



Facultad
de
Ciencias

Estudio del consumo eléctrico como indicador económico

Trabajo final de la asignatura
Data Life Cycle del

MÁSTER DE DATA SCIENCE

Autores: Adrià Nova
Raúl Santos
Samuel Valer Nasta
Abel González
Antonio Coll
María González-Estéfani
Enero - 2022

Índice

1. Introducción	2
2. Descripción general del proyecto	2
2.1. Objetivos y resultados esperados	2
2.2. Requisitos y requerimientos técnicos	2
2.3. Cobertura geográfica y temporal	3
3. Gestión del proyecto	3
3.1. Logframe	3
3.2. Paquetes de trabajo y diagrama de Gantt	4
3.2.1. WP0: Presupuestos, legislación y coordinación.	6
3.2.2. WP1: Difusión inicial y toma de datos.	8
3.2.3. WP2: Preprocesado y curación de los datos.	9
3.2.4. WP3: Análisis.	11
3.2.5. WP4: Soluciones y resultados.	12
3.2.6. WP5: Publicación y preservación.	13
3.2.7. WP6: Publicación y promoción.	15
3.3. Plan de gestión de datos	16
4. Procesado de datos	19
4.1. Limpieza y curación de los datos	19
4.2. Plan de preservación de datos	22
4.3. Metadatos	23
5. Análisis de datos	25
5.1. Metodología	25
5.2. Resultados	25
5.2.1. Tendencias anuales y contexto macroeconómico.	25
5.2.2. Perfiles diarios y patrones de consumo.	27
6. Conclusiones	33

1. Introducción

En el marco económico actual el precio de la luz es un factor económico diferencial debido a su alto impacto sobre las economías domésticas, esta afirmación por parte de nuestro compañero Abel desencadenó en una inquietud grupal que se intentó solventar mediante una búsqueda para determinar la relación del consumo eléctrico y la situación laboral en un país. Puesto que la búsqueda realizada no arrojó resultados algunos decidimos realizar este estudio.

La plenitud de datos disponibles y la calidad de estos también influyeron a la hora de determinar la realización de este estudio puesto que agencias como Eurostat resultan ser medios fiables.

2. Descripción general del proyecto

2.1. Objetivos y resultados esperados

El objetivo principal de este estudio es determinar si existe una correlación entre el consumo eléctrico y una serie de magnitudes que caracterizan la economía de un país. Estas magnitudes son:

- **Paro:** este indicador fue elegido debido a que si un país tiene menos paro la situación económica debería de ser favorable.
- **PIB:** se eligió esta magnitud puesto que mide directamente la situación económica de un país.
- **Población:** según el criterio grupal si un país tiene más población que otro debería de tener también un consumo eléctrico mayor.

El análisis se va a realizar sobre los 15 países más representativos de Europa debido a la calidad de los datos que aportan sus instituciones.

Sin haber realizado el análisis ni visto los datos nuestro grupo espera encontrar una correlación directa entre las variables seleccionadas y los consumos eléctricos.

2.2. Requisitos y requerimientos técnicos

Se hace evidente la necesidad de ordenadores con cierta capacidad de computo, aunque no es necesario que sea excesiva. Es posible montar un servidor

para ofrecer a los expertos e ingenieros del proyecto un herramienta para tener un almacenamiento común y preservarlos. Otra opción sería utilizar servicios de computo y almacenamiento como AMS (Amazon Web Service).

2.3. Cobertura geográfica y temporal

En cuanto a la cobertura geográfica, como se ha advertido, se trata de países europeos, estos son: **Alemania, Austria, Bélgica, Dinamarca, España, Francia, Irlanda, Italia, Luxemburgo, Noruega, Países Bajos, Portugal, Reino Unido, República Checa, y Suecia.** Se han tomado estos países debido a su, posiblemente, mayor disponibilidad a la hora de proporcionar datos. A su vez, el estudio pretende estimar correlaciones entre indicadores económicos y consumo eléctrico en Europa. Tratar de extrapolar estos datos a países de otros continentes como América, África o Asia supone una generalidad necesaria difícilmente alcanzable.

Respecto a la cobertura temporal, esta supone una toma de datos desde el año 2015 hasta el año 2019, ambos años incluidos. Se han escogido estos años debido a su proximidad a la actualidad, pero sin verse afectado por otros factores como la pandemia mundial de la COVID-19, la cual supondría comportamientos extraños en los datos. Es decir, outliers que esconderían las posibles conclusiones que tratan de generalidad comportamientos.

3. Gestión del proyecto

Respecto a la gestión de este proyecto, el itinerario va de la mano de las etapas fundamentales que siguen los datos, desde su toma, hasta su preservación y disponibilidad. Por ello, como se aprecia en los siguientes, el eje del desarrollo es la obtención de los datos, el procesamiento previo el cual incluye limpieza y curación de los mismos, su análisis, su preservación y, finalmente, su difusión para permitir su disponibilidad. Se ofrece a continuación el logframe que resume las ideas principales del proyecto.

3.1. Logframe

Los logframe refieren a un marco lógico que resume la planificación, motivaciones y dependencias en las que sustenta un proyecto, a la vez que define el éxito del mismo. En este caso, se tiene el siguiente logframe.

Estudio del consumo eléctrico diario de una población como indicador de la economía de un país

	Objetivos	Éxito	Verificación	Condiciones
Meta	Conseguir categorizar el grado de desarrollo de un país a través de su consumo eléctrico	Modelo funcional: Clasificar un nuevo país como desarrollado o no desarrollado	Comparar con indicadores ya existentes Alguna entidad use el modelo	Interés económico y necesidad de caracterizar con pocos datos
Propósito	Encontrar la correlación entre el consumo eléctrico de un país y distintas variables económicas	Existencia de correlación	Sustituir el uso de otras variables	Obtención de los datos a partir de repositorios en abierto (Eurostat, ENTSO-E)
Resultado	Análisis descriptivo de los datos junto a la clasificación de los países	Modelo validado: Correcta clasificación	Publicar el proyecto en Zenodo y asistir a una convención con el fin de exponer los resultados	Aceptación de la integridad de los resultados y su difusión
Componentes	Datos Modelo Grupo de trabajo	Presupuesto Calendario	Reconocimiento por parte de un portal de datos abiertos o una entidad privada	Equipo interdisciplinar de análisis de datos

Figura 2: Logframe del proyecto.

3.2. Paquetes de trabajo y diagrama de Gantt

Continuando con la gestión del proyecto, comenzaremos observando los distintos paquetes de trabajo que incluirá nuestro trabajo. Esta herramienta nos permite realizar un desglose de las distintas actividades por las que pasará el proyecto, resumiendo en una simple panorámica todo el proceso.



Figura 3: Paquetes de trabajo del proyecto

Como podemos observar en la figura anterior, el proyecto puede dividirse en siete paquetes claramente diferenciados, seis pertenecientes al proyecto per se y uno inicial, llamado paquete 0, de planificación.

