

AgriClimate Analytics

Selecció i Justificació del Conjunt de Dades

Projecte de Visualització de Dades

Eric Jiménez Barril

Desembre 2025

*“Les dades no menteixen: el clima està canviant,
i l'agricultura n'és testimoni i protagonista”*

Context Fictici i Motivació Acadèmica

Nota preliminar: Aquest document es presenta en un context *fictici* amb finalitat exclusivament acadèmica. L'escenari empresarial descrit serveix com a marc narratiu per justificar la selecció de dades i el disseny de visualitzacions, però no representa una empresa real ni projectes professionals en curs. Tot i que el discurs desenvolupat hauria de disposar d'un volum molt més elevat d'informació i una infraestructura analítica robusta per ser possible, en realitat el conjunt de dades utilitzat és limitat i condicionat per l'abast del projecte acadèmic, així com per les fonts públiques obertes disponibles, però la idea era tenir una motivació per impulsar el projecte.

L'Escenari: AgriClimate Analytics

Imaginem que som els fundadors d'**AgriClimate Analytics**, una startup de consultoria agrotecnològica especialitzada en intel·ligència climàtica per al sector agrícola. La nostra proposta de valor consisteix a proporcionar a cooperatives agràries, grans productors i institucions governamentals una plataforma d'anàlisi que integri dades climàtiques, productives i socioeconòmiques per prendre decisions informades sobre:

- **Selecció de cultius resilents** adaptats a escenaris climàtics futurs.
- **Gestió de recursos** (aigua, fertilitzants, pesticides) optimitzada segons tendències climàtiques.
- **Avaluació d'inversions** en infraestructures de regadiu i tecnologia agrícola.
- **Predicció de rendiments** considerant variables ambientals i econòmiques.
- **Anàlisi de vulnerabilitat** de regions agrícoles davant el canvi climàtic.

Aquest projecte de visualització constitueix el nostre *proof-of-concept* per demostrar als inversors i clients potencials que disposem de les capacitats tècniques i analítiques necessàries per transformar dades brutes en insights accionables.

1. Justificació de la Selecció del Conjunt de Dades

1.1 Motivació Personal i Professional (Fictícia)

La decisió de crear AgriClimate Analytics neix d'una experiència personal: el meu avi era agricultor de secà a l'interior de Catalunya i, durant els últims anys de la seva activitat, va patir les conseqüències directes de sequeres recurrents i onades de calor que van diezmar les collites. Aquesta vivència m'ha impulsat a aplicar les meves competències en ciència i anàlisi de dades per abordar un problema real i urgent: **com pot l'agricultura adaptar-se al canvi climàtic de manera sostenible i rendible**.

Des d'un punt de vista professional, he identificat una oportunitat de mercat: les dades climàtiques i agràries existeixen (moltes d'elles de lliure accés), però estan **fragmentades**, provenen de fonts heterogènies i requereixen competències tècniques avançades per integrar-les i interpretar-les. AgriClimate Analytics vol ser el pont entre aquestes dades i els decisors del sector agrícola.

1.2 Selecció de les Fonts de Dades

Per construir el nostre *proof-of-concept*, hem integrat dades de sis fonts institucionals d'alta credibilitat:

1. **FAOSTAT** (FAO - Organització de les Nacions Unides per a l'Alimentació i l'Agricultura):
<https://www.fao.org/faostat/en/#data>
2. **Climate Change Indicators Dashboard** (Fons Monetari Internacional):
https://climatedata.imf.org/datasets/4063314923d74187be9596f10d034914_0
3. **Climate Change Knowledge Portal** (Banc Mundial):
<https://climateknowledgeportal.worldbank.org/download-data>
4. **World Development Indicators** (Banc Mundial):
<https://databank.worldbank.org/source/world-development-indicators>
5. **GDP Data** (Datahub.io):
<https://datahub.io/core/gdp>
6. **Countries Geographic Coordinates** (Google Developers Public Data):
https://developers.google.com/public-data/docs/canonical/countries_csv

Aquestes fonts garanteixen **rigor metodològic, cobertura global i geoespacial, actualització regular i accés obert**, elements essencials per a un projecte amb vocació d'escalabilitat.

