



Pràctica 1: Web Scraping

INTEGRANTS GRUP: Àngels Calvet i Mirabent i Albert Estadella Valls

EXERCICI 1: Context

Explicar en quin context s'ha recollit la informació. Explicar per què el lloc web triat proporciona aquesta informació.

Des de les últimes dècades s'han pogut percebre canvis ambientals relacionats amb el canvi climàtic com per exemple, temperatures extremes, precipitacions més intenses, creixement de zones desertificades, escurçament de les estacions intermitges, augment de la temperatura del mar, desaparició progressiva dels glacials entre d'altres. Tots aquests canvis poden tenir un gran impacte en l'agricultura, la nostra forma de relacionar-nos amb el planeta, els quals es poden estudiar.

Per tal d'estudiar aquests canvis, primer és necessari recopilar dades ambientals mesurades en diferents estacions meteorològiques. En el nostre cas, estem interessats en realitzar aquest estudi en el territori de Catalunya, per tant, les dades seran extretes de la pàgina web del Servei Meteorològic de Catalunya, també conegut com 'meteocat', la qual proporciona dades diàries capturades desde estacions meteorològiques distribuïdes a tot el territori català.

EXERCICI 2: Títol

Definir un títol que sigui descriptiu pel dataset.

Dades ambientals de l'Alt Penedès recollides entre el 2017 i 2021.

EXERCICI 3: Descripció del dataset

Desenvolupar una descripció breu del conjunt de dades que s'ha extret. És necessari que aquesta descripció tingui sentit amb el títol escollit.

El dataset generat conté diferents variables meteorològiques registrades diàriament per les estacions situades al Alt Penedès (Sant Sadurní d'Anoia, Sant Martí Sarroca, Canaletes, la Granada i Font Rubí) entre els anys 2017 i 2021. Entre aquestes variables representades, trobem per exemple la temperatura mitjana, la temperatura mínima i màxima juntament amb l'hora del dia en que es va assolir, la humitat i la irradiació solar global.

No totes les estacions estan preparades per recollir la mateixa informació, és a dir, no totes estan equipades amb els mateixos instruments. I aquests instruments, no sempre es troben operatius, poden deixar de recolectar dades per avaries puntuals. Per tant, en el dataset, es poden donar casos de registres amb no totes les seves variables informades, per exemple Font Rubí proporciona dades sobre la pressió atmosfèrica mitjana, en canvi la resta no.

Tenir en compte també que no totes les estacions han estat instal·lades en les mateixes dades, per tant, es pot donar el cas de no disposar de dades per a tots els dies que comprenen el període de busqueda desitjat per a una o més estacions meteorològiques. En concret, en el dataset generat, per a l'estació meteorològica situada a Sant Sadurní d'Anoia, únicament es disposen de dades a partir del 07/05/2021, data en la qual l'estació va ser operativa.



EXERCICI 4: Representació gràfica

Dibuixar un esquema o diagrama que identifiqui el dataset visualment i el projecte escollit.