DIAGRAMA DE GANTT DEL PROYECTO

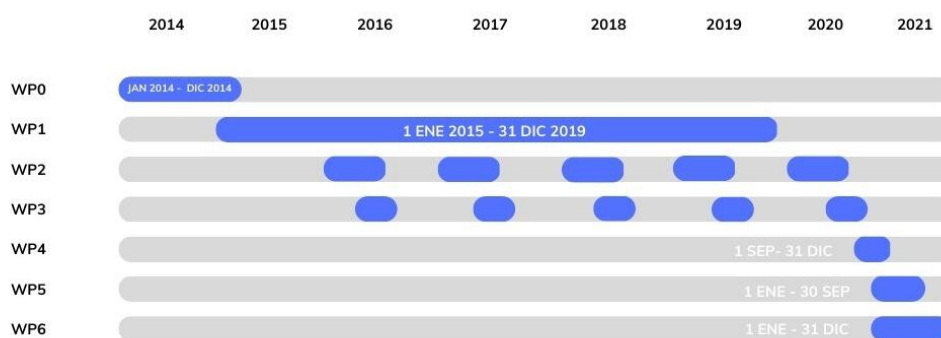


Figura 4

Acompañaremos la descripción de los distintos paquetes de trabajo con un diagrama de Gantt, herramienta que permite visualizar fácilmente el tiempo que ocupará cada paquete. La figura anterior se corresponde con el diagrama general, en la que vemos cómo este proyecto se extiende desde 2014 hasta

2021. Dentro de este periodo, los años comprendidos entre 2015 y 2019 serán los objeto de análisis.

3.2.1. WP0: Presupuestos, legislación y coordinación.

El paquete de trabajo inicial, Presupuesto, legislación y coordinación, consiste en realizar todas las gestiones sobre las que se sustenta el proyecto, tales como la solicitud de presupuestos o de los distintos permisos para acceder a algunos datos. Realizadas todas las tareas, el proyecto queda de la siguiente forma:



Figura 5: Asignación del presupuesto y planificación del proyecto

Donde podemos observar cómo los salarios y, especialmente, los convenios con empresas europeas ocupan la mayor parte del presupuesto.

[H]

Work Package 0	
Presupuesto, legislación y coordinación	
Comienzo del WP: 01/01/2014	Finalización del WP: 31/12/2014
Coordinador: Administrativo	
Participantes Data Scientist Economista	
Objetivos O1. Presupuestar el proyecto. O2. Organizar plazos y plan de trabajo.	
Descripción del trabajo Tarea 0.1 Cálculos y solicitud de presupuestos. Tarea 0.2 Diseño de plantilla e inventario, reparto de tareas Tarea 0.3 Fijar plazos de hitos, tareas e informes. Tarea 0.4 Solicitud y concesión de permisos para acceder a datos privados.	
Entregables Entregable 0.1 Plantilla e inventario.Fecha: 31/03/2014. Entregable 0.2 Presupuesto. Fecha: 31/06/2014. Entregable 0.3 Informe final. Entrega: 30/7/2022. Entregable 0.4 Solicitud de permisos. Fecha: 01/10/2014	



Figura 6: Diagrama de Gantt del paquete de trabajo 0

En cuanto a la duración de este paquete, comprobamos que se concentra en 2014. Las tres primeras tareas se realizan durante la primera mitad del año, mientras que la solicitud de permisos, al tratarse de un proceso sensiblemente más lento, lo ocupa en su totalidad.

3.2.2. WP1: Difusión inicial y toma de datos.

Realizados los trámites anteriores, estamos en disposición de comenzar el trabajo mediante la Difusión inicial y toma de los datos. Como bien indica el nombre del paquete, el objetivo en esta etapa es dar a conocer nuestro proyecto al mismo tiempo que recopilamos los datos a analizar. La primera tarea está a cargo del community manager, responsable de dirigir una campaña en redes sociales en la que explicar los objetivos e ir actualizando del progreso. Por otra parte, team manager, psicólogo y economistas son los responsables de ponerse en contacto con los distintos organismos para recoger los datos.

Work Package 1	
Difusión inicial y toma de datos	
Comienzo del WP: 01/01/2015	Finalización del WP: 31/12/2019
Coordinador: Community manager	
Participantes Team manager Psicólogo Economistas	
Objetivos O.1 Dar a conocer el proyecto que vamos a realizar O.2 Toma de los datos a analizar	
Descripción del trabajo Tarea 1.1 Campaña en redes. Tarea 1.2 Emisión de encuestas y contacto con empresas. Tarea 1.3 Contacto con censos. Tarea 1.4 Consultar indicadores económicos.	

WP - 1 DIFUSIÓN INICIAL Y TOMA DE DATOS



Figura 7: Diagrama de Gantt del paquete de trabajo 1.

Esta tarea ocupará todo el tiempo de tratamiento de datos en el proyecto, es decir, desde enero de 2015 hasta diciembre de 2019.

3.2.3. WP2: Preprocesado y curación de los datos.

Seguimos con el segundo paquete de trabajo, Preprocesado y curación de datos, en el que el objetivo será manipular los datos obtenidos en la etapa anterior hasta que estos sean aptos para el análisis. Esta tarea será llevada a cabo por los perfiles más técnicos de la plantilla: el cloud engineer y los científicos de datos.

Work Package 2	
Preprocesado y curacion de datos	
Comienzo del WP: 01/01/2016	Finalización del WP: 31/07/2020
Coordinador: Cloud engineer	
Participantes Data Scientists	
Descripción del trabajo Tarea 0.1 Contraste de fuentes de datos. Tarea 0.2 Creación de repositorio e integración de los datos. Tarea 0.3 Preprocesado y curación de datos multifuente. Tarea 0.4 Elaboración inicial de metadatos.	
Objetivos O1. Obtener unos datos adecuados para empezar a trabajarlos.	
Entregables Entregable 0.1 Integración de los datos en el repositorio.Fecha: 30/04. Entregable 0.2 Datos preparados para su análisis. Fecha: 30/06. Entregable 0.3 Metadatos iniciales. Entrega: 31/7.	



Figura 8: Diagrama de Gantt del paquete de trabajo 2.

Por su parte, este paquete deberá realizarse de manera periódica durante la primera mitad del año entre 2016 y 2019, justo después de obtener los datos

iniciales.

3.2.4. WP3: Análisis.

A continuación tendríamos el paquete de análisis, etapa del trabajo que consiste en estudiar la relación entre las variables seleccionadas y la construcción del perfil diario de consumo eléctrico en los países seleccionados. Los encargados serán los mismos que durante el procesado con la inclusión de los economistas.

Work Package 3	
Análisis	
Comienzo del WP: 23/05/2016	Finalización del WP: 31/07/2020
Coordinador: Cloud engineer	
Participantes Data scientists Economistas	
Objetivos O.1 Tratar los datos obtenidos y obtener información de ellos	
Descripción del trabajo Tarea 1.1 Hipótesis inicial. Tarea 1.2 Relación entre variables y estadística descriptiva.	
Entregables Entregable 0.1 Elaboración del análisis de los datos anuales. Fecha: 30/06.	

WP - 3 ANÁLISIS



Figura 9: Diagrama de Gantt del paquete de trabajo 3.

Dado que es una tarea más amable que la anterior, se realizará en menos tiempo. Más concretamente, se completarla entre los meses de junio y julio de 2016 a 2020, inmediatamente después de obtener los datos curados. Además, resaltar que la hipótesis inicial se fórmula una única vez en mayo de 2016 y que cada año comprobamos si esta se cumple.

3.2.5. WP4: Soluciones y resultados.

Continuamos con el cuarto paquete, Soluciones y resultados, con el que podemos dar por terminado el tratamiento de datos. Los encargados serán los mismos que en el anterior y su objetivo será obtener conclusiones del análisis previo y escribir los informes necesarios para su posterior publicación.

Work Package 4	
Soluciones y resultados	
Comienzo del WP: 01/09/2020	Finalización del WP: 31/12/2020
Coordinador: Cloud engineer	
Participantes Data Scientists Economistas	
Objetivos O1. Evaluación de resultados y preparación de informes.	
Descripción del trabajo Tarea 0.1 Interpretación de resultados y comprobación de hipótesis. Tarea 0.2 Elaboración de informes (internos y externos) (informe semestral)	
Entregables Entregable 0.1 Entrega de los informes pertinentes. Fecha: 31/12/2020.	



