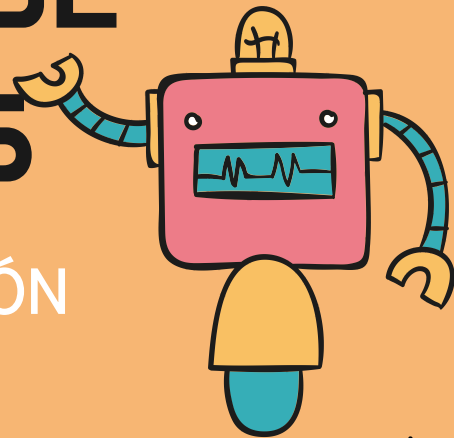
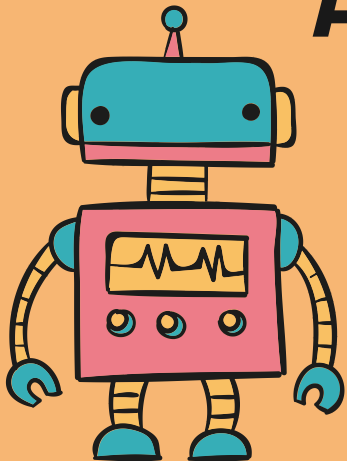


ARMADO Y USO DE BASE DE DATOS

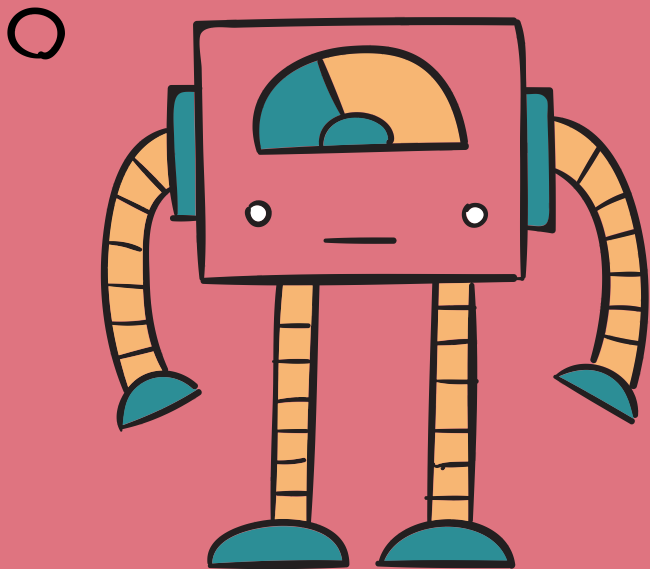
PARA PROYECTOS DE INVESTIGACIÓN



02.

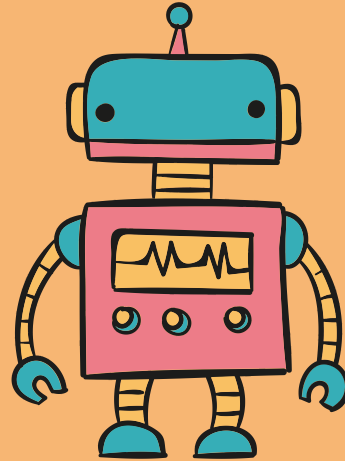
Diseño y Estructura de Bases de Datos

Repaso TP1 · Filtros · Tablas dinámicas. Clasificación de variables · Diseño conceptual. Fuentes primarias y secundarias. Caso: Base de datos de soja. Caso: BD cabras – fuente secundaria mal estructurada. Tarea asincrónica

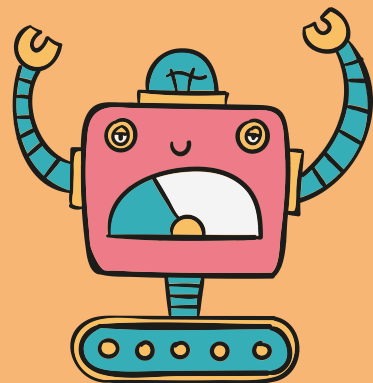


¿CÓMO LES FUE CON LA TAREA?

