

# Visualització de Dades

## Pràctica - Part 1: Selecció del conjunt de dades

Marc Bracons Cucó

**Joc de dades seleccionat:** Consolidated Open Source Global Development Dataset (COSGDD)[1], 1421 registres, 32 variables.

### 1. Justificació

La tria d'aquest dataset té una doble motivació, per la banda personal, tinc interès en comprendre com factors econòmics, socials, polítics i ambientals s'interrelacionen i contribueixen al benestar i la felicitat de la població. A nivell professional, haver trobat un joc de dades que integri tots aquests paràmetres de diferents àmbits representa un repte al fer-ne un anàlisi complex i multidimensional. A més, tindre tota aquesta informació en una mateixa font em permetrà centrar-me en fer preguntes rellevants i la creació de visualitzacions significatives.

*“El món és complicat. La felicitat no depèn únicament dels diners, també depèn de la felicitat, la justícia, les relacions o fins i tot de com de net és l'aire”.*

- Usuari AKHILCHHH de Kaggle

### 2. Rellevància

El conjunt de dades és actual, la seva última actualització va ser fa 20 dies (la data de redacció d'aquest document és el 16 de Desembre de 2024).



AKHILCHHH · UPDATED 20 DAYS AGO

Figura 1. Última actualització del joc de dades

A més, combina fonts obertes contemporànies (World Happiness Report [2], The Economist Democracy Index[3] i d'altres fonts, donant un punt de vista rellevant sobre qüestions crítiques com el benestar humà, la sostenibilitat ambiental, la qualitat i la transparència política, etc. que poden ser d'interès no només per al ciutadà mitjà, sinó que també per a polítics, investigadors i organitzacions internacionals.

### 3. Complexitat del dataset

Es tracta d'un joc de dades d'alta complexitat, ja que conté dades de més de 150 països al llarg de diversos anys, acumulant més de mil registres. També inclou desenes de variables que mesuren aspectes socials, econòmics, ambientals, polítics i de benestar. Hi ha variables numèriques (GDP per capita, emissions, esperança de vida, ...), categòriques (tipus de governança política), escales (çindex de democràcia, llibretat per prendre decisions), així com indicadors derivats i mètriques obtingudes a partir de les variables anteriors.

Aquesta diversitat evita la simplicitat, permeten l'obtenció de visualitzacions i anàlisis complexes, correlacions entre diversos paràmetres i creació de narratives sobre el desenvolupament global.

### 4. Originalitat del joc de dades

Aquest dataset pot ser considerat original, ja que ofereix diversos punts de vista sobre una mateixa qüestió. Les mateixes dades tractades per dos analistes diferents donaran com a resultats visualitzacions completament diferents. Es tracta d'una compilació que integra diferents fonts públiques i variables generades per oferir noves perspectives. A més, el fet que tinguem mètriques derivades a partir d'indicadors ja existents (com la combinació de la llibertat i la riquesa o l'equilibri ambiental), permet anar més enllà de l'exploració bàsica i implica un alt potencial per desenvolupar visualitzacions interessants i abordar preguntes no convencionals.

### 5. Qüestions a respondre

Com ja em comentat, la gran varietat de punts de vista que ofereix el joc de dades permet abordar qüestions de tot tipus. Algunes d'elles poden ser:

- Com es relaciona la felicitat amb el nivell econòmic i la llibertat?
- Quin és l'impacte de la governança en la percepció de corrupció i la confiança?

- Existeix algun trade-off entre el desenvolupament econòmic i la sostenibilitat ambiental?
- Com varia la felicitat al llarg del temps en funció de canvis econòmics?

Aquests qüestions ja han sigut abordades en altres estudis, com per exemple en el World Happiness Report[2] o en el Democracy Index[3], però la gran quantitat de paràmetres donarà un punt de vista diferent, donant més èmfasi en els resultats d'estudis posterior o potser, fins i tot, arribant a altres conclusions.

A continuació hi ha un diccionari de variables que facilitarà la comprensió del joc dades i permetrà mantindre una estructura de pensament ordenada a l'hora de respondre a les qüestions plantejades.

Variable	Significat	Tipus	Fet a estudiar (F) o Dimensió (D)
Country	Nom del país	Categòrica	D
Year	Any de referència	Numèrica	D
Life Ladder	Escala felicitat	Numèrica	F
Log GDP per capita	PIB per càpita en escala log.	Numèrica	F
Social Support	Proporció de persones amb suport social	Numèrica	F
Healthy life expectancy at birth	Esperança de vida saludable al néixer	Numèrica	F
Freedom to make life choices	Grau de llibertat per pendre decisions	Numèrica	F
Generosity	Generositat	Numèrica	F
Perceptions of corruption	Percepció de corrupció	Numèrica	F
Positive affect	Proporció de persones que experimenten emocions positives	Numèrica	F
Negative affect	Proporció de persones que experimenten emocions negatives	Numèrica	F

Regime Type	Tipus de règim polític	Categòrica	D
Democracy Index	Índex de democràcia	Numèrica	F
Biofuel elec per capita	Producció d'electricitat per càpita a partir de biocombustibles	Numèrica	F
Coal elec per capita	Producció d'electricitat per càpita a partir de carbó	Numèrica	F
Coal production	Producció de carbó	Numèrica	F
Gas production	Producció de gas	Numèrica	F
Low carbon electricity	Producció d'electricitat baixa en emissions de carboni	Numèrica	F
Oil production	Producció de petroli	Numèrica	F
Renewables production	Producció total d'energies renovables	Numèrica	F
Non renewables production	Producció total d'energies no renovables	Numèrica	F
Total production	Producció total d'energia	Numèrica	F
Food supply (kcal per capita per day)	Subministrament d'aliments en quilocalories per persona i dia	Numèrica	F
Food supply (Protein g per capita per day)	Proteïna disponible per persona i dia	Numèrica	F
Food supply (Fat g per capita per day)	Greix disponible per persona i dia	Numèrica	F
Deaths	Nombre de morts	Numèrica	F
Total Emissions	Emissions totals de gasos d'efecte hivernacle	Numèrica	F

Human Development Index	Índex de desenvolupament Humà	Numèrica	F
Rule_of_Law_Index	Índex de l'estat de dret	Numèrica	F
Median Age	Edat mitjana de la població	Numèrica	F
Urban Population	Percentatge de població urbana	Numèrica	F
Tax Revenue	Ingressos fiscals com a percentatge del PIB	Numèrica	F

## 6. Bibliografia

1. Consolidated Open Source Global Development Dataset. Usuari AKHILCHH. Novembre de 2024. <https://www.kaggle.com/datasets/akhilchhh/cosgdd>.
2. World Happiness Report 2024. <https://worldhappiness.report/>.
3. Democracy Index Report. 2023. <https://www.eiu.com/n/campaigns/democracy-index-2023/>.