Esteban Hernandez Agudelo

Conocimiento preguntas

1. ¿Qué es Pandas y para qué sirve?

Pandas es una librería que se instala con el gestor de librerías de Python llamado pip (es el npm/yarn de Python solo que en Python se le denominan librerías y no paquetes), esta librería se basa en el análisis de medianos volúmenes de información y también nos ofrece la opción de seleccionar esta información como deseemos, por ejemplo puedo coger toda la información o campos, porque sí, es un modelo muy similar al de las bases de datos, puedo seleccionar n cantidad de campos o también puedo seleccionar 1 campo en especifico y hacer con él lo que queramos, como por ejemplo extraer esa información para utilizarla en nuestra aplicación.

2. ¿Cuál es la función principal de Pandas en el análisis de datos y cómo se compara con otras bibliotecas en Python?

Como ya se respondió en la pregunta anterior, Pandas nos permite ‘jugar’ con un tamaño de información mediano(Tiene sus límites con el denominado big data ya que puede generar errores por falta de memoria) y así lograr utilizarla en nuestro proyecto, el punto fuerte de pandas contra otras bibliotecas de Python es que esta librería está siendo usada por mas de la mitad de los programadores de Python, por lo tanto tiene un buen soporte y un buen feedback, además de que pandas es utilizado por su fácil uso a diferencia de NumPy.

3. Describe que es un DataFrame en Pandas

La respuesta rápida sería, un DataFrame es una tabla de SQL pero en pandas, pero para ser mas especifico, un DataFrame es el esquema principal en donde se guarda la información , esta cuenta con filas y columnas (por eso si comprendes el comportamiento de una tabla de una base de datos vas a entender instantáneamente que es un DataFrame).

4. Describe la diferencia entre una Serie y un DataFrame en Pandas

Una Serie es una columna y el conjunto de estas forma un DataFrame, por ejemplo: en un DataFrame Vehículo tengo la serie de placa[‘TXT-192’,’GPS-192’, …] y la serie de modelo[‘Mazda’,’Bmw’,’Toyota’, …], por separadas se llaman series pero si las junto se llamaría un DataFrame,, como dato final cada fila de la serie y por tanto del DataFrame también tiene un índice, así que al concepto de serie le podemos sumar ‘y además se representa como si fuera un arreglo, ya que como estos, cada ítem de una serie cuenta con un índice para así poder identificarlos de una manera rápida ’.

5. ¿Qué es la indexación en Pandas y por qué es importante en el análisis de datos?

Indexación es una manera de buscar o filtrar campos de una serie y esto es lo que nos permite hacer procedimientos con la información especifica

6. Que es el concepto de filtrado de datos con Pandas y proporcione un ejemplo práctico.

Un filtrado de información es solo tomar la información que necesito para un caso en especifico

Pongo un ejemplo, tengo un DataFrame en el que almaceno vehículos y tengo las series Placa, Marca, Color y año Fabricación y quiero obtener los registros cuyo color sea rojo.

print(Vehiculos.query(‘Color== “rojo”’))

En donde vehículos es el nombre que le dimos al momento de importar pandas y usar el DataFrame Vehiculos

El resultado esperado serian los registros que cumplan con esa condición.