



# Introducción al Análisis de Datos

Clase 1

Prof. Natalia Lucero





# ¿Qué es el Análisis Exploratorio de Datos? (E.D.A)

Es el proceso determinado por el científico [John W. Turkey](#) para su tratamiento estadístico de muestras recogidas durante la investigación.

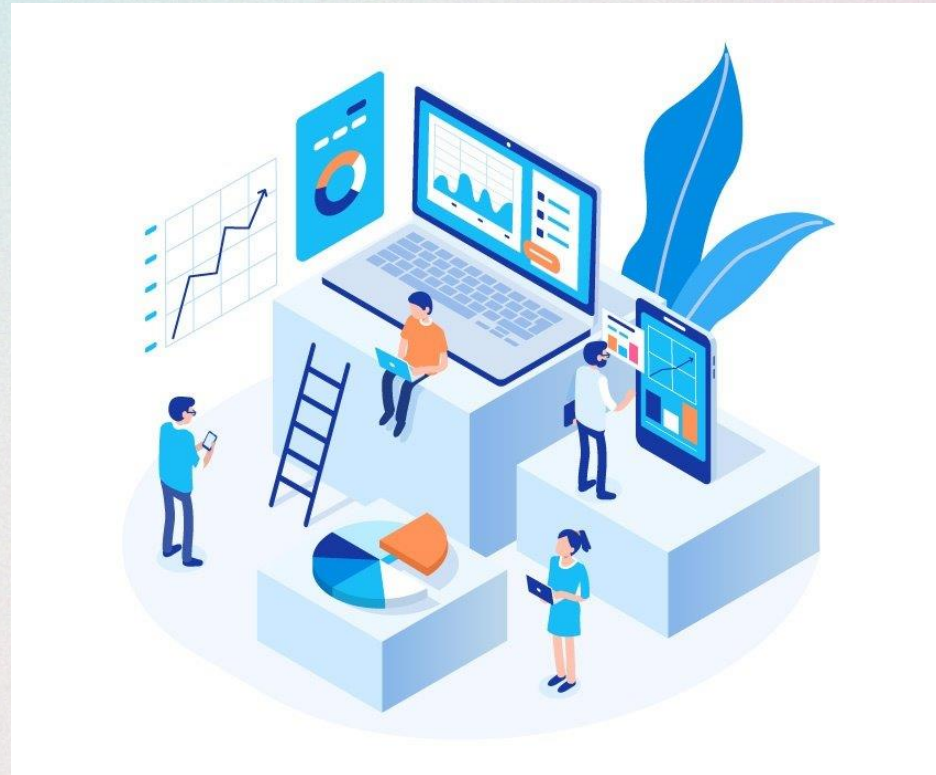


# El Análisis Exploratorio de Datos (EDA)

Es el proceso de conocer en detalle, darle sentido a los datos, visualizarlo por primera vez y ver como se comportan, cuantas variables existen, de que tipo son, indagar en ellos haciendo preguntas. En este proceso vamos a determinar como tratarlos, interrogarlos para obtener las respuestas necesarias y de esa manera transformarlo en información útil para la empresa o para resolver el problema al que te enfrentas.



# Los procesos básicos para un E.D.A





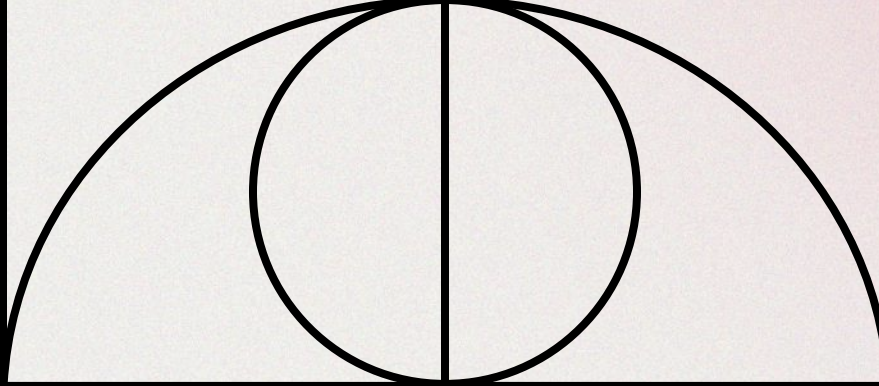
- **Transformar los datos:** Nos sirve para saber qué hacer frente a valores nulos, faltantes o datos atípicos. Además de establecer si hay necesidad de reducir la dimensionalidad de datos.
- **Visualizar:** Utilizar alguna herramienta para hacer una representación gráfica de los datos.
- **Analizar e Interpretar:** Analizar e interpretar los datos a través de diferentes visualizaciones.
- **Documentar:** Documentar todas las gráficas y estadísticas generadas



