

December 17, 2024

VISUALITZACIÓ DE DADES — Pràctica I

El focus principal de les dades que s'utilitzaran té a veure amb el món de l'escalada. Provenen de dues fonts:

1. 3 datasets [Zaragoza, 2021] sobre escalada que abasten: dades sobre escaladors (demogràfiques, número de vies completades...), dades sobre rutes (país, grau...) i dades sobre equivalències en els grau de les rutes. Ens referirem a ells com *climbers*, *routes* i *grades* respectivament.
2. 1 dataset [Tamošauskas, 2024] de correspondències entre els codis de país i les coordenades. Ens referirem a ell com a *countries*

Justificació

L'interès per a l'elecció d'aquestes dades és purament personal. Recentment he estat profundament atret pel món de l'escalada i el fet de poder elaborar una pràctica, el resultat de la qual sigui quelcom que analitzaria per gust, em resulta motivador.

Rellevància

Tenint en compte què tant desenvolupat està el món de l'escalada i que la majoria de sectors arreu ja han estat descoberts i explotats, dades de fa 3 anys són virtualment actuals. Dades com aquestes són importants per la gent que escala. Normalment aquest tipus d'informació, es passa de boca en boca o està repartida en guies separades; tenir un conjunt de dades centralitzat i utilitzable és de gran rellevància pel col·lectiu. Es tracta de forma igualitària als gèneres femení i masculí però ho tracta de forma binària. Opino que, probablement, no és excloent de forma deliberada. És una conseqüència resultant de que la intersecció dels conjunts "gent que escala i reporta informació sobre les vies" i "gent que no s'adequa en la binarització de gènere" és tant ínfima que amb les dades que tenim pot no haver-hi cap exemple.

Complexitat

Cada conjunt de dades utilitzat té unes característiques diferents. Els dos conjunts troncats (*climbers* i *routes*) tenen el número de registres i tipus de dades contemplats a la taula 1.

Per la seva banda, *countries*, conté dades geoespacials (longitud i latitud) de tots els països, que podrem utilitzar per *mapejar* les rutes amb les seves respectives valoracions fetes pels escaladors.

La complexitat sembla adequada: milers de registres amb desenes de variables a considerar amb tipus qualitatius, quantitatius i geoespacials.

Originalitat

Degut a ser unes dades pertanyents a un nínxol molt concret, les dades de base no han estat gairebé usades prèviament. En la font d'on s'han extret, apareix que algú ha elaborat un *Exploratory Data Analysis* amb algunes visualitzacions mostrant, per exemple, la distribució d'algunes variables; és tot el que hi ha. No

Tabla 1: Climbers and Routes

Climbers (10927 registres)		Routes (55858 registres)	
Name	Type	Name	Type
user_id	int64	name_id	int64
country	object	country	object
sex	int64	crag	object
height	int64	sector	object
weight	int64	name	object
age	float64	tall_recommend_sum	int64
years_cl	int64	grade_mean	float64
date_first	object	cluster	int64
date_last	object	rating_tot	float64
grades_count	int64		
grades_first	int64		
grades_last	int64		
grades_max	int64		
grades_mean	float64		
year_first	int64		
year_last	int64		

hi ha literatura que la utilitzi. Més enllà de la originalitat del conjunt *perse*, l'ampliació de les dades que li aporta una dimensió geoespacial no és quelcom que s'hagi fet prèviament. Almenys, de forma pública i/o correctament citada.

Prèvia a la visualització

La visualització serà original i aportarà una nova dimensió a les dades, podent filtrar visualment per país. Considero que les dades que he escollit són adequades pel fi que tindrà la visualització. No hi ha visualitzacions existents que compleixin la funció d'aquesta.

Diccionari de variables

Separarem el diccionari pels 4 conjunts de dades:

Tabla 2: Diccionari de variables

Nom de la variable	Descripció	Fet a estudiar
climbers		
user_id	Identificador únic de l'escalador	
country	País de l'escalador (ISO Alpha-3)	X
sex	Sexe de l'escalador (codificat)	X
height	Alçada de l'escalador (cm)	X
weight	Pes de l'escalador (kg)	X
age	Edat de l'escalador	X
years_cl	Anys d'experiència en escalada	
date_first	Data de la primera activitat registrada	
date_last	Data de l'última activitat registrada	
grades_count	Nombre total de vies realitzades	
grades_first	Grau de dificultat de la primera via	X
grades_last	Grau de dificultat de l'última via	X
grades_max	Grau màxim assolit	X
grades_mean	Grau mitjà de dificultat	X
year_first	Any de la primera activitat registrada	
year_last	Any de l'última activitat registrada	
Vies d'escalada (routes)		
name_id	Identificador únic de la via	
country	País on es troba la via (ISO Alpha-3)	X
crag	Zona d'escalada (conglomerat de sectors)	
sector	Sector específic dins de la zona d'escalada	
name	Nom de la via	
tall_recommend_sum	Recompte de recomanacions pels escaladors alts	
grade_mean	Grau mitjà de dificultat estimat	X
cluster	Clúster o agrupació de vies	
rating_tot	Valoració total de la via	
grades		
grade_id	Identificador únic del grau	
grade_fra	Grau en el sistema francès	
countries		
Country	Nom del país	X
Alpha-2 code	Codi ISO Alpha-2 del país	
Alpha-3 code	Codi ISO Alpha-3 del país	
Numeric code	Codi numèric ISO del país	
Latitude (average)	Latitud mitjana del país o zona	X
Longitude (average)	Longitud mitjana del país o zona	X

Referències

- [Tamošauskas, 2024] Tamošauskas, T. (2024). country. <https://gist.github.com/tadast/8827699>.
- [Zaragoza, 2021] Zaragoza, J. (2021). Climb dataset. <https://www.kaggle.com/datasets/jordizar/climb-dataset>.

Submitted by Alexandre Sánchez on 17 de diciembre de 2024.