

ACTIVIDAD 6
PROYECTO FINAL
BASE DE DATOS

WALTER GARCÍA SALAZAR
CÓDIGO: 100107707

Profesor
IVÁN TRUJILLO

CORPORACIÓN UNIVERSITARIA IBEROAMERICANA
FACULTAD DE INGENIERÍA
INGENIERÍA DE SOFTWARE VIRTUAL
BOGOTÁ D.C.
2023

MANEJO DE BASE DE DATOS

HERRAMIENTAS

- Python 3.X
- Pandas

INTRODUCCIÓN

Pandas es una poderosa biblioteca en Python diseñada para facilitar la manipulación y análisis de datos. A continuación, se detallan algunas de las características clave y conceptos importantes asociados con Pandas:

1. Estructuras de Datos Principales:

a. DataFrame:

- Es una estructura bidimensional tabular con etiquetas en las filas y columnas.
- Se puede pensar en él como una hoja de cálculo o una tabla SQL.
- Permite almacenar y manipular datos heterogéneos de manera eficiente.

```
import pandas as pd
```

```
# Crear un DataFrame
```

```
data = {'Nombre': ['Alice', 'Bob', 'Charlie'],  
        'Edad': [25, 30, 35],  
        'Ciudad': ['A', 'B', 'C']}
```

```
df = pd.DataFrame(data)
```

b. Series:

- Es una estructura unidimensional etiquetada, similar a un array o lista en Python.
- Puede ser vista como una columna de un DataFrame.

```
# Crear una Serie
```

```
nombres = pd.Series(['Alice', 'Bob', 'Charlie'], name='Nombre')
```

2. Operaciones Básicas:

a. Exploración de Datos:

- `head()`, `tail()`: Muestra las primeras o últimas filas.
- `info()`: Proporciona información sobre el DataFrame.
- `describe()`: Ofrece estadísticas descriptivas.

```
print(df.head())  
print(df.info())  
print(df.describe())
```

b. Acceso y Manipulación:

- Acceder a columnas o filas mediante etiquetas o índices.
- Filtrar datos según condiciones.
- Añadir o eliminar columnas.

```
# Acceder a una columna  
nombres = df['Nombre']
```

```
# Filtrar datos  
mayores_de_30 = df[df['Edad'] > 30]
```

```
# Añadir una nueva columna  
df['Sexo'] = ['F', 'M', 'M']
```

```
# Eliminar una columna  
df = df.drop('Ciudad', axis=1)
```

3. Entrada y Salida de Datos:

a. Lectura y Escritura de Archivos:

- Pandas es compatible con varios formatos, como CSV, Excel, SQL, JSON, HDF5, etc.

```
# Leer desde un archivo CSV  
df = pd.read_csv('archivo.csv')
```

```
# Escribir a un archivo CSV
```

```
df.to_csv('nuevo_archivo.csv', index=False)
```

4. Operaciones Avanzadas:

a. Agrupación y Agregación:

- `groupby()`: Agrupa datos por categorías.
- `agg()`: Aplica funciones de agregación.

```
# Agrupar datos
```

```
por_ciudad = df.groupby('Ciudad').mean()
```

b. Fusionar y Concatenar:

- `concat()`: Combina DataFrames.
- `merge()`: Fusiona DataFrames basándose en columnas específicas.

```
# Fusionar DataFrames
```

```
df1 = pd.DataFrame({'A': [1, 2], 'B': [3, 4]})
```

```
df2 = pd.DataFrame({'A': [5, 6], 'B': [7, 8]})
```

```
resultado = pd.concat([df1, df2])
```

5. Funcionalidades Especiales:

a. Manejo de Fechas y Tiempos:

- Pandas proporciona funciones y estructuras específicas para trabajar con fechas y tiempos.

```
# Crear un rango de fechas
```

```
fechas = pd.date_range('2023-01-01', '2023-01-10')
```

```
# Crear una Serie de tiempo
```

```
serie_tiempo = pd.Series(range(10), index=fechas)
```

PROBLEMA

Refacil es una plataforma de tipo marketplace que posee una funcionalidad de autorregistro, la cual permite que un usuario gestione su información y cree su propio comercio en la plataforma. Este registro tiene varios pasos para la creación del usuario, antes de que finalice se conoce como Lead

- teléfono
- datos básicos
- otp
- password
- ubicación
- terminado

Estos pasos permiten saber en qué stage de creación se encuentra el usuario, en este problema importaremos la información correspondiente a 1 mes de registros para poder hacer el cargue, conversión y análisis de la data.

La plataforma está disponible para registro en el siguiente link:

<https://autoregistro.refacil.co/register>

El notebook que tiene el desarrollo del proyecto es:

https://colab.research.google.com/drive/12D70G0KMbw-K32cUq-nxUDQ_KYJ8TS5C?usp=sharing

El video que explica el proceso de desarrollo se encuentra en este link de youtube:

<https://youtu.be/dZBUss8IUWI>

BIBLIOGRAFÍA

Algar Díaz, M. J. & Fernández de Sevilla Vellón, M. (2019). Introducción práctica a la programación con Python. Universidad de Alcalá.

<https://elibro.net/es/lc/unapepec/titulos/124259>

Arévalo Ovalle, D. (2021). Métodos numéricos con Python. 1. Politécnico Grancolombiano.

<https://elibro.net/es/lc/unapepec/titulos/218585>