# ¿Cómo hacer un análisis exploratorio de datos?







# **Por qué deberías realizar un Análisis Exploratorio de Datos?**

**Organizar y entender las variables**

**Lograremos identificar los distintos tipos de variables, las categorías a las que pertenecen y por consiguiente el tipo de análisis que podemos realizar con ellas.**





## • Establecer relaciones entre variables

- Lograr establecer relaciones entre las variables, por ejemplo, ver como una variable se modifica si la grafico junto a otra,

## • Encontrar patrones ocultos en los datos.

- Encontrarás patrones que de no haber hecho las visualizaciones y la exploración de los datos no hubieran sido visibles a simple vista.
- Este será el punto donde buscaremos encontrar los **data insights**



## **Agudarte a escoger el modelo correcto para la necesidad correcta**

Al empezar con un conjunto de datos podemos pensar que elegiremos algún modelo que prediga X cosa, pero no obstante sin antes identificar que variables sirven, cuáles no sirven, como se encuentran relacionadas, cuáles son las variables que mejor se ajustan a algún tipo de modelo u otro, lo más probable es que al construir el modelo tengamos un resultado deficiente.

El hacer el EDA nos ayudará a encontrar el modelo correcto para la necesidad correcta.

## **Aguda a la toma de decisiones basadas en datos**

La toma de decisiones basadas en los datos será uso de hechos, métricas y datos para guiar decisiones estratégicas que se alineen con las metas, objetivos y las iniciativas de nuestra organización.



# ¿Cuáles son los pasos para realizar un análisis exploratorio de datos?

- **PASO 1: Hacer preguntas**
- Una vez visualizados los datos, debemos empezar a preguntarnos sobre ellos.
- ¿Que te gustaría encontrar?
- Que quisieras saber de los datos?
- Cuál es la razón para actuar el análisis?
- Una vez tengamos nuestras preguntas, podremos generar una guía para nuestro proceso de exploración de datos



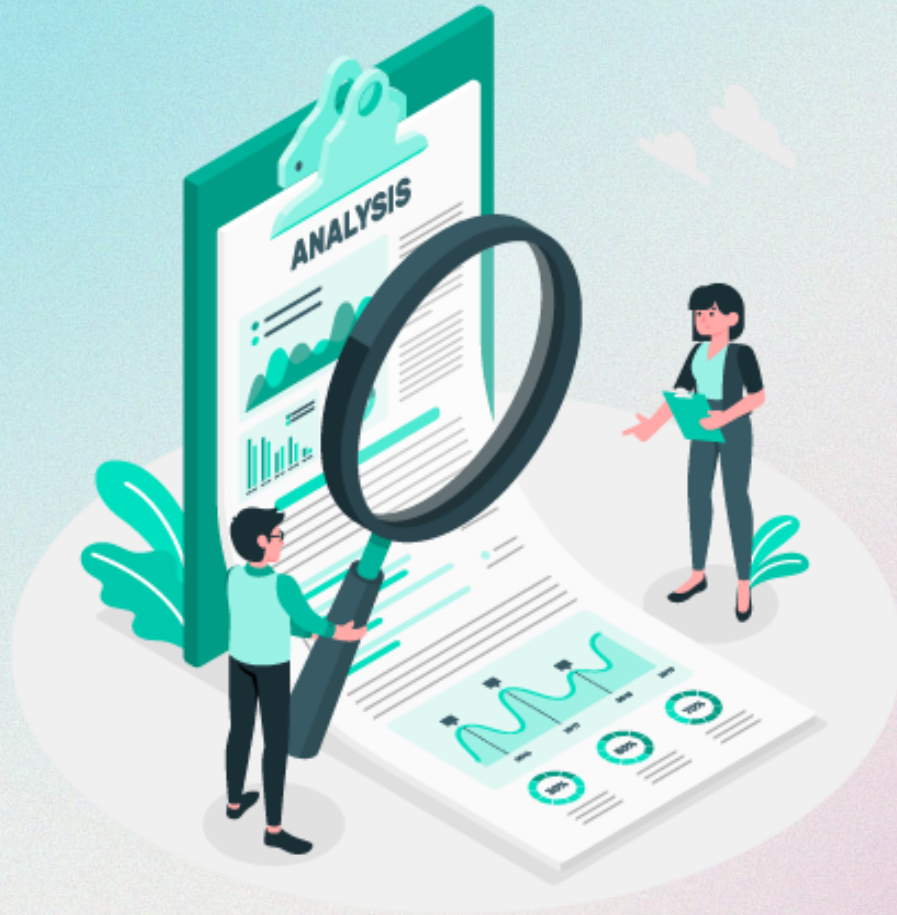
## PASO 2: Determinar el tamaño de los datos

