/questions/36663/under-what-circumstance-is-lemmatization-not-an-advisble-step-when-working-with>

[Consultado: 10/11/2022]

<https://www.analyticsvidhya.com/blog/2021/06/vader-for-sentiment-analysis/>

[Consultado: 10/11/2022]

<https://www.imf.org/external/datamapper/PCPIEPCH@WEO/RUS?zoom=RUS&highlight=RUS>

[Consultado: 10/11/2022]

<https://www.klipfolio.com/resources/articles/what-is-a-key-performance-indicator>

[Consultado: 14/12/2022]

<https://data.oecd.org/price/inflation-cpi.htm>

[Consultado: 10/11/2022]

<https://data.oecd.org/gdp/quarterly-gdp.htm#indicator-chart>

[Consultado: 10/11/2022]

<https://data.oecd.org/unemp/unemployment-rate.htm>

[Consultado: 10/11/2022]

<https://www.kpi.org/kpi-basics/>

[Consultado: 14/12/2022]

<https://economipedia.com/definiciones/producto-interior-bruto-pib.html>

[Consultado: 14/12/2022]

<https://www.cs.umd.edu/users/mvz/handouts/gqm.pdf>

[Consultado: 15/12/2022]

[https://www.ecb.europa.eu/ecb/educational/explainers/tell-me-more/html/what\\_is\\_inflation.en.html](https://www.ecb.europa.eu/ecb/educational/explainers/tell-me-more/html/what_is_inflation.en.html)

[Consultado: 15/12/2022]

<https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.interpolate.html>

[Consultado: 15/12/2022]

<https://www.consilium.europa.eu/en/infographics/impact-sanctions-russian-economy/#:~:text=The%20chart%20shows%20that%2C%20in,3.4%25%20according%20to%20the%20IMF>

[Consultado: 15/12/2022]

[https://pandas.pydata.org/docs/user\\_guide/merging.html](https://pandas.pydata.org/docs/user_guide/merging.html)

[Consultado: 15/12/2022]

<https://www.ibm.com/cloud/learn/unsupervised-learning#:~:text=Unsupervised%20learning%2C%20also%20known%20as,the%20need%20for%20human%20intervention>

[Consultado: 15/12/2022]

<https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1>  
<https://iterationinsights.com/article/cluster-analysis-in-power-bi/>

[Consultado: 15/12/2022]

[https://en.wikipedia.org/wiki/List\\_of\\_countries\\_and\\_dependencies\\_by\\_population](https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population)

[Consultado: 10/01/2023]

<https://www.indec.gob.ar/indec/web/Nivel4-Tema-3-9-47#:~:text=La%20estimaci%C3%B3n%20preliminar%20del%20producto,mismo%20per%C3%ADodo%20del%20a%C3%B1o%20anterior>

[Consultado: 10/01/2023]