



ANÁLISIS DE LA TEMPERATURA Y SU RELACIÓN CON LA PRODUCCIÓN DE MAÍZ EN COLOMBIA

JAVIER ALONSO DEL SASO
ANA GONZALEZ GUERRA
SILVIA LOPEZ MONZO
DAVID MONTERO LOAIZA

UIMP
Universidad Internacional
Menéndez Pelayo

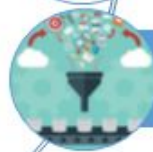
UC
UNIVERSIDAD
DE CANTABRIA



Descripción general del proyecto



Data Management Plan (DMP)



Curación de los datos y ETL



Plan de preservación



Análisis de los datos



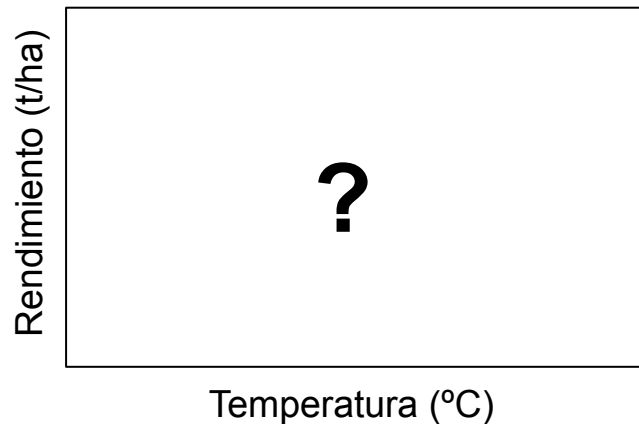
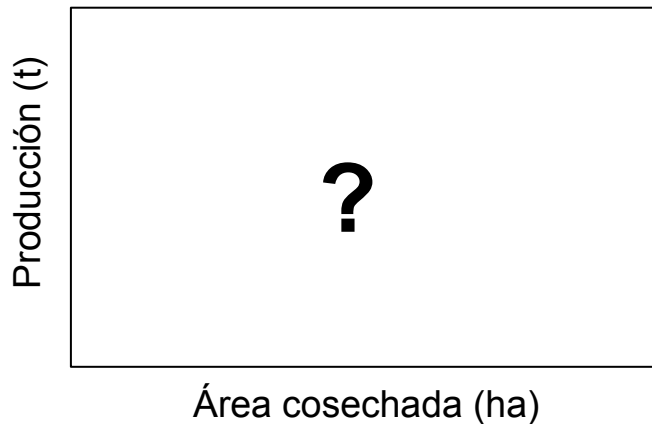
Conclusiones



Descripción general del proyecto

Objetivos:

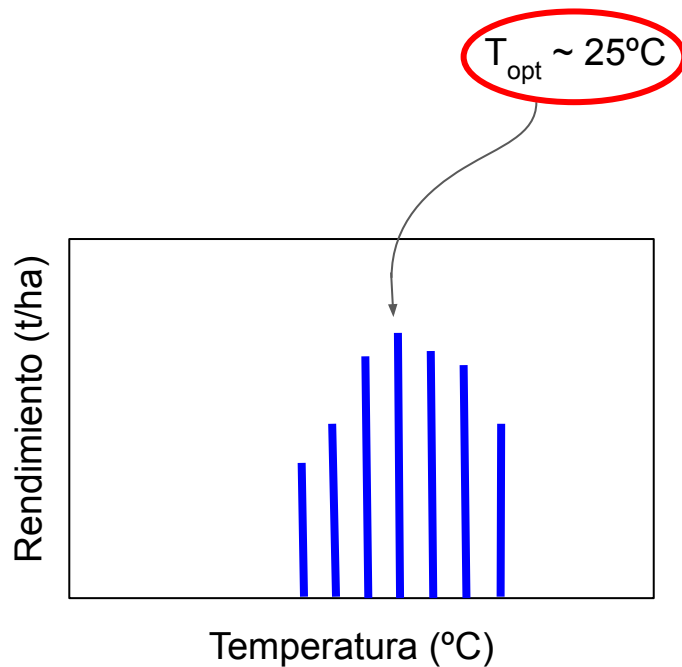
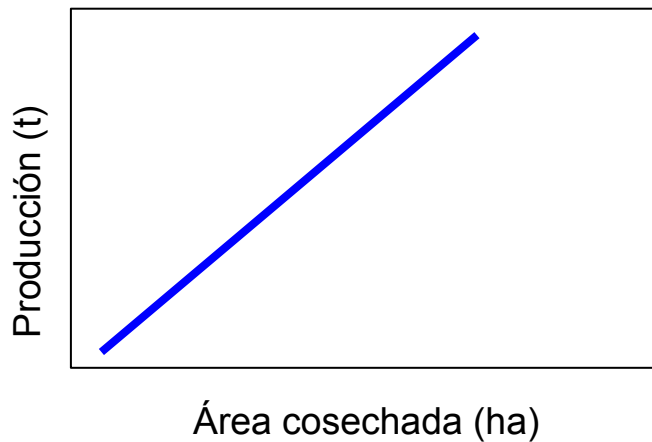
- Área cosechada (ha) VS producción (t)
- Temperatura (°C) VS rendimiento (t/ha)