- ¿Cuántas observaciones existen?
- ¿Cuántas variables hay?
- ¿Necesito todas las observaciones?
- ¿Necesito todas las variables?





# PASO 3: Categorizar las variables



¿Cuántas variables categóricas existen?

¿Cuántas variables continuas existen?

¿Cómo puedo explorar cada variable dependiendo de su categoría?



## PASO 4: Limpieza y validación de datos



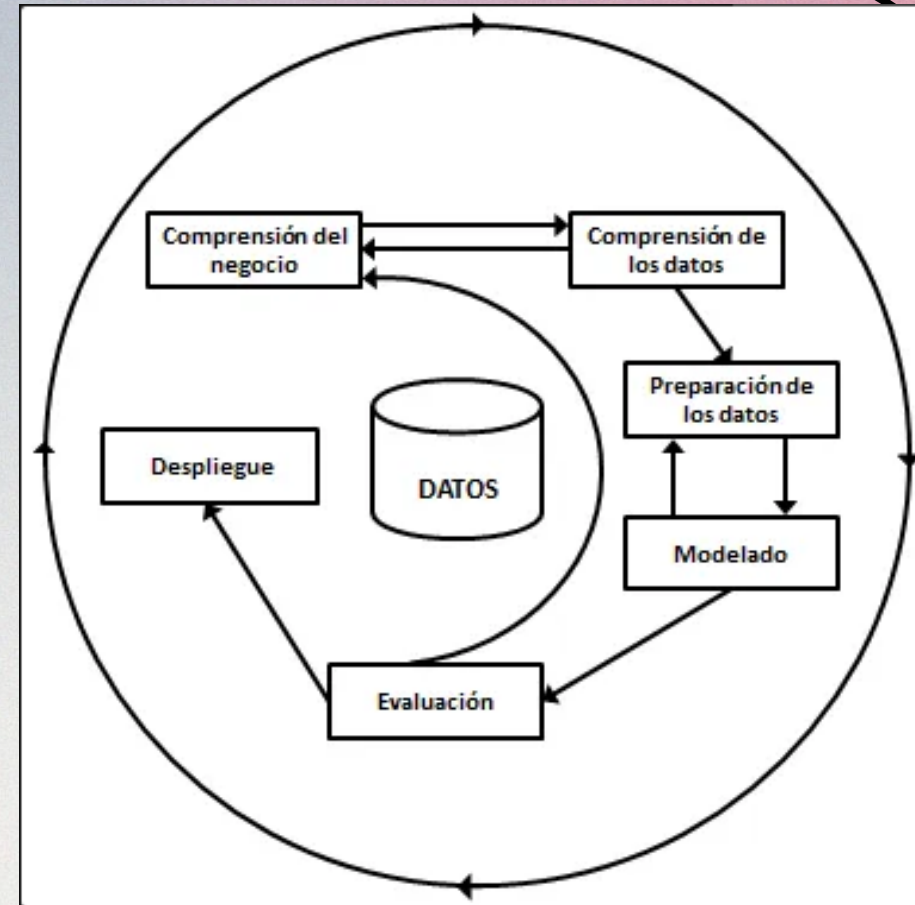


# PASO 5: Establecer relaciones

- ¿Existen algún tipo de relación entre mi variable  $X$  y  $Y$ ?
- ¿Qué pasa si ahora considero a la variable  $Z$  en el análisis?
- ¿Qué significa que las observaciones se agrupen?
  - ¿Qué significa el patrón que se observa?



Este proceso es cíclico. A pesar de que pueda parecer infinito, este proceso en algún momento debe salir del ciclo y continuar para obtener algún tipo de valor





## En algún momento debes romperlo y continuar





# Programación Orientada a Objetos

```
Inicio  
Leer N // número de datos  
// llamada a procedimientos:  
Leer lista de números( N, lista )  
calcular promedio( promedio, N, lista )  
calcular desviación estándar( desviación estándar, promedio, N, lista )  
escribir desviación estándar  
Fin
```



# ¿QUÉ ES LA POO?

Es un paradigma de la programación y se basa justamente en objetos.

