

Part I: selecció del conjunt de dades

Carlos romero matarin | Visualització de dades -UOC

1.[10%] Justifiqueu breument la vostra selecció, sigui per motius personals o professionals.

He triat aquest conjunt de dades per les meves motivacions personals i professionals, que van sorgir de la meua investigació prèvia en el grau d'economia, en l'àmbit de l'energia durant el meu treball final de grau (TFG, <https://ddd.uab.cat/record/283044>). Com podem observar la volatilitat del mercat elèctric i el seu impacte significatiu a l'economia global i les polítiques de sostenibilitat està a l'orde del dia des de la guerra d'Ucraïna i l'Europa afecta per les energies fòssils.

Ham aquest conjunt tinc com a objectiu ampliar la meua comprensió del dinàmic mercat energètic en explorar com interactuen i afecten els preus de l'electricitat els factors externs com el clima i el tipus de generació, etc. Aquest enfocament està en línia amb la meua trajectòria acadèmica i el meu compromís amb la creació d'estratègies energètiques.

A més, en la meua feina actual al sector de les telecomunicacions, observo directament com les variacions en els preus de l'electricitat afecten els costos operatius, cosa que destaca la importància d'anticipar i administrar aquestes variacions. L'elecció d'aquest conjunt de dades com una oportunitat per aplicar i expandir el meu coneixement en contextos pràctics i de gran rellevància es recolza en l'enllaç entre la formació acadèmica i la meua experiència professional.

2.[10%] La rellevància del conjunt de dades en el context. Són dades actuals? Tracten un tema important per algun col·lectiu concret? S'ha tingut en compte la perspectiva de gènere?

Crec que aquest conjunt de dades és extremadament rellevant per al context actual per diverses raons importants. Per començar, les dades són actuals (dels anys passats: 2021, 2022 i 2023, el 2024 com no està complet no he volgut afegir-lo, però es pot afegir ràpidament) s'inclouen dades de preus de l'electricitat, la generació i informació meteorològica detallada. Això facilita una anàlisi en temps real de les tendències i proporciona una base sòlida per preveure futurs canvis i adaptacions necessàries al sector energètic.

Podríem dir que el valor d'aquest conjunt de dades transcendeix grups específics, podem veure que les dades són fonamentals per a aquelles persones encarregades d'elaborar polítiques, per a les companyies del sector energètic que han d'adaptar les seves capacitats de producció i preus per adaptar-se a les variacions al mercat i la necessitat. A més, crec el que els que més poden utilitzar les dades són els investigadors i acadèmics que estudien els efectes del canvi climàtic a l'economia energètica, aquest conjunt de dades és de gran interès.

3.[25%] La complexitat (mesura, variables disponibles, tipus de dades, etc.). Heu de tenir de l'ordre de milers de registres mínim. I ha de tenir un mínim de l'ordre de desenes de variables. Combina dades categòriques i quantitatives? Inclou altres tipus de dades? Evita els conjunts excessivament simples.

Crec que la complexitat d'aquest conjunt de dades el fa ideal per a una anàlisi exhaustiva i detallada de les dinàmiques del mercat energètic i dels impactes climàtics. Aquí enumero 2 dels punts que crec que es pot destacar més:

- Volum de dades: El conjunt actual conté 1.095 registres, cadascú dels registres representa les dades consolidades d'un dia. Cal dir que a principi, les dades de preus estaven disponibles en intervals horaris, però per reduir el conjunt i tindre dades diàries el que he realitzat és fer una mitjana diària de mesuraments, això de cara a optimitzar la gestió de la gran quantitat de dades. Aquesta transformació manté l'essència i la utilitat de les dades per a l'anàlisi temporal i de tendència, cosa que permet una gestió més efectiva del volum de dades sense comprometre la capacitat d'obtenir insights significatius.
- Variables: El conjunt de dades que tractaré conté 27 variables diferents, com dates meteorològiques, dades diversos tipus de generació d'energia i dades dels preus. Podem veure la presència de variables tan categòriques (com l'any, el més i el dia) com les quantitatives (com la temperatura mitjana, el preu de l'electricitat i la generació per tipus d'energia), el que ens pot facilitar l'ús de diversos mètodes estadístics i de modelització predictiva com també la visualització en diferents tipus d'infografies.