Descripción general del proyecto

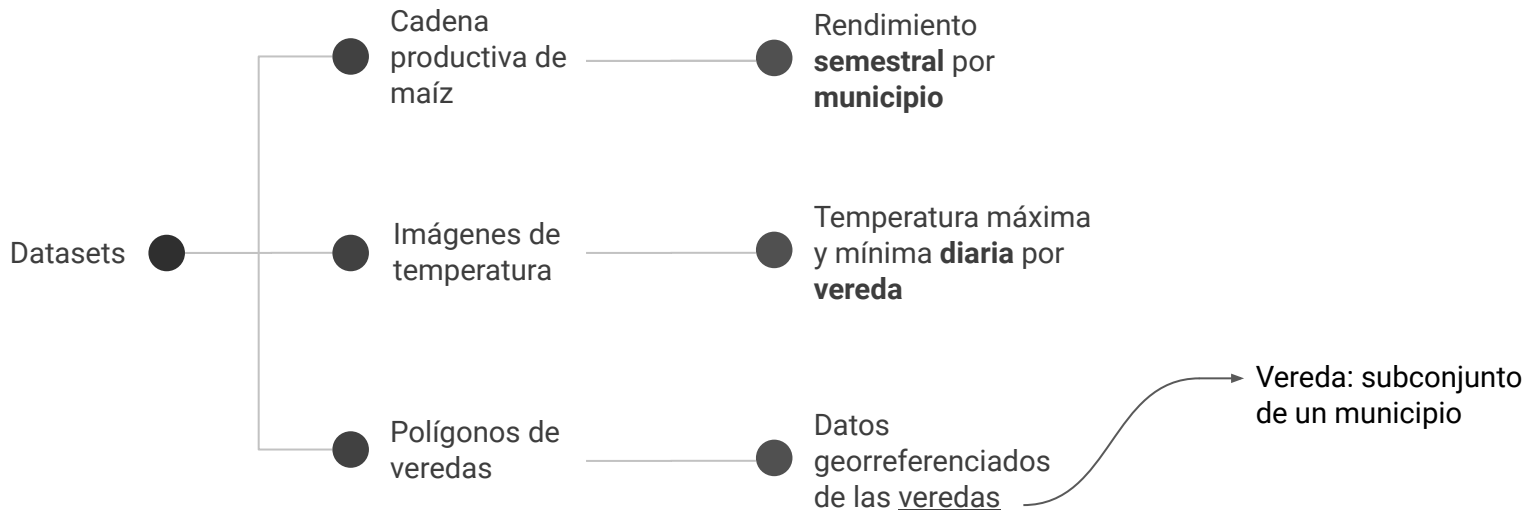
- ¿Qué esperamos obtener?





Descripción general del proyecto

- ¿Qué requisitos necesitamos?





Descripción general del proyecto

Requerimientos técnicos



Personal cualificado

Calendario de siembra

Análisis químico del suelo



Estaciones meteorológicas



Imágenes vía satélite



Región	Período de siembra	Referencia
ZAE1	Abril; septiembre	(modificado de FAO[2006])
ZAE2	Marzo	
ZAE3	Marzo; septiembre	
ZAE4;ZAE5,ZAE6	Febrero; septiembre	



Descripción general del proyecto

¿Por qué este proyecto y no otro?



