

Tecnológico de Monterrey - Campus Estado de México
Data Mining
Asignación 1

Asignaciones para cada equipo asignado a un dashboard Coca Cola FEMSA.

1. Considerando la importancia de la etapa de entendimiento de negocio en cualquier proyecto de analítica de datos o desarrollo de software se solicita a cada equipo de trabajo:
 - a. Realizar el proceso de valoración de la situación con el fin de contar con información detallada de los reportes de Power BI que serán actualizados por medio de los procedimientos y nuevos elementos de datos a ser desarrollados. Para cada tabla de los reportes de Power BI existentes:
 - i. Describir el esquema de la tabla indicando:
 1. Nombre del campo.
 2. Posición del campo.
 3. Tipo de datos físico (p.e. Text, Numbre, Int64, Date).
 4. Tipo de datos general (p.e. text, number, date).
 5. Admite valores nulos.
 6. Nota: El siguiente artículo describe cómo obtener el esquema de una tabla de Power BI
<https://blog.crossjoin.co.uk/2016/01/19/the-table-schema-function-in-power-bim/>
 - ii. Elaborar un diagrama entidad/relación con todas las tablas del reporte.
2. Considerando la importancia de la etapa de entendimiento de datos en cualquier proyecto de analítica de datos o desarrollo de software se solicita a cada equipo de trabajo:
 - a. Realizar el proceso de recolección de los datos origen por medio de la carga de datos en la herramienta de analítica (Notebook del workspace de Azure Data Bricks).
 - b. Elaborar un reporte de recolección de datos listando para cada tabla origen:
 - i. Dirección del servidor.
 - ii. Nombre del servidor.
 - iii. Nombre de la tabla.
 - iv. Método de conexión.
 - v. Listar problemas encontrados en la conexión a la tabla origen.
 - c. Realizar el proceso de descripción de datos con el fin de contar con una apreciación de alto nivel de las propiedades de los datos adquiridos.
 - d. Elaborar un reporte de descripción de datos en donde se incluya,
 - i. por cada tabla:
 1. Esquema, incluyendo nombre de la tabla, así como nombre, tipo de datos y si admite valores nulos o no cada columna.
 - a. Nota: Emplear la función printSchema() de pyspark
<https://spark.apache.org/docs/latest/api/python/reference/api/pyspark.sql.DataFrame.printSchema.html>

2. Cantidad de registros y columnas de cada tabla.
 3. Columnas de identidad de los renglones de la tabla.
- ii. un diagrama de entidad/relación con cada una de las tablas y las relaciones que tienen entre sí.
3. Considerando la importancia de contar con una visión unificada del avance del proyecto elaborar una tabla de resumen disponible por medio de tecnologías Web con del estatus del desarrollo de cada dashboard indicando:
 - a. Nombre del KPI.
 - b. Estatus: a tiempo, con adelanto respecto al cronograma, con retraso respecto al cronograma.
 - c. Porcentaje de avance respecto al número de tareas.
 - d. Fecha estimada de finalización.
 - e. Los 3 riesgos de mayor atención (mayor valor en su ponderación de impacto versus probabilidad de ocurrencia).

Envío de la asignación.

1. Elaborar documento PDF con la respuesta a la asignación:
 - a. Indicar nombre del dashboard.
 - b. Indicar nombre y matrícula de los integrantes del equipo.
 - c. Responder a los puntos 1 y 2 de la asignación dentro del cuerpo del documento PDF.
 - d. Incluir en el documento PDF el enlace a la tabla Web con la respuesta al punto 3.
2. Enviar el documento PDF al correo electrónico marciano_moreno@outlook.com
3. Fecha límite: domingo 26 de abril.