

# Pràctica 1

Tipologia i cicle de vida de les dades

Estat de la neu a les pistes

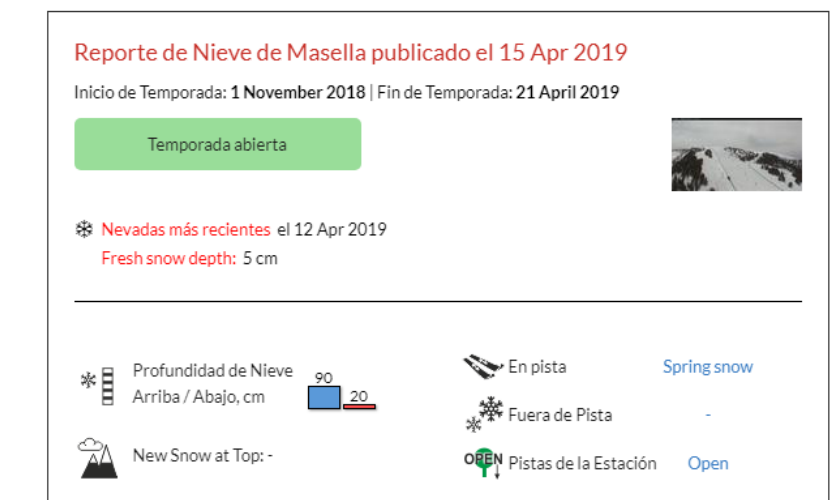
**Web Scraping**

*Jordi Orriols / Jordi Aballó*

1)

Per començar la pràctica ens agradaria contextualitzar que ambdós participants de la practica ( Jordi Aballó i Jordi Orriols) som uns fans incondicionals del esquí i en general dels esports de neu. Per tant, hem trobat molt interessant recol·lectar informació sobre un tema que tant ens agrada. Abans de que arribi el cap de setmana, entre els aficionats dels esports de neu, es molt comú observar tant les previsions de neu com saber el estat de neu actual de les pistes, per així saber a quina estació anar.

En el nostre cas, tenim diverses estacions a Catalunya i també a Andorra i França que ens cauen força a prop. El que normalment es pot fer es entrar a veure el part de neu de cada estació a la seva pagina principal, tot i que aquest procés pot ser tediós el que normalment es fa es entrar o be a infonieve o a snowforecast i mirar els raports de neu:



*Captura de pantalla del part de neu de Masella a Snow-forecast.com*

Tal i com podem comprovar ens surt tota la informació destacada, tipus de neu, estat de les pistes, noves nevades si n'hi ha hagut i la profunditat de neu màxim i mínim de la estació. Per tant hem decidit que podria ser molt útil fer una recol·lecta d'aquesta informació ja que es podria schedulejar i programar per que es fes cada setmana i anar veient l'evolució històrica per saber quina es la millor època per anar a esquiar de cara l'any que be o simplement per tenir la informació a més fàcil accés.

Si bé poder estem ja a les acaballes de la temporada, vam creure que seria interessant seguir per aquest camí, a més que aquesta web no ens prohibeix de manera explicita la recaudació d'informació com es el cas d'altres web que ens vam plantejar amb anterioritat com una de títols musicals.

2)El dataset s'anomena EstadoNieve, ja que la informació que recopilem a nivell de estació es el estat actual de la neu.

3) Per aconseguir les dades, hem hagut de anar entrant a la mateixa web però amb diferents extensions:

<https://es.snow-forecast.com/resorts/Masella/snow-report>

tal i com podem observar, en aquest cas el hyperlink correspon a la estació d'esquí de Masella, amb la qual cosa, en aquesta web hem extret la informació corresponent a Masella i després hem saltat a una altra estació (per tant a una altra pàgina web) d'on hem extret exactament la mateixa informació. El data set ens aporta tota la informació que necessitaríem per tal de valorar a quina estació volem esquiar, ja que tindríem les cotes de neu, centímetres acumulats així com la condició de la neu i si hi ha hagut alguna nevada recent.

https://es.snow-forecast.com/resorts/Masella/snow-report

Snow Alerts

Up to 20% off Ski Passes

Up to 50% off Equipment Hire

Ski Lesson Promo Ski Instructors

20% OFF your Ski Passes

BOOK NOW

Advertise your business on this resort page. Find out how Click to Sign Up

Nuestro reporte de nieve para Masella ofrece actualizaciones diarias sobre las condiciones de nieve, profundidades de nieve, condiciones en pista y fuera de pista y la cantidad de remontes abiertos. El último reporte de nieve para Masella a continuación se actualizó el 15 Apr 2019. Reporte de Nieve se proporcionan regularmente a lo largo de la temporada de esquí gracias a nuestra red de gestores de estación de esquí, el Skiclub de Gran Bretaña y Skiresort Service International GmbH. Además del actual Reporte de condiciones de esquí, también ofrecemos webcams (con un archivo de 4 semanas), las observaciones actuales en vivo de estación meteorológicas cercanas y los datos históricos de nieve para Masella.

