



TECNOLÓGICO NACIONAL DE MÉXICO

INSTITUTO TECNOLÓGICO DE TIJUANA SUBDIRECCIÓN ACADÉMICA

Departamento de Sistemas y Computación

EXAMEN

Carrera: Ingeniería En Sistemas Computacionales/ Tecnologías de la información/ Informática

Materia: Datos Masivos

Unidad (es) a evaluar: Unidad 1

Catedrático: Jose Christian Romero Hernandez

Grupo: BDD-1704

Tipo de examen: Practico

Firma del maestro:

Período: **Agosto-Diciembre 2021**

Salón:

Fecha:

Calificación:

Alumno: _____ No. Control: _____

Instrucciones

Responder las siguientes preguntas con Spark DataFrames y Scala utilizando el “CSV” Netflix_2011_2016.csv que se encuentra en la carpeta de spark-dataframes.

1. Comienza una simple sesión Spark.
2. Cargue el archivo Netflix Stock CSV, haga que Spark infiera los tipos de datos.
3. ¿Cuáles son los nombres de las columnas?
4. ¿Cómo es el esquema?
5. Imprime las primeras 5 columnas.
6. Usa describe () para aprender sobre el DataFrame.
7. Crea un nuevo dataframe con una columna nueva llamada “HV Ratio” que es la relación que existe entre el precio de la columna “High” frente a la columna “Volumen” de acciones negociadas por un día. Hint es una operación
8. ¿Qué día tuvo el pico mas alto en la columna “Open”?
9. ¿Cuál es el significado de la columna Cerrar “Close” en el contexto de información financiera, explíquelo no hay que codificar nada?
10. ¿Cuál es el máximo y mínimo de la columna “Volumen”?

11. Con Sintaxis Scala/Spark \$ conteste los siguiente:

- a. ¿Cuántos días fue la columna “Close” inferior a \$ 600?
- b. ¿Qué porcentaje del tiempo fue la columna “High” mayor que \$ 500?
- c. ¿Cuál es la correlación de Pearson entre columna “High” y la columna “Volumen”?
- d. ¿Cuál es el máximo de la columna “High” por año?
- e. ¿Cuál es el promedio de columna “Close” para cada mes del calendario?

Instrucciones de evaluación

- Tiempo de entrega 28 de septiembre 2021
- Al terminar poner el código y la documentación con su explicación en el branch correspondiente de su github, así mismo realizar su explicación de la solución en su google drive en documento de google (Portada, Introducción, Desarrollo, etc).
- Finalmente defender su desarrollo en un video de 6-8 min explicando su solución y observaciones, este servirá para dar su calificación de esta práctica evaluatoria, este video debe subirse a youtube para ser compartido por un link público (Utilicen google meet con las cámaras encendidas y graben su defensa para elaborar el video).

Happy Coding :) !