Figura 10: Diagrama de Gantt del paquete de trabajo 4.

De nuevo se realizará periódicamente entre 2016 y 2020, esta vez durante el último cuatrimestre de cada año.

3.2.6. WP5: Publicación y preservación.

El siguiente paquete es el referente a la publicación del proyecto. Durante su duración se elabora el informe definitivo del trabajo, además de distintos artículos para revistas. Por otra parte, también reformateamos los datos para facilitar su acceso al público y elaboramos los metadatos finales.

Work Package 5	
Publicación y preservación	
Comienzo del WP: 01/01/2021	Finalización del WP: 30/09/2021
Coordinador: Cloud Engineer	
Participantes Data Scientists Economistas	
Objetivos O1. Obtención de conclusiones definitivas. O2. Elaboración de artículos. O2. Preparación de datos y metadatos para el acceso al público.	
Descripción del trabajo Tarea 0.1 Informe final - resultados definitivos. Tarea 0.2 Elaboración de artículos Tarea 0.3 Reformato de datos. Tarea 0.4 Elaboración de metadatos para la publicación.	
Entregables Entregable 0.1 Informe definitivo.Fecha: 31/03/2014. Entregable 0.2 Artículos. Fecha: 30/09/2021.	



Figura 11: Diagrama de Gantt del paquete de trabajo 5.

Estas tareas se realizan a lo largo de 2021, siendo la escritura de artículos la tarea que más tiempo nos ocupa.

3.2.7. WP6: Publicación y promoción.

Finalmente, terminamos el proyecto con el sexto paquete: Publicación y promoción. Esta tarea se resume en la promoción del proyecto y su posterior difusión por diferentes medios: revistas, congresos, presentaciones, etc.

Work Package 6	
Publicación y promoción	
Comienzo del WP: 01/01/2021	Finalización del WP: 31/12/2021
Coordinador: Community manager	
Participantes Data Scientists Economistas Cloud Engineer Asesores Administrativo Team manager	
Objetivos O1. Presentación del proyecto final.	
Descripción del trabajo Tarea 0.1 Promoción del proyecto. Tarea 0.2 Publicación de resultados (artículos, revistas, repositorios...). Tarea 0.3 Congresos, presentaciones.... Tarea 0.4 Solicitud y concesión de permisos para acceder a datos privados.	
Entregables Entregable 0.1 Publicación de artículos, revistas, etc. Fecha: 31/12/2021. Entregable 0.2 Presupuesto. Fecha: 31/06/2014. Entregable 0.3 Informe final. Entrega: 30/7/2022. Entregable 0.4 Solicitud de permisos. Fecha: 01/10/2014	

WP - 6 PUBLICACIÓN Y PROMOCIÓN

2021



Figura 12: Diagrama de Gantt del paquete de trabajo 6.

Al igual que el paquete anterior, este se realiza durante 2021. Por un lado, la promoción lo ocupa en su totalidad, mientras que la publicación de los resultados definitivos se completa en el mes de diciembre.

3.3. Plan de gestión de datos

Por lo que respecta a los datos los diversos paquetes de trabajo llevarán a cabo su gestión. Sin embargo, pueden reflejarse aquí ciertas consideraciones que son tenidas en cuenta a la hora de gestionar los datos.

- **Recopilación de datos:**

La recopilación de datos es una parte crucial del proyecto, por lo que es importante asegurarse de utilizar fuentes confiables. Particularmente, en cuanto al P.I.B. (Producto Interior Bruto) como indicador económico puede obtenerse a partir de portales como Eurostats, el cual es un repositorio de alta fiabilidad. La tasa de desempleo, esta se trata de un valor típicamente recogido por entidades de estadística de cada país. Por poner algún ejemplo, podría considerarse Instituto Nacional de Estadística (INE), en España; Institut National de la Statistique et des Études Économiques (INSEE), en Francia; Office for National Statistics (ONS), en Reino Unido. En cuanto a la población puede hacerse lo propio acudiendo a organismos como Registro civil (España), Service de l'Etat Civil (Francia), General Register Office (Reino Unido)

o Standesamt (Alemania). Respecto al consumo eléctrico de cada país podrían consultarse los datos a las principales redes de distribución de energía eléctrica, estas son: Red Eléctrica Española (REE), Réseau de Transport d'Electricité (RTE), TransnetBW, ... Se trata de una labor no trivial debido a la dependencia con cada organismo de cada país.

- Limpieza de datos:

La limpieza de datos es un paso importante para asegurar la calidad de los datos y su adecuación para el análisis. Se realizará un análisis exploratorio de los datos para identificar posibles errores o valores faltantes, y para corregirlos. Además, se asegurará que los datos estén en el mismo formato y tenga la misma unidad de medida, lo que permitirá un análisis más preciso. Este proceso acaba siendo la mayor parte de la gestión de los datos, y probablemente, este proyecto no sea la excepción. Destacar la gran variedad de fuentes de datos necesarias para recopilar todos los datos (distintos países y organizaciones) por tanto, habrá una gran labor de unificación de los datos, teniendo que lidiar con las diferencias y errores aportados por cada fuente. Es posible hacer uso de herramientas como OpenRefine, el propio python, pandasProfiling, ...

- Metadatos: Resulta de utilidad resumir el contenido y estructura del dataset ya listo para analizar mediante metadatos adecuados. Esto resulta de utilidad para su manejo y utilización, incluso para los propios data scientist del proyecto. Podrían escribirse en el formato DublinCore, siendo este un de los estándares más utilizados.

- Análisis de datos:

Una vez que los datos han sido recopilados y limpiados, se llevará a cabo el análisis. Se realizará un estudio de la correlación entre los indicadores económicos y sociales y el consumo eléctrico de los países. Se mostrarán visualmente gráficos y matrices de correlación, para representar los datos de manera clara y concisa. El objetivo del análisis es identificar patrones y tendencias en los datos y responder a la pregunta de investigación: ¿Qué relaciones existen entre los indicadores económicos y sociales y el consumo eléctrico de los países?

- Interpretación de resultados:

Una vez que se han obtenido los resultados del análisis, es necesario interpretarlos. Se discutirán los hallazgos más relevantes y se plantearán posibles explicaciones para las relaciones encontradas. Cabe destacar

que, realmente, son pocas magnitudes estando todas ellas influidas por otros muchos factores. Sin embargo, se aspira a encontrar cierta tendencia o relaciones entre ella debido a los numerosos países considerados.

- **Conclusión:**

Se resumirán los hallazgos más importantes y se presentarán las conclusiones finales del proyecto. A su vez, se destacarán implicaciones de los resultados y se plantearán posibles líneas de investigación futura.

- **Preservación y difusión:**

Estos datos, de acuerdo a los principios F.A.I.R. de los datos (findable, accesible, interoperable, reusable), serán publicados en portales abiertos para su uso por cualquier persona como Zenodo. Esto permite su preservación, difusión y accesibilidad, pudiendo descargarlo de manera programática.

4. Procesado de datos

El procesamiento de datos es un proceso preliminar durante la minería de datos. El objetivo es aplicar procesos sobre los datos brutos para transformarlos en datos que tengan formatos que sean más fáciles para utilizar las técnicas deseadas. En la realidad, los datos frecuentemente no están limpios, faltan valores clave, contienen inconsistencias y suelen mostrar ruido, conteniendo errores y valores atípicos. Sin un buen preprocesamiento de datos, estos errores harán disminuir la calidad de la minería de datos. Para el caso que nos ocupa, hemos tenido que hacer un conjunto de operaciones en los distintos datasets para poder amoldarlos al estudio del consumo eléctrico diario de una población como indicador de la economía de un país.