2. Rellevància del Conjunt de Dades en el Context

2.1 Actualitat de les Dades

Les dades integrades cobreixen períodes històrics extensos (1961-2024) però inclouen informació actualitzada fins a 2024, la qual cosa les fa extremadament rellevants en el context actual de:

- **Crisi climàtica accelerada:** L'any 2024 ha estat un dels més càlids registrats. Les dades de temperatura i precipitació del Banc Mundial (sèrie CRU TS 4.09) proporcionen sèries temporals fins a l'actualitat.
- **Inseguretat alimentària global:** La FAO actualitza anualment les seves estadístiques de producció, emissions i ús del sòl. Les dades integrades inclouen informació fins a 2023-2024 segons la variable.
- **Transició agroecològica:** Els indicadors del Banc Mundial sobre ús de fertilitzants, regadiu i percentatge de terra agrícola són clau per avaluar la sostenibilitat del model productiu actual.

2.2 Importància per al Col·lectiu Agrícola

Aquest conjunt de dades és especialment rellevant per a diversos col·lectius:

1. **Petits i mitjans agricultors:** Necessiten entendre com les tendències climàtiques afectaran els seus cultius tradicionals i quines alternatives tenen.
2. **Cooperatives agràries:** Han de planificar inversions col·lectives (regadiu, maquinària, assegurances) basant-se en projeccions climàtiques i econòmiques.
3. **Administracions públiques:** Requereixen informació per dissenyar polítiques agràries adaptatives i ajuts focalitzats en regions vulnerables.
4. **Sector financer:** Bancs i asseguradores necessiten avaluar riscos climàtics en inversions agrícoles.
5. **Investigadors i ONG:** Treballen en projectes de desenvolupament rural sostenible i adaptació climàtica.

2.3 Perspectiva de Gènere

La integració de dades ha tingut en compte explícitament la **perspectiva de gènere**:

- Les dades de FAOSTAT inclouen indicadors d'ocupació agrícola desagregats per sexe (*Employment_Indicators_Agriculture_E_Sexs.csv*), permetent analitzar la participació de dones en el sector.
- El dataset final `agri_country_year_df_processed.csv` conté les variables:
 - `female_population_%`: Percentatge de dones en la població total.
 - `female_share_of_employment_in_agriculture`: Proporció de dones ocupades en agricultura.
- Aquesta desagregació permet visualitzar **desigualtats de gènere** en l'accés a l'ocupació agrícola i estudiar com el canvi climàtic pot afectar diferencialment dones i homes en contextos rurals.
- És especialment rellevant en països en desenvolupament on les dones són responsables del 60-80% de la producció d'aliments però tenen menys accés a recursos (terra, crèdit, tecnologia).

3. Complexitat del Conjunt de Dades

3.1 Dimensions del Dataset

El projecte ha generat **dos datasets finals processats** amb dimensions considerables:

Dataset	Registres		Variables
agri_country_year_df_processed.csv	10.653		31
agri_production_and_prices_df_processed.csv	617.291		7
Total	627.944	35 úniques (3 comparties)	

Taula 1: Dimensions dels datasets finals processats

Amb **més de 627.000 registres** i **35 variables úniques**, el conjunt de dades compleix àmpliament el requisit de complexitat per a un projecte de visualització avançat.

3.2 Tipologia de Variables

El conjunt de dades combina múltiples tipus de variables, oferint riquesa analítica:

3.2.1 Variables Categòriques

- **Geogràfiques:** Country, IS03 (196 països diferents)
- **Temporals:** Year (1961-2024, 64 anys)
- **Productives:** Commodity (més de 300 cultius i productes ramaders)

3.2.2 Variables Quantitatives Contínues

- **Ambientals:**
 - surface_temperature_change_celsius: Canvi de temperatura respecte la línia base
 - prec_mm_per_year: Precipitació anual en mm
 - mean_annual_tas_deg_celsius: Temperatura mitjana anual
- **Superfícies (en hectàrees):**
 - land_area_ha, agricultural_land_ha, cropland_ha
 - permanent_meadows_and_pastures_ha, area_harvested_ha
- **Emissions (en kilotones CO₂ eq):**
 - emissions_soils_CO2eq_AR5_kt
 - emissions_rice_cultivation_agricultural_CO2eq_AR5_kt
 - emissions_synthetic_fertilizers_CO2eq_AR5_kt
 - emissions_crop_residues_CO2eq_AR5_kt
 - emissions_total_livestock_CO2eq_AR5_kt
 - emissions_total_agriculture_CO2eq_AR5_kt
- **Inputs agrícoles:**

