**2° Entrega Proyecto Final: Obtención de insights a partir de visualizaciones**

**Proyecto Final DS 📚 💻 📊 🧪**

*Camino hacia la transición energética - Una perspectiva desde el mercado eléctrico*



Comunidad Coderhouse

Curso Data Science

Profesor responsable: Ruiz, Jorge

Tutora: Ruscitti, Aldana

Alumno: Risculese, Francisco

Consigna

1. Entrenar y optimizar diversos modelos de Machine Learning para resolver una problemática específica, detectada en la instancia de entrega anterior. Siempre bajo el objetivo de poder utilizar modelos de Machine Learning para resolver el problema de una industria o negocio.

Objetivos generales para la pre-entrega

* Obtener datos de diversas fuentes como APIs o Bases de datos públicas para luego analizarlos mediante el lenguaje Python con el fin de contestar una pregunta de interés para una industria, negocio o proyecto personal. Se deberán utilizar datasets complejos implementando técnicas avanzadas para la limpieza y adquisición de datos.

Objetivos específicos para la pre-entrega

* Estructurar un problema en función de múltiples pero simples preguntas/hipótesis a responder.
* Importar datos crudos de APIs o bases de datos usando Python.
* Limpiar y transformar los datos para permitir un posterior análisis.
* Contar una historia mediante el análisis exploratorio de datos.

Requisitos base

* Un notebook (Colab o Jupyter) que debe contener:
  + Abstracto con motivación y audiencia: Descripción de alto nivel de lo que motiva a analizar los datos elegidos y que audiencia se podrá beneficiar de este análisis.
  + Preguntas/hipótesis a responder: Lista de preguntas que se busca responder mediante el análisis de datos. Bloques de código donde se importan los datos desde una API o base de datos pública y los guarda en un archivo local csv o json. El estudiante puede luego de descargar los datos, comentar este bloque de código.
  + Análisis exploratorio de datos (EDA): Análisis descriptivo de los datos mediante visualizaciones y herramientas estadísticas.
* Una presentación (PDF; PowerPoint o Google Slides) que debe contener:
  + Abstracto con motivación y audiencia: Descripción de alto nivel de lo que motiva a analizar los datos elegidos y que audiencia se podrá beneficiar de este análisis.
  + Resumen de metadata: resumen de los datos a ser analizados es decir, número de filas/columnas, tipos de variables, etc.
  + Preguntas hipótesis que queremos responder: Lista de preguntas que se busca responder mediante el análisis de datos.
  + Visualizaciones ejecutivas que responden nuestras preguntas: utilización de gráficos que responden las preguntas de interés de nuestro proyecto.
  + Insights: resumen de hallazgos del proyecto. Consolidar las respuestas a las preguntas/hipótesis que se han ido contestando con las visualizaciones.

Sugerencias

* Es conveniente retomar el dataset trabajado en la primera pre entrega y enriquecerlo (e.g joins, y creación de nuevas columnas) con información proveniente de APIs públicas siempre que se pueda con el fin de practicar las nuevas habilidades adquiridas.
* Se recomienda retomar la metodología de trabajo y reutilizar algoritmos ya entrenados, de ser necesario.

Requisitos extra

* Subir el proyecto a Github

Dont’s

* Utilizar jerga demasiado técnica en la presentación (recordar que la audiencia de la misma son roles ejecutivos).
* Sobrecargar las diapositivas.
* Realizar una presentación con más de 12 slides de extensión.

La notebook ha ido desarrollándose, desde su diseño inicial, mediante los siguientes apartados:

| **2° Entrega Proyecto Final** | | |
| --- | --- | --- |
| **Nombre e instancia de entrega** | **Estado** | **Objetivos y tareas realizadas** |
| Descarga de datos desde APIs públicas (Entrega N° 6) | Aprobada | Continuación sobre lo trabajado en la Primera Entrega (o pre entrega) del Proyecto Final.  Creación de una notebook donde se seleccionará una API de interés, luego se creará una API key.  Finalmente, se extraerá la información para ser almacenada en un DataFrame (se sugiere que esta información complemente o enriquezca el dataset elegido en el desafío “Elección de potenciales Datasets e importe con la librería Pandas”) |
| Data Wrangling (Pre-Entrega N° 7) | Aprobada | Continuación sobre lo trabajado en el desafío “Descarga de datos desde APIs públicas”.  Creación de un notebook donde se desarrollará la limpieza de los datos elegidos para el Proyecto Final, a tener en cuenta: técnicas vistas en clase para el tratamiento de valores duplicados, nulos y outliers con su respectiva justificación. |
| Data Storytelling (Pre-Entrega N° 8) | Aprobada | Continuación de lo trabajado en el desafío “Data Wrangling”.  Creación de un notebook donde se desarrolle una narrativa que permita dar respuesta a las preguntas/hipótesis formuladas para el proyecto final. |
| Obtención de Insights (Pre-Entrega N° 9) | En revisión | Continuación de lo trabajado en el desafío “Data Storytelling”.  Creación de un notebook donde se puedan observar las fases de análisis uni y bivariado, que junto con el trabajo previo realizado permitan obtener insights que ayuden a dar respuesta a las preguntas problema del Proyecto Final. |

*Tabla 1: Componentes de la 2° Entrega del Proyecto Final*

Adicionalmente, se realizó una presentación ejecutiva para exponer la obtención de insights a partir del EDA realizado, y los pasos previos.