- ¿Pudieron transcribir la base de viales completa?
- ¿Cuántos errores de nomenclatura encontraron?
- ¿Cómo codificaron los datos faltantes?
- ¿Qué decisiones de codificación tomaron y documentaron?



edad	sexo	fecha_lesion	hora_lesion	dia	lugar	Zona	condicion_lesionado	tipo_accidente	tipo_vehiculo
7	MASCULINO	40484	0,9375	Martes	Calle	urbana	Conductor	Caída del vehículo	Bicicleta
11	MASCULINO	40493	0,90625	Jueves	Calle	urbana	Pasajero	Caída del vehículo	Camión-Camioneta
20	MASCULINO	40495	0,208333333	Sábado	Ruta	rural	Conductor	Vuelco	Automóvil
27	MASCULINO	40496	0,8125	Domingo	Ruta	rural	Conductor	Colisión vehículo	Automóvil
31	FEMENINO	40495	0,6875	Sábado	Calle	urbana	Conductor	Caída del vehículo	Bicicleta
18	MASCULINO	40499	0,625	Miércoles	Ruta	rural	Conductor	Colisión con objeto fijo	Automóvil
20	MASCULINO	40503	0,1875	Domingo	Ruta	rural	Conductor	Colisión vehículo	Moto / ciclomotor
30	MASCULINO	40503	0,3125	Domingo	Ruta	rural	Conductor	Caída del vehículo	Bicicleta
22	MASCULINO	40503	0,1875	Domingo	Calle	urbana	Conductor	Caída del vehículo	Moto / ciclomotor
28	MASCULINO	40503	0,75	Domingo	Ruta	rural	Conductor	Vuelco	Automóvil
45	MASCULINO	40503	0,75	Domingo	Ruta	rural	Pasajero	Vuelco	Automóvil
25	MASCULINO	40503	0,552083333	Domingo	Ruta	rural	Pasajero	Colisión con objeto fijo	Automóvil
7	FEMENINO	40503	0,5625	Domingo	Ruta	rural	Pasajero	Colisión con objeto fijo	Automóvil
11	FEMENINO	40509	0,854166667	Sábado	Calle	urbana	Conductor	Caída del vehículo	Moto / ciclomotor
15	MASCULINO	40510	0,125	Domingo	Ruta	rural	Conductor	Caída del vehículo	Moto / ciclomotor
5	FEMENINO	40512	0,895833333	Martes	Calle	urbana	Pasajero	Caída del vehículo	Bicicleta
1	MASCULINO	40513	0,875	Miércoles	Calle	urbana	Pasajero	Caída del vehículo	Bicicleta
27	FEMENINO	40515	0,572916667	Viernes	Calle	urbana	Conductor	Caída del vehículo	Moto / ciclomotor
8	FEMENINO	40516	0,5	Sábado	Calle		Conductor	Caída del vehículo	Bicicleta
25	MASCULINO	40516	0,916666667	Sábado	Calle	urbana	Conductor	Caída del vehículo	Moto / ciclomotor
17	MASCULINO	40517	0,71875	Domingo	Ruta	rural	Conductor	Vuelco	Camión-Camioneta
17	FEMENINO	40517	0,875	Domingo	Calle	urbana	Conductor	Caída del vehículo	Moto / ciclomotor
19	FEMENINO	40517	0,875	Domingo	Calle	urbana	Pasajero	Caída del vehículo	Moto / ciclomotor
17	MASCULINO	40523	0,177083333	Sábado	Calle	urbana	Pasajero	Caída del vehículo	Moto / ciclomotor
19	MASCULINO	40523	0,177083333	Sábado	Calle	urbana	Conductor	Caída del vehículo	Moto / ciclomotor