- `pesticide_agricultural_use_t / pesticide_agricultural_use_kt`
- `fertilizers_per_area_of_cropland_kg_ha`
- `fertilizers_per_capita_kg_cap`
- `fertilizers_per_value_of_agri_prod_g_Intdollar`

- **Producció:**

- `production_kt`: Producció en kilotones
- `yield_kg_ha`: Rendiment per hectàrea

- **Econòmiques:**

- `GDP_usd`: PIB en dòlars americans
- `agriculture_share_gdp_percent`: Pes de l'agricultura en el PIB
- `agriculture_annual_growth_percent`: Creixement anual del sector

- **Demogràfiques:**

- `total_population_No`
- `female_population_%, rural_population_%`
- `female_share_of_employment_in_agriculture`
- `total_share_of_employment_in_agriculture`

- **Geogràfiques (coordenades):**

- `latitude, longitude`

3.3 Complexitat de les Relacions

La complexitat del dataset no resideix només en el volum, sinó en les **relacions multidimensionals** que permet explorar:

1. **Sèries temporals:** Evolució de qualsevol variable al llarg de 64 anys (1961-2024).
2. **Anàlisi geogràfica:** Comparació entre 196 països amb coordenades per crear mapes interactius.
3. **Correlacions clima-producció:** Relacionar temperatura/precipitació amb rendiments de cultius específics.
4. **Anàlisi d'emissions:** Descomposició de les fonts d'emissions agrícoles per país i any.
5. **Eficiència de recursos:** Estudiar la relació entre inputs (fertilitzants, pesticides) i outputs (producció, rendiments).
6. **Dimensions socioeconòmiques:** Vincular variables agràries amb PIB, ocupació i demografia.
7. **Perspectiva de gènere:** Analitzar com les tendències climàtiques i econòmiques afecten diferencialment homes i dones.

3.4 Processos de Transformació i Enginyeria de Variables

Durant el procés de neteja i integració (documentat a `data_integration.html` i `data_cleaning.html`), s'han aplicat transformacions avançades:

- **Imputació de valors nuls:** Aplicació de KNN Imputer per mantenir la coherència espacial i temporal.
- **Agregació de fertilitzants:** Suma de N, P₂O₅ i K₂O per crear una variable composta.
- **Normalització d'emissions:** Agregació de diferents fonts d'emissions agrícoles.
- **Enriquiment geogràfic:** Addició de latitud/longitud per cada país.
- **Conversió d'unitats:** Harmonització de tones, kilotones, hectàrees, etc.
- **Creació de ràtios:** Variables derivades com fertilitzants per hectàrea, per càpita i per valor de producció.

Aquesta complexitat en el preprocessament garanteix que els datasets finals siguin **robustos, coherents i analíticament útils**.

3.5 Originalitat en la Gestió de Valors Faltants: Imputació Adaptada

Un dels aspectes més originals i tècnicament sofisticats del projecte rau en l'estratègia multicapa d'imputació de valors faltants. Enlloc d'aplicar una única tècnica genèrica, hem desenvolupat un enfocament adaptatiu que considera la naturalesa de cada variable i la qualitat de les dades disponibles.

3.5.1 Anàlisi Prèvia: Decisió Informada sobre Imputació vs. Eliminació

Abans d'imputar qualsevol variable, s'ha realitzat una anàlisi exhaustiva de qualitat de dades:

- **Percentatge de valors nuls:** Variables amb més del 80% de valors faltants (com l'indicador 21111 d'ocupació) han estat *eliminades* perquè la imputació introduiria més soroll que valor.
- **Consistència interna:** Variables amb incoherències lògiques (per exemple, on $\text{total} \neq \text{female} + \text{male}$) han estat corregides o eliminades segons la gravetat.
- **Continuïtat temporal:** Variables amb discontinuïtats extremes no s'han imputat amb models predictius per evitar crear tendències artificials.

Aquest procés de *tria conscient* demostra una comprensió profunda de les limitacions dels mètodes d'imputació i garanteix que només s'apliquen on són realment beneficiosos.

3.5.2 Tècniques d'Imputació Aplicades

Per a les variables amb qualitat suficient, s'han aplicat tècniques d'imputació **adaptades al context**:

1. Interpolació Temporal Lineal per País

Per a variables econòmiques i demogràfiques amb evolució gradual (PIB agrícola, percentatge de terra irrigada, ocupació), s'ha aplicat interpolació lineal *dins de cada país*. Això garanteix que:

- Les tendències històriques es mantenen.