Etapas del proyecto

# 📜 Abstracto

La estimación del precio de la energía eléctrica es un desafío crucial en el contexto de la sostenibilidad y la transición energética, temas de gran relevancia en la actualidad. Para abordar este desafío, es fundamental comprender en detalle el comportamiento de la variable a estimar, identificar a los actores clave en el mercado energético y analizar las complejidades que influyen en este sector.

El incremento en los precios de los combustibles fósiles, así como su dependencia en muchos países, ha resultado en un aumento significativo en el precio de la energía eléctrica. Sin embargo, esta situación ha impulsado el interés en soluciones energéticas renovables a corto plazo. En España, el gobierno está acelerando la aprobación de proyectos que se centran en la sostenibilidad y la protección del medio ambiente. Esto incluye la adopción de tecnologías avanzadas y promueve aspectos positivos como la autogestión y el autoconsumo de energía.

# 🎯 Objetivo

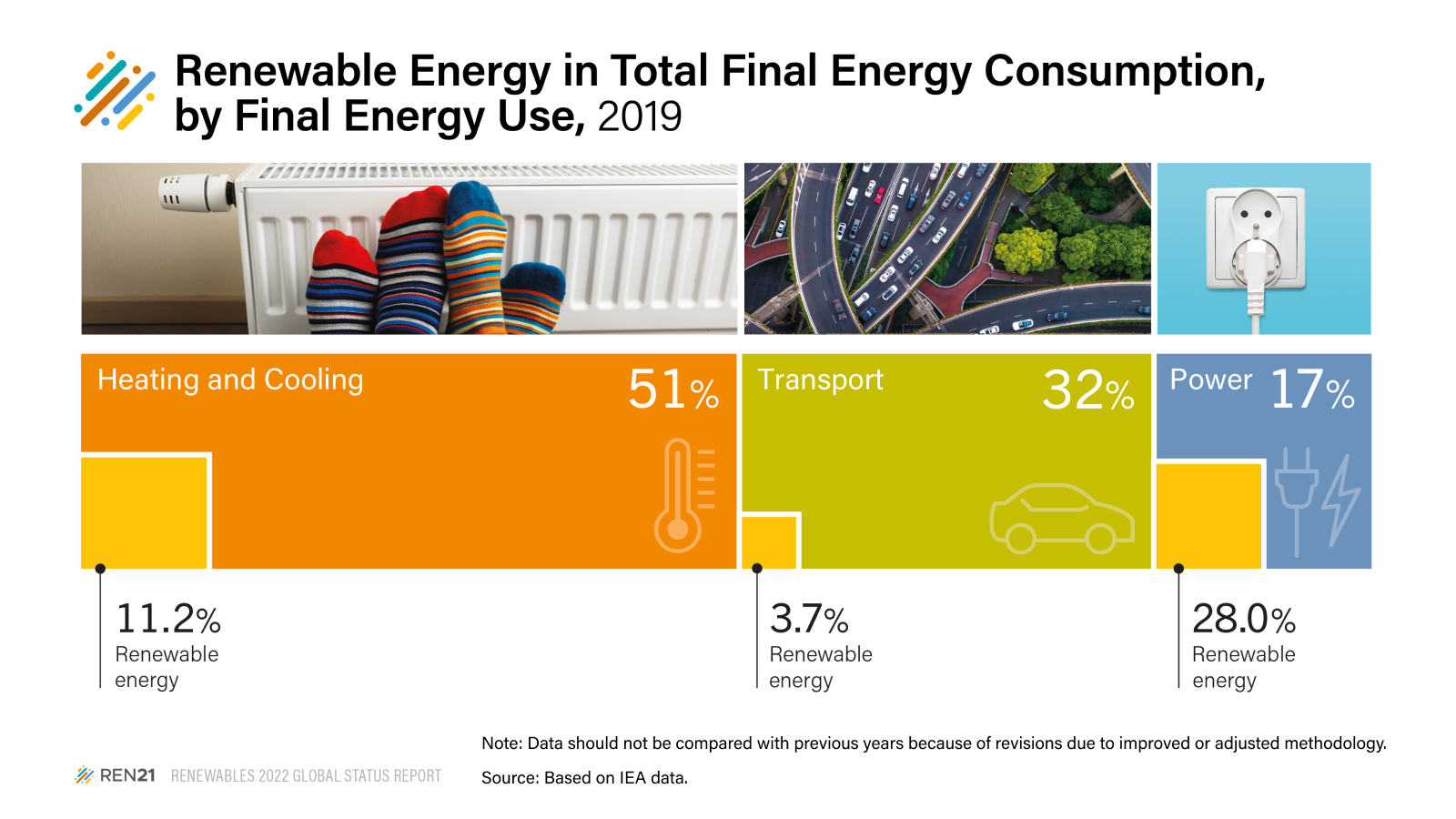
Predicción del precio de la energía eléctrica en España, mediante modelos de regresión, para el período 2015-2018 empleando un dataset obtenido del website "Kaggle".

