**Explicación de solución punto 2**

Por: Sebastian Carmona Estrada

Fecha: 06/02/2024

Prueba: **MANIPULACIÓN DE DATOS**

“Cargar un data set, del archivo Excel Master Data, únicamente las siguientes columnas: ➢ Nombre visible Agente ➢ AGENTE (OFEI) ➢ CENTRAL (dDEC, dSEGDES, dPRU…) ➢ Tipo de central (Hidro, Termo, Filo, Menor) ➢ Seleccionar los registros que pertenecen al agente EMGESA ó EMGESA S.A. y adicionalmente que el Tipo de Central sea ‘H’ o ‘T’. ➢ Cargar el archivo dDEC1204.TXT que viene por Central. ➢ Realizar el merge de los dos data sets por Central. ➢ Calcular la suma horizontal de todas las horas para cada planta. ➢ Seleccionar solamente los registros de las plantas cuya suma horizontal sea mayor que cero. ➢ Los resultados deben ser entregados en un dataset. ➢ Enviar junto con la tabla resultante el código utilizado. ➢ Explicar el paso a paso en un archivo de texto (.doc o .pdf).”

**Solución:**

**Texto

Descripción generada automáticamente**

**Explicación**:

1. Para este código, se crea una función llamada *custom\_parser()* para recolectar y organizar la información requerida.
2. La función *custom\_parser()* tiene un parámetro el cual corresponde a la ruta del archivo a procesar.
3. Línea 10: Se abre el archivo haciendo uso de *with().* Aquí se especifica el tipo de codificación del archivo. En este caso, corresponde a una codificación tipo “utf-8”. Además, se abre el archivo en modo lectura “r”.
4. Línea 11: Se crea un diccionario en el cual se va a guarda la información requerida de forma estructurada.
5. Línea 12 – 15: Se crean las llaves correspondientes a los datos deseados, en este caso, “Agente”, “Planta” y 24 columnas correspondiente a la “Hora\_n”, con n entre 1 y 24.
6. Línea 17: Se realiza un ciclo el cual itera en cada una de las líneas del documento procesado, utilizando la variable “row” como la variable de iteración.

**Explicación:**

1. **Función excel\_parser(file\_path):**
   1. Esta función se encarga de leer el archivo de Excel especificado en la ruta file\_path.
   2. Utiliza Pandas para cargar el archivo Excel en un DataFrame (ds).
   3. La selección de columnas se realiza utilizando una lista predefinida de nombres de columnas (‘columns’), que incluye el nombre del agente, la planta y otros detalles relevantes.
   4. Luego, filtra las filas para seleccionar solo aquellas donde el nombre del agente es "EMGESA" y el tipo de central es hidroeléctrica o térmica.
   5. Devuelve el DataFrame resultante.
2. **Función txt\_parser(file\_path):**
   1. Esta función procesa un archivo de texto para crear un DataFrame.
   2. Genera dinámicamente una lista de nombres de columnas basada en las horas del día.
   3. Utiliza Pandas para leer el archivo de texto, especificando la codificación y sin considerar la primera fila como encabezado.
   4. Devuelve el DataFrame resultante.
3. **Función merge\_datasets(ds1, ds2):**
   1. Esta función combina dos DataFrames (“ds” y “ds2”) basándose en la columna común 'CENTRAL (OFEI)'.
   2. Utiliza la función pd.merge() de Pandas para realizar la combinación.
   3. Devuelve el DataFrame resultante después de la combinación.
4. **Función sum\_h(ds):**
   1. Esta función calcula la suma de las horas en el DataFrame recibido como entrada (`ds`).
   2. Crea una nueva columna llamada 'Total' que almacena la suma de las horas para cada fila.
   3. Filtra las filas donde la suma de las horas es mayor que cero, es decir, donde se registraron actividades.
   4. Devuelve el DataFrame resultante.
5. **Función main():**
   1. En esta función principal, se realiza el flujo principal del programa.
   2. Se llama a las funciones `excel\_parser()` y `txt\_parser()` para cargar los datos de los archivos Excel y de texto, respectivamente.
   3. Luego, se llama a `merge\_datasets()` para combinar los conjuntos de datos.
   4. Después, se llama a `sum\_h()` para sumar las horas y filtrar los resultados.
   5. inalmente, se guarda el DataFrame resultante en un nuevo archivo de Excel.

Este código proporciona una solución eficiente para procesar, combinar y resumir datos de diferentes fuentes, siguiendo un flujo claro y organizado.