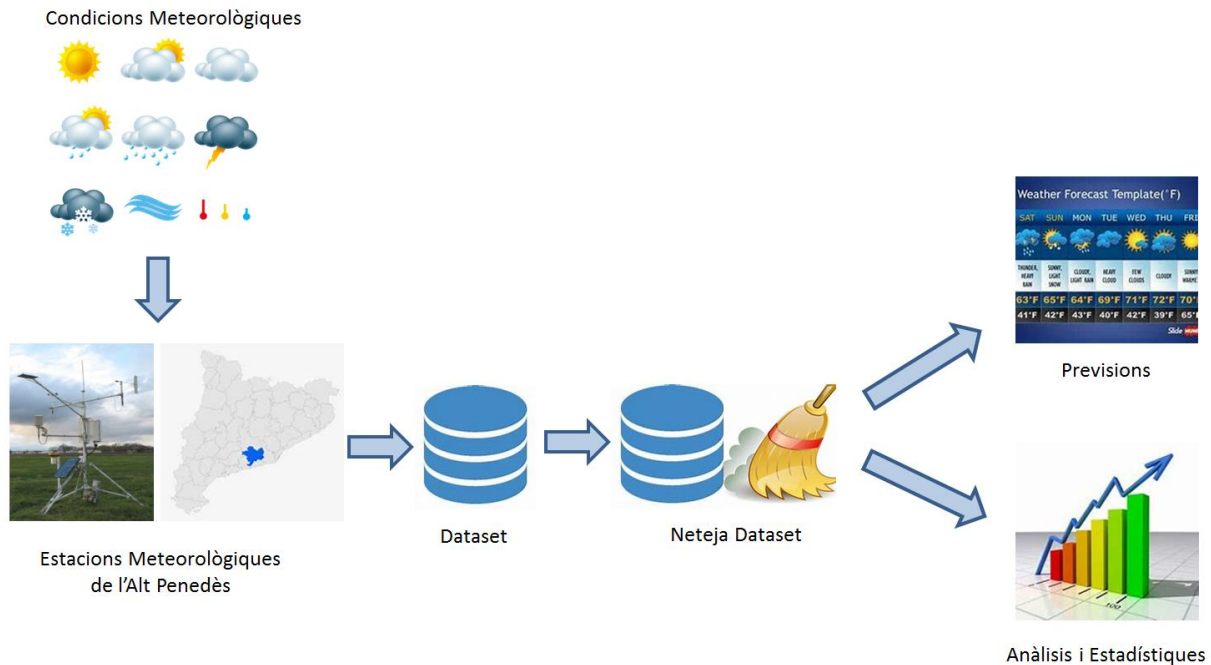


Figura 1: Esquema obtenció i ús de les dades.

Nota: Imatges extretes del cercador 'bing' filtrant per llicència 'Llibre para compartir y usar'.

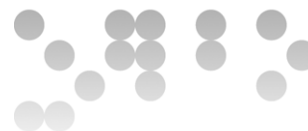
EXERCICI 5: Contingut

Explicar els camps que inclou el dataset, el període de temps de les dades i com s'ha recollit.

El dataset inclou les dades recollides entre l'1 de gener del 2017 i el 31 de desembre de 2021 de totes les estacions existents de l'Alt Penedès. Aquestes estacions es troben situades a Sant Sadurn d'Anoia, Sant Martí Sarroca, Canaletes, la Granada i Font Rubí.

Es recullen les següents característiques en el dataset:

- **Estació meteorològica:** Nom de l'estació meteorològica (coincideix amb la localització).
- **Dia:** dia en el qual s'han recollit les dades en format aaaa.mm.dd.
- **Temperatura mitjana:** temperatura mitjana registrada durant el dia.
- **Temperatura màxima:** temperatura màxima registrada durant el dia.
- **HORA d'obtenció de Temperatura màxima:** hora en la qual s'ha registrat la temperatura màxima en format hh:mm TU. On TU significa Temps Universal, per tant, cal sumar una hora en horari d'hivern i dues en horari d'estiu per passar a l'hora oficial.
- **Temperatura mínima:** temperatura mínima registrada durant el dia.



- **HORA d'obtenció de Temperatura mínima:** hora en la qual s'ha registrat la temperatura mínima en format hh:mm TU. On TU significa Temps Universal, per tant, cal sumar una hora en horari d'hivern i dues en horari d'estiu per passar a l'hora oficial.
- **Humitat relativa mitjana:** humitat relativa mitjana en % registrada durant el dia.
- **Precipitació acumulada:** quantitat d'aigua precipitada en mil·límetres. Per tant, equival a la quantitat d'aigua caiguda en litres per metre quadrat.
- **Ratxa màxima del vent:** velocitat màxima del vent en km/h registrada durant el dia i direcció en graus en la qual aquesta es va produir.
- **HORA d'obtenció de Ratxa màxima del vent:** hora en la qual s'ha registrat la màxima velocitat del vent en format hh:mm TU. On TU significa Temps Universal, per tant, cal sumar una hora en horari d'hivern i dues en horari d'estiu per passar a l'hora oficial.
- **Irradiació solar global:** potència per unitat de superfície (MJ/m^2) rebuda del sol en forma de radiació electromagnètica durant el període de temps d'un dia.
- **Pressió atmosfèrica mitjana:** registre referent al càlcul de la pressió atmosfèrica mitjana obtinguda durant el dia en hPa.

Les dades recopilades en el dataset contenen tant el valor numèric com la seva unitat, és a dir, són alfanumeriques. Per tant, abans de ser utilitzades han de ser preprocessades (separar els valors numèrics de les seves unitats). Destacar que en el cas d'estudi, tots els registres es troben en les mateixes unitats, excepte la ratxa màxima del vent el qual la direcció en graus pot variar.