Es una forma en la que vamos a poder optimizar nuestras aplicaciones.





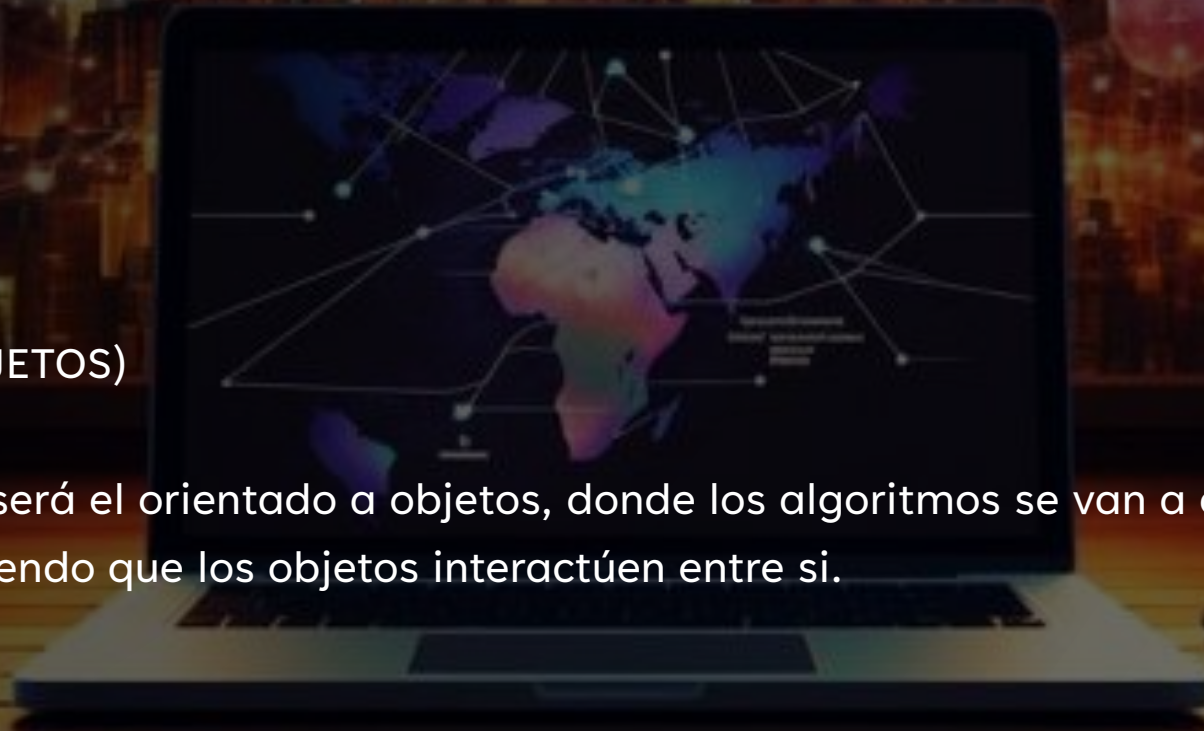
# FUNCIÓN DEL MODELO Y DISEÑO ORIENTADO A OBJETOS

En la informática vamos a tener muchos modelos para la resolución de problemas

Tenemos 4 categorías :

- Lenguaje Imperativos
- Funcionales
- Declarativos
- POO (ORIENTADA A OBJETOS)

El que veremos nosotros será el orientado a objetos, donde los algoritmos se van a expresar definiendo objetos y haciendo que los objetos interactúen entre si.





# ¿QUÉ SON LOS OBJETOS?

Estos objetos son cosas que deben ser manipuladas y existen en el mundo real:

- . Edificios
- Artefactos
- Seres humanos





# Los objetos pueden tener :

- Sus Características dentro de este paradigma de programación
- Sus Funcionalidades, ósea que es lo que va a hacer un objeto.





# Los objetos van a tener dos características

–**Datos:** van a ser las características que este objeto va a tener, también llamados atributos.