- No es generen valors impossibles (per exemple, percentatges negatius).
- Els valors imputats són coherents amb el context nacional.

2. Regressió Lineal per Tendències Temporals

Per a sèries productives (producció de cultius, superfície collida), quan hi ha almenys dos anys amb valors positius, s'ha aplicat regressió lineal basada en l'any. Els valors predits es limiten a ≥ 0 per evitar resultats físicament impossibles.

3. Imputació Basada en Proporcions Anuals per Commodity

Aquesta és la tècnica més **innovadora** del projecte. Per a variables productives com A (superfície), P (producció) i Y (rendiment), on existeix una relació matemàtica coneguda:

$$P = A \times \frac{Y}{1000}$$

S'ha implementat un sistema que:

- Utilitza les mitjanes anuals per cultiu (μ_A^{kt} , μ_P^{kt} , μ_Y^{kt}) per mantenir proporcions coherents.
- Imputa valors faltants mantenint la relació dimensional entre les tres variables.
- Garanteix consistència física: si falta A però tenim P i Y , calculem $A = \frac{P \times 1000}{Y}$.

Aquesta tècnica és especialment potent perquè respecta les **lleis agronòmiques fonamentals** i no genera valors arbitraris.

4. Normalització de Proporcions Complementàries

Per a variables que han de sumar una constant (per exemple, `rural_population+urban_population = 1`), s'ha aplicat una normalització proporcional per garantir consistència lògica:

$$\text{rural_normalitzat} = \frac{\text{rural}}{\text{rural} + \text{urban}}$$

5. Imputació per KNN (K-Nearest Neighbors)

Per a variables amb dependències espacials (temperatura, precipitació), s'ha considerat l'ús de KNN Imputer que utilitza la similitud entre països (basada en latitud, longitud i altres indicadors climàtics) per imputar valors faltants. Aquest mètode és especialment adequat perquè països geogràficament propers tendeixen a tenir patrons climàtics similars.

3.5.3 Resultat: Dataset Complet i Coherent

Després d'aplicar aquesta estratègia multicapa, els datasets finals presenten:

- **0% de valors nuls** en variables productives.
- **Consistència dimensional** entre variables relacionades.
- **Respecte per les lleis físiques i agronòmiques.**
- **Transparència metodològica:** Cada decisió d'imputació està documentada al notebook `data_cleaning.html`.

Aquesta aproximació demostra que la imputació de dades no és un procés mecànic, sinó una tasca que requereix **comprensió del domini**, **creativitat metodològica** i **rigor analític**. És un dels elements més originals i valuosos del projecte.

4. Originalitat del Conjunt de Dades

4.1 Evitació de Datasets Clàssics

El projecte s'allunya deliberadament de datasets sobreutilitzats (Titanic, Iris, MNIST, COVID-19) per treballar amb dades **reals, actuals i amb impacte social**. La temàtica de l'**adaptació climàtica de l'agricultura** és:

- **Pertinent:** El canvi climàtic i la seguretat alimentària són reptes globals del segle XXI.
- **Poc explotada en visualització:** Tot i existir dades, hi ha poques visualitzacions integrades que combinin clima, producció i economia agrícola.
- **Interdisciplinària:** Connecta ciències ambientals, economia, agronomia i anàlisi de dades.

4.2 Integració Multi-Font: L'Originalitat com a Procés

L'originalitat principal del projecte resideix en la **integració de sis fonts institucionals independents**:

1. **FAOSTAT:** 8 categories de dades (Emissions, Ocupació, Cobertura del Sòl, Fertilitzants, Ús del Sòl, Pesticides, Població, Producció i Preus).
2. **IMF Climate Data:** Indicadors globals de temperatura superficial.
3. **World Bank Climate Portal:** Sèries temporals de temperatura i precipitació amb resolució 0.5° (CRU TS 4.09).
4. **World Bank Development Indicators:** Variables econòmiques i estructurals del sector agrícola.
5. **Datahub GDP:** Contextualització macroeconòmica.
6. **Google Developers Public Data:** Coordenades geogràfiques (latitud/longitud) per a visualitzacions cartogràfiques i anàlisis espacials.

Cap d'aquestes fonts, per separat, permet respondre les preguntes que plantegem. La integració és el que genera valor afegit.