Repaso de Conceptos

ATRIBUTO
COLUMNA

FILA

REGISTRO

FILTROS

TABLA

edad	sexo	fecha_lesion	hora_lesion	dia	lugar	Zona	condicion_lesionado	tipo_accidente	tipo_vehiculo
7	MASCULINO	40484	0,9375	Martes	Calle	urbana	Conductor	Caída del vehículo	Bicicleta
11	MASCULINO	40493	0,90625	Jueves	Calle	urbana	Pasajero	Caída del vehículo	Camión-Camioneta
20	MASCULINO	40495	0,208333333	Sábado	Ruta	rural	Conductor	Vuelco	Automóvil
27	MASCULINO	40496	0,8125	Domingo	Ruta	rural	Conductor	Colisión vehículo	Automóvil
31	FEMENINO	40495	0,6875	Sábado	Calle	urbana	Conductor	Caída del vehículo	Bicicleta
18	MASCULINO	40499	0,625	Miércoles	Ruta	rural	Conductor	Colisión con objeto fijo	Automóvil
20	MASCULINO	40503	0,1875	Domingo	Ruta	rural	Conductor	Colisión vehículo	Moto / ciclomotor
30	MASCULINO	40503	0,3125	Domingo	Ruta	rural	Conductor	Caída del vehículo	Bicicleta
22	MASCULINO	40503	0,1875	Domingo	Calle	urbana	Conductor	Caída del vehículo	Moto / ciclomotor
28	MASCULINO	40503	0,75	Domingo	Ruta	rural	Conductor	Vuelco	Automóvil
45	MASCULINO	40503	0,75	Domingo	Ruta	rural	Pasajero	Vuelco	Automóvil
25	MASCULINO	40503	0,552083333	Domingo	Ruta	rural	Pasajero	Colisión con objeto fijo	Automóvil
7	FEMENINO	40503	0,5625	Domingo	Ruta	rural	Pasajero	Colisión con objeto fijo	Automóvil
11	FEMENINO	40509	0,854166667	Sábado	Calle	urbana	Conductor	Caída del vehículo	Moto / ciclomotor
15	MASCULINO	40510	0,125	Domingo	Ruta	rural	Conductor	Caída del vehículo	Moto / ciclomotor
5	FEMENINO	40512	0,895833333	Martes	Calle	urbana	Pasajero	Caída del vehículo	Bicicleta
1	MASCULINO	40513	0,875	Miércoles	Calle	urbana	Pasajero	Caída del vehículo	Bicicleta
27	FEMENINO	40515	0,572916667	Viernes	Calle	urbana	Conductor	Caída del vehículo	Moto / ciclomotor
8	FEMENINO	40516	0,5	Sábado	Calle		Conductor	Caída del vehículo	Bicicleta
25	MASCULINO	40516	0,916666667	Sábado	Calle	urbana	Conductor	Caída del vehículo	Moto / ciclomotor
17	MASCULINO	40517	0,71875	Domingo	Ruta	rural	Conductor	Vuelco	Camión-Camioneta
17	FEMENINO	40517	0,875	Domingo	Calle	urbana	Conductor	Caída del vehículo	Moto / ciclomotor
19	FEMENINO	40517	0,875	Domingo	Calle	urbana	Pasajero	Caída del vehículo	Moto / ciclomotor
17	MASCULINO	40523	0,177083333	Sábado	Calle	urbana	Pasajero	Caída del vehículo	Moto / ciclomotor
19	MASCULINO	40523	0,177083333	Sábado	Calle	urbana	Conductor	Caída del vehículo	Moto / ciclomotor

Clasificación de Variables:

Variables Cualitativas



Nominal

Atributos sin un orden,
puede ser binomial



Ordinal

Atributos con un orden
jerarquico/escala de valores

Variables Cuantitativas



Discreta

Conteos



Continua

Cualquier número
entero

Diseño Conceptual de la BD



Objetivo?

**Unidad de
observación?**

**Variables
respuesta?**

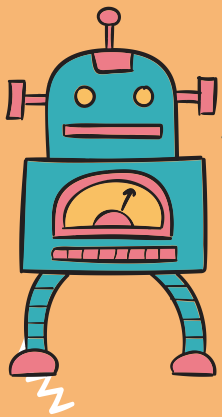
**Variables
explicatorias?**

- Persona/paciente
- Parcela/transecta
- Casa/barrio/país
- Ventas/ganancias
- Producción

