

# parcial\_1

March 9, 2024

## 1 Examen Parcial 1

### 1.1 Instrucciones

Vamos a generar un analisis utilizando un conjunto de datos de factores de riesgo para la enfermedad cardíaca cardiovascular.

Solo debemos escribir el codigo que se nos pida en las celdas que digan `## tu codigo aqui`

### 1.2 Cargando la data

#### 1.2.1 1. Cargamos el set de datos

Cargamos la data que se encuentra en el archivo CSV `heart_data.csv`, para esto utilizamos `pandas`.

Este conjunto de datos contiene información detallada sobre los factores de riesgo de enfermedad cardiovascular. Incluye información sobre edad, sexo, altura, peso, valores de presión arterial, niveles de colesterol, niveles de glucosa, hábitos de tabaquismo y consumo de alcohol de más de 70 mil personas. Además, indica si la persona es activa o no y si tiene alguna enfermedad cardiovascular. Este conjunto de datos proporciona un gran recurso para que los investigadores apliquen técnicas modernas de aprendizaje automático para explorar las posibles relaciones entre los factores de riesgo y las enfermedades cardiovasculares que, en última instancia, pueden conducir a una mejor comprensión de este grave problema de salud y diseñar mejores medidas preventivas.

Imprimimos un sample de la data

```
[11]:
```

	index	id	age	gender	height	weight	ap_hi	ap_lo	cholesterol	gluc	\
0	0	0	18393	2	168	62.0	110	80	1	1	
1	1	1	20228	1	156	85.0	140	90	3	1	
2	2	2	18857	1	165	64.0	130	70	3	1	
3	3	3	17623	2	169	82.0	150	100	1	1	
4	4	4	17474	1	156	56.0	100	60	1	1	

	smoke	alco	active	cardio
0	0	0	1	0
1	0	0	1	1
2	0	0	0	1
3	0	0	1	1
4	0	0	0	0

### 1.2.2 2. Cargamos el diccionario de datos

El diccionario de datos se encuentra en el archivo JSON `diccionario_datos.json`

Utilizar un ciclo for para imprimir los valores de el diccionario de datos.

**Ayuda:** utilizar iterar sobre `data_dict.items()` [ejemplo](#)

Age: Age of participant, its in days (integer)  
Gender: Gender of participant (male/female).  
Height: Height measured in centimeters (integer)  
Weight: Weight measured in kilograms (integer)  
Ap\_hi: Systolic blood pressure reading taken from patient (integer)  
Ap\_lo: Diastolic blood pressure reading taken from patient (integer)  
Cholesterol: Total cholesterol level read as mg/dl on a scale 0 - 5+ units( integer). Each unit denoting increase/decrease by 20 mg/dL respectively.  
Gluc: Glucose level read as mmol/l on a scale 0 - 16+ units( integer). Each unit denoting increase Decreaseby 1 mmol/L respectively.  
Smoke: Whether person smokes or not(binary; 0= No , 1=Yes).  
Alco: Whether person drinks alcohol or not(binary; 0 =No ,1 =Yes ).  
Active: whether person physically active or not( Binary ;0 =No,1 = Yes ).  
Cardio: whether person suffers from cardiovascular diseases or not(Binary ; 0-no , 1=yes ).

### 1.2.3 Analisis de datos

Como podemos ver en el diccionario de datos la variable `age` esta en dias. Vamos a tranformar esta variable en años y generar ciertos analisis. En la siguiente celda se guarda un arreglo de Numpy con las edades.

