**Proyectos II, integración, calidad y análisis exploratorio de datos**

**Segunda presentación:**

**Integración de datos**

**Introducción**

Antes de cada HITO de presentación, debemos rellenar estas fichas y presentarlas a través de PoliformaT, en la tarea que indiquen los profesores. Cada equipo de trabajo presenta las mismas fichas. Sólo será necesario que las suba uno de los componentes del equipo.

Nombres y apellidos de los autores:

| Álvaro | Prado Expósito |
| --- | --- |
| Óscar | García Martínez |
| Aleixandre | Tarrasó Sorní |
| Víctor | Mañez Poveda |
| Jorge | Muedra Vela |

**2. Interés y alcance del proyecto.**

* 1. **Explica el objetivo principal de tu proyecto ¿qué presenta este estudio?**

El objetivo principal del proyecto es comprender el funcionamiento de los sistemas energéticos europeos; haciendo hincapié en España e investigando cómo actúan ante una variación de su demanda u oferta energética. Partiendo de estas variables pretendemos realizar un modelo que consiga predecir el precio de la energía.

* 1. **Explica para qué y para quién podría ser de utilidad este estudio**

Este estudio puede resultar de interés para los gobiernos de los países analizados ya que abarca distintos aspectos del funcionamiento de los sistemas energéticos, con análisis que ayudan a comprender la situación de los mismos.

Las compañías eléctricas también pueden beneficiarse de nuestro estudio, usando nuestras técnicas de visualización y análisis como el primer paso para estudiar la situación energética de un país.

Y por último, también interesaría a las agencias de medioambiente, ONG de renovables, y empresas petrolíferas, ya que pueden ver la evolución durante los años de los tipos de energía más utilizados, los cambios entre países, y hacer posibles predicciones sobre estos.

* 1. **¿Por qué piensas que es novedoso? ¿has visto estudios similares?**

El interés que tiene este estudio respecto a otros estudios similares es que no solo analiza la estructura energética anualmente, sino que aumenta el nivel de granularidad para conocer cómo afecta la estacionalidad a los sistemas energéticos.

Los informes que hemos encontrado con temática similar los adjuntamos a continuación:

Gómez Aguirre, M., & Castro Loaiza, J. (2017). *Crecimiento económico, consumo de energía eléctrica y comercio. Un análisis de causalidad para México 1968-2013*.

*La contribución de las redes eléctricas a la transición energética*. (2018, febrero 25). Deloitte Spain. https://www2.deloitte.com/es/es/pages/strategy/articles/contribucion-redes-electricas-transicion-energetica.html

Guido, L. (s. f.). *REDES DE ENERGÍA E “ INNOVACIÓN ” TECNOLÓGICA EN EUROPA . EL CASO DE F RANCIA*. Edu.ar. Recuperado 8 de marzo de 2024, de http://sedici.unlp.edu.ar/bitstream/handle/10915/59982/Documento\_completo.7703.pdf-des.pdf-PDFA.pdf?sequence=1

* 1. **Alcance (objetivos definitivos del proyecto)**

**Define los objetivos del análisis de datos de tu proyecto**. Se deben presentar 5 objetivos de análisis.

1. CLASIFICACIÓN. La meta es clasificar a los diversos países miembros de la Unión Europea según los datos que podemos encontrar en los balances finales publicados. Contamos con un dataset proporcionado por Eurostat que en cada hoja ofrece el balance energético de un país.

Para iniciar el análisis, realizaremos un diagrama de Sankey donde visualizaremos cómo se distribuye la energía en un país. Es decir, podremos ver cuánta de la energía consumida es propia o importada, y de qué fuente proviene; y a la vez saber cuánta de la energía generada es exportada o se mantiene en el país.

A partir de esta información, podemos observar qué variables varían más entre países y realizar histogramas a partir de ellas. Clasificando a los países por los valores en estas variables.

Asimismo, podemos apreciar que variables varían más (tienen más variabilidad) en el tiempo y realizar gráficos de líneas o tendencias a partir de ellas. Clasificando a los países por su tendencia en estas variables.

Lo más importante es extraer conclusiones fiables sobre la situación energética en cada país y comprender las razones que han llevado a esta situación.

1. CLUSTERING. Para este objetivo nos centramos en España. La idea es realizar un clustering sobre las CCAA de España según la generación, la demanda y la diferencia entre ambas. El gránulo del estudio es trimestral.