Reporte de Nieve de Masella publicado el 15 Apr 2019

Inicio de Temporada: 1 November 2018 | Fin de Temporada: 21 April 2019

Temporada abierta

Nevadas más recientes el 12 Apr 2019

Fresh snow depth: 5 cm

Profundidad de Nieve Arriba / Abajo, cm

90 20

New Snow at Top: -

En pista Spring snow

Fuera de Pista -

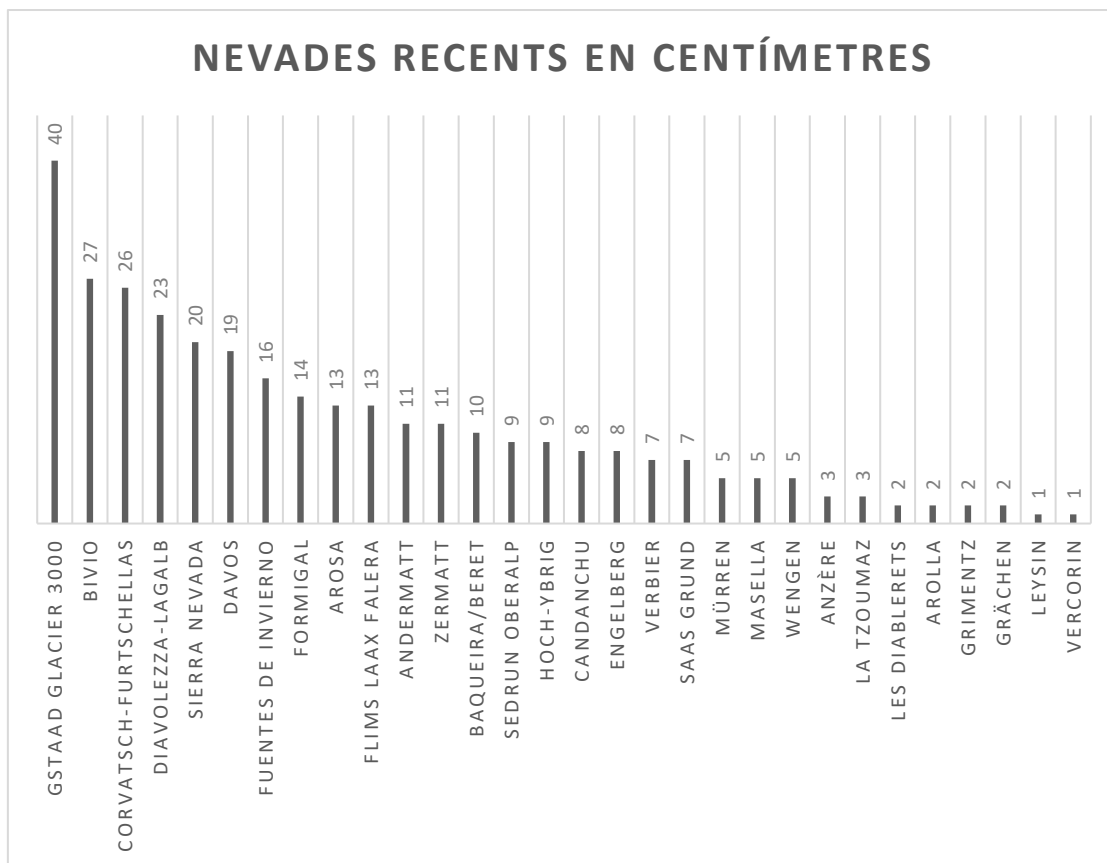
OPEN Pistas de la Estación Open

Reporte de Estación de Esquí: 13 de 27 remontes abiertos: 69 of 141 km of pistas open. Our

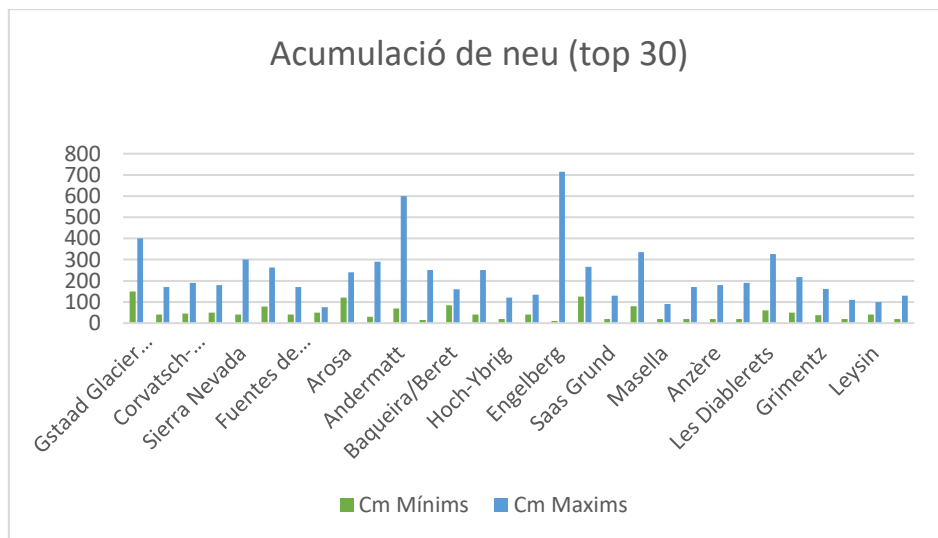
*En Vermell podem observar les variables que hem scrapejat i guardat en el dataset*