# 🏭 Contexto comercial

Se debe conocer de antemano que el aumento de los precios de los combustibles fósiles, como el gas natural y el carbón, ha provocado un aumento de los precios de la energía eléctrica. Este aumento es especialmente notable en los países que dependen en gran medida de los combustibles fósiles para generar electricidad.

Las energías renovables (hidroeléctricas, solar o eólica) son una fuente de energía más barata y fiable que los combustibles fósiles.

Además, su aplicación y campo de acción ha ocupado un buen espacio en el sector de la energía para electrificación, según datos de consumo medidos en el año 2019.



*Figura 1: EE.RR. en el Total del Consumo Final de Energía, según el uso final de esta energía*

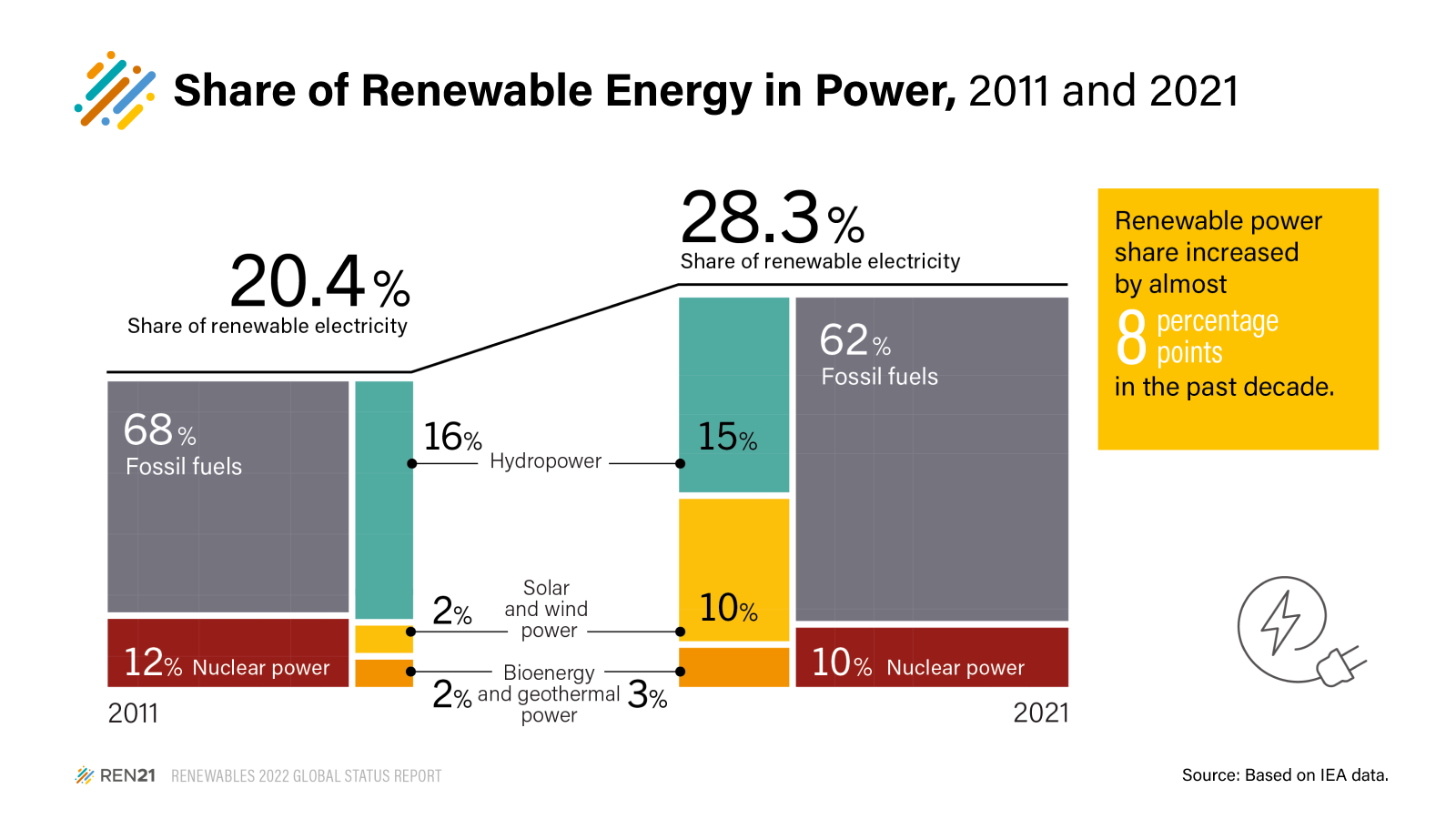
En materia de costos, las energías solar y eólica han disminuido significativamente su valor en los últimos años, lo que las ha hecho más competitivas con respecto a las fuentes de energía convencionales (combustibles fósiles). El aumento de la participación de estas energías limpias en la matriz energética apoya a la reducción de los precios de la energía eléctrica.