Las conclusiones del objetivo se centrarán en determinar qué comunidades autónomas tienen un comportamiento energético similar. A grandes rasgos, trataremos de discernir qué CCAA destacan por tener una generación de energía superior a su demanda, y cuáles tienen una actividad económica tal como para necesitar ayuda de otras comunidades para suplir su demanda energética.

1. VISUALIZACIÓN. Para apoyar el objetivo anterior vamos a realizar una visualización de los datos. Generamos un mapa de España por CCAA coloreadas por el indicador económico escogido (que puede ser Empleo Industrial, IPI o PIB per Cápita), la generación o la demanda asociada a dicho mes en la CCAA.

Visualmente podemos apreciar si es cierto que las comunidades autónomas con mayor actividad económica tienden a tener déficit energético, y apreciamos que CCAA ayudan a suplir ese déficit.

1. REGRESIÓN. Realizamos una regresión múltiple para generar un modelo que prediga el precio del día posterior. Usaremos como país estudiado y de referencia España, ya que es el más completo, con más variables de recolección de energía, y mayor número de datos para el estudio; aunque planeamos ampliar el estudio a más países.

El modelo será entrenado con datos diarios desde el 01-01-2011 al 20-03-2019. Trabaja con el precio del día siguiente como variable respuesta; y la demanda energética y generación por fuente de energía entre otras como variables predictoras.

1. EVOLUCIÓN. La meta es estudiar la evolución del uso de cada fuente de energía a lo largo del tiempo en cada comunidad autónoma. Para llevar a cabo el análisis, se realizaran gráficos de tendencias por fuente de energía, y se explicarán las tendencias en cada región con el contexto que han vivido en la última década (es desde cuando tenemos datos).