En el meu TFG destaco el següent de les variables més important:

- *Any, més i dia: són variables quantitatives discretes que indiquen la data de cada observació. Aquestes variables són rellevants perquè ens permeten estudiar l'estacionalitat i la tendència del preu de l'electricitat. A més, són variables disponibles al portal de dades obertes de l'operador del mercat elèctric (OMIE).*
- *tmed, tmin i tmax: són variables quantitatives contínues que mesuren la temperatura mitjana, mínima i màxima en graus Celsius. Aquestes variables són rellevants perquè afecten la demanda i la generació d'energia, especialment la solar i l'eòlica. A més, són variables disponibles a 17 el portal de dades obertes de l'Agència Estatal de Meteorologia (AEMET).*
- *prec: és una variable quantitativa contínua que mesura la precipitació a*

mil·límetres. Aquesta variable és rellevant perquè influeix en la generació hidroelèctrica, que és una de les fonts més barates i flexibles. També és una variable disponible al portal de dades obertes de l'AEMET.

- *Diferents tipus de generació d'energia: són variables quantitatives contínues que mesuren la generació elèctrica en GWh de cada font d'energia (eòlica, solar, hidràulica, etc.). Aquestes variables són rellevants perquè cada font té un cost i una capacitat diferents d'adaptar-se a les variacions de la demanda. Són variables disponibles al portal de dades obertes de REE.*
- *Generació total: és una variable quantitativa contínua que mesura la generació elèctrica total a GWh. Aquesta variable és rellevant perquè ha de ser igual a la demanda més les pèrdues en el transport i distribució. És una variable disponible al portal de dades obertes de REE.*
- *Preu: és una variable quantitativa contínua que mesura el preu mitjà diari de l'electricitat en euros. Aquesta variable és la dependent o d'estudi, ja que volem analitzar com influeixen les altres variables en comportament. És una variable disponible al portal de dades obert d'OMIE.*

TFG, Hacia un Futuro Energético Predecible: Aplicando Machine Learning para Prever los Precios de la Electricidad, pag. 16-17

4. [25%] L'originalitat. Es valora no repetir els conjunts de dades clàssiques o molt treballades [Links to an external site](#). Ni temes ja molt tractats (p. ex. Covid-19, trànsit, criminalitat...) Podeu combinar o millorar el conjunt de dades. En el primer cas, enriquir el conjunt de dades amb altres de diferents per donar un enfocament nou. En el segon cas, generant noves mètriques o indicadors amb les variables existents mitjançant transformacions. Hi ha altres visualitzacions basades en aquest conjunt de dades? És una evolució o una actualització d'un conjunt anterior? Heu enriquit un conjunt de dades ja existent?

El conjunt de dades que he estret es destaca per la seva originalitat, que prové tant de la naturalesa de les fonts com de la metodologia utilitzada per a la seva preparació.

Fonts de dades diverses i pertinents

Tres fonts principals són essencials per obtenir una visió completa i precisa del mercat energètic:

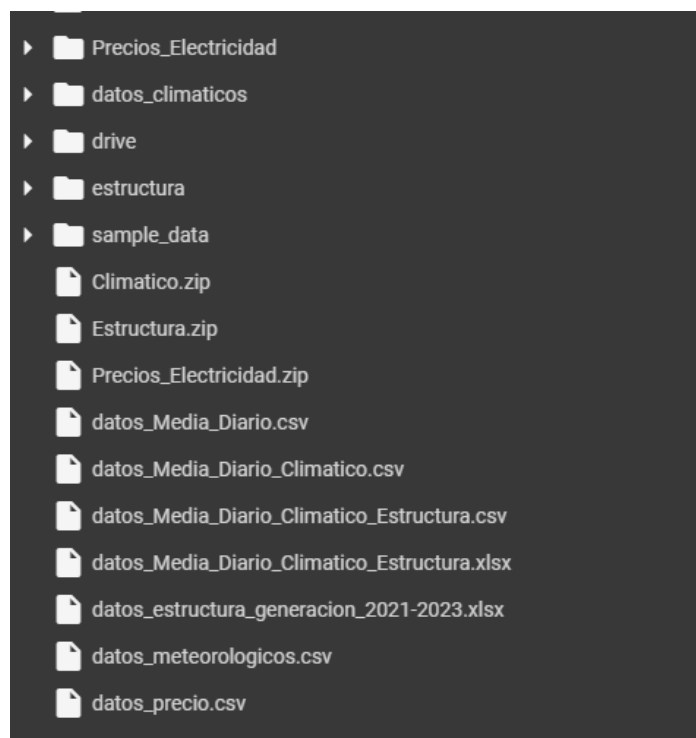
- OMIE (Operador del Mercat Ibèric d'Electricitat): Proporciona informació sobre els preus horaris del mercat diari i infradiari. Aquestes dades són essencials per analitzar les fluctuacions del mercat i comprendre com els preus canvien amb els canvis a l'oferta i la demanda.
- L'Agència Estatal de Meteorologia (AEMET): Proporciona informació detallada sobre les condicions meteorològiques, com temperatura, humitat, velocitat del vent, radiació solar, etc. Aquests elements són essencials per estudiar com el clima afecta el consum i la generació d'energia.
- Xarxa Elèctrica d'Espanya (REE): Proporciona informació sobre la producció d'energia organitzada per tipus de tecnologia (hidràulica, eòlica, solar, tèrmica...). Aquestes dades són essencials per avaluar com la barreja de generació afecta el mercat i per investigar l'impacte de les energies renovables en l'estabilitat i la sostenibilitat del sistema energètic.