La determinación del precio sigue una secuencia de actividades que involucra a todos los actores de la cadena de negocio (explicado a continuación), pero para entenderlo desde una perspectiva clara desde el precio: se debe conocer que aquellas energías poco flexibles que ofertan su producción a costos bajos (solar y eólica) o aquellas que ofertan su producción de forma constante y a plena potencia, como la nuclear, podrán disminuir con sus ofertas el precio de este servicio. Esto se justifica porque las tecnologías mencionadas ofertan de manera necesaria toda la energía que producen, ya que no poseen la capacidad de almacenar y gestionar sus recursos (son, como se dijo, poco flexibles).

Por otro lado, aquellas que ofertan su generación con costos de producción elevados (las que emplean combustibles fósiles), hacen que el precio suba, y pueden participar con mucha más flexibilidad frente a los requerimientos de la demanda. Un comportamiento similar ocurre con las centrales hidroeléctricas que pueden gestionar de forma práctica el recurso hídrico, pero con la diferencia que poseen un costo de obtención del recurso nulo, al ser obtenido del medio ambiente.

En España, el precio de la electricidad se ha multiplicado por cuatro en los últimos años. Este aumento es debido a la creciente dependencia de los combustibles fósiles para generar electricidad. Por ello, el gobierno español está acelerando la transición a las energías renovables. Fuentes gubernamentales proponen aumentar la participación de las energías renovables en la matriz energética española del 40% al 74% para 2030.

Como España, muchos países se involucran en este movimiento acelerado y los cambios en un plazo de 10 años, a nivel mundial, están tomando cada vez más impronta, sobre todo en el sector de la electrificación.



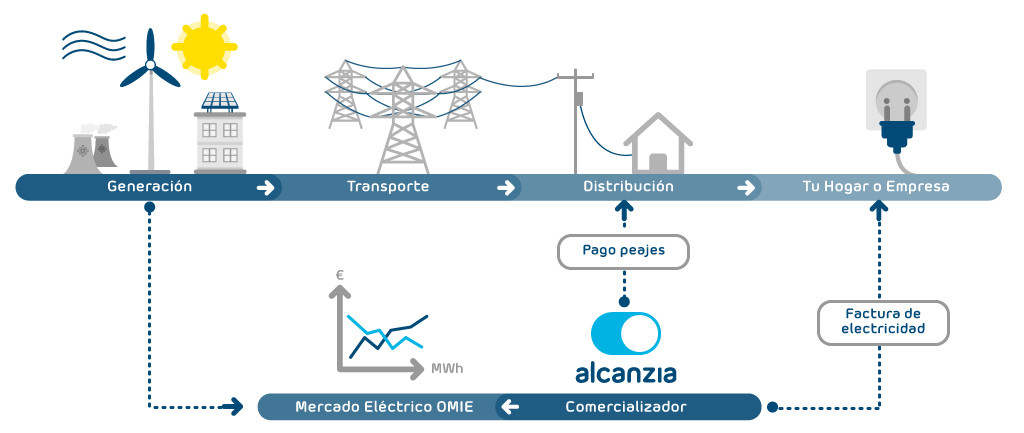
*Figura 2: Porcentaje de participación de las EE.RR. en el sector de la electricidad.*

# 💭 Problema comercial

Como científico de datos, se desempeñará la labor en estimar el precio de la energía eléctrica para la empresa OMIE. Es obligatorio conocer el comportamiento del mercado eléctrico español, y el de sus actores participantes. Siendo el OMIE el operador del Mercado Ibérico de Energía, que opera tanto en Portugal como en España: este tiene diversas funciones en todo el proceso de intercambio de energía, regulando el proceso e interviniendo en ciertas circunstancias, por ejemplo en acuerdos entre los comercializadores (compradores de energía) y los generadores (vendedores de energía).

En forma resumida, participa como un actor terminal el sector de los generadores, que son los destinados a producir la energía eléctrica, sea generada por fuentes convencionales de energía (como la nuclear o los combustibles fósiles) o por fuentes renovables (destacándose la hidroeléctrica y la eólica).

Luego, la matriz de transportistas, distribuidores y comercializadores, por separado, brindarán las condiciones necesarias para que el servicio sea trasladado, por medio de las redes de alta-media-baja tensión, al otro actor terminal que forma parte del negocio: el sector consumidor.



*Figura 3: Estructura del mercado español de la energía eléctrica*

Interrogantes principales

⚡ Variable precio. ¿Qué resultados se observan, haciendo un análisis estadístico descriptivo e inferencial, sobre la variable target?