No menos importante es la preservación de estos datos para un posible uso en un futuro para otros investigadores, comunidades, etc. Además, establecemos unos metadatos sobre los datos que aportan información adicional y de gran relevancia sobre los mismos.

4.1. Limpieza y curación de los datos

Una vez hemos conseguido obtener los datos en crudo de los acuerdos con las distribuidoras energéticas, los datos del producto interior bruto, población y paro por país de los sistemas públicos ‘open data’ procedemos a la aplicación de las técnicas de preprocesamiento de los datos.

En nuestro estudio hemos considerado el consumo eléctrico para 15 países europeos (Alemania, Austria, Bélgica, Dinamarca, España, Francia, Irlanda, Italia, Luxemburgo, Noruega, Países Bajos, Portugal, Reino Unido, República Checa y Suecia), así como los datos de PIB, paro y población para 5 años (2015-2019). Para llevar a cabo el preprocesamiento hemos usado Open Refine para limpiar y transformar los datos en los formatos deseados aplicando las siguientes técnicas:

- **Datos del producto interior bruto (PIB):** En un primer momento, los datos de PIB se descargan con formato ‘tsv’. Todos los datos forman parte de una misma columna, ya que están separados por comas y tabuladores simultáneamente, de modo que separamos la columna en varias para diferenciar el PIB por años.
- Eliminamos aquellas columnas generadas de forma predeterminada y que se encuentran vacías de información ‘freq’, ‘unit’, ‘na.item’. También eliminamos las columnas para los años 2013, 2014, 2020 y 2021 por

conveniencia. Encontramos que las columnas de los años tienen un espacio al final y que dificultan su manipulación, de modo que decidimos quitarlo para que sea más homogéneo.

- Seguido de ello, renombramos la columna que contiene países a 'País'. Asimismo, eliminamos las filas que tienen observaciones que no son de nuestro interés, como países de los que no tenemos todos los datos, respecto de otros datasets. Por ejemplo, tenemos los datos del PIB de Polonia, pero no forma parte de nuestro conjunto de países para el análisis o no tenemos datos de paro, población o consumo eléctrico.

Datos_PIB

País	2015	2016	2017	2018	2019
AT	344269.2	357608.0	369361.9	385274.1	397169.5
BE	416701.4	430085.3	445050.1	460050.8	478645.0
CH	625532.7	621765.1	615776.3	614304.4	644443.2
DE	3026180.0	3134740.0	3267160.0	3365450.0	3473260.0

Figura 13: Muestra de los datos del producto interior bruto (PIB, millones de euros) de cada país entre 2015 y 2019.[\(1\)](#)

- **Datos consumo eléctrico por país:** Los datos de consumo eléctrico vienen formateados por hora, por media hora o por cuarto de hora, dependiendo de la recogida de los datos de cada país. Por ejemplo, Alemania recoge el consumo eléctrico en megavatios cada cuarto de hora. Se han separado el día y la hora de cada consumo, adaptando las horas para conseguir un ajuste entre la hora inicio y final de cada día.
- Se ha generado un archivo JSON con los cambios realizados en el dataset del consumo eléctrico de Alemania para poder ser replicado para el resto de datasets de los diferentes países que conforman nuestro análisis, automatizando el proceso.

alemania

Día	Hora Inicio	Hora Final	load
2015-01-01	00:00	00:15	41917.0
2015-01-01	00:15	00:30	41518.0
2015-01-01	00:30	00:45	41179.0
2015-01-01	00:45	01:00	40756.0

Figura 14: Muestra de los datos de consumo de Alemania (MW/15min) entre 2015 y 2019.(2)

- **Datos de población:** : Los datos de población se encuentran en millones de habitantes. Se ha aplicado el mismo procedimiento que para los datos PIB, ya que nos encontramos con el mismo formato ‘tsv’.

datos_poblacion

pais	2015	2016	2017	2018	2019
austria	8584926	8700471	8772865	8822267	8858775
belgica	11237274	11311117	11351727	11398589	11455519
republica checa	10538275	10553843	10578820	10610055	10649800

Figura 15: Muestra de los datos de población (millones de habitantes) por país entre 2015 y 2019.(3)

- **Datos de desempleo:** Los datos de desempleo es la tasa entre la población parada (16 a 74 años) entre la población activa (16 a 74 años). Encontramos un dato faltante para Reino Unido que se obtiene a partir de otra fuente de datos llamada ‘Statista’.

paro_paises

pais	2015	2016	2017	2018	2019
austria	6.1	6.5	5.9	5.2	4.8
belgica	8.7	7.9	7.2	6.0	5.5
republica checa	5.1	4.0	2.9	2.2	2.0

Figura 16: Muestra de los datos de desempleo (tasa de parados %) por país entre 2015 y 2019. (4)

Una vez limpiados los datasets de consumo eléctrico, PIB, población y paro, transponemos las filas y columnas de cada dataset mediante un script de Python con el objetivo de tener en las columnas todos los países y en las filas todos los años. Para el consumo eléctrico se ha aplicado la media de consumo eléctrico por país entre 2015 y 2019. También se han mantenido los datasets de consumo eléctrico originales, ya que a lo largo del análisis analizamos el promedio del consumo eléctrico por hora para cada país para cinco años, además de aplicar otro tipo de análisis. Una vez tenemos los datasets transpuestos, juntamos el PIB, paro, población y consumo por país.

A nivel de script se han seguido los siguientes pasos:

- Extraer los datos de paro por país con el nuevo formato
- Extraer los datos de PIB por país con el nuevo formato
- Extraer los datos de población por país con el nuevo formato
- Iterar sobre los archivos de consumo eléctrico de cada país, agrupar los datos en columnas [PIB, Paro, Poblacion, Consumo] y guardar el nuevo archivo para ser usado en el análisis de datos.

4.2. Plan de preservación de datos

Uno de los retos más importantes de cualquier ciencia es facilitar el aprendizaje tanto de los humanos como de las máquinas. Gracias a los principios FAIR, podemos generar datos siguiendo la filosofía de los datos en abierto, estableciendo una base común que aumenta el increíble potencial del dato abierto.

Una vez el trabajo ha concluido, nuestros datos han sido subidos a un repositorio en abierto llamado Zenodo siguiendo los principios FAIR como se detallan a continuación:

- **FINDABLE:** Para facilitar la búsqueda, la citación y reutilización, es importante disponer de metadatos, estos se describen en el siguiente apartado, utilizando indicadores persistentes con el fin de identificar nuestros datos y facilitar su citación.
- **ACCESSIBLE:** Los datos generados en nuestro trabajo se encuentran en el repositorio en abierto Zenodo, creando una accesibilidad que pueden ser recuperados por sus identificadores mediante protocolos estandarizados de comunicación (abiertos, gratuitos e implementados universalmente).
- **INTEROPERABLE:** Los formatos en los que se encuentran nuestros datos son integrables con otras aplicaciones. Todos los archivos se encuentran en formato 'csv'.
- **REUSABLE:** La información contenida en el repositorio esta correctamente explicada, detallando el contenido del documento, los pasos seguidos en el tratamiento de los datos y añadiendo la información extra necesaria. También se ha explicado cómo se han formado los datos.

Los datos del trabajo no únicamente se encuentran publicados en Zenodo, sino que también se han guardado en un servidor privado, de modo que en caso que el repositorio no respondiera, estos se preservan en un servidor externo, reinsertando los datos del servidor privado al repositorio Zenodo cuando vuelva a estar operativo.