Pequeños agricultores

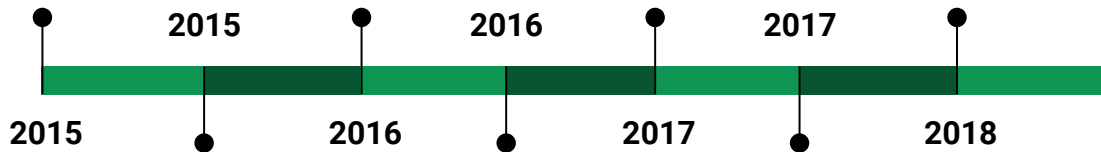


Temperatura óptima para la producción

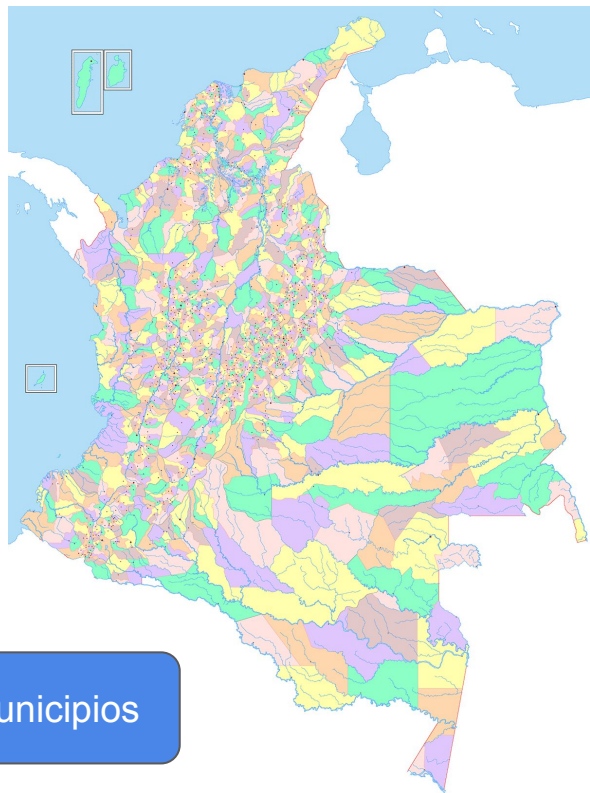


Descripción general del proyecto

Cobertura temporal: semestral



Cobertura geográfica: municipios





Descripción general del proyecto

Descripción de las fuentes de datos

Dataset	Base de datos
Producción de maíz	Gobierno de Colombia
Imágenes de temperatura	Land Processes Distributed Active Center (NASA)
Polígonos de veredas	Geoportal del Departamento Administrativo Nacional de Estadística de Colombia



Data Management Plan (DMP)



Findable

Accesible

Interoperable

Re-usable



Data Management Plan (DMP)

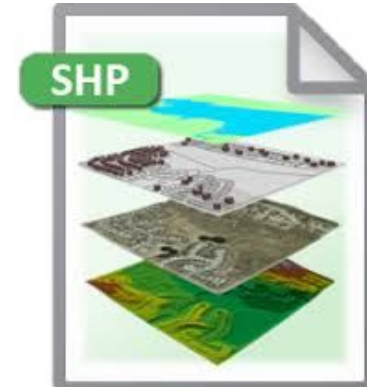
Tipos y formatos de datos



- Producción de maíz en Colombia.
- Datos generados.



- Imágenes de temperatura georreferenciadas.



- Polígonos de veredas en Colombia.



Data Management Plan (DMP)

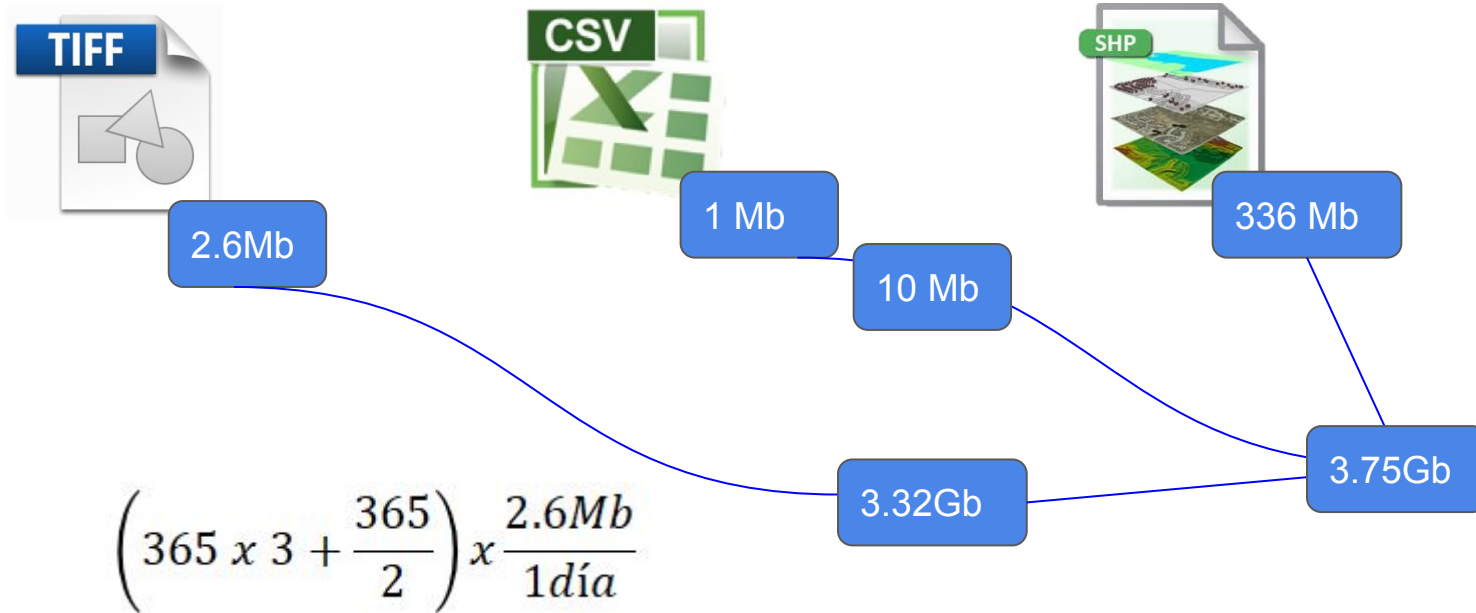
¿Usaremos datos ya existentes? ¿Cómo?