⚡ Evolución temporal para el precio. ¿Cómo se comporta el precio de la energía eléctrica a lo largo del período considerado? ¿Cómo se lleva a cabo la descomposición de la serie temporal, y qué se puede determinar realizando ésto? ¿Qué sucede con la estacionalidad de la serie temporal?

⚡ Generación energética y precio de la electricidad. ¿Cómo se da la evolución en el aporte de cada fuente de generación energética en un período anual? ¿Y cómo se da la evolución en el aporte de cada fuente de generación energética en un período diario? ¿Qué relación se puede obtener si se incluye la variable precio a estas visualizaciones?

⚡ Valores atípicos de la variable precio. ¿Qué análisis se puede realizar teniendo en cuenta los valores outliers?

⚡ Generación de energía por ubicación geográfica. Si el conjunto de datos incluye información sobre la ubicación de las plantas de generación de energía, ¿cómo varía la generación de energía en diferentes regiones geográficas? ¿Qué se obtiene de un análisis multivariado?

Interrogantes complementarios

⚡ Evolución de la demanda de energía en el país. ¿Cómo evoluciona la situación de esta variable para los años de estudio? ¿Qué comportamientos se muestran con respecto a su tendencia? ¿Qué se puede concluir respecto a un período semanal de consumo?

⚡ Evolución temporal para el precio. ¿Cómo se comporta el precio de la energía eléctrica analizando cada año en particular?

⚡ ¿Qué relación se puede establecer, a grandes rasgos, entre la variable demanda y el precio fijado de la energía eléctrica?

⚡ Relaciones con la variable precio. ¿Qué relación se puede establecer, a grandes rasgos, entre la variable target y las demás features?

⚡ Generación energética. ¿Qué se puede observar cuando se analiza la distribución de valores en un histograma para cada variable?

⚡ Valores atípicos de la variable precio. ¿Qué comportamiento presenta la variable target, analizando sus valores atípicos?

⚡ Valores atípicos de las variables climatológicas. ¿Qué comportamientos presentan las variables, analizando sus valores atípicos? ¿Qué toma de decisión se puede realizar con el análisis?

⚡ Generación energética. ¿Cuál es el aporte absoluto de cada fuente de energía (renovable o convencional) a su generación total? ¿Qué comportamiento exhibe cada fuente?

⚡ Proporción en la generación. ¿Cuál es la proporción porcentual de generación de energía proveniente tanto de fuentes renovables como de fuentes convencionales?

⚡ Impacto ambiental. ¿Existe alguna correlación entre la generación de energía a partir de fuentes fósiles y las emisiones de gases de efecto invernadero u otros impactos ambientales?

⚡ Potencia instalada en el sector de generación energética. ¿Cómo se distribuye la cantidad de potencia instalada para las diversas tecnologías de producción de energía eléctrica?

# 👨‍💻 Contexto analítico

Se cuentan con dos datasets, bajo un nombre genérico que identifica a ambos. Estos, están en formato .csv.

De estos dos, uno se utilizó para desarrollar los trabajos de desafíos entregables iniciales, el dataset restante contiene información adicional y complementaria que se utilizó cuando fue necesario aplicar el Data Enrichment, dentro del Data Wrangling (etapa de enriquecimiento de la data).

Además, se contó con información adicional, para ampliar el alcance del proyecto, y así contar con información en proyectos, datos históricos e infografías.

* Dataset general: *Hourly energy demand generation and weather*. El dataset general contiene, en un período de 4 años, datos sobre el consumo, generación, precio y climatología en España.
  + [energy\_dataset.csv](https://www.kaggle.com/datasets/nicholasjhana/energy-consumption-generation-prices-and-weather?select=energy_dataset.csv): Contiene datos sobre consumo y generación, recopilados desde el [portal público del ENTSOE-E](https://transparency.entsoe.eu/dashboard/show) (European Network of Transmission System Operators) para los operadores de red de transporte TSO (Transmission Service Operator). La declaración de los precios fue obtenida por el TSO de España: Red Eléctrica España.
  + [weather\_features.csv](https://www.kaggle.com/datasets/nicholasjhana/energy-consumption-generation-prices-and-weather?select=weather_features.csv): La información climatológica fue comprada como parte de un proyecto personal, desde el [Open Weather API](https://openweathermap.org/api), para las 5 ciudades más grande del país, y empleadas para este proyecto.
* [Website del OMIE (ente regulador del precio)](https://www.omie.es/es/market-results/daily/daily-market/hourly-power-technologies)
* [Website de la Red Eléctrica de España (ente del transporte energético)](https://demanda.ree.es/visiona/seleccionar-sistema)
* [Centro de descargas del CNIG](https://centrodedescargas.cnig.es/CentroDescargas/busquedaRedirigida.do?ruta=PUBLICACION_CNIG_DATOS_VARIOS/aneTematico/Espana_Principales-centrales-electricas_2014-2016_mapa_15048_spa.zip)
* [Informes anuales de la REE](https://www.sistemaelectrico-ree.es/)
* [Bibliografía adicional consultada](https://drive.google.com/drive/folders/1OHJ3Zt4fFWxlzVgdGE1oD7PcfASJOYl6?usp=sharing)
* [Valor atípico - Wikipedia](https://es.wikipedia.org/wiki/Valor_at%C3%ADpico)