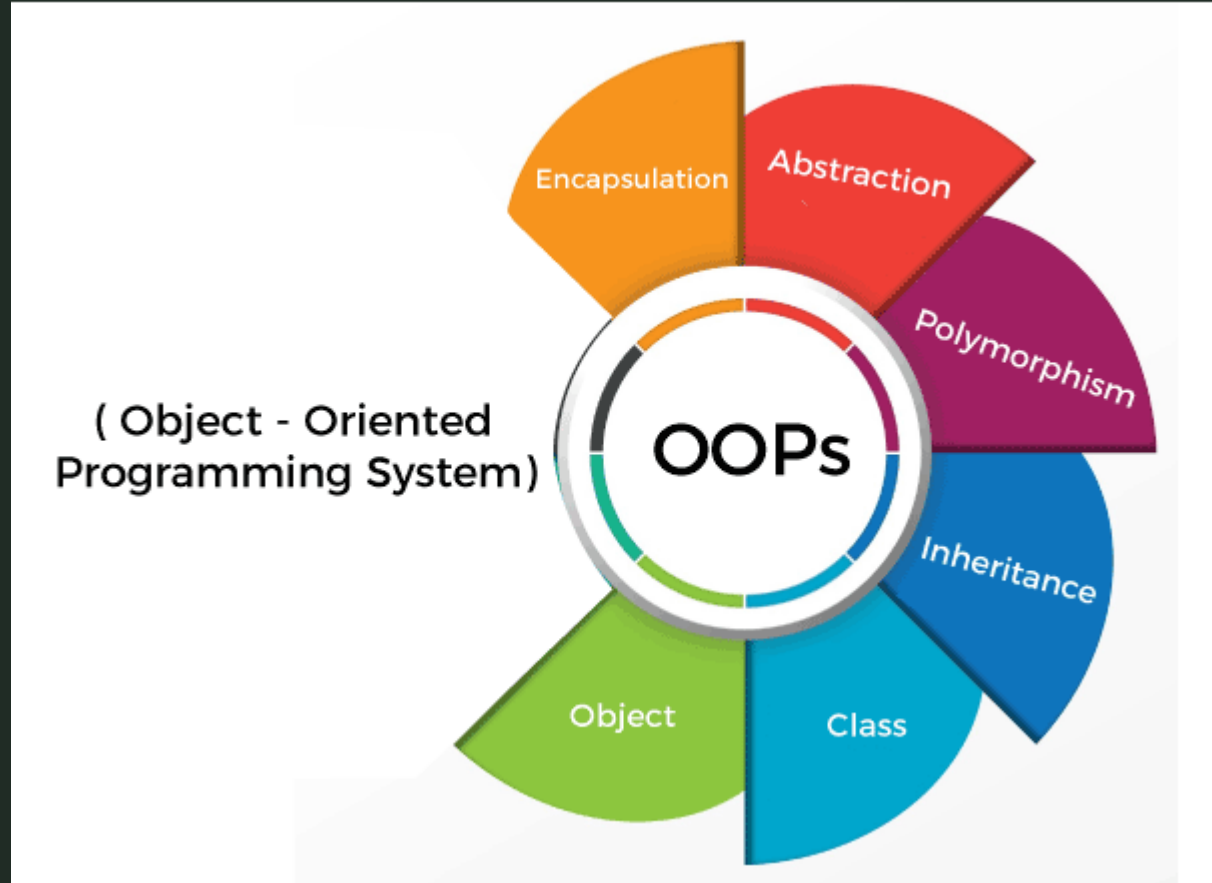
–**Funcionalidad:** se refiere a lo que va a estar haciendo el objeto. En código se lo llamaría métodos de una clase.





# Ventajas de trabajar con POO

- Es muy útil al trabajar en equipo ya que es muy fácil poder adaptar nuestro proyecto a nuevas actualizaciones y también les permite organizar mejor los códigos para añadir nuevas funcionalidades.





# Una clase : son los moldes de los objetos

- Es una representación de un código de lo que es la POO
- Se podría decir que es como un molde o una plantilla de un objeto, si se obtiene una galleta , cada galleta tendrá sus propias particularidades como chispas de chocolate o trocitos de nuez.