Paquete de trabajo	Responsable(s)
Cultivo de <i>Zea mays</i>	
Contratación de personal	Junta directiva del proyecto
Sembrado ZAE1	Agricultores de la región geográfica correspondiente
Sembrado ZAE2	Agricultores de la región geográfica correspondiente
Sembrado ZAE3	Agricultores de la región geográfica correspondiente
Sembrado ZAE4	Agricultores de la región geográfica correspondiente
Sembrado ZAE5	Agricultores de la región geográfica correspondiente
Sembrado ZAE6	Agricultores de la región geográfica correspondiente
Cosechado ZAE1	Agricultores de la región geográfica correspondiente
Cosechado ZAE2	Agricultores de la región geográfica correspondiente
Cosechado ZAE3	Agricultores de la región geográfica correspondiente
Cosechado ZAE4	Agricultores de la región geográfica correspondiente
Cosechado ZAE5	Agricultores de la región geográfica correspondiente
Cosechado ZAE6	Agricultores de la región geográfica correspondiente
Recogida de datos de producción	Agricultores y división de Data Science del proyecto
Estaciones meteorológicas	
Contratación del servicio con el gobierno de Colombia	Junta directiva del proyecto
Recogida de datos de temperatura	División de Data Science del proyecto
Obtención de imágenes de satélite de Colombia	
Contratación del servicio con el gobierno de Colombia	Junta directiva del proyecto
Procesado de las imágenes por municipios	División de Data Science del proyecto
Integración de los datos de producción, localización geográfica y temperatura	
Procesado de los datos en R	División de Data Science del proyecto
Procesado de los datos en Python	División de Data Science del proyecto
Análisis de los resultados obtenidos	
Elección de la mejor región para la producción	División de Data Science del proyecto
Publicación del informe desarrollado	
Redacción del informe	División de Data Science del proyecto
Data Management Plan	División de Data Science del proyecto
Creación de los metadatos correspondientes	División de Data Science del proyecto
Presentación de la convocatoria a la comisión del Horizonte 2020	Junta directiva del proyecto
Desarrollo de licencia y derechos de uso sobre los datos creados	
Definición de los términos de uso	Junta directiva del proyecto
Adquisición de una licencia que se adecue a los términos	Junta directiva del proyecto
Deposito de los datos generados	
Depósito de los datos e informe en github	División de Data Science del proyecto
Depósito de los datos e informe en el repositorio del gobierno de Colombia	División de Data Science del proyecto



Data Management Plan (DMP)

Tamaño esperado de los datos





Data Management Plan (DMP)

FAIR data

Findable: identificadores

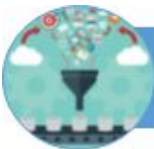


Accesible: repositorios

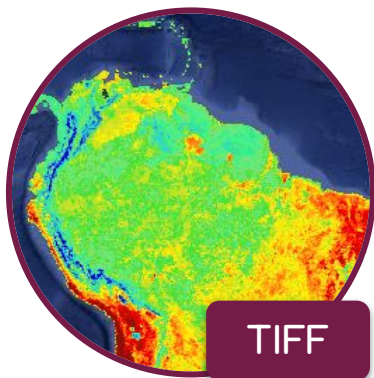
Reusable



Interoperable: vocabulario estándar de metadatos



Curación de los datos y ETL



Temperatura
Superficial

Datos ráster
georreferenciados

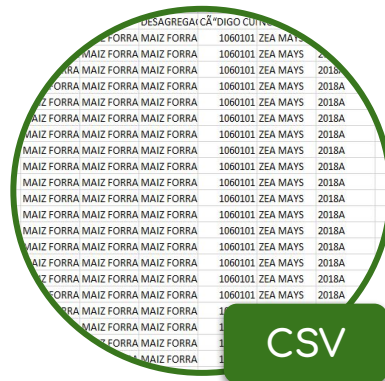
[1 imagen diaria]



Cobertura Espacial de
Veredas

Datos vectoriales
georreferenciados

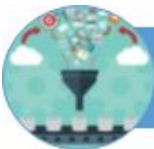
[datos únicos]



Datos de Producción
de Maíz

Datos tabulares.