4.3. Metadatos

Una vez con los datos finales ya limpios y agrupados en un dataset general, añadimos los metadatos para describirlos. Estos metadatos los sacamos tanto de los diferentes comentarios que hemos hecho anteriormente como de datos ya establecidos, como el título o los autores.

Al igual que en la gestión del proyecto, suponemos que el proyecto se ha tenido que empezar de cero, pero mantenemos que nuestro equipo ha sido el ejecutor del proyecto, y que Eurostat ha sido la fuente en donde se han obtenido y contrastado los datos iniciales.

Los metadatos son los siguientes:

- **Título:** Estudio del consumo eléctrico como indicador económico.
- **Creadores:** Adrià Nova, Raúl Santos, Samuel Valer Nasta, Abel González, Antonio Coll de San Simón, María González-Estéfani.
- **Subject:** consumo eléctrico, Europa, PIB, población, desempleo.
- **Descripción:** estudio de la correlación entre variables de carácter económico y demográfico y el consumo eléctrico de un país.
- **Editor:** Zenodo.
- **Colaboradores:** gobiernos de distintos países de Europa.
- **Fecha:** diciembre 2021.
- **Tipo:** artículo de investigación, dataset.
- **Formato:** PDF (artículo), CSV (dataset).
- **Identificador:** AD01.
- **Fuente:** Eurostat.
- **Idioma:** español.
- **Relación:** Datos PIB de Eurostat.
- **Cobertura:** 15 países seleccionados de 2015 a 2019.
- **Derechos:** Universidad de Cantabria.

Se ha seguido el esquema de metadatos Dublin Core (5) y se ha creado un archivo *xml* para adjuntarlos junto al dataset.

5. Análisis de datos

5.1. Metodología

- **Contexto macroeconómico de los países.** Representación de tendencias anuales y análisis de correlaciones.

Tendencia anual de los principales indicadores socio-económicos de cada país desde 2015 hasta 2020.

Obtención de las correlaciones existentes entre los valores anuales de los 4 indicadores para cada país.

Análisis de correlaciones entre los principales indicadores de cada país.

- **Patrón de consumo de una sociedad.** Análisis de perfiles diarios de consumo energético.

Caracterización de los perfiles diarios de consumo energético para enero, abril y agosto.

Cálculo y comparación del perfil diario de consumo en abril entre países. (Promedio de un perfil diario de consumo energético para un día de abril sobre 5 años).

Cálculo de la matriz de correlaciones para el conjunto de perfiles diarios de los países.

5.2. Resultados

5.2.1. Tendencias anuales y contexto macroeconómico.

Tendencias socio-económicas entre 2015 y 2019. El primer resultado obtenido ha sido el promedio anual de cada indicador desde 2015 hasta 2019. Como podemos ver a continuación en la figura 17, el comportamiento general es muy similar entre los países. En su gran mayoría los países presentan tendencias de paro a la baja, mientras que el PIB y la población han aumentado. Por otro lado, el promedio anual de consumo tiene un comportamiento más variado.

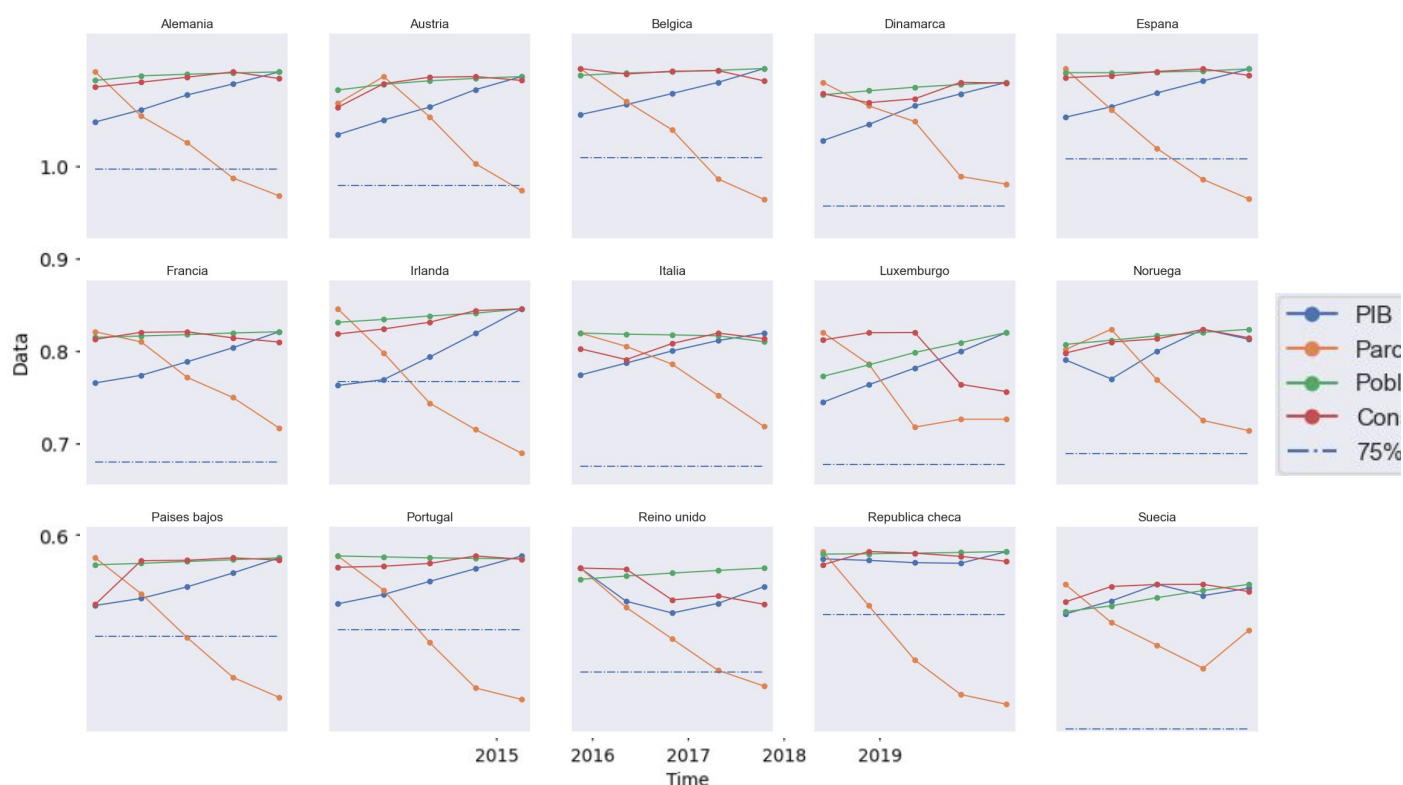


Figura 17: Representación de las tendencias de los principales indicadores socio-económicos de los 15 países entre 2015 y 2019. Todas las variables se encuentran normalizadas para su representación en conjunto.

Vemos que para la mayoría de los países el consumo eléctrico fue en aumento a lo largo de los años considerados. No obstante, algunos países como Francia, Luxemburgo y Reino Unido se salen de esta norma.

Correlaciones entre indicadores. A continuación, para comparar los comportamientos de estas variables entre países a lo largo de este periodo temporal, se ha obtenido la matriz de correlaciones entre variables para cada país. En general sobre todo se observan correlaciones fuertes del consumo con PIB y población para la mayoría de países.