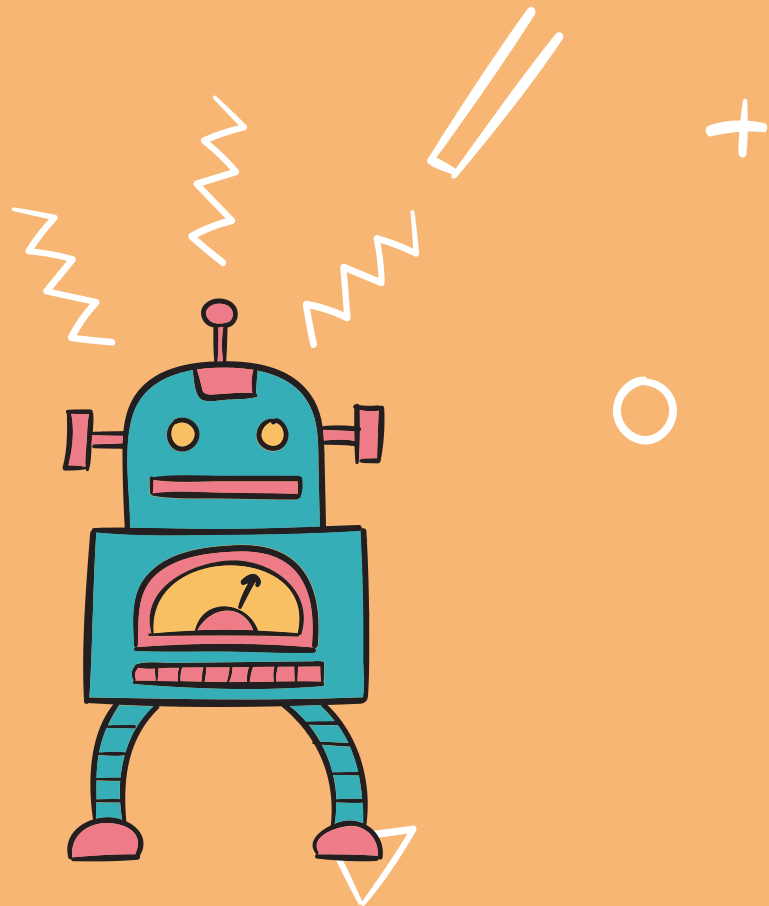
- Tipo de variable dependiente:
- Cualitativa nominal
 - Cualitativa ordinal
 - Conteos/discreta
 - Cuantitativa continua

- Variables independientes:
- Factores
 - Niveles

Responder estas cuatro preguntas antes de recolectar datos ahorra enormes problemas después



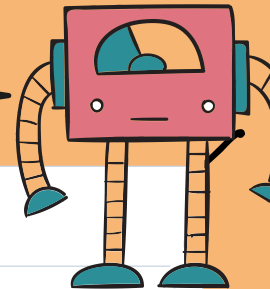
**AHORA LES TOCA
CONTAR A UDS.
SOBRE SUS
PROYECTOS Y
BASES DE DATOS**



Normalización de bases de datos

Evitar redundancia y asegurar consistencia

Normalizar es organizar los datos de manera que cada hecho se registre una sola vez, en el lugar correcto, de manera consistente. Es el principio detrás del "Tidy Data".



Sin redundancia

X

El mismo dato no se repite en múltiples columnas o filas. Si cambia un valor, solo se cambia en un lugar.

Sin ambigüedad

X

Cada valor tiene un significado claro y único. No mezclar "M", "masc." y "MASCULINO" en la misma columna.

Sin valores calculados

X

Los totales, promedios y resúmenes no van en la tabla de datos — se calculan después con funciones o tablas dinámicas.

Formato largo

✓

Preferir filas adicionales sobre columnas adicionales. Una fila por observación, una columna por variable.



Diccionario de datos y libro de códigos⁺

La documentación que hace que una BD sea comprensible para cualquiera

Un diccionario de datos describe cada variable de la base: su nombre, tipo, unidad de medida y los valores posibles o códigos que puede tomar. Es el manual de lectura de la base de datos.