Preprocessament i Enriquiment de Dades

Per tindre un bon conjunt de dades he realitzat un codi en Python utilitzant Google Colab (enllaç:

<https://colab.research.google.com/drive/1PRw6T5mnsrdtnI2SouzjfvmVpQ9NniR?usp=sharing>) que mostra detalladament la manera de preparar dades, alguns aspectes importants són:

- Estructuració i Organització de Dades: Creació de directoris específics per a cada tipus de dada (preus, clima i estructura) facilita un maneig sistemàtic i organitzat. Garanteix que tots els conjunts de dades estiguin separats i fàcilment accessibles.



- Automatització: L'ús de descàrregues automàticament dades de fonts fiables com OMIE i AEMET optimitza el procés d'obtenció de dades i permet un flux de treball més eficient i fiable.
- Neteja i Preparació de Dades: El procés de neteja realitat ha fet que les dades estiguin preparades per a una anàlisi, alguns dels mètodes són eliminar files i columnes innecessàries, normalitzar els formats i corregir valors atípics. També és molt important la conversió de dades de diversos fitxers en un format unificat i unificació per la clau Dia, mes i any.
- Innovació en la integració de dades: Gràcies a la combinació de dades de generació d'energia, preus i meteorologia en una sola taula de dades permet fer anàlisis complexes que exploren les interaccions entre el clima, la demanda d'energia i els preus del mercat.
- Desenvolupament de Noves Mètriques: Mètriques com a mitjanes de preus diaris (realitzat des de les dades originals de preus diaris i horaris) i les anàlisis per grups (d'energia, generació total, fonts renovables, ...), que són bàsics per a la visualització i el modelatge predictiu en estudis.

5. [30%] Les qüestions que respondreu amb la visualització de dades, tenen en compte els punts anteriors? Han estat plantejades en altres visualitzacions o en altres projectes? Són adequades per al conjunt de dades escollit? En aquest punt, elaboreu un diccionari de les variables, el seu significat i si és un fet a estudiar o una dimensió que el mesura, us pot ajudar.

Totes les consideracions en les preguntes prèvies es prenen en compte en la visualització de dades de la pràctica. Les preguntes específiques que se cerquen respostes són:

- **Com afecten les condicions meteorològiques els preus de l'electricitat?**
Aquesta pregunta té en compte les dades meteorològiques (temperatura, precipitació i altres proporcionades per AEMET) com les dades de preus. L'objectiu és examinar com els canvis climàtics afecten els preus del mercat.
- **Hi ha una correlació entre el tipus de generació d'energia i el preu de l'electricitat?**
És objectiu descobrir com la diversificació de fonts de generació d'energia afecta la variabilitat dels preus fent servir les dades de generació proporcionades per REE per les diferents tecnologies (hidràulica, eòlica, solar, etc.).
- **És possible reconèixer patrons temporals (diaris, mensuals, anuals) en l'ús i la producció d'energia i similitud dels preus?**
L'objectiu és identificar tendències estacionals o anomalies la producció d'energia a través de l'anàlisi de dades diàries, com també similitud de preus estacionals.

Metodologia i aplicació del conjunt de dades

Com he indicat el conjunt de dades és molt apropiat per a aquestes preguntes perquè conté informació àmplia i detallada que permet una anàlisi completa dels factors que afecten el mercat energètic, indicar que:

- Cal indicar que per facilitar-ne l'anàlisi i millorar la visualització de tendències a llarg termini, s'han afegit dades de preus en mitjanes diàries i s'han afegit camps com dia, mes, any.
- Les dades de generació i meteorologia brinden una imatge completa de com els factors externs i les decisions de producció afecten els preus.