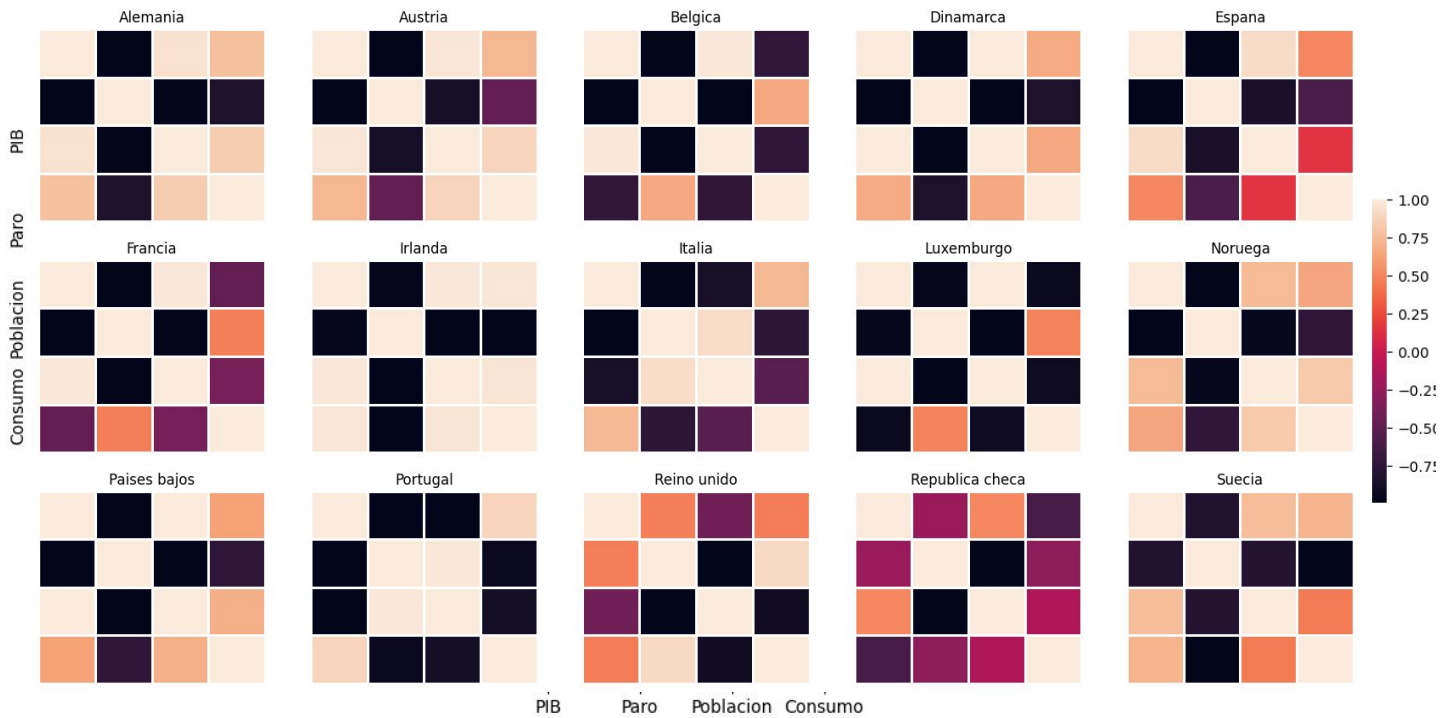


Figura 18: Matriz de correlaciones para las variables de Consumo, Población, PIB y Paro en los 15 países analizados.

En la figura 18 lo primero que podemos observar es que el patrón más habitual es el seguido por países como Alemania, Austria, Irlanda, Noruega o Dinamarca en los que todas las variables presentan correlaciones fuertes entre ellas (blanco) y correlacionan inversamente con el paro (negro).

En contraposición a esto algunos de los países mostrados presentan grandes descorrelaciones del consumo con algunos de los indicadores y un comportamiento más anómalo. Entre ellos podemos destacar Reino Unido, España, Francia y especialmente República Checa; el país que presenta una mayor descorrelación del consumo eléctrico en general.

También podemos destacar otros comportamientos como Portugal, Italia y Bélgica en los que la población correlaciona directamente con el paro.

5.2.2. Perfiles diarios y patrones de consumo.

Comparación de perfiles en distintas estaciones. En la segunda parte del análisis se ha caracterizado el perfil de consumo energético diario estacional para tres países de ejemplo. Se ha calculado el perfil diario para los

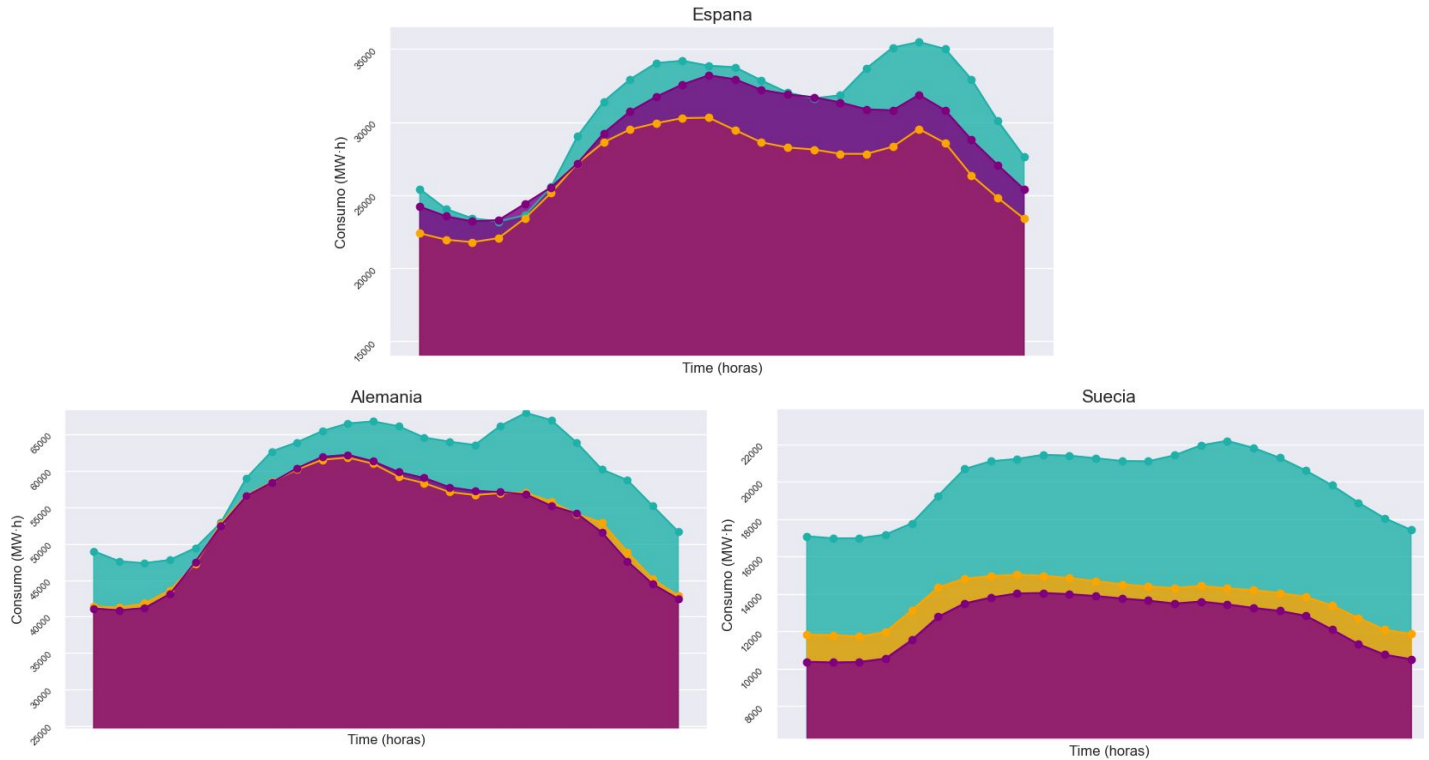


Figura 19: Perfiles de consumo diario para Alemania, España y Suecia en los meses de enero, abril y agosto.

