ACTIVIDAD SESIÓN APACHE SPARK

Imagina que trabajas para una empresa de análisis de mercado y tu tarea es estudiar las preferencias de los consumidores en Sudamérica en cuanto a los modelos de teléfonos inteligentes. Para esto, se te ha entregado un conjunto de datos que contiene información sobre las marcas y modelos de teléfonos más vendidos en distintos países de Sudamérica, junto con la edad de los compradores y las características del teléfono (como la cantidad de memoria RAM, la capacidad de la batería, el precio, etc.).

INSTRUCCIONES

1.- I**nstalación y configuración de PySpark (1 punto)**

- Configura correctamente el entorno de PySpark y crea una SparkSession con el nombre AnalisisTelefonos.

2.- Carga de datos (1 punto)

- Carga el archivo CSV proporcionado (con el nombre telefonos_sudamerica.csv) en un DataFrame de PySpark. Asegúrate de que el archivo contenga los encabezados.

3.- Exploración inicial de los datos (1 punto)

- Muestra las primeras 10 filas del DataFrame y realiza una inspección básica de los tipos de datos de cada columna. ¿Existen valores nulos o erróneos en alguna columna?

4.- Filtrado de datos (2 puntos)

- Filtra el DataFrame para obtener únicamente los teléfonos vendidos en **Brasil**.
- Luego, filtra esos datos para obtener solo los teléfonos de la marca **Samsung**.

5.- Operaciones de agrupamiento y agregación (2 puntos)

- Agrupa los datos por **país** y calcula la **venta promedio** de los teléfonos (promedio de precio).
- Agrupa los datos por **marca** y calcula el **número de teléfonos vendidos** por cada marca.

6.- Análisis por rango de edad (1 punto)

- Crea una nueva columna en el DataFrame que agrupe a los compradores por **rango de edad** (por ejemplo, 18-25 años, 26-35 años, 36-50 años, 51+ años).

- Agrupa los datos por este nuevo rango de edad y muestra el **promedio de precio de los teléfonos vendidos** para cada rango de edad.

7.- Análisis de correlación (1 punto)

- Calcula la correlación entre las columnas memoria_ram y precio. ¿Qué tipo de correlación existe entre estas dos variables?

8.- Filtrado por características del teléfono (1 punto)

- Filtra los teléfonos que tengan una **memoria RAM mayor a 6 GB** y una **batería mayor a 4000 mAh**. ¿Cuántos teléfonos cumplen con esta condición?

9.- Guardar los resultados (1 punto)

- Guarda el DataFrame resultante de los filtros y agregaciones anteriores en un nuevo archivo CSV llamado resultados_analisis.csv.

INSTRUCCIONES ADICIONALES:

- Puntos totales = 10 puntos.
- Comprimir el archivo en formato .zip o .rar.
- Subir el archivo a la plataforma.