EXERCICI 6: Agraïments

Presentar el propietari del conjunt de dades. És necessari incloure cites d'anàlisis anteriors o, en cas de no haver-n'hi, justificar aquesta cerca amb anàlisis similars. Justificar quins passos s'han seguit per actuar d'acord amb els principis ètics i legals en el context del projecte.

Les dades han estat obtingudes de la web '<https://www.meteo.cat/>' que pertany del Servei Meteorològic de Catalunya.

Navegant per la direcció web '<https://www.meteo.cat/wpweb/divulgacio/publicacions/>' podem trobar el recull de publicacions realitzades amb les seves pròpies dades. Entre aquestes publicacions, destaquen algunes relacionades amb canvis ambientals com:

- 10 anys del torrid estiu del 2003 (13/08/2013)
'https://www.meteo.cat/wpweb/divulgacio/publicacions/efemerides/2003-06-01_10-anys-del-torrid-estiu-del-2003/'
- 30 anys dels catastròfics aiguats de novembre de 1983 (08/11/2013)
'https://www.meteo.cat/wpweb/divulgacio/publicacions/efemerides/1983-11-07_30-anys-dels-catastrofics-aiguats-de-novembre-de-1983/'
- La normalitat de batre rècords (24/05/2016)
'<https://www.meteo.cat/wpweb/divulgacio/publicacions/miscellania/normalitat-batre-records/>'
- Principals trets climàtics del 27 de setembre (10/09/2015)
'<https://www.meteo.cat/wpweb/divulgacio/publicacions/miscellania/principals-trets-climatic-dels-27-de-setembre/>'



No obstant això, es tracta d'anàlisis que no es centren ni en la zona de l'Alt Penedès ni en els últims anys. Per tant, hem realitzat un petit *state-of-the-art* d'estudis sobre els impactes ambientals en cultius com les vinyes:

- Impacto del cambio climático en el viñedo en España
Resco Sánchez, Pablo (2012). *'Impacto del cambio climático en el viñedo en España'* [Treball Final de Màster] Universidad Politécnica de Madrid. https://ceigram.upm.es/wp-content/uploads/2020/07/TFM_Pablo-Resco.pdf
- Estrategias de adaptación al cambio climático en el viñedo de la cuenca mediterránea: El caso de la Rioja
Teodoro Lasanta, Carlos Baroja-Sáenz, Melani Cortijos-López, Estela Nadal-Romero, Ignacio Martín, Enrique García-Escudero (2021). *'Estrategias de adaptación al cambio climático en el viñedo de la cuenca mediterránea: El caso de la Rioja'* Cuadernos de Investigación Geografica. <https://publicaciones.unirioja.es/ojs/index.php/cig/article/view/5062>

Tot i això, destacar com els estudis tenen un abast diferent al que podem obtenir a partir de les dades del nostre dataset.

Revisant els aspectes legals que podem trobar a la web (<https://www.meteo.cat/wpweb/avis-legal/#int>), podem concloure que les dades mentre no siguin modificades, desnaturalitzades, es citi la font i la última actualització, aquestes poden ser utilitzades. Tot i això, en el punt 6 de l'apartat 'Propietat intel·lectual i industrial de Meteocat' no queda clara aquesta autorització d'ús de dades. Per resoldre aquest dubte, s'ha seguit inspeccionant la web i s'ha pogut observar que permeten la descarrega de les dades per API's per estudiants, centres de recerca i/o administracions (<https://apidocs.meteocat.gencat.cat/documentacio/acces-ciutada-i-administracio/>).

A més a més, es poden extreure les mateixes dades des de la web de transparència de Catalunya en format CSV o JSON, a la qual podem accedir a través de web del meteocat i et proporciona una llicència d'ús d'aquestes dades.

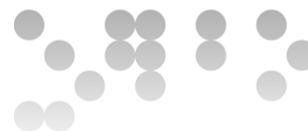
EXERCICI 7: Inspiració

Explicar per què és interessant aquest conjunt de dades i quines preguntes es pretenen respondre. És necessari comparar amb les anàlisis anteriors presentades a l'apartat 6.