En el gráfico presentado (figura 19) puede observarse para empezar la gran diferencia existente entre los perfiles obtenidos para España y Suecia. También vemos grandes diferencias en el volumen de consumo en función de la estación del año.

Por un lado, en la serie temporal del caso de España predominan dos picos de consumo que coinciden con las franjas horarias de mayor actividad laboral e industrial (5-14h , 17-21h) resultando en una mayor concentración del consumo diario . Por el otro, vemos como el perfil en el caso de Suecia es mucho mas llano e independiente de la hora. Se consume de forma más constante y repartida a lo largo del día.

Esto nos indica que es posible que el consumo eléctrico tenga una mayor dependencia con la actividad laboral en el caso de España, y por tanto podría ser un peor indicador para el caso de Suecia.

Además podemos ver cómo la variación del perfil diario en función del mes del año es mucho más notable para Suecia, mientras que en el caso de España esta variación es menor, viéndose menos condicionada por la estación. Suecia presenta una diferencia del consumo diario máximo de en torno al 30 %

entre verano e invierno y en el caso de España esta diferencia es del 10 % aproximadamente.

Esto puede ser indicador de que el consumo energético Sueco es mucho más dependiente de las condiciones climatológicas así como de la época del año.

Finalmente, como podemos ver Alemania, aun asemejándose más al caso de España, se encuentra en un punto intermedio entre ambos comportamientos. Presenta dos picos bastante suavizados en su perfil diario, similares al caso de España, y al mismo tiempo las variaciones de volumen de consumo no son tan significativas entre estaciones.

Perfil diario de abril promediado sobre 5 años. Vistas las diferencias, mostramos los perfiles de consumo para el mes de abril de los 15 países estudiados. El perfil de consumo diario es una buena forma de caracterizar el patrón de consumo de una sociedad (en este caso los 15 países de estudio) y que permite extraer algunas conclusiones respecto al uso de la energía eléctrica y su adecuación como indicador socio-económico. Se ha elegido el mes de abril, para evitar la influencia estacional en el análisis del consumo.

El consumo energético a lo largo del día es una variable que como hemos podido observar es dependiente del PIB, el Paro y la Población en muchos de los casos estudiados. Además hay que tener en cuenta que esta variable también es dependiente del tiempo (hora del día) y se verá influenciada por el horario de luz de cada país.

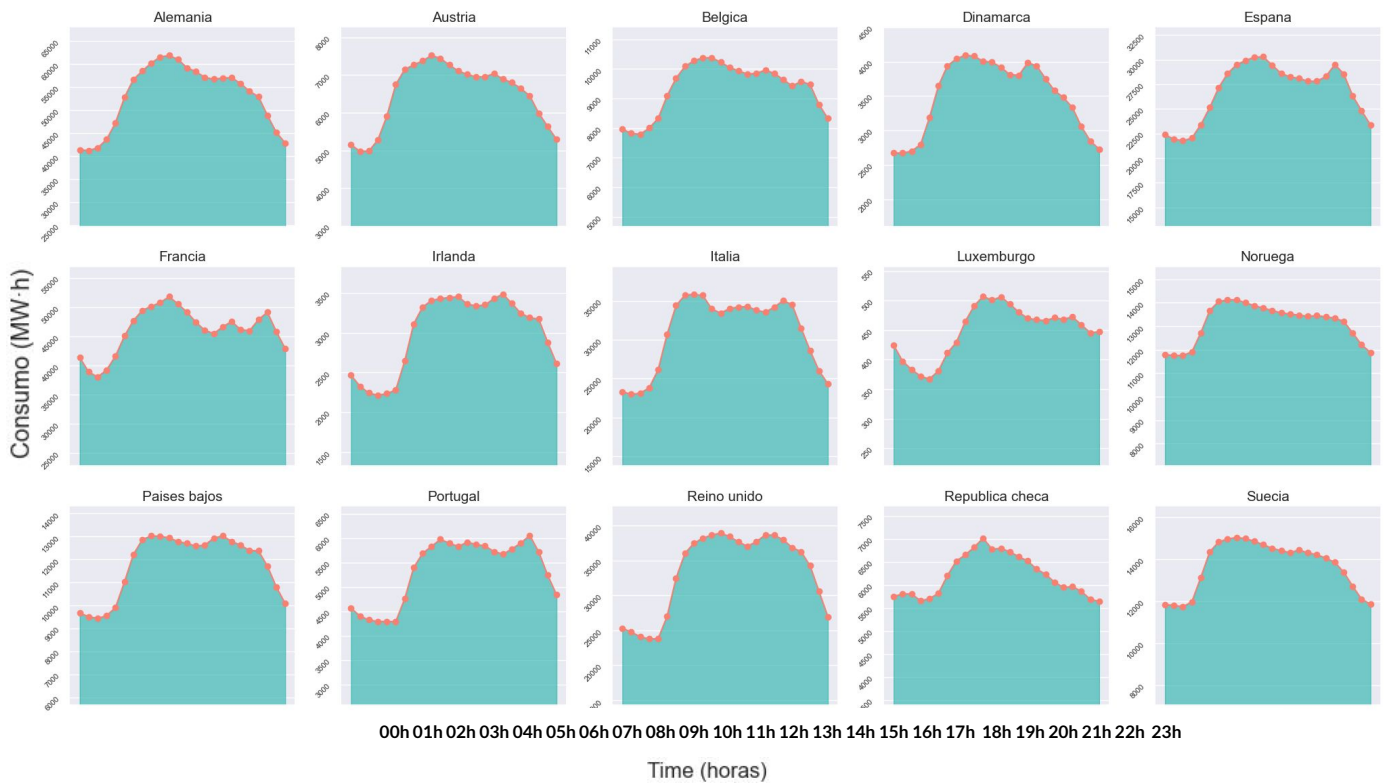


Figura 20: Perfil de consumo diario para el mes de abril promediado sobre 5 años (2015 - 2019).

En la siguiente figura 20 se muestra el perfil primaveral obtenido promediando a lo largo de los 5 años de estudio. Cabe destacar la similitud existente entre los países más próximos geográficamente. Esto puede deberse en parte a la similitud horaria entre estos países así como similitudes culturales entre ellos.

Principalmente podemos distinguir dos bloques diferenciados y un tercer bloque de países intermedio:

- Para empezar, observamos como los países del norte de Europa como Noruega, Suecia y Países Bajos, poseen un consumo más suave a lo largo del día y más independiente del horario laboral.
- Por otra parte, países situados en Europa occidental como España, Portugal o Italia, presentan un perfil mucho más dependiente de las horas del día con grandes picos por la mañana y por la tarde en horarios asociados a la actividad laboral e industrial.

Como tercer caso, apreciamos que en el caso de la mayoría de países el perfil se presenta con picos de consumo suavizados que indican una dependencia

más débil con la hora del día. También cabe destacar el caso de Republica Checa, el país que presenta el perfil de consumo energético más anómalo de los 15 con un perfil suave y un solo pico central.

Correlación entre los perfiles diarios de los países. Finalmente, en el último apartado se ha calculado la matriz de correlaciones para los perfiles mostrados anteriormente. De esta forma podemos comparar en mayor detalle la correlación existente entre las series temporales del consumo energético de cada país.