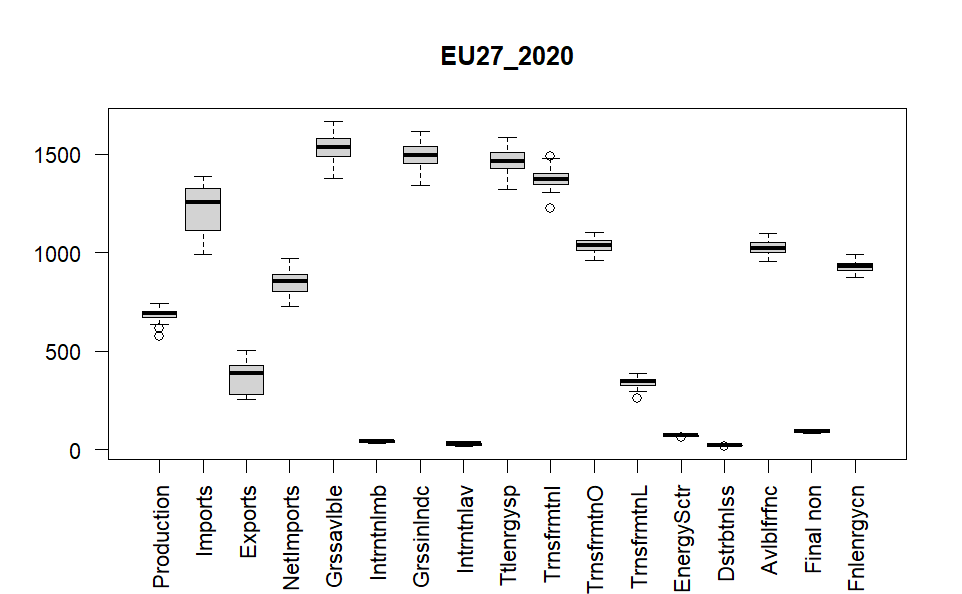
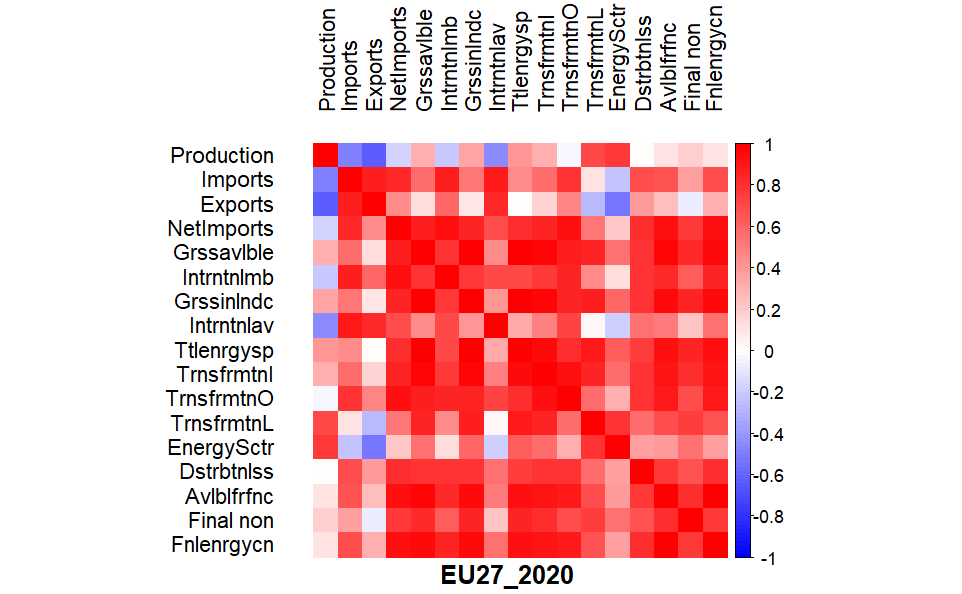
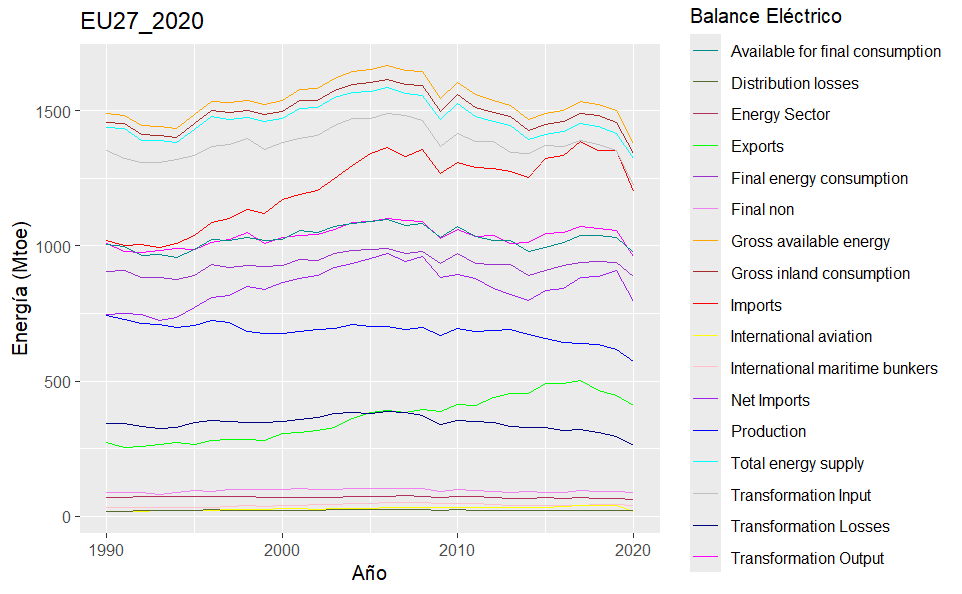
**3. Calidad y Análisis exploratorio.**

Como se puede observar, los datos estudiados son diferentes a los definidos en el primer hito. Esto se debe a que en aquel momento pensábamos en desarrollar los cinco objetivos sobre una gran base de datos, pero debido a la difícil aplicación de esta idea, hemos decidido idear objetivos centrados en bases de datos más específicas. En cambio, muchas de las variables definidas en el primer hito se seguirán utilizando para el cuarto objetivo.

