

MÀSTER DE FORMACIÓ PROFESSIONAL

INTEL·LIGÈNCIA ARTIFICIAL I BIG DATA

TITULACIÓ OFICIAL FP

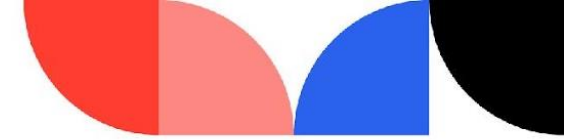
**Combinar diferentes
fuentes y tipos de
datos.**

**Análisis con
estadísticas
descriptiva.**

**Clase
2**

Objetivo

El objetivo de la clase de hoy es, a través del uso de estadísticas descriptivas, se haga una primera aproximación al entendimiento del dato y su tratamiento. Para luego, empezar a combinar diferentes tipos de datos y ver qué insights se pueden crear desde la indagación inicial.



Estructura de la classe

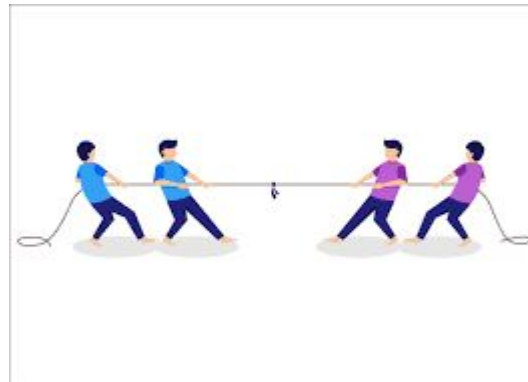
Parte 1:

Teoría y entendimiento técnico.



Parte 2:

Puesta en práctica en grupos de trabajo.



Parte 3:

Exposición de resultados por grupos de trabajo.



Estadística Descriptiva

La estadística descriptiva se ocupa de **describir, resumir y organizar** los datos de manera efectiva.

Nos ayuda a entender las **características básicas** de un conjunto de datos mediante medidas numéricas simples y gráficos.

Los conceptos principales incluyen la **media, mediana, moda, varianza, desviación estándar**, entre otros.



Media (Promedio)

Suma de todos los valores dividido por el número de observaciones.

$$\text{Media}(\mu) = \frac{1}{n} \sum_{i=1}^n x_i$$

Ejercicio: Sacar la media de la edad de la clase.



Mediana

El **valor central** en un conjunto de datos ordenados. Si el número de observaciones es par, es el promedio de los dos valores centrales.

Ejercicio: Sacar la mediana de la edad de la clase.



Moda

El valor o valores que aparecen con mayor **frecuencia** en un conjunto de datos.

Ejercicio: Sacar la moda de la edad de la clase.



Varianza

Varianza: Medida de la dispersió que indica qué tanto varían los datos respecto a la media.

$$\text{Varianza}(\sigma^2) = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

Ejercicio: Sacar la varianza de la edad de la clase.



Desviación estándar

Desviación estándar: Raíz cuadrada de la varianza. Indica cuánto se dispersan los datos en torno a la media.

$$\text{Desviación estándar}(\sigma) = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2}$$

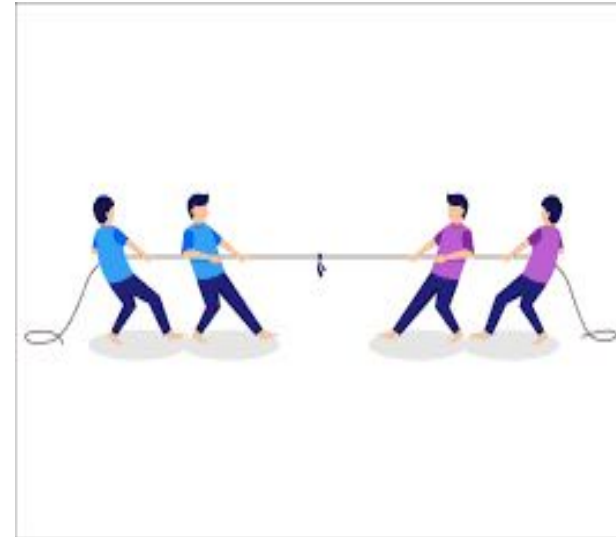
Ejercicio: Sacar la desviación estándar de la edad de la clase.



Ejemplo Práctico

<https://www.kaggle.com/code/carolinamorenostucom/cardiogoodfitness-descriptive-statistics>

Conceptos que se trabajan: Combinar diferentes tipos de datos y hacer inferencia básica con estadísticas descriptivas

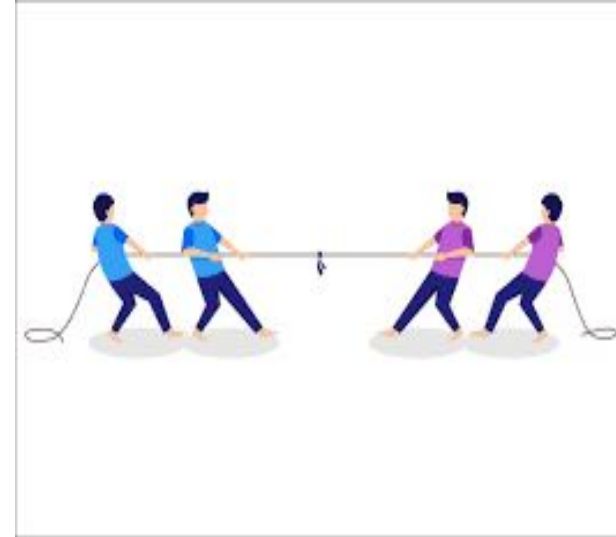


Caso Práctico

<https://www.kaggle.com/code/carolinamorenostucom/cardiogoodfitness-descriptive-statistics>

Conceptos que se trabajan: Poder a prueba conocimiento adquirido

En los grupos de trabajo explorar la base de datos desde python o desde excel y preparar un story telling de los datos usando los conceptos claves.





Gracias