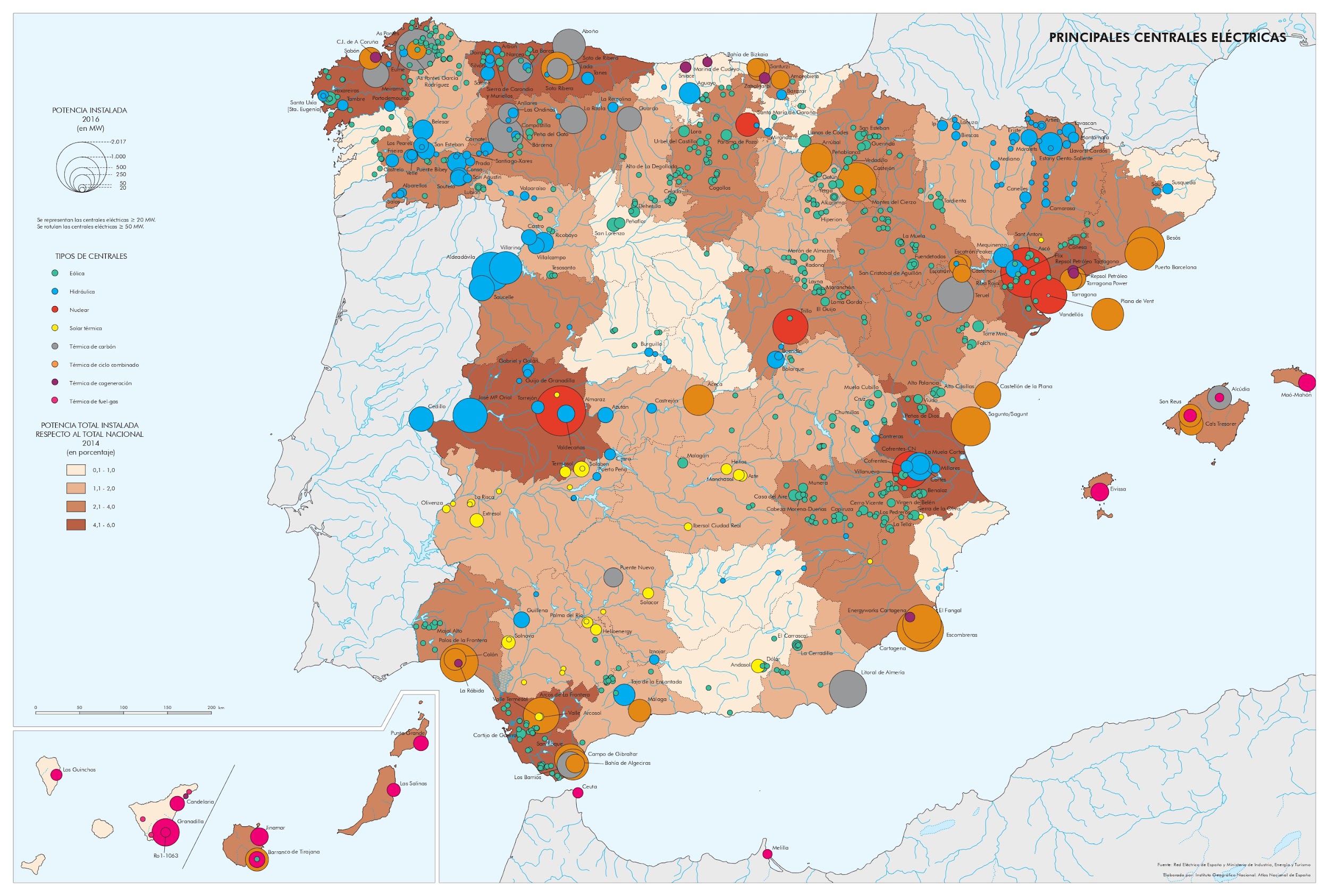
Se puede realizar un breve análisis descriptivo de las variables que contiene el primer dataset, detallando sus unidades de medida, y colocando una breve descripción con respecto al recurso energético que figure. Todas las medidas son para el país de España, a nivel nacional.