4.3 Enriquiment i Transformació

A més de la integració, el projecte ha enriquit les dades originals:

- **Geolocalització:** Addició de coordenades geogràfiques (latitud/longitud) obtingudes de Google Developers Public Data per permetre visualitzacions cartogràfiques. A partir del codi ISO3 de cada país, s'han assignat coordenades del centroid geogràfic, proporcionant una localització aproximada del "centre" de cada país. Aquesta informació és essencial per crear mapes interactius, anàlisis espacials i correlacions geogràfiques que relacionin clima, producció i ubicació física dels països.
- **Agregacions temporals:** Creació de mitjanes mòbils i tendències per suavitzar la variabilitat interanual.
- **Indicadors compostos:**

- Ràtio emissions totals / PIB agrícola (intensitat de carboni)
 - Ràtio fertilitzants / rendiments (eficiència productiva)
 - Índex de vulnerabilitat climàtica (combinant temperatura, precipitació i dependència econòmica de l'agricultura)
- **Desagregació per gènere:** Explotació de variables específiques per estudiar desigualtats.

4.4 Revisió de Visualitzacions Existents

S'ha realitzat una recerca de visualitzacions similars:

- Our World in Data: Té visualitzacions sobre emissions agrícoles i temperatura global, però no integra producció de cultius específics ni desagregació per gènere.
- FAO Data Portal: Ofereix gràfics estàtics per variable, però no permet anàlisis creuades amb dades climàtiques del Banc Mundial.
- Climate Change Knowledge Portal: Se centra en indicadors climàtics, però no inclou dades productives ni econòmiques.

En conclusió, no he pogut trobar cap plataforma pública que integri aquestes cinc fonts amb el nivell de detall i riquesa de variables que hem aconseguit. Aquesta és la nostra proposta de valor diferencial.

4.5 Dataset com a Evolució

El nostre conjunt de dades pot considerar-se una evolució de datasets previs:

- És una actualització de projectes anteriors de la FAO i el Banc Mundial, incorporant dades fins a 2024.
- És un enriquiment d'aquests datasets individuals, creant un producte integrat.
- És una adaptació específica per al cas d'ús d'AgriClimate Analytics.

5. Qüestions que Respondrem amb la Visualització

5.1 Formulació de Preguntes d'Investigació

Basant-nos en els punts anteriors (rellevància, complexitat, originalitat), hem identificat cinc grans preguntes d'investigació que guiaran el disseny de les visualitzacions:

5.1.1 Pregunta 1: Com ha evolucionat la relació entre temperatura global i producció agrícola?

Adequació al dataset: El dataset conté sèries temporals de temperatura (`surface_temperature_change_celsius`, `mean_annual_tas_deg_celsius`) i producció de més de 300 cultius (`production_kt`, `yield_kg_ha`) durant 1961-2024.

Variables implicades:

- *Fet a estudiar:* Rendiments agrícoles (`yield_kg_ha`)
- *Dimensió que el mesura:* Temps (`Year`), Temperatura (`mean_annual_tas_deg_celsius`), Cultiu (`Commodity`), País (`Country`)

Originalitat: Tot i existir visualitzacions sobre temperatura global i producció agrícola per separat, poques mostren la correlació específica per cultiu i regió geogràfica. La nostra visualització permetrà identificar quins cultius són més vulnerables i en quines regions.

5.1.2 Pregunta 2: Quines regions són més vulnerables al canvi climàtic des d'una perspectiva agroalimentària?

Adequació al dataset: Disposem de coordenades geogràfiques, indicadors climàtics, dependència econòmica de l'agricultura (`agriculture_share_gdp_percent`) i emissions agrícoles.

Variables implicades:

- *Fet a estudiar:* Vulnerabilitat climàtica (índex compost)
- *Dimensions que la mesuren:*
 - Canvi de temperatura (`surface_temperature_change_celsius`)
 - Variació de precipitació (`prec_mm_per_year`)
 - Dependència econòmica (`agriculture_share_gdp_percent`)
 - Intensitat d'emissions (`emissions_total_agriculture_CO2eq_AR5_kt`)
 - Població rural (`rural_population_%`)

Originalitat: Crearem un **índex de vulnerabilitat climàtica agrícola** que no existeix en altres visualitzacions, combinant dimensions ambientals, econòmiques i socials.