# Una clase tiene tres partes importantes:

Nombre de la clase

Atributos (variables)

Métodos (Funciones)





# Nombre de la clase

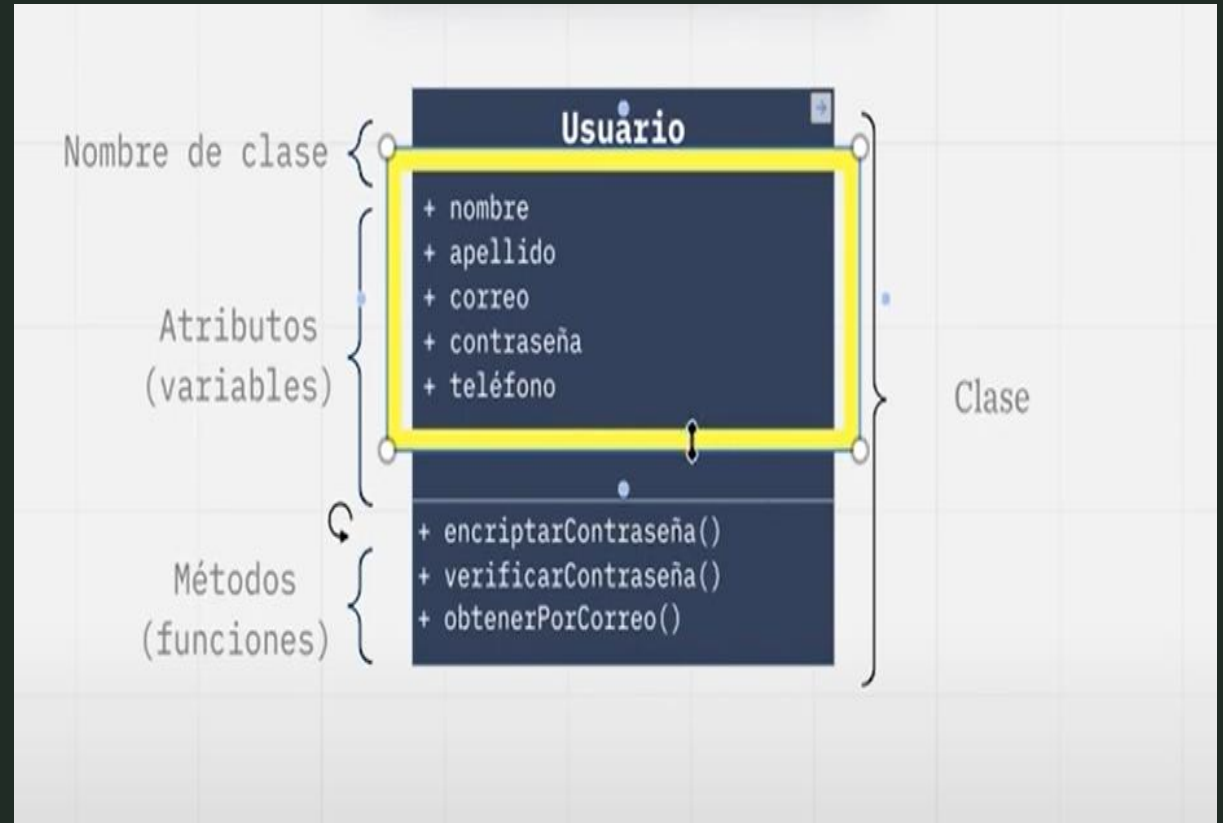
- Es muy importante para una buena práctica de programación.
- Los nombres de las clases deben representar su propósito.
- Debe describir su razón de clase





# Datos: (Atributos)

- Son las características de una clase , en programación serían variables.
- Los datos deben ser lo óptimos posibles.





# Método: (Funciones)

- Es lo que nuestra clase va a estar usando.
- Por ejemplo : validando un usuario y su contraseña





# Ejemplo en JavaScript

```
class Usuario {  
    constructor(nombre, apellido, correo, contraseña, telefono) {  
        this.nombre = nombre;  
        this.apellido = apellido;  
        this.correo = correo;  
        this.contraseña = contraseña;  
        this.telefono = telefono;  
    }  
  
    encriptarContraseña() {  
        // Lógica para encriptar la contraseña  
    }  
  
    verificarContraseña(contraseña) {  
        // Lógica para verificar si la contraseña ingresada es correcta  
    }  
  
    obtenerPorEmail(correo) {  
        // Lógica para obtener un objeto Usuario a partir de su correo electrónico  
    }  
}
```



# Ejemplo en Python

```
class Usuario:
    def __init__(self, nombre, apellido, correo, contraseña, telefono):
        self.nombre = nombre
        self.apellido = apellido
        self.correo = correo
        self.contraseña = contraseña
        self.telefono = telefono

    def encriptarContraseña(self):
        # encriptar la contraseña
        pass

    def verificarContraseña(self, contraseña):
        # verificar la contraseña del usuario
        pass

    def obtenerPorEmail(self, correo):
        # obtener usuario por email
        pass
```