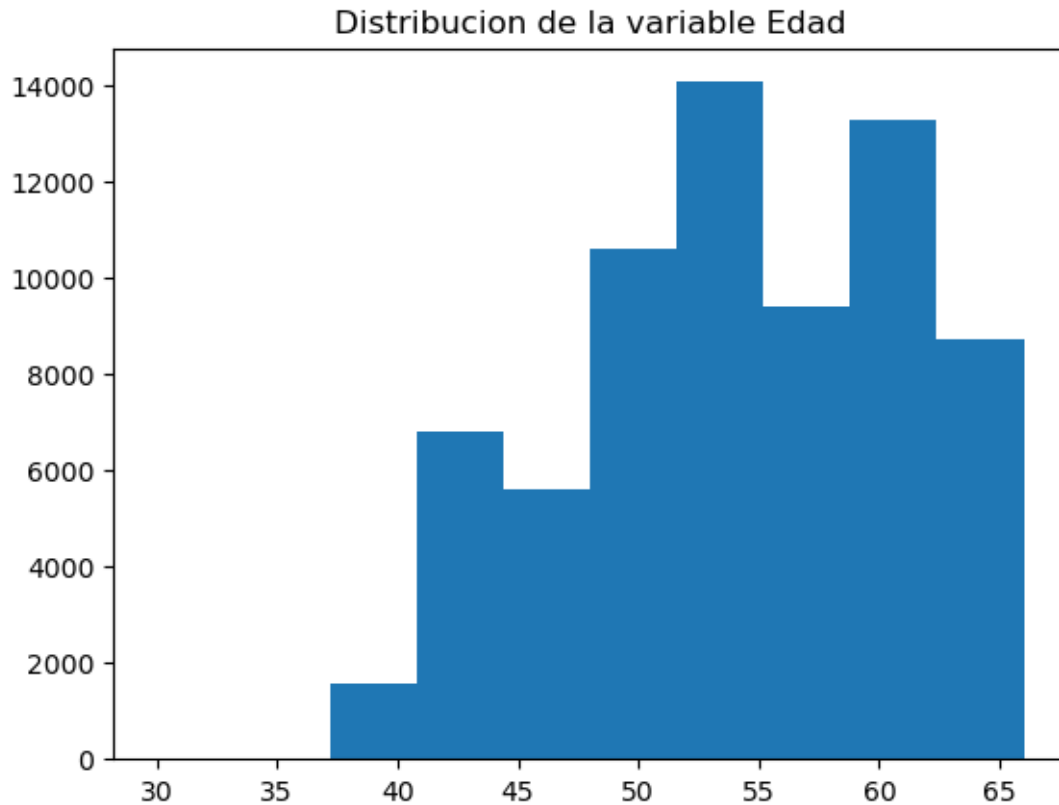
La variable `edad_dias` es un arreglo de numpy con las edades en dias. Ahora debemos transformar estas edades en años y guardar estas edades en una variable llamada `edad`, muy importante la edad en años debe ser un numero entero.

**Ayuda:** El operador `/` en numpy hace una operacion elemento por elemento **Ayuda:** Para aproximar un arreglo de numpy podemos usar la funcion de numpy `rint()`, ver [ejemplos](#)

Calculemos la media de la variable edad

**Ayuda:** `numpy.mean()`

[36]: 54.0779



Vamos a cargar los datos de las siguientes variables en arreglos de numpy. - Alco - Active - Cardio