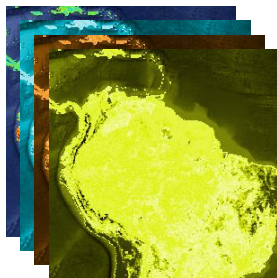
[temporalidad semestral
por municipio]



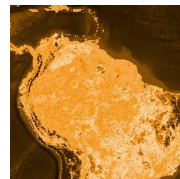
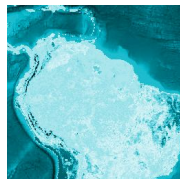
Curación de los datos y ETL

Conversión a
Celsius

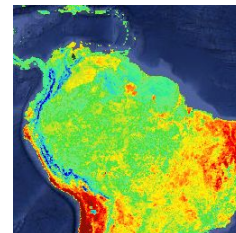
Diaria



Múltiples sub-imágenes
(bandas) [diaria]



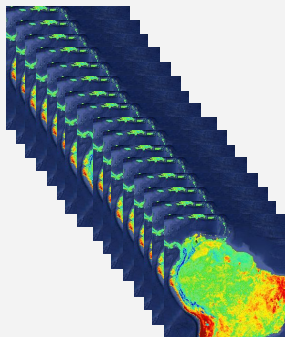
$((T_{min} + T_{max}) / 2) - 273.15$



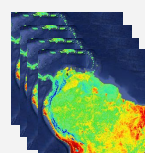
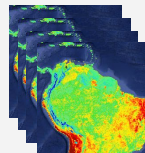
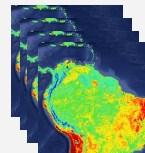
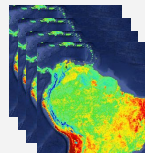
Temperatura media °C
[diaria]

Agregación de
imágenes

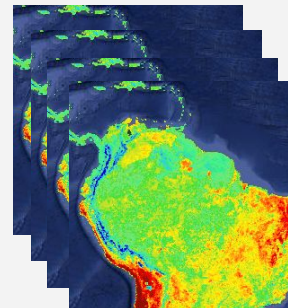
Semestral



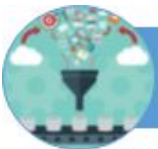
Imágenes de temperatura
media diaria



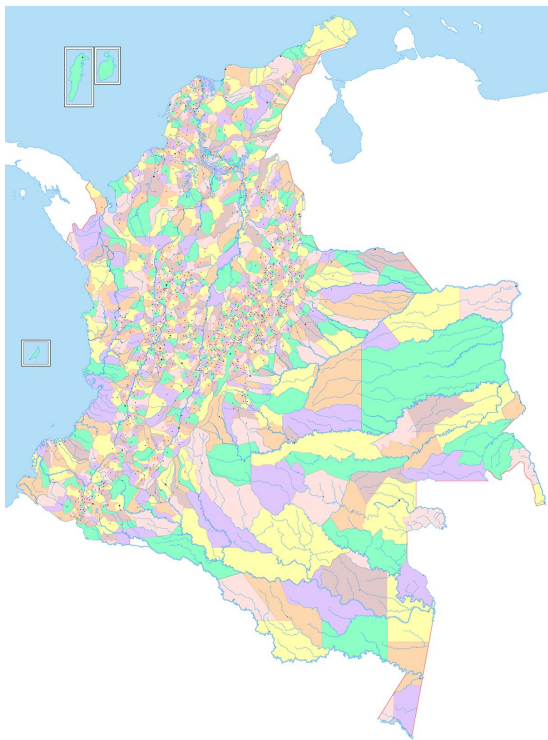
Agregación semestral
(promedio)