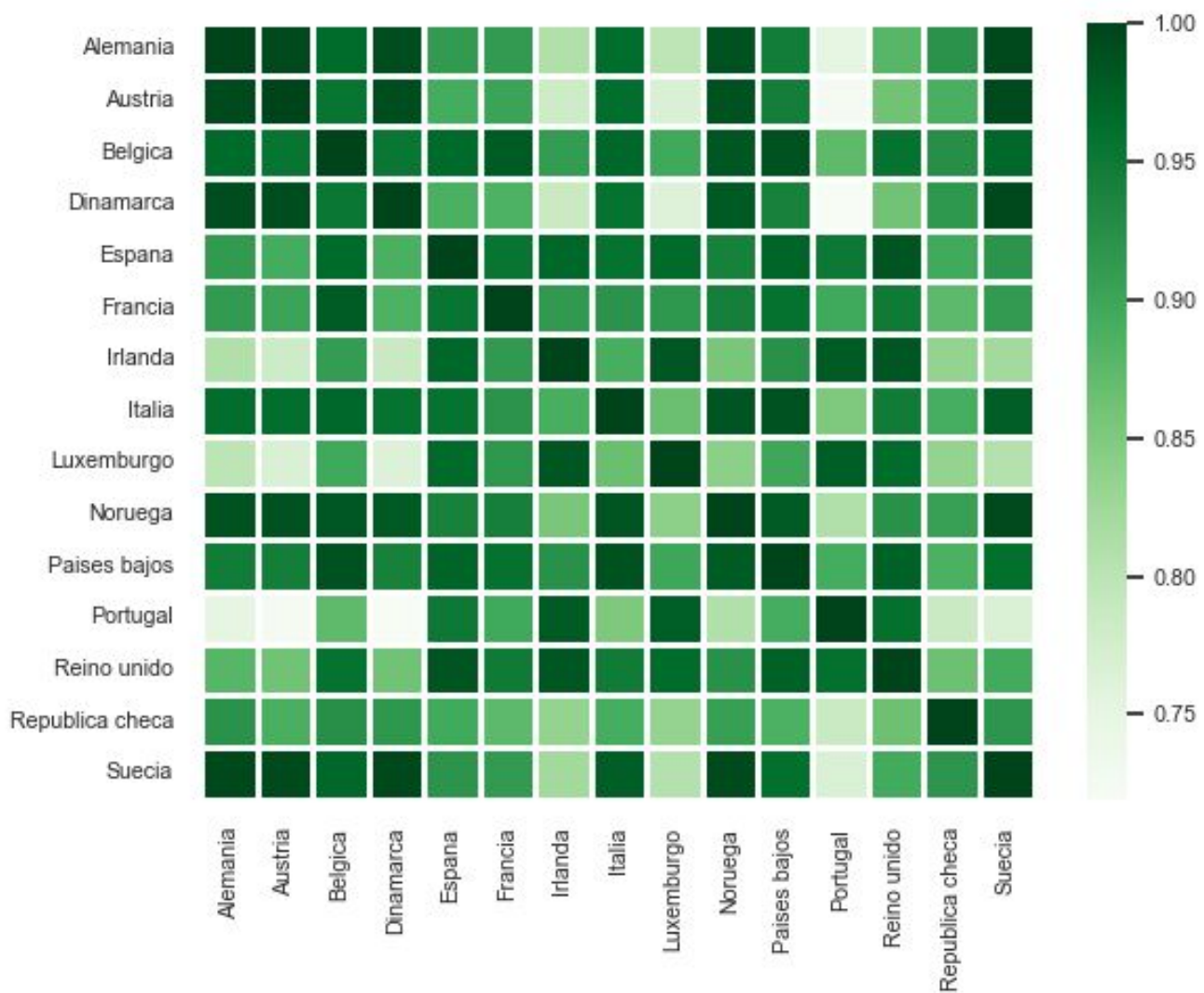


Figura 21: Matriz de correlaciones de los perfiles diarios de consumo de los 15 países de estudio. El rango de valores de correlación es de 0 a 1.

En la figura 21 mostrada, en primer lugar se aprecia una correlación alta generalizada entre países, sobre todo entre aquellos que son cercanos geográfica o culturalmente. Esto es de esperar y debemos fijar la mirada en los casos de correlaciones débiles entre países.

De nuevo en base a lo observado en 21 podemos dividir el conjunto de países en 3 grupos diferenciados.

Principalmente, observamos como Portugal, Luxemburgo e Irlanda son los países cuyo perfil diario descorrelaciona más con el resto, con correlaciones próximas a 0 en algunos casos, mientras que correlacionan fuertemente entre ellos. La serie temporal de estos países descorrelaciona especialmente con la de Alemania, Austria y Dinamarca (República Checa, Suecia y Noruega en menor medida), países cuya correlación es fuerte entre sí.

Estos dos grupos son los que presentan mayores diferencias en su patrón de consumo diario y el resto se encuentran en un término más intermedio con tendencias mas o menos débiles hacia alguno de los dos conjuntos mencionados.

En base a lo observado podemos concluir que Portugal, Luxemburgo e Irlanda presentan un patrón de consumo diario bastante alejado del resto de países europeos considerados en este estudio.

6. Conclusiones

Una vez realizado el análisis de los datos, podemos volver a nuestra hipótesis inicial y ver cómo de acertada era. Esta hipótesis inicial era que existía una relación entre nuestra variable objetivo (el consumo eléctrico) y las magnitudes que caracterizan la economía de un país (el paro, PIB y la población), y ciertamente podemos concluir que esta hipótesis es cierta, pues las correlaciones vistas en la figura 18 son cercanas a ± 1 .

Pero al mismo tiempo, fijándonos en estas correlaciones se puede ver que el tipo de relación que hay entre las variables no es la misma para todos los países, pues en algunos casos la relación es directa mientras que en otros es inversa, e incluso en ciertas ocasiones la correlación es baja. De aquí podemos concluir también que no podemos considerar a todos los países de la misma forma y meterlos todos en el mismo saco.

Esta diferencia entre países es también notable en el perfil diario 20, donde los picos de mañana, tarde y noche son más distinguibles en ciertos países, que concluimos que se deba a la zona geográfica como se ha mencionado anteriormente. Aún así, estos perfiles son altamente similares para todos los países como se ve tanto en los propios perfiles diarios como en la Figura 21, pues todas las correlaciones son superiores a 0,7.

Así, finalmente tenemos que existe una relación entre el consumo eléctrico de un país y algunas magnitudes que miden la economía de un país, pero no podemos deducir un modelo general para predecir a futuro con sólo estos datos debido a la disparidad en el tipo de relación entre estas variables.

Referencias

- [1] EUROSTAT, “Gdp and main components (output, expenditure and income),” Last Update: 01/01/2023. [Online]. Available: https://ec.europa.eu/eurostat/databrowser/view/nama_10_gdp/default/table?lang=en
- [2] F. RAUCENT, “Five years of hourly electricity consumption of 15 european countries,” 2021. [Online]. Available: <https://www.kaggle.com/datasets/francoisrauent/western-europe-power-consumption?select=es.csv>
- [3] E. (Eurostat), “Population by age and sex,” Last Update: 01/06/2022. [Online]. Available: https://ec.europa.eu/eurostat/databrowser/view/demo_pjan/default/table?lang=en
- [4] EUROSTAT, “Total unemployment rate,” Last Update: 01/01/2023. [Online]. Available: <https://ec.europa.eu/eurostat/databrowser/view/tps00203/default/table?lang=en>
- [5] D. U. Board, “Dcml metadata terms,” 2020, last accessed 05 February 2023. [Online]. Available: <https://www.dublincore.org/specifications/dublin-core/dcml-terms/>