



SECRETARÍA DE  
INNOVACIÓN

# CIENCIA DE DATOS



SECRETARÍA DE  
INNOVACIÓN



# Agenda

## Sesión 4/18

### **Recopilación de datos y limpieza.**

- Un dato
- ¿Qué es recopilación de datos?
  - Para que utilizamos la recopilación de información
  - Técnicas de recopilación de datos
- ¿Qué es la limpieza de los datos?
  - Como realizar una limpieza de datos
- Importancia del Análisis y de la Recopilación de datos
- API de twitter

## El dato

**Un dato** es un conjunto que pueden existir en cualquier tipo de formato pero no tienen un significado por si mismos.

**La información** son datos procesados que son útiles para proporcionar respuestas a preguntas como: Quien, Que, Donde y Como.

**El conocimiento** es la colección analítica y la aplicación de la información para lograr un objetivo significativo, proporcionando respuestas a las preguntas Como y Porque.

## El dato

### **Dato:**

Rojo

### **Información:**

Rojo en un semáforo

### **Conocimiento:**

Si el semáforo está en rojo no puedo cruzar la calle porque podría causar un accidente.

## ¿Qué es recopilación de datos?

Es el proceso de obtener y medir información sobre variables específicas en un sistema establecido, que luego permite responder preguntas relevantes y evaluar resultados.

La recopilación de datos es un componente de la investigación en todos los campos de estudio, incluidas las ciencias físicas y sociales, las humanidades y los negocios, etc.

Para que utilizamos la recopilación de información

La recopilación precisa de datos es esencial para mantener la integridad de la investigación.

La selección de instrumentos de recolección de datos apropiados (existentes, modificados o recientemente desarrollados) e instrucciones claramente delineadas para su uso correcto reducen la probabilidad de errores.

Para que utilizamos la recopilación de información

### **Problemas de integridad de datos**

Hay dos enfoques que pueden proteger la integridad de los datos y asegurar la validez científica de los resultados del estudio:

**Garantía de calidad:** todas las acciones realizadas antes de la recopilación de datos

**Control de calidad:** todas las acciones realizadas durante y después de la recopilación de datos



## La limpieza de datos

Es el acto de descubrimiento y corrección o eliminación de registros de datos erróneos de una tabla o base de datos. El proceso de limpieza de datos permite identificar datos incompletos, incorrectos, inexactos, no pertinentes, etc. y luego substituir, modificar o eliminar estos datos sucios ("data duty").

Después de la limpieza, la base de datos podrá ser compatible con otras bases de datos similares en el sistema.

## La limpieza de datos

### **Motivación**

Datos que son falsos o incorrectos pueden crear conclusiones falsas y dirigir a inversiones.

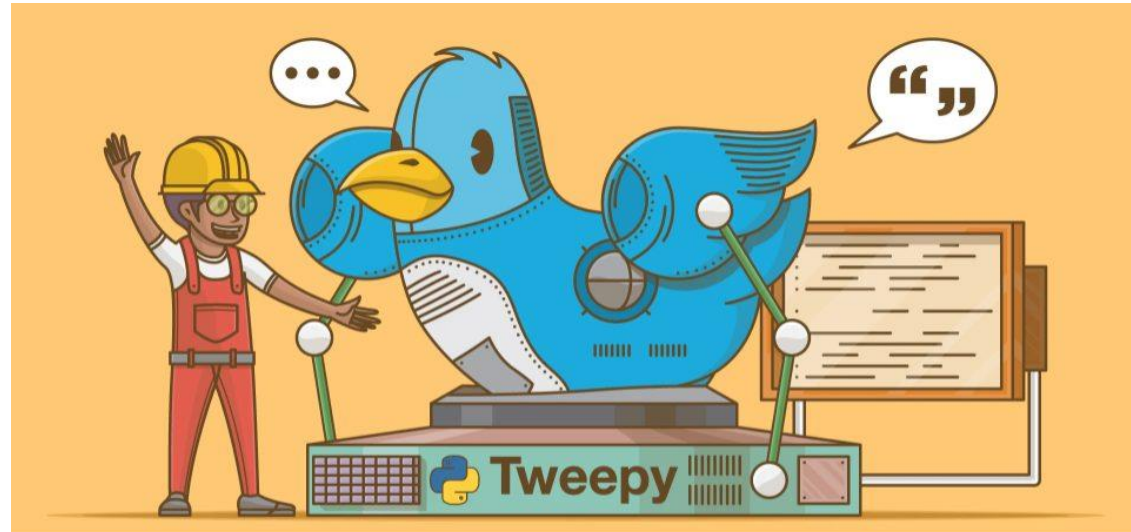
Si un gobierno quiere analizar datos de un censo de población será muy importante que los datos sean fiables para evitar tomar decisiones fiscales erróneas. En ámbitos laborales de forenses, contabilidad e investigación de fraudes usan la limpieza de datos para preparar los datos y se hace antes de mandarlos a un almacén de datos.



# Importancia de Análisis Recopilación de datos y limpieza

<https://www.youtube.com/watch?v=dOqDdGuDrOo>

## Ejemplo de API de Tweepy



**Suite de bots de Twitter en Python con Tweepy**

## Análisis de Sentimientos

Un análisis del sentimiento en redes sociales te dice cómo se sienten las personas sobre tu marca en el mundo digital.

En lugar de hacer un simple recuento de menciones o comentarios, el análisis de sentimiento engloba emociones y opiniones.

Este proceso implica recopilar y analizar información de las publicaciones que las personas comparten sobre tu marca en redes sociales.

# RESUMEN DE SESIÓN



SECRETARÍA DE  
INNOVACIÓN