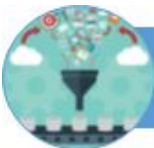


Imágenes de temperatura
media semestral

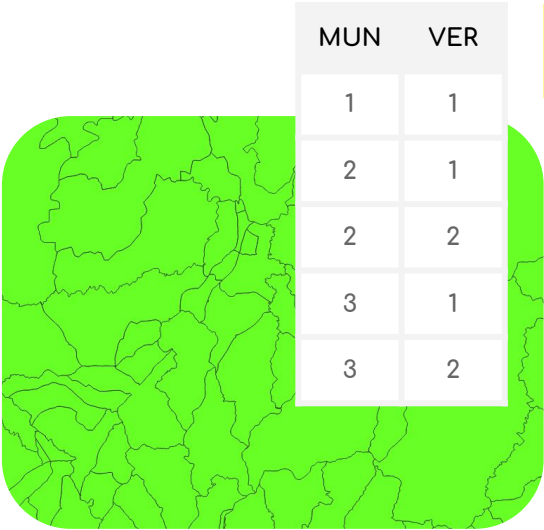


Curación de los datos y ETL



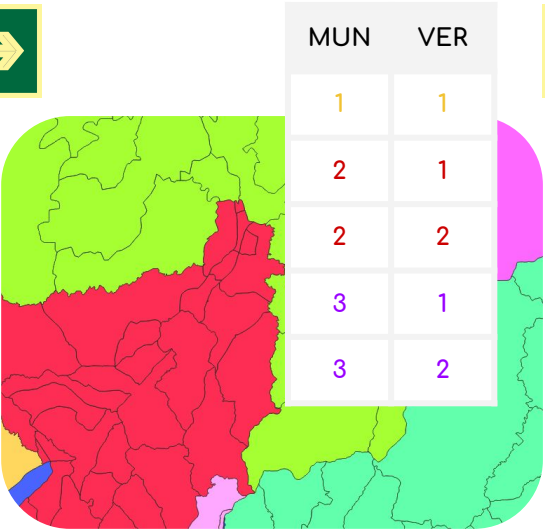


Curación de los datos y ETL



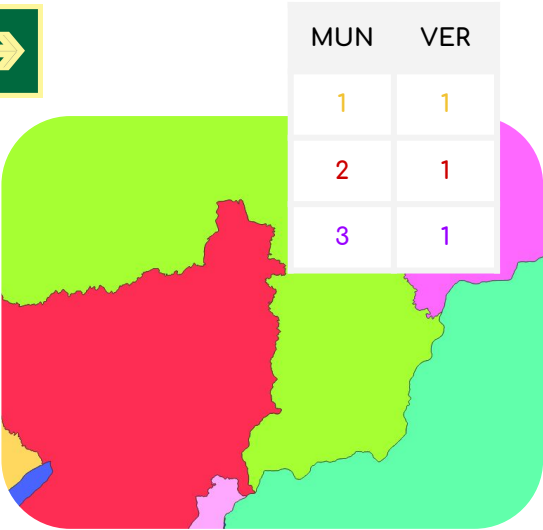
Veredas

Datos vectoriales de veredas con atributos



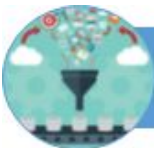
Veredas Clasificadas por Municipio

Datos vectoriales agregados por atributos

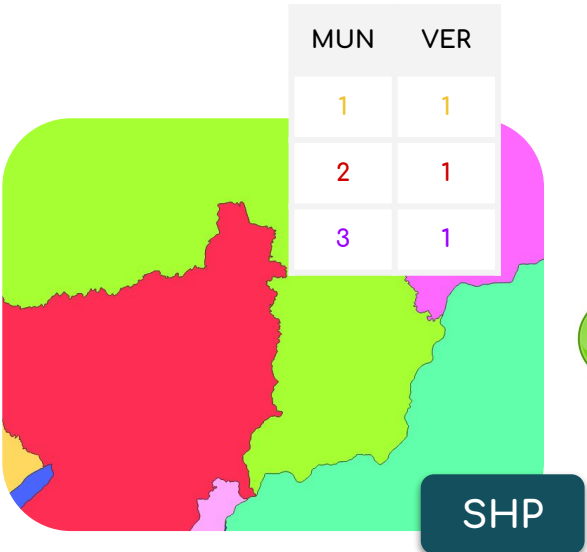


Municipios

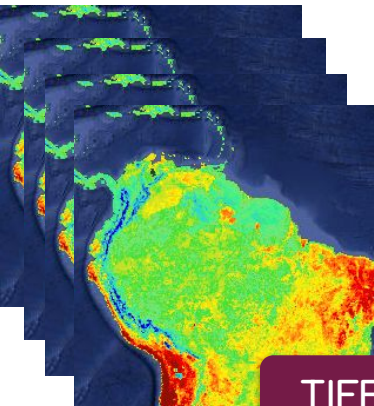
Veredas DISUELTAS por el atributo de municipio



