

Prueba técnica – Ingeniero de Datos

Generalidades:

- Plantea tus respuestas en base a tu experiencia / conocimientos.
- Puedes establecer los supuestos que consideres para completar la prueba.
- Se presume que, en tu rol, tienes la autonomía para tomar decisiones sobre que herramientas utilizar.
- Prepara una presentación / archivos que contenga las respuestas de las secciones detalladas.

Parte 1

Encontrarás adjunto 2 archivos csv correspondientes a una las aplicaciones de créditos y el historial crediticio de dichos usuarios respectivamente. Se le ha solicitado su apoyo como Ingeniero de Datos para la creación de un dataset que facilite algunas variables relevantes para algunos análisis y entrenamiento de modelos predictivos.

Se necesita que, a la tabla de aplicaciones, se le añadan las siguientes columnas:

- El Estatus que el usuario tenía 3 meses antes de hacer la aplicación. En caso de no tener estatus disponible, colocar “No disponible”
- El saldo promedio de los 6 meses previos a la fecha de la aplicación.

Para esta tarea se necesita que cree un proceso que tome las 2 bases como recursos y haciendo uso ya sea de Python o SQL, pueda realizar las uniones y transformaciones necesarias. Finalmente se espera que este proceso funcione para futuros clientes que apliquen.

Notas:

- La tabla de aplicaciones está a nivel de usuario, es decir que cada registro corresponde a un usuario único
- La tabla de récord crediticio muestra el estatus y saldos de los clientes a en ciertos rangos de fechas específicas, en este caso son cortes mensuales, pero su proceso debe de funcionar, aunque esta tabla tuviese otros niveles de granularidades (diarias, semanales, etc)

Parte 2 - Extracción de datos haciendo uso de APIs

Usando <https://pokeapi.co/> extraer la información relacionada (especie, base_experience, habilidades, ataques (moves), entre otros) a los siguientes Pokemons:

1. Bulbasaur
2. Charmander
3. Pidgeotto

Como resultado, cree los siguientes Datasets / Tablas / Dataframes

1