| 'time' | Es el tiempo medido desde el 31 de diciembre del 2014 hasta el 31 de diciembre del 2018, en escala diaria, y en intervalos de 1 hora. Siguiendo el estándar del CET (Central European Time). |
| --- | --- |
| 'generation biomass' | Es la generación eléctrica a partir del recurso renovable biomasa - Unidades en MW |
| 'generation fossil brown coal/lignite' | Es la generación eléctrica a partir del combustible fósil Lignito/carbón marrón - Unidades en MW |
| 'generation fossil coal-derived gas' | Es la generación eléctrica a partir del combustible fósil gas de carbón - Unidades en MW |
| 'generation fossil gas' | Es la generación eléctrica a partir del gas fósil natural - Unidades en MW |
| 'generation fossil hard coal' | Es la generación eléctrica a partir del carbón fósil duro o Antracita - Unidades en MW |
| 'generation fossil oil' | Es la generación eléctrica a partir de fueloil - Unidades en MW |
| 'generation fossil oil shale' | Es la generación eléctrica a partir de Esquistos bituminosos fósiles (Oil shale) - Unidades en MW |
| 'generation fossil peat' | Es la generación eléctrica a partir del material fósil Turba - Unidades en MW |
| 'generation geothermal' | Es la generación eléctrica renovable a partir de la energía geotérmica - Unidades en MW |
| 'generation hydro pumped storage aggregated' | Es la generación eléctrica renovable a partir del recurso hídrico, considerando un almacenamiento mediante un sistema de bombeo añadido - Unidades en MW |
| 'generation hydro pumped storage consumption' | Es la generación eléctrica renovable a partir del recurso hídrico, considerando un almacenamiento mediante un sistema de bombeo añadido - Unidades en MW |
| 'generation hydro run-of-river and poundage' | Es la generación eléctrica renovable a partir del recurso hídrico, considerando una central hidroeléctrica fluyente/de pasada - Unidades en MW |
| 'generation hydro water reservoir' | Es la generación eléctrica reanovable a partir del recurso hídrico, considerando un reservorio/embalse - Unidades en MW - |
| 'generation marine' | Es la generación eléctrica renovable a partir del recurso del mar - Unidades en MW - |
| 'generation nuclear' | Es la generación eléctrica a partir de la energía nuclear -Unidades en MW |
| 'generation other' | Es la generación eléctrica considerando otras fuentes energéticas, no consideradas en este listado de variables - Unidades en MW |
| 'generation other renewable' | Es la generación eléctrica renovable considerando otras fuentes energéticas, no consideradas en este listado de variables - Unidades en MW |
| 'generation solar' | Es la generación eléctrica renovable empleando la energía solar - Unidades en MW |
| 'generation waste' | Es la generación eléctrica a partir del aprovechamiento energético de los residuos - Unidades en MW |
| 'generation wind offshore' | Es la generación eléctrica renovable a partir del recurso eólico marino, u "offshore" - Unidades en MW |
| 'generation wind onshore' | Es la generación eléctrica renovable a partir del recurso eólico terrestre, u "onshore" - Unidades en MW |
| 'forecast solar day ahead' | Es la previsión del día siguiente al considerado, sobre la generación eléctrica renovable empleando la energía solar - Unidades en MW |
| 'forecast wind offshore day ahead' | Es la previsión del día siguiente al considerado, sobre la generación eléctrica renovable a partir del recurso eólico marino, u "offshore" - Unidades en MW |
| 'forecast wind onshore day ahead' | Es la previsión del día siguiente al considerado, sobre la generación eléctrica renovable a partir del recurso eólico terrestre, u "onshore" - Unidades en MW |
| 'total load forecast' | Es la previsión del día siguiente al considerado, sobre la demanda eléctrica que tiene el país - Unidades en MW |
| 'total load actual' | Es la demanda eléctrica que tiene el país - Unidades en MW |
| 'price day ahead' | Es la previsión del día siguiente al considerado, sobre el precio de la energía eléctrica que tiene el país - Unidades en €/MWh |
| 'price actual' | Es el precio de la energía eléctrica que tiene el país - Unidades en €/MWh |

*Tabla 2: Análisis descriptivo de las variables (features) contenidas en el primer dataset*

El siguiente [documento en formato .xlsx](https://docs.google.com/spreadsheets/d/1HJvQQjM-gGTbYf9eCyegBNbutm3htHRKweu1HQsPEmk/edit?usp=sharing) hace una revisión de algunos parámetros importantes, referidos a las 5 principales Comunidades Autónomas (CC.AA.) del país ibérico. Se busca que a la hora de tratar y analizar los datos en la notebook ya se cuente con una base de información estadística sobre parámetros como población, infraestructura, generación eléctrica y consumo eléctrico.

Un ejemplo de la información recolectada se presenta en la siguiente imagen:



*Figura 4: Información complementaria sobre las comunidades autónomas españolas, con respecto al mercado eléctrico.*

# ⚙️ Data Wrangling

Las etapas realizadas sobre el Data Wrangling fueron las siguientes:

| **Etapas del Data Wrangling** | |
| --- | --- |
| **Etapa aplicada a la data** | **Descripción** |
| Descubrimiento | "Comprensión de los datos, su estructura, tipología y cantidad. También lo es conocer por qué una compañía los utiliza y cómo. Ésto sirve para tomar decisiones posteriores con un rumbo claro". |
| Estructuración | "Estandarizar el formato de los datos, esto tiene dependencia directa en diversas fuentes u orígenes: así los datos estarán en diferentes formatos y estructuras". |
| Limpieza | "Eliminación de los datos que no brinden información extra como los duplicados, datos faltantes, etc. Este paso logrará estandarizar el formato de las columnas (float, datatimes, etc)." |
| Enriquecimiento | "Etapa referida a agregación de datos extra (de fuentes externas) que complementen a los ya existentes, para agregar información extra al análisis. En algunos casos se habla de la creación de variables resumen, en forma adicional". |
| Validación | “Es muy importante para los equipos, asegurarse que los datos son precisos y que la información no se alteró durante el proceso. Esto significa asegurar la fiabilidad, credibilidad y calidad de los datos limpios debido a que van a utilizarse para tomar decisiones.“ |
| Publicación | “Una vez que los datos están validados, se pueden compartir para su uso, realizar análisis exploratorios, entrenar modelos y tomar decisiones. Se entiende como un producto final que se entrega para ser usado.” |