Curación de los datos y ETL



Municipios
Datos vectoriales
georreferenciados



Temperatura Media Semestral
Datos ráster
georreferenciados

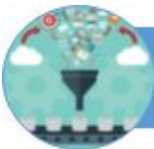


Nuevas columnas de temperatura media por municipio

MUN	VER	T2015A	T2015B
1	1	20.5	20.8
2	1	21.6	22.1
3	1	21.3	21.8

CSV

Temperatura Media Semestral por Municipio
Datos tabulares



Curación de los datos y ETL

Una columna para el
PERIODO y otra para la
TEMPERATURA



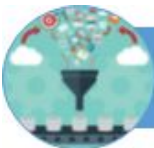
MUN	VER	T2015A	T2015B
1	1	20.5	20.8
2	1	21.6	22.1
3	1	21.3	21.8



Temperatura Media
Semestral por
Municipio

Datos tabulares

✓	✗	✓	✓
MUN	VER	PER	TEMP
1	1	2015A	20.5
1	1	2015B	20.8
2	1	2015A	21.6
2	1	2015B	22.1
3	1	2015A	21.3
3	1	2015B	21.8



Curación de los datos y ETL

JOIN



PK

MUN	PER	TEMP
1	2015A	20.5
1	2015B	20.8
2	2015A	21.6
2	2015B	22.1
3	2015A	21.3
3	2015B	21.8

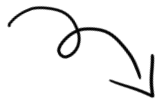
Temperatura media
semestral por municipio

PK

MUN	PER	PROD	ÁREA
1	2015A	130.4	2001
1	2015B	130.5	1987
2	2015A	123.6	3987
2	2015B	124.0	3876
3	2015A	126.1	1234
3	2015B	126.6	1345

Producción semestral por
municipio

Nueva tabla uniendo
Producción y Temperatura
por municipio y semestre



MUN	PER	PROD	ÁREA	TEMP
1	2015A	130.4	2001	20.5
1	2015B	130.5	1987	20.8
2	2015A	123.6	3987	21.6
2	2015B	124.0	3876	22.1
3	2015A	126.1	1234	21.3
3	2015B	126.6	1345	21.8

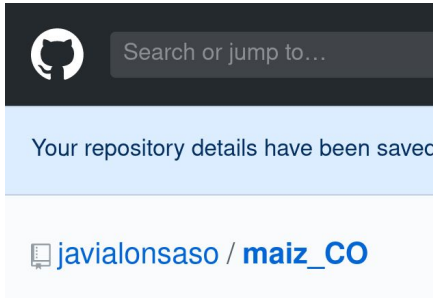
Temperatura media y
producción semestral por
municipio

CSV



Plan de preservación

- Nivel 1: La nube





Plan de preservación

- Nivel 2: HDD

Interfaz SATA



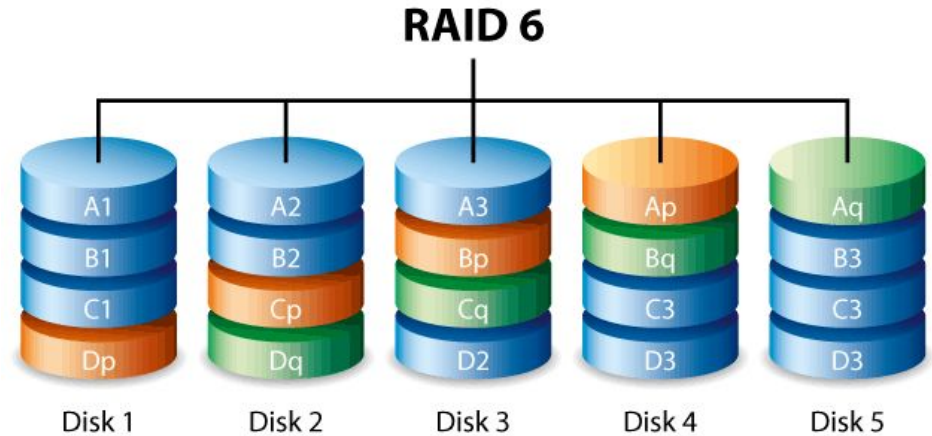
- 0.05€/GB
- Gasto eléctrico: 8-5W



Plan de preservación

- Nivel 2: RAID VI

- Alta velocidad de lectura
- Datos protegidos





Plan de preservación

- Nivel 3: Cintas magnéticas

LTO-1

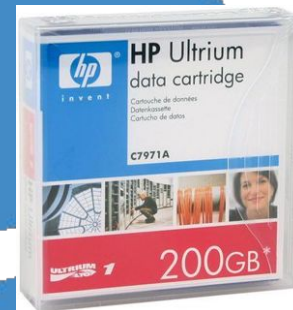
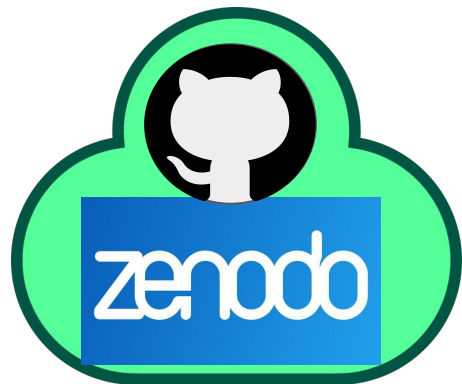