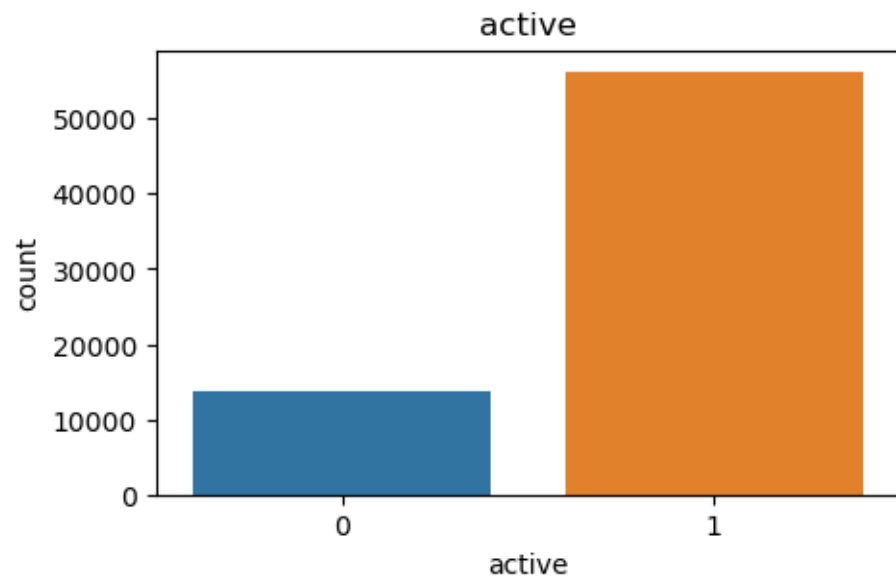
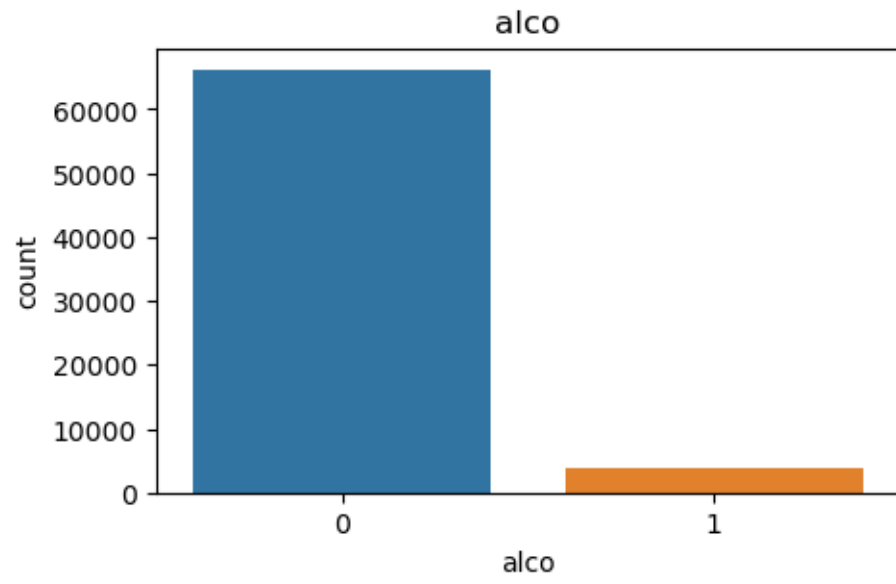
Como ya tenemos el diccionario de datos podemos consultar ahí que significa cada una. Consultemos el diccionario para entender mejor que son, debemos imprimir la explicación de la variable, las variables que están listadas en esta celda son las llaves del diccionario:

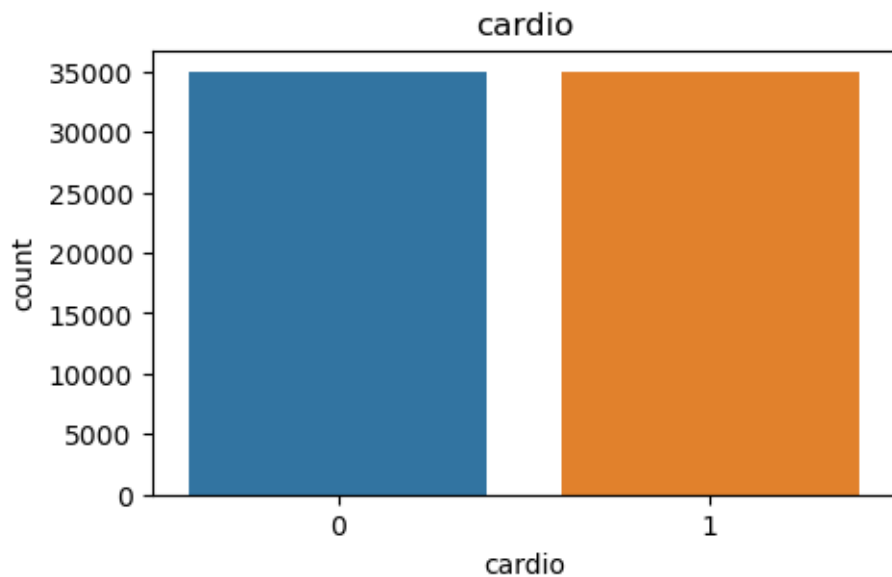
Vamos a utilizar las variables `alco`, `active`, `cardio` para responder las siguientes preguntas: - ¿Cuántas personas en este conjunto de datos beben alcohol? - ¿Cuántas personas tienen actividad física? - ¿Cuántas personas sufren enfermedades cardiovasculares?

**Ayuda:** podemos utilizar un ciclo `for`

**Respuestas:** Escribe en markdown las respuestas de las preguntas

En la siguiente celda hay un código que va a graficar los conteos con esto vamos a poder validar los resultados.





### 1.3 Analisis de datos

A continuación vamos a graficar los conteos de las variables, pero ahora vamos a ver los conteos agrupando los datos por una enfermedad cardiovascular. Esta parte es solo de ejecutar el código y escribir una conclusión en Markdown