5.1.3 Pregunta 3: Com ha evolucionat la intensitat de carboni de l'agricultura a nivell global?

Adequació al dataset: Tenim emissions desagregades per font (`emissions_soils_CO2eq_AR5_kt`, `emissions_synthetic_fertilizers_CO2eq_AR5_kt`, etc.) i valor econòmic del sector (`agriculture_share_gdp_GDP_usd`).

Variables implicades:

- *Fet a estudiar:* Intensitat de carboni (emissions / PIB agrícola)
- *Dimensions que la mesuren:*

- Emissions totals (`emissions_total_agriculture_CO2eq_AR5_kt`)
- Descomposició per font (sòls, arrossars, fertilitzants, ramaderia)
- PIB agrícola (calculat com $\text{GDP_usd} \times \text{agriculture_share_gdp_percent}$)
- Evolució temporal (`Year`)

Originalitat: Aquesta mètrica no està disponible en visualitzacions públiques de la FAO ni del Banc Mundial. Permetrà identificar països que estan *desacoblant* el creixement agrícola de les emissions.

5.1.4 Pregunta 4: Quina és la relació entre ús de fertilitzants, rendiments i emissions?

Adequació al dataset: Disposem de variables d'input (`fertilizers_per_area_of_cropland_kg_ha`), output (`yield_kg_ha`) i externalitat negativa (`emissions_synthetic_fertilizers_CO2eq_AR5_kt`).

Variables implicades:

- *Fet a estudiar:* Eficiència productiva i ambiental
- *Dimensions que la mesuren:*
 - Input: `fertilizers_per_area_of_cropland_kg_ha`
 - Output: `yield_kg_ha` (per cultiu)
 - Externalitat: `emissions_synthetic_fertilizers_CO2eq_AR5_kt`
 - Context econòmic: `agriculture_annual_growth_percent`

Originalitat: Permetrà visualitzar la **frontera d'eficiència** entre països que aconseguixen alts rendiments amb baixos inputs i emissions (model sostenible) vs. països amb alta intensitat d'inputs i emissions (model intensiu).

5.1.5 Pregunta 5: Com afecta el canvi climàtic a la participació de les dones en l'agricultura?

Adequació al dataset: Variables de gènere (`female_share_of_employment_in_agriculture`, `female_population_%`) combinades amb indicadors climàtics i productius.

Variables implicades:

- *Fet a estudiar:* Desigualtat de gènere en ocupació agrícola
- *Dimensions que la mesuren:*
 - `female_share_of_employment_in_agriculture`
 - `total_share_of_employment_in_agriculture`
 - `surface_temperature_change_celsius`
 - `agriculture_share_gdp_percent`
 - `rural_population_%`

Originalitat: És una pregunta **poc explorada** en visualitzacions públiques. Permetrà identificar si les regions amb major impacte climàtic presenten també major desigualtat de gènere en l'accés a l'ocupació agrícola.

Taula 2: Diccionari de Variables dels Datasets Finals

Variable	Significat	Rol
Country	Nom del país	Dimensió
ISO3	Codi ISO3 del país	Dimensió
Year	Any de registre (1961–2024)	Dimensió
latitude	Latitud del país	Dimensió
longitude	Longitud del país	Dimensió
Commodity	Tipus de cultiu	Dimensió
land_area_ha	Superfície total del país (ha)	Dimensió
agricultural_land_ha	Terra agrícola total (ha)	Dimensió
cropland_ha	Terra cultivada (ha)	Dimensió
permanent_meadows_and_pastures_ha	Pastures permanents (ha)	Dimensió
area_harvested_ha	Superfície collida per cultiu (ha)	Dimensió
surface_temperature_change_celsius	Canvi de temperatura respecte a la línia base (°C)	Fet
mean_annual_tas_deg_celsius	Temperatura mitjana anual (°C)	Fet
prec_mm_per_year	Precipitació anual (mm)	Fet
pesticide_agricultural_use_t	Ús de pesticides (tones)	Fet
pesticide_agricultural_use_kt	Ús de pesticides (kilotones)	Fet
fertilizers_	Fertilitzants per...	
per_area_of_cropland_kg_ha	hectàrea (kg/ha)	Fet
per_capita_kg_cap	càpita (kg/persona)	Fet
per_value_of_agri_prod_g_Intdollar	unitat de valor de producció (g/Int\$)	Fet
emissions_-_CO2eq_AR5_kt	Emissions en kt de CO ₂ eq	
soils	dels sòls agrícoles	Fet
rice_cultivation_agricultural	del cultiu d'arròs	Fet
synthetic_fertilizers	per fertilitzants sintètics	Fet
crop_residues	de residus de cultius	Fet
total_livestock	totals de la ramaderia	Fet
total_agriculture	totals de l'agricultura	Fet
production_kt	Producció del cultiu (kilotones)	Fet
yield_kg_ha	Rendiment per hectàrea (kg/ha)	Fet
total_population_No	Població total	Dimensió
female_population_%	Percentatge de dones a la població	Dimensió
rural_population_%	Percentatge de població rural	Dimensió