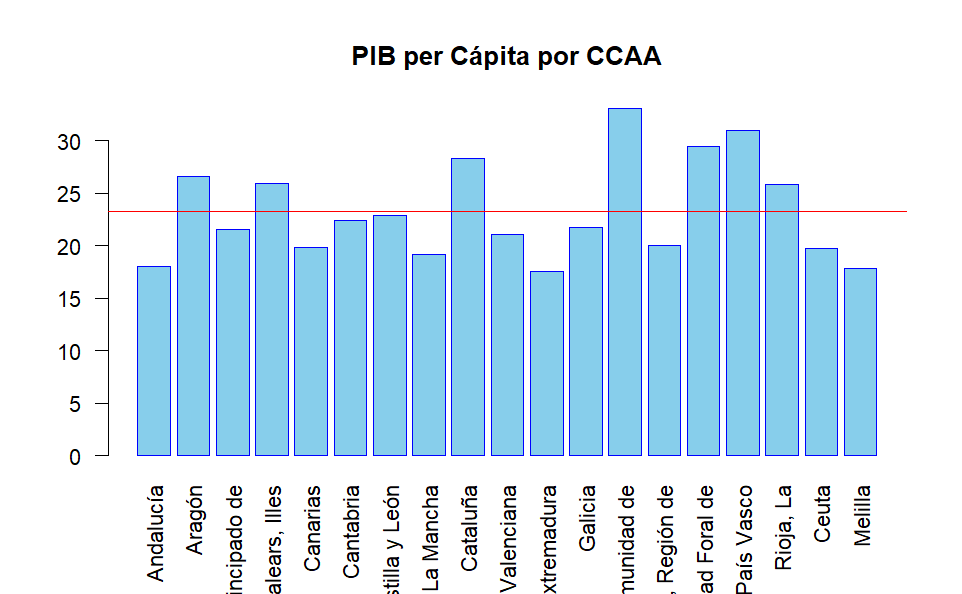
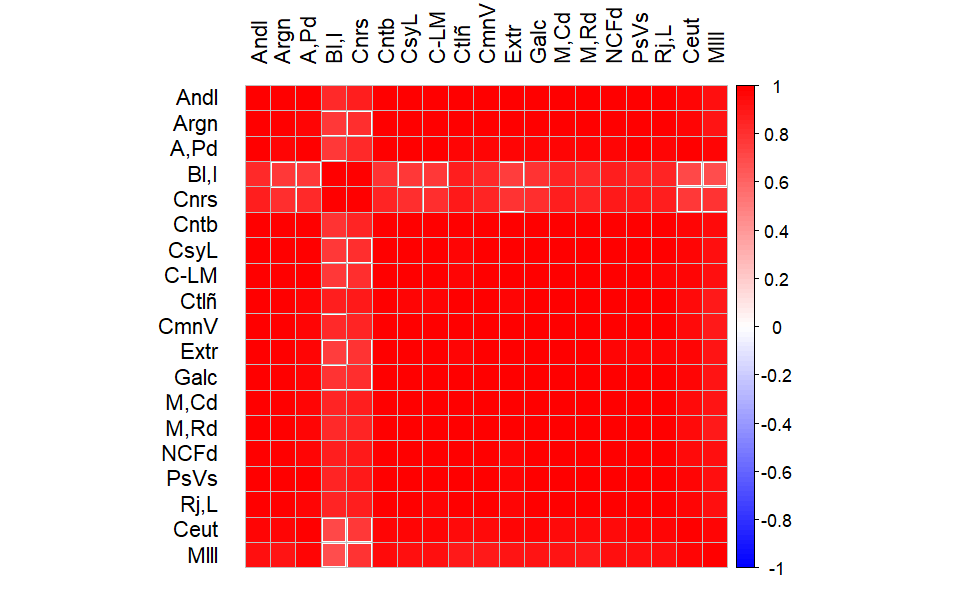
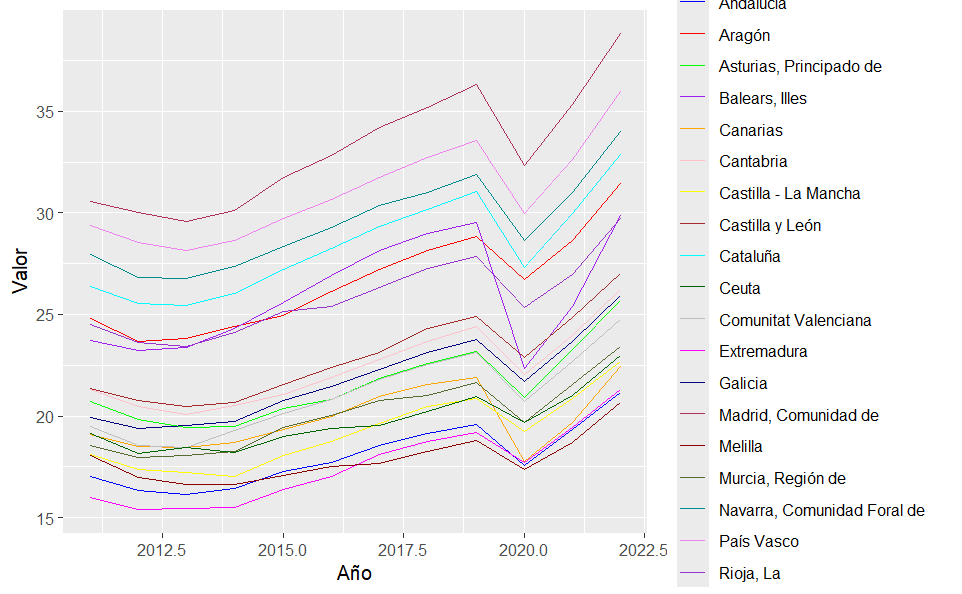
El objetivo 1 se fundamenta en dos bases de datos publicadas por Eurostat a las que no ha sido necesario aplicar ninguna transformación.

