

## ***Taller 03 Ciencia de datos***

### **Dataset y Hojas de Calculo**

Un dataset es un conjunto de datos. En el contexto del análisis de datos, un dataset es generalmente una colección de datos organizados de manera específica para facilitar su acceso y análisis. Pueden contener números, palabras, medidas o incluso imágenes, según el área de estudio o el propósito para el que se han recopilado.

### **Objetivos**

- Realizar actividades iniciales con dataset, para identificar sus partes, características de los datos, fuentes, formato y tipos de datos.
- Relacionar algunas herramientas con el uso de dataset relacionadas a la ciencia de datos.
- Identificar los elementos que deben tenerse en cuenta antes de la manipulación de un dataset.
- Aprender a utilizar funciones básicas en una hoja de cálculo para manipular dataset.

### **Ejercicio 1:**

Descargar un dataset público de internet (Recomendado Dataset Iris- ): un link opcional:

<https://www.kaggle.com/datasets/vikrishnan/iris-dataset>

Otros dataset: recoemndados

<https://www.kaggle.com/datasets/thegoanpanda/brain-tumor-stage-based-recurrence-patterns>

<https://www.kaggle.com/datasets/muhammad0subhan/fruit-and-vegetable-disease-healthy-vs-rotten>

<https://www.kaggle.com/code/alexisbcook/titanic-tutorial>

- a) Descargar la Metadata y/o leer toda la información acerca del dataset: Metadata  
De que se trata esa información  
Columnas y filas

Tipos de datos

Cantidad de registros

Formato etc...

- b) Abrir el DS en block de notas
- c) Abrir el DS en Excel

## **Ejercicio 2:**

### **Ejercicios Hojas de calculo**

- a) Abrir hoja de calculo – Datos asignados ( con texto, numero, fecha)  
Utilizar teclas de desplazamiento (Flechas, control, Shift...Pgdown, PgUp..)  
Ir al final de los datos  
    Identificar cuantas filas hay  
        Cuantas columnas  
Ir a una celda directamente ir a la c3000  
    Ir a la b900  
    Ir a la: a1  
Formatear tipos de datos  
    Numéricos con 2 decimales  
    Fecha con formato mes-día-año  
    Comunas de nombres o empresa en minuscula

#### **b) de 2 columnas numéricas calcular de cada una:**

- PROMEDIO
- MEDIANA:
- MÁXIMO: .
- MÍNIMO:
- CONTAR:
- SUMA:
- MODA

### **C) Operaciones**

#### **FILTROS**

Crear filtros en columnas buscando un nombre específico (una sola columna)

    Describe lo que hace el filtro

Crear un filtro por 2 columnas

    Describe lo que hace el filtro

## ORDENACION

Tome la hoja de datos y ordene ascendente por el primer campo alfabético. – **Revisas**

Tome la hoja y ordenela por un campo numérico: **Revisas**

Ordene por 2 campos simultáneos, Uno numérico y otro de texto - **Revisas**

### d) REEMPLAZAR

Todas las “ñ”, reemplazarlas por “ng”

Modificar los puntos en el nit “.” Por un guion “-”

Las celdas que no tengan NIT, llenarlo con “SIN DEFINIR”

### e) Copiar las el 20% de filas en otra hoja

Convertir los datos de esa nueva hoja en tabla

## Operaciones de búsqueda en Excel

### \*Funciones de búsqueda

1. VLOOKUP (Buscar y reemplazar valor): Busca un valor en una tabla y devuelve el valor correspondiente en una columna específica.

2. INDEX-MATCH (Buscar y reemplazar valor): Busca un valor en una tabla y devuelve el valor correspondiente en una columna específica.

3. SEARCH (Buscar texto): Busca un texto dentro de una celda o rango de celdas.

4. FIND (Buscar texto): Busca un texto dentro de una celda o rango de celdas.

### Funciones de reemplazo

1. REPLACE (Reemplazar texto): Reemplaza un texto por otro en una celda o rango de celdas.

2. SUBSTITUTE (Reemplazar texto): Reemplaza un texto por otro en una celda o rango de celdas.

3. LEN (Longitud de texto): Devuelve la longitud de un texto en una celda o rango de celdas.

### #LARGO

### Ejemplos

#### 1. Buscar y reemplazar un valor:

Supongamos que tienes una tabla con los nombres de los empleados y sus respectivos departamentos. Quieres buscar el nombre "John" y reemplazarlo por "Juan".

=VLOOKUP(A2,B:C,2,FALSE) # BUSCARV, BUSCAR

Donde A2 es la celda con el nombre "John", B:C es la tabla con los departamentos, y 2 es la columna que contiene los departamentos.

## 2. Buscar un texto:

Supongamos que tienes una celda con un texto y quieres buscar la palabra "Hola".

=SEARCH("Hola",A1)

Donde A1 es la celda con el texto.

## 3. Reemplazar un texto:

Supongamos que tienes una celda con un texto y quieres reemplazar la palabra "Hola" por "Adiós".

=REPLACE(A1,"Hola","Adiós") #REEMPLAZAR

Donde A1 es la celda con el texto.

## Referencias:

<https://keepcoding.io/blog/que-son-datasets/>

<https://archive.ics.uci.edu/dataset/53/iris>

<https://thedataschools.com/que-es/data-set/>

<https://dominaexcel.org/lo-basico-que-debes-saber-de-excel/>

<https://academiaaprenderhaciendo.com/introduccion-a-excel/#:~:text=Conceptos%20Clave%3A%201%201.%20Celdas%20y%20Rangos%3A%20,F%C3%B3rmulas%20y%20Funciones%3A%203%203.%20Hojas%20de%20C%C3%A1lculo%3A>

Fecha Creación	Enero 25 2024
Responsable	Plinio Neira Vargas
Revisado por	Sonia Escobar
Fecha Revisión	Febrero 10 2024