Com s'ha comentat en apartats anteriors, als canvis ambientals generen un impacte sobre la nostra vida, un exemple clar seria els efectes que té sobre l'agricultura. Per exemple, un increment de pluges torrencials o pedregades destrossen collites, canvis de temperatura poden fer variar processos de maduració de fruits i per tant, s'han d'adaptar les èpoques de collita. En considerar que aquest canvis aniran en augment, es genera una motivació per estudiar la tendència ambiental dels últims anys amb l'objectiu d'intentar predir cap a on aniran els canvis i així estar més previnguts en les mesures d'adaptació.

Sabem que un dels cultius afectats és la vinya. Per aquests, les temperatures extremes (valors propers o per sota 0° i superiors als 40°) i l'increment de les pluges torrencials poden arribar a suposar la pèrdua de la collita. Per aquest motiu, es volen estudiar les variacions climatològiques que es produeixen a la zona de l'Alt Penedès, una de les zones més representatives del sector vinícola a Catalunya. Algunes de les preguntes a les que es podran donar resposta són:

- Evolució de les temperatures en l'època de creixement de la vinya (temperatures màximes i mínimes).
- Anàlisi de les precipitacions en la zona per detectar possibles necessitat d'adaptacions en el reg i/o mesures per minimitzar l'impacte de les pluges torrencials.



En comparació amb les anàlisis presentades en l'apartat 6, destacar que els anàlisis que parteixen de la mateixa font de dades a la presentada en el dataset de la pràctica, no fan referència específicament a la zona de l'Alt Penedès ni busquen relacionar la variació dels valors ambientals amb els efectes a la vinya.

Respecte els anàlisis més centrats en l'efecte del canvi climàtic sobre la vinya, observem com les dades provenen de diferents fonts a les presentades en el nostre dataset i com l'estudi 'Estrategias de adaptación al cambio climático en el viñedo de la cuenca mediterránea: El caso de la Rioja' està més orientat en donar a conèixer les conseqüències i les possibles solucions i no pas en la detecció de canvis en el temps per poder plantejar accions que minimitzin les conseqüències negatives del canvi climàtic.

EXERCICI 8: Llicència

Seleccionar una d'aquestes llicències pel dataset resultant i justificar el motiu de la seva selecció:

- Released Under CC0: Public Domain License
- Released Under CC BY-NC-SA 4.0 License
- Released Under CC BY-SA 4.0 License
- Database released under Open Database License, individual contents under Database Contents License
- Other (specified above)
- Unknown License

La llicència per al codi i el data set serà CC BY-NC-ND, d'acord amb el termes establerts a Creative Commons.

Els motius que porten a escollir aquesta llicència són els següents:

- Les dades del dataset poden ser copiades i/o compartides ja que provenen de dades públiques fàcilment visualitzables a la web del meteoat i fàcilment descarregables des de la web de transparència.
- Es decideix no permetre usos comercial ja que des de la font de dades original, s'especifica que les seves dades no poden fer-se servir per usos comercials.
- No es permet modificar les dades, per tant tampoc distribuir modificacions d'aquestes, donat que aquest és un dels requisits de la web d'on s'han extret les dades per generar el dataset.

EXERCICI 9: Codi

Adjuntar al repositori Git el codi amb el qual s'ha generat el dataset, preferiblement en Python o, alternativament, en R.

El codi es pot trobar penjat al repositori Git a través de l'enllaç:

- <https://github.com/acalvet0/meteoAltPenedes>

Tenir en compte que aquest codi està pensat per recuperar les dades desitjades per l'usuari en funció de les estacions meteorològiques i el període de temps. Per generar el mateix dataset que es pot trobar en format CSV a Zenodo, s'han d'utilitzar els següents paràmetres:

- Estacions: Sant Sadurní d'Anoia, Sant Martí Sarroca, Canaletes, la Granada i Font Rubí.
- Data inicial: 01.01.2017
- Data final: 31.12.2021



EXERCICI 10: Dataset

Publicar el dataset obtingut() en format CSV a Zenodo amb una breu descripció. Obtenir i adjuntar l'enllaç del DOI.*

El dataset el podem trobar en format CSV a Zenodo. Podem accedir a ell des de el següent enllaç del DOI:

- <https://doi.org/10.5281/zenodo.6435640>

TAULA CONTRIBUCIONS

Contribucions	Signatura
Investigació prèvia	 
Redacció de les respostes	 
Desenvolupament del codi	 

Taula 1: Contribucions treball