- Capacidad: 200 GB
- Coste: <20 €
- Gasto energético: 0 W



Plan de preservación

- Sistema de backups:



backup_inc01.adf

backup_inc02.adf

backup_inc03.adf

backup_inc04.adf

...

backup_full01.spf



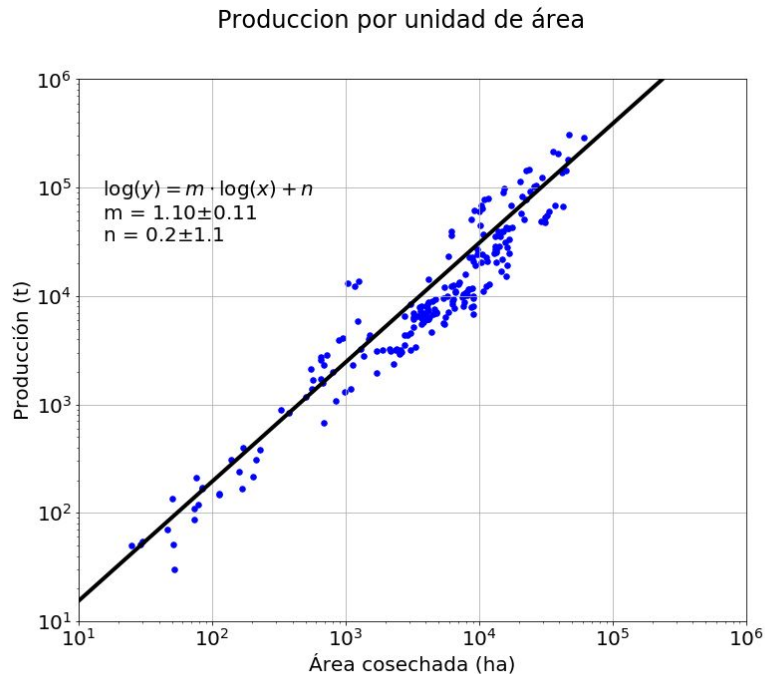
Análisis de los datos

- Relación entre área cosechada (x) y producción (y)

$$y \propto x$$

$$\log(y) = m \cdot \log(x) + n$$

“Cuanta más área, más producción”





Análisis de los datos

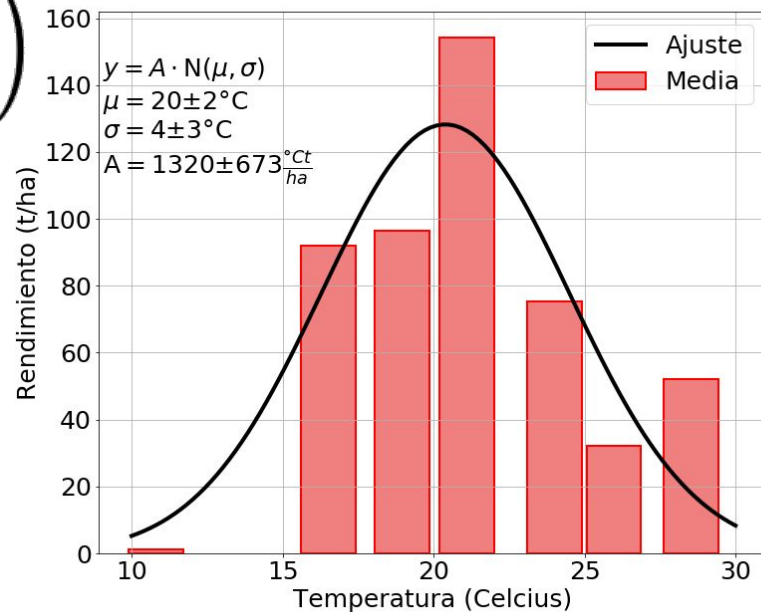
Rendimiento según la temperatura

$$y = A \cdot N(\mu, \sigma)$$

Temperatura óptima: $20 \pm 2^\circ\text{C}$

Producción: ≈ 131 toneladas

Rendimiento de la tierra





Conclusiones

- La temperatura influencia el rendimiento de la producción de maíz
- El tamaño de la cosecha no afecta a la producción
- Los datos en abierto permitirán futuros proyectos a seguir desarrollando el problema



¡Gracias
por su atención!



It ain't much, but it's honest work

INFORMACIÓN ADICIONAL



Análisis de los datos

- Representación de puntos

Rendimiento de la tierra