*Tabla 3: Etapas del Data Wrangling*

Para más información, consultar el [desafío entregable](https://docs.google.com/document/d/1Y31aSFICOLzXZ9wg_a__UCHI8ciB9HSqsHsOVz1adOk/edit?usp=sharing) sobre Data Wrangling.

# 📊 Exploratory Data Analysis (EDA)

Para llevar a cabo el EDA, se plantean las hipótesis para todos los interrogantes propuestos, luego, se resuelven en la [notebook (archivo .ipynb)](https://colab.research.google.com/drive/1yr1xNR986l1mvEOrFUz_Ns7rx1mBr4ep?usp=drive_link).

# 🧩 Feature Selection (FS)

Se consulta en bibliografía recomendada sobre algunas técnicas para reducir la dimensionalidad del dataset.

Debido a las siguientes razones, es crucial aplicar estas técnicas antes de seguir a la parte de construcción y alimentación del modelo de ML:

* Entrenamiento de algoritmos a mayores velocidades.
* Reducción en la complejidad del modelo, y hacer más fácil su interpretación.
* Mejora en la precisión del modelo.
* Búsqueda en la mejora con respecto al fenómeno del *overfitting*, donde el modelo memoriza el conjunto de entrenamiento en lugar de capturar las relaciones subyacentes en los datos.

Se aplicarán, inicialmente, los métodos/técnicas de Filtro (*Filter* *Methods*). Esto se debe, fundamentalmente, a que se buscará una mayor rapidez y ausencia en materia de entrenamiento de modelos, con respecto al uso de otros métodos, como por ejemplo los Envolventes (*Wrapper Methods*) y los Integrados (*Embedded Methods*).

Como en muchos apartados de este proyecto, se buscará adentrarse en el resto de las técnicas que ofrecen los demás métodos, tanto los Envolventes como los Integrados.

Las técnicas de filtro empleadas son las siguientes:

* Fisher’s Score.
* Coeficientes de correlación.
* Umbral de varianza.
* SelectKBest.

# 📋 Algoritmos de Machine Learning (ML) empleados

Se ha hecho una revisión bibliográfica de los modelos para problemas de regresión vistos en la cursada, de menor complejidad y más empleados, estos son:

* Regresión lineal.
* Regresión KNN.
* Regresión polinómica no lineal - Grado 2.
* Regresión polinómica no lineal - Grado 3.
* Promedio Móvil Simple (SMA).
* Promedio Móvil Exponencial (EMA).
* Árboles de decisión.

Dentro de la carpeta compartida, el ya citado [documento en formato .xlsx](https://docs.google.com/spreadsheets/d/1HJvQQjM-gGTbYf9eCyegBNbutm3htHRKweu1HQsPEmk/edit#gid=1644567813), contiene información donde se visualizan las diferentes evaluaciones de los algoritmos vistos, guiadas por las métricas empleadas y contemplando algunos factores tenidos en cuenta.

Por ejemplo, entre las tablas N°1 y N°3, se comparan las métricas de los modelos antes y después de aplicar un proceso de FS.

# 💶 Conclusiones

* Se hizo una revisión bibliográfica y puesta en práctica de los modelos - para problemas de regresión - vistos en la cursada.
* Modelos como Árboles de decisión y KNN presentan un mejor rendimiento referido sólamente a las métricas evaluadas (MSE, MAE y R cuadrado). De forma general, y a grandes rasgos, no hay problemas excesivos de overfitting en los resultados de las métricas.
* Los modelos que siguen medias móviles -SMA o EMA- no dependen de las features seleccionadas y fijan su predicción según los valores de la target.
* Sólo el modelo KNN presentó una mayor variabilidad en el resultado de las métricas, al no estandarizar las variables y al aplicar el proceso de FS, siendo un dato a tener en cuenta. El resto de las variables se comportó, relativamente, de igual forma.
* Sin embargo, parece ser que el desempeño de los modelos (de forma global) performa negativamente, aplicando el proceso de feature selection mediante métodos de filtro.
* Por lo tanto, se motiva a seguir investigando y ahondar en otros métodos para la toma de decisión más eficiente en materia de selección de variables.

Por ejemplo, se debería adentrar en la puesta en marcha de un modelo ARIMA, pero requiere de un análisis previo con respecto a la serie temporal utilizada.