nom Estació	nevades recents (cm)	Condió de la neu en pista	estat de la estació	Cm Mínims	Cm Maxims
Gstaad Glacier 3000	40	Excellent	Excellent	150	400
Bivio	27	Powder	Open	40	170
Corvatsch- Furtschellas	26	Packed powder or groomed snow	Open	45	190
Diavolezza-Lagalb	23	Packed powder or groomed snow	open	50	180
Sierra Nevada	20	Spring snow	Open	40	300
Davos	19	Spring snow	Open	79	263
Fuentes de Invierno	16	Spring snow	open	40	170
Formigal	14	Spring snow	Open	50	75
Arosa	13	Old corn snow, Hard	Open	120	240
Flims Laax Falera	13	Packed powder or groomed snow	Open	30	290
Andermatt	11	Packed powder or groomed snow	Open	69	600
Zermatt	11	Packed powder or groomed snow	Open	15	250
Baqueira/Beret	10	Spring snow	Open	85	160
Sedrun Oberalp	9	Packed powder or groomed snow	closed	40	250
Hoch-Ybrig	9	Powder	Closed	20	120
Candanchu	8	Spring snow	Open	40	135
Engelberg	8	Powder	Open	10	715
Verbier	7	Packed powder or groomed snow	Open	125	265

4) Tal i com podem veure a la imatge, aquesta es la forma que tindria el nostre data set, que es la informació extreta de la primera captura. Amb aquestes dades, fàcilment podríem aplicar qualsevol visualització que ens permeti una comparativa entre estacions:



Aquí gràcies a la informació extreta podem veure quines han sigut les estacions en les quals recentment ha caigut més neu, per tant ens podrien semblar més atractives per anar a esquiar.



Per altre banda, també podem realitzar aquesta visualització, on es poden veure els nivells màxims i mínims de neu acumulada a cada estació, d'on podem veure que poder alguns ha nevats molt però hi haurà zones de la estació que tindrà poca neu, o per altre banda que ha nevats menys però es més uniforme al llarg de totes les cotes de la estació. Aquests gràfics han estat fets amb el dataset que hem escrapejat, però si féssim un scraping periòdic podríem obtenir visualitzacions molt més interessants.

5) El dataset es compon de tota la informació mostrada en la captura anterior, es a dir del nom de la estació ( string), el link de la pagina d'on hem fet el scrapping ( link), La quantitat de neu nova que ha caigut ( nevades en cm), la data en la que esta efectuat el informe ( datetime), el tipus de neu que hi ha a la pista, l'estat de la estació ( open or close), i les cotes màximes i mínimes de neu que es troben a la estació. El període de recol·lecció de dades es diària, es a dir cada dia es pot executar el script i obtenir nova informació, el camp del timestamp esta present amb lo qual si anéssim executant el script cada dia podríem anar veient la evolució de les cotes de neu, tot hi que per la practica i per la rellevància de les dades seria la mateixa i per això nomes ho hem executat un cop.

6) Les dades son publiques ja que les publiquen les estacions de esquí a la seva pagina web, en aquest cas el equip de Snow forecast el que fa es recopilar totes les dades i posar-les en un lloc en comú afegint-hi els seus propis models de predicció meteorològica. La empresa esta fundada i basada al regne unit.

Per altre banda, ells mateixos t'ofereixen la opció d'embedir part de les seves dades amb diversos widgets a la teva pagina web tal i com es mostra al següent link: <https://www.snow-forecast.com/pages/feed>

7) Aquest dataset, si es segueix recopilant a diària per exemple ens podria respondre quina es la millor època per anar a esquiar, quines estacions la neu aguanten mes, quants dies tarda la neu en desfer-se, en definitiva qualsevol pregunta referent a les estacions d'esquí i qualitat de neu. Preguntes molt interessants per qualsevol esquiador, preguntes que es voldrien fer i a vegades no sabem a qui ferles i en aquest cas al posseir la informació, visualment ens podríem contestar les preguntes. De cara al any que bé pensem deixar una petita màquina de amazon que vagi corrent el escript a diari per tal de compilar aquestes dades per el seu posterior anàlisis.

8) Pel que fa als drets i llicencies de la informació, nosaltres creiem que es tracta de una Database released under Open Database License, individual contents under Database Contents License, tot hi que tal i com podem observar a la seva web, no en podem fer cap us comercial:

*You may download extracts, of any page(s) from our site for your personal reference and you may draw the attention of others within your organisation to material posted on our site.*

Per tan ens donen el consentiment de recopilar la informació que ells mateixos exposen.

Contribucions	Signa
Recerca prèvia	JO/JA
Redacció de les respostes	JO/JA
Desenvolupament codi	JO/JA