Los objetivos 2, 3 y 5 trabajan con bases de datos muy similares; tablas con datos mensuales sobre generación y demanda energética por comunidades autónomas. Estos datos se obtienen mediante una API ofrecida por la Red Eléctrica Española. El objetivo 4 trabaja con datos extraídos de las redes eléctricas de España, Francia, Portugal, Dinamarca y Estonia que muestran la estructura de la generación y demanda a nivel estatal. Para España los hemos extraído de la API y para el resto hemos descargado datasets ofrecidos por sus respectivas redes eléctricas. La extracción de datos está en el archivo **extraccion.Rmd**.

En el primer objetivo se analiza una base de datos que contiene el balance energético de cada país de la UE, incluido el total de todos los países miembros. Para poder realizar un análisis exploratorio sobre estos datos se ha desarrollado una función que para cada balance energético realiza un gráfico de líneas, una matriz de correlación y un gráfico de caja y bigotes. De este modo hemos podido realizar un análisis exploratorio en el archivo adjunto **objetivo\_1.Rmd**; concluyendo que las variables siguen diferentes tendencias dependiendo del contexto nacional e histórico y que los datos tienen la calidad suficiente como para ser añadidos en el proyecto.

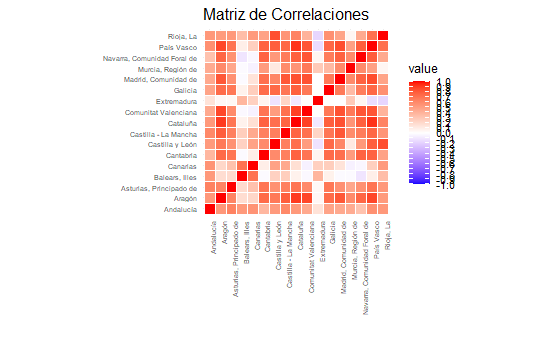
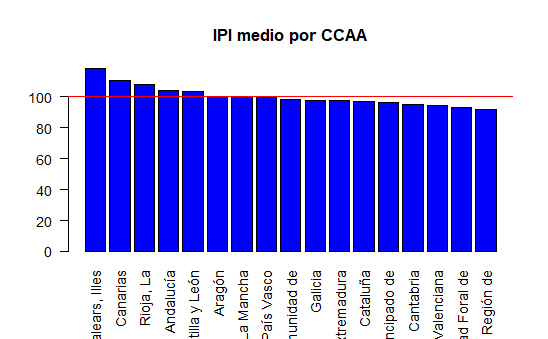
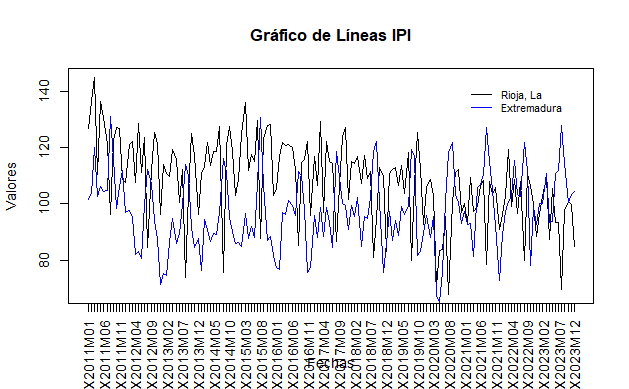