Comparació amb altres estudis

En aquest apartat, es presenta una revisió de la literatura relacionada amb el mercat energètic i preus, amb èmfasi en l'anàlisi de preus del mercat elèctric. Alguns estudis que han estudiat les variables són:

- Com article de referència podem veure "Electricity price forecasting: A review of the state-of-the-art with a look into the future. International Journal of Forecasting," de Rafal Weron. En aquest article l'autor crea un marc a la predicció de preus de l'electricitat utilitzant diferents tècniques d'aprenentatge profund i de machine learning.
- Un altre document pioner i important en el camp de la predicció de preus és el llibre "Price forecasting using an integrated approach" de Rafal Weron. On se'ns exposa diferents temes i aspectes importants del mercat i fa diferents explicacions de les variables.

Innovació a la visualització

Algunes de les visualitzacions que tinc pensades incloure en la pràctica són:

- Mapes de calor: Aplicant mapes amb funcions interactives per explorar diferents anys o períodes, mostraran la variabilitat del preu i la generació segons la temporada.
- Gràfics en línia: Per examinar tendències, amb opcions de zoom i enfocament per detallar.
- Els diagrames de dispersió: Que relacionen directament els canvis en els preus amb les variables meteorològiques i inclouen línies de tendència per mostrar-ne la correlació.
- Els gràfics de barres empilades: Que ens mostren la composició de la generació d'energia per diferents fonts al llarg del temps i ofereixen interactivitat per triar fonts particulars.
- Visualització de flux de dades: Per poder analitzar el moviment d'energia entre diverses fonts de generació, mostrant dinàmiques en temps real o simulades.

Diccionari de les variables

Aquí també deixo el glossari de totes les variables:

Variable	Significat	Unitat	Tipus de Dada	Classificació
Any	Any en què es van registrar les dades	Any	Discreta	Dimensió
Mes	Mes de l'any de les dades registrades	Mes	Discreta	Dimensió
Dia	Dia del mes corresponent a cada registre	Dia	Discreta	Dimensió
Data	Data completa de cada registre en format YYYY-MM-DD	Data	Discreta	Dimensió
tmed (Temperatura Mitjana)	Temperatura mitjana diària en graus Celsius	°C	Contínua	Variable de Estudi
prec (Precipitació)	Quantitat de precipitació registrada en mil·límetres	mm	Contínua	Variable de Estudi
tmin (Temperatura Mínima)	Temperatura mínima aconseguida cada dia en graus Celsius	°C	Contínua	Variable de Estudi
tmax (Temperatura Màxima)	Temperatura màxima registrada cada dia	°C	Contínua	Variable de Estudi
Hidràulica	Energia generada per fonts hidràuliques	GWh	Contínua	Variable de Estudi
Turbinació bombament	Energia generada per turbinació de bombament	GWh	Contínua	Variable de Estudi
Nuclear	Energia produïda per plantes nuclears	GWh	Contínua	Variable de Estudi
Carbó	Energia generada usant carbó	GWh	Contínua	Variable de Estudi
Fuel + Gas	Energia generada per combustió de fuel i gas	GWh	Contínua	Variable de Estudi
Motors dièsel	Energia generada per motors dièsel	GWh	Contínua	Variable de Estudi
Turbina de gas	Producció energètica mitjançant turbines de gas	GWh	Contínua	Variable de Estudi
Turbina de vapor	Producció energètica mitjançant turbines de vapor	GWh	Contínua	Variable de Estudi
Cicle combinat	Energia generada per plantes de cicle combinat	GWh	Contínua	Variable de Estudi
Hidroeòlica	Energia generada combinant recursos hídrics i eòlics	GWh	Contínua	Variable de Estudi
Eòlica	Energia produïda per turbines eòliques	GWh	Contínua	Variable de Estudi

Solar fotovoltaica	Energia generada per panells solars fotovoltaics	GWh	Contínua	Variable de Estudi
Solar tèrmica	Producció energètica mitjançant tecnologia solar tèrmica	GWh	Contínua	Variable de Estudi
Altres renovables	Energia produïda per altres fonts renovables	GWh	Contínua	Variable de Estudi
Cogeneració	Energia generada per sistemes de cogeneració	GWh	Contínua	Variable de Estudi
Residus no renovables	Energia produïda a partir de residus no renovables	GWh	Contínua	Variable de Estudi
Residus renovables	Energia generada a partir de residus renovables	GWh	Contínua	Variable de Estudi
Generació total	Suma total d'energia generada	GWh	Contínua	Variable de Estudi
Preu	Preu mitjana diària de l'electricitat en euros per MWh	€/MWh	Contínua	Variable Principal de Estudi