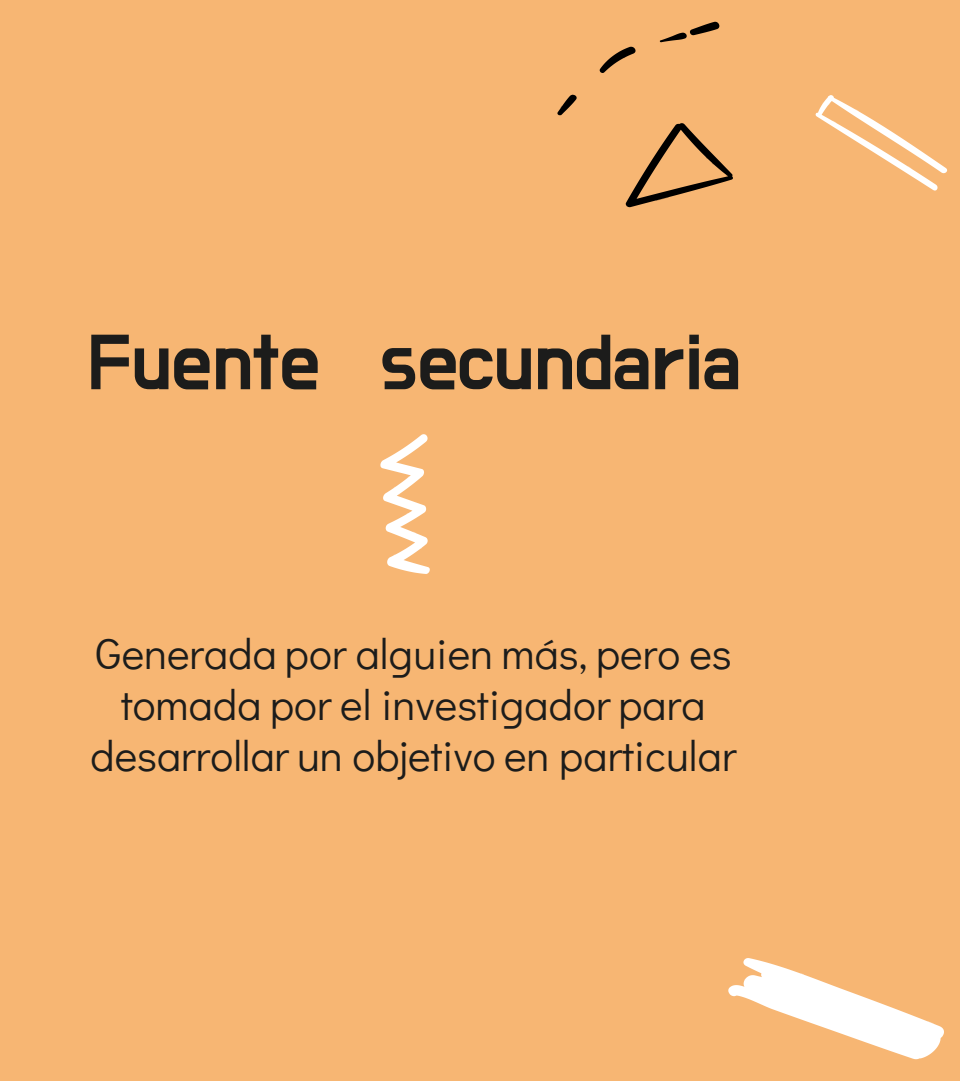
Variable	Tipo	Descripción	Valores / Códigos
<i>sexo</i>	Nominal	Sexo biológico del lesionado	MASCULINO / FEMENINO
<i>edad</i>	Discreta	Edad en años cumplidos	0 – 120 (entero)
<i>zona</i>	Nominal	Tipo de zona del accidente	urbana / rural
<i>tipo_accidente</i>	Nominal	Clasificación del tipo de accidente	Vuelco / Caída del vehículo / ...
<i>hora_lesion</i>	Continua	Hora del accidente	HH:MM (formato 24 hs)



Fuente primaria



Generada por el propio investigador
para un objetivo específico



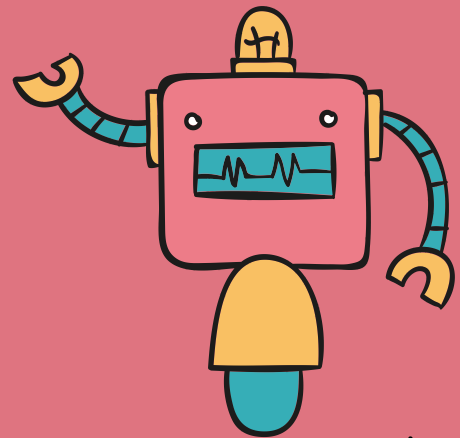
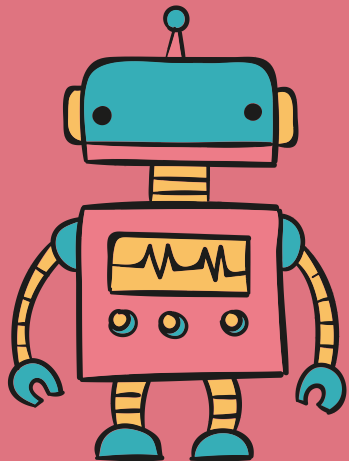
Fuente secundaria



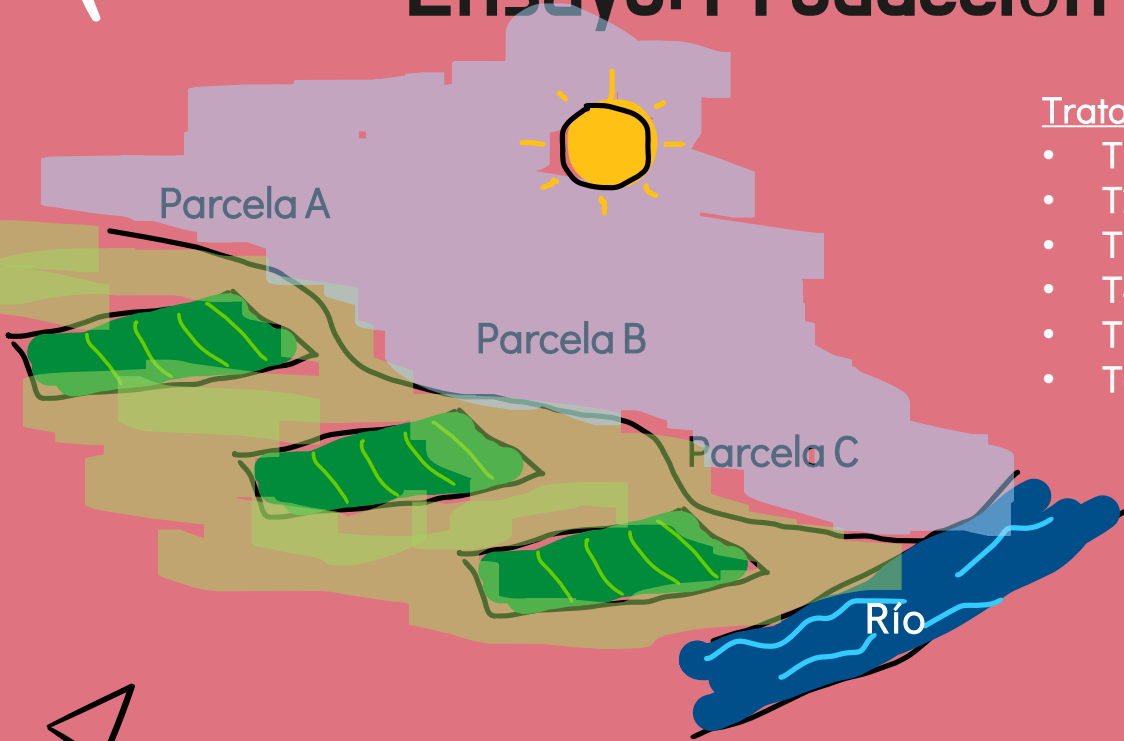
Generada por alguien más, pero es
tomada por el investigador para
desarrollar un objetivo en particular

Fuente Primaria

Ejemplo: Cultivo de soja



Ensayo: Producción de Soja

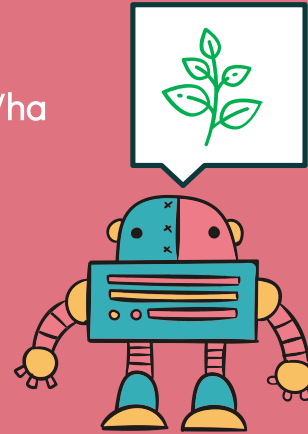


Tratamientos:

- T1: Testigo sin fertilizar
- T2: Superfosfato triple 60 kg/ha
- T3: Max Phos 50 kg/ha
- T4: Max Phos 80 kg/ha
- T5: Microstar PZ 20kg/ha
- T6: Microstar PZ 30kg/ha

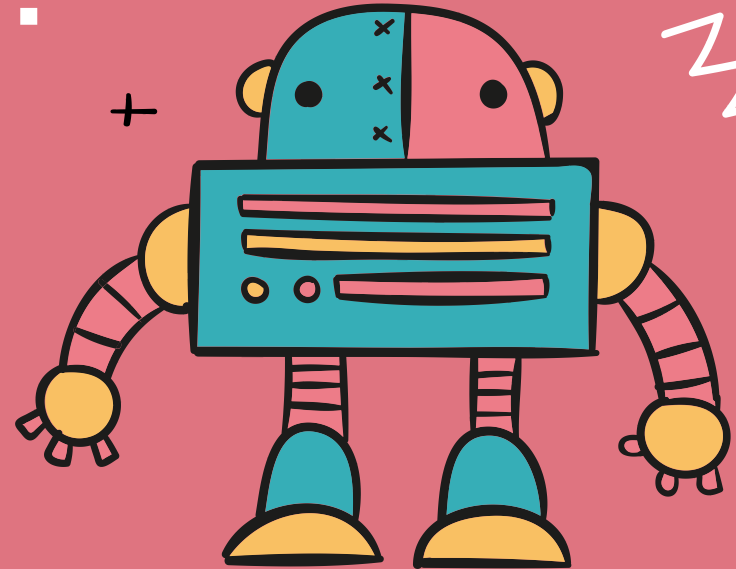
Indicadores:

- Rendimiento(kg/parcela)
- Numero de vainas/parcela
- Peso de 1000 semillas (gr)



¿Cómo armar la Base de Datos?

- Objetivo?
- Unidad de observación?
- Variables respuesta?
- Variables explicatorias?

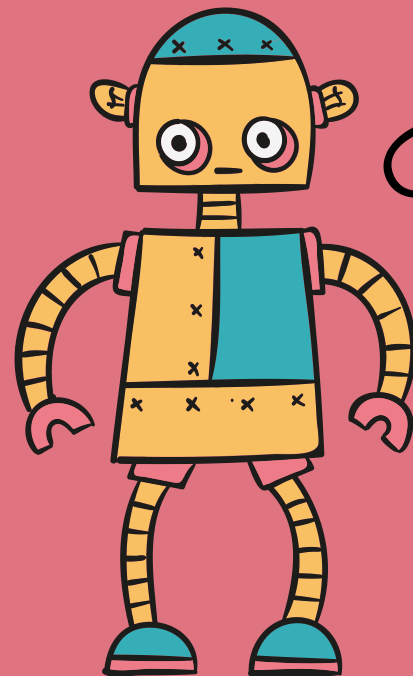


¿Cómo la organizo?

	Columna A	Columna B	Columna...	Columna n
Fila 1				
Fila 2				
Fila 3				
Fila 4				
Fila...				
Fila n				

Objetivo a alcanzar

Tratamiento	Parcela	Rendimiento (kg/ha)	Numero de vainas/parcela	Peso de 1000 semillas (gr)
T1	A	2476	7008	166
T1	B	2897	7987	173,8
T1	C	4075	9833	182
T2	A	3221	8972	173
T2	B	2909	8014	181,3
T2	C	4171	11511	181,6
T3	A	2692	7588	176,2
T3	B	3341	8384	183,6
T3	C	3798	10906	175,9
T4	A	3125	8493	179,5
T4	B	3245	8839	183,4
T4	C	3486	10167	174,1
T5	A	3197	8814	174,3
T5	B	3510	10114	173,5
T5	C	4363	11865	183,7
T6	A	2945	8087	163,9
T6	B	3293	9594	171,7
T6	C	4351	11823	180,1



Revisar tarea en el aula virtual

TP2_Actividades_Asincronicas_Unidad2