Dado que en el segundo y el tercer objetivo se estudian diversos indicadores económicos hemos debido realizar un análisis exploratorio sobre ellos. En el archivo **ipc+pib.Rmd** que ha sido enviado en los adjuntos, se ha realizado el análisis exploratorio sobre el IPC, el PIB y el PIB per Cápita; llegando a la conclusión de que únicamente el PIB per Cápita tiene el interés suficiente como para ser añadido en el estudio. El IPC ha sido descartado debido a su poca variabilidad entre CCAA. A la hora de decidir entre el PIB per cápita y el PIB escogimos la primera opción puesto que la consideramos un mejor indicador de la riqueza de una región y cruzar esta información con la generación y demanda energética es más interesante. Para este análisis se han utilizado histogramas, gráficos de líneas y matrices de correlación.

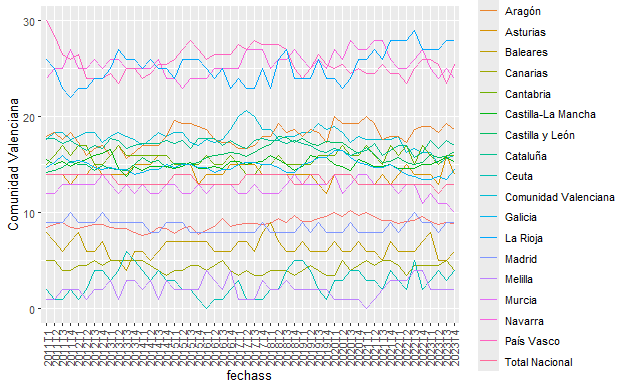
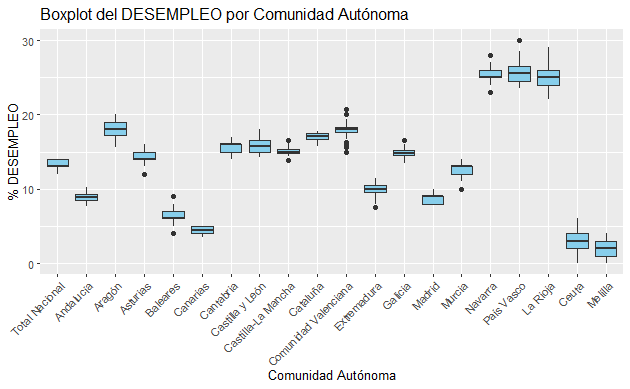


Además, también hemos analizado indicadores como el IPI, el desempleo y el empleo industrial en los archivos **markdown\_ipi.Rmd**, **markdown\_desempleo.Rmd** y **Empleo\_industria.Rmd** respectivamente. Las tres variables han resultado tener la variabilidad suficiente cómo para ser incluidas en el estudio, cómo demuestran los siguientes gráficos. Se ha seguido la misma estructura de análisis que en **ipc+pib**.

-Gráficos para el IPI:

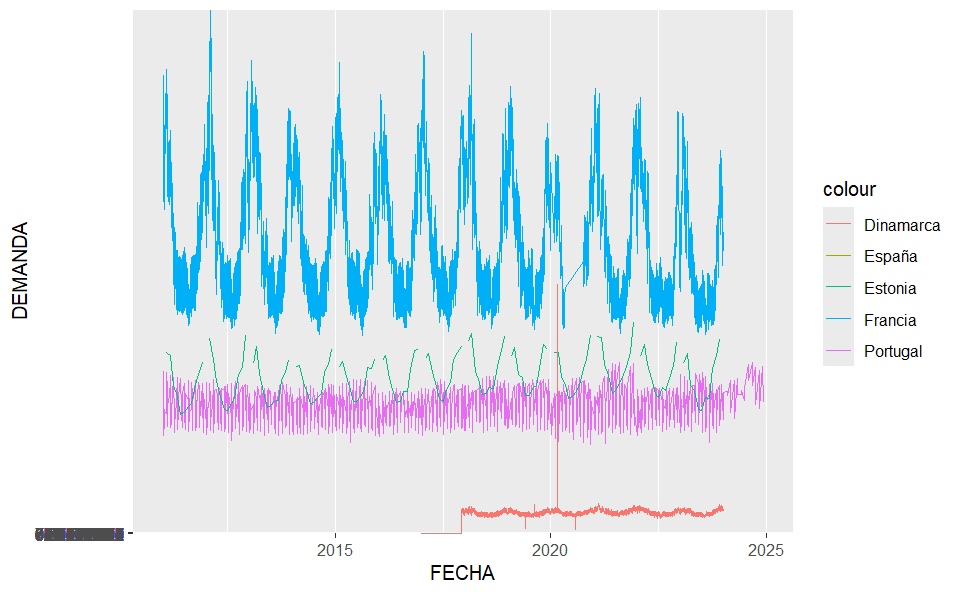
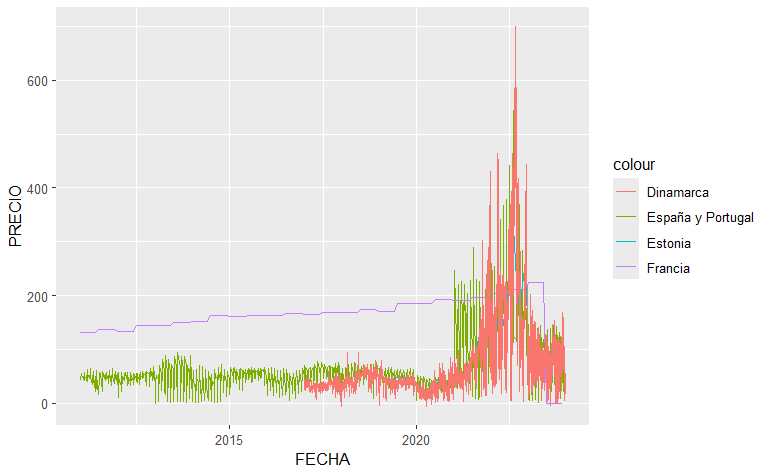
****  

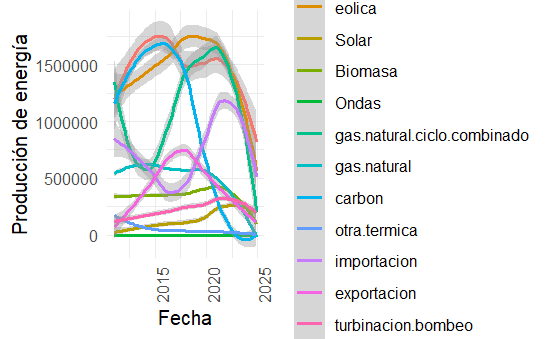
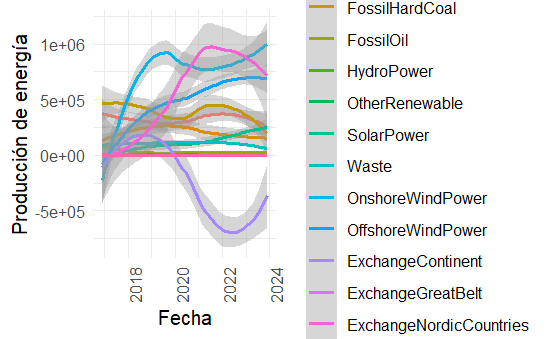
-Gráficos de empleo y desempleo:



En el cuarto objetivo nos adentramos en el análisis de varias bases de datos provenientes de diferentes países. Estas bases de datos contienen información sobre el precio, la demanda y el consumo energético. Realizamos un análisis exploratorio detallado que abarca diversas dimensiones.

En primer lugar, estudiamos la demanda de energía a lo largo del tiempo en los países bajo estudio. Además, examinamos la evolución de los precios de la energía a lo largo del mismo período.

También nos sumergimos en los diferentes tipos de generación de energía utilizados por cada país, presentando los porcentajes correspondientes a cada uno. También hemos comenzado a analizar la regresión del precio de la energía utilizando el precio y la demanda del día anterior, así como el coeficiente de colaboración de los distintos tipos de demanda.

La variabilidad de la generación y demanda energética nos sirve para confirmarlas como variable de interés para otros objetivos.

El análisis exploratorio de este objetivo ha resultado satisfactorio y nos confirma que la regresión puede ofrecer buenos resultados.