(Continuació de la pàgina anterior)

Variable	Significat	Rol
share_of_employment_in_agriculture	(%)	
female_	Dones ocupades al sector agrícola	Fet
total_	Ocupació agrícola sobre el total	Fet
GDP_usd	PIB del país (USD)	Dimensió
agriculture_share_gdp_percent	Pes de l'agricultura al PIB (%)	Fet
agriculture_annual_growth_percent	Creixement anual del sector agrícola (%)	Fet

5.2 Adequació de les Preguntes al Dataset

Les cinc preguntes formulades compleixen els següents criteris d'adequació:

1. **Cobertura de variables:** Cada pregunta utilitza entre 5 i 10 variables del dataset, aprofitant la seva riquesa.
2. **Combinació de dimensions:** Totes les preguntes combinen dimensions temporals, geogràfiques i temàtiques.
3. **Tipus de dades:** S'aprofiten variables categòriques (país, cultiu) i quantitatives (temperatura, emissions, rendiments).
4. **Complexitat analítica:** Les preguntes requereixen agregacions, càlculs de ràtios, correlacions i creació d'índexs compostos.
5. **Originalitat:** Cap d'aquestes preguntes ha estat plantejada amb aquest nivell d'integració en visualitzacions públiques existents.

Conclusions

Aquesta primera part del projecte de visualització de dades s'ha concebut des del principi amb una vocació d'aplicabilitat real, encara que s'emmarca en un escenari fictici amb finalitat acadèmica. La selecció del conjunt de dades ha estat guiada per cinc principis:

1. **Rellevància social:** Abordar el repte del canvi climàtic en l'agricultura, un problema urgent i global.
2. **Complexitat tècnica:** Integrar més de 627.000 registres i 38 variables de sis fonts institucionals (que originalment eren prop de 5 milions amb més de 60 columnes).
3. **Originalitat:** Crear un producte analític que no existeix en plataformes públiques.
4. **Perspectiva de gènere:** Incorporar una perspectiva de gènere en l'anàlisi.
5. **Potencial de visualització:** Formular preguntes que permetin crear visualitzacions innovadores i informatives per la pròxima part.

El *proof-of-concept* d'AgriClimate Analytics demostra que disposem de les capacitats tècniques per transformar dades brutes en intel·ligència accionable per al sector agrícola. Les visualitzacions que desenvoluparem en la següent fase del projecte tindran el potencial de ser utilitzades per cooperatives agràries en la planificació de cultius, administracions públiques en el disseny de polítiques adaptatives, investigadors en l'estudi de la vulnerabilitat climàtica o inclús sector financer en l'avaluació de riscos agrícoles.

Referències de les Fonts de Dades

- [1] FAO. *FAOSTAT Statistical Database*. <https://www.fao.org/faostat/en/#data>. Accés: Desembre 2025.
- [2] International Monetary Fund. *Climate Change Indicators Dashboard*. https://climatedata.imf.org/datasets/4063314923d74187be9596f10d034914_0. Accés: Desembre 2025.
- [3] World Bank Group. *Climate Change Knowledge Portal - Historical Data*. <https://climateknowledgeportal.worldbank.org/download-data>. Sèrie CRU TS 4.09 (1901-2024). Accés: Desembre 2025.
- [4] World Bank Group. *World Development Indicators*. <https://databank.worldbank.org/source/world-development-indicators>. Accés: Desembre 2025.
- [5] Datahub.io. *Country, Regional and World GDP - Gross Domestic Product*. <https://datahub.io/core/gdp>. Accés: Desembre 2025.
- [6] Google Developers. *Countries Geographic Coordinates - Public Data*. https://developers.google.com/public-data/docs/canonical/countries_csv. Accés: Desembre 2025.