



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-1704TI9A

Tipo de examen: Practico

Firma del maestro:

Período: **Septiembre_Enero
2021**

Salón:

Fecha:

Calificación:

Alumno: Marco Antonio Rodriguez Medrano

No. Control: 17210635

Instrucciones

Responder las siguientes preguntas con Spark DataFrames utilizando el "CSV" Netflix_2011_2016.csv.

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 entre el precio de la columna "High" frente a la columna "Volume" de acciones negociadas por un día. (Hint: Es una operación de columnas).
8. ¿Qué día tuvo el pico mas alto en la columna "Close"?
9. Escribe con tus propias palabras en un comentario de tu codigo. ¿Cuál es el significado de la columna Cerrar "Close"?
10. ¿Cuál es el máximo y mínimo de la columna "Volume"?

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

- Hint: Basicamente muy parecido a la session de dates, tendran que crear otro dataframe para contestar algunos de los incisos.
- 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 4 días
- Al terminar poner el codigo y la documentación con su explicación en el branch correspondiente de su github, asi mismo realizar su explicación de la solución en su google drive en documento de google (Portada, Introduccion, Desarrollo, etc).
- Finalmente defender su desarrollo en un video de 8-10 min explicando su solución y obsevaciones, este servira para dar su calificación de esta practica evaluatoria, este video debe subirse a youtube para ser compartido por un link publico (Utilicen google meet con las camaras encendidas y graben su defensa para eleborar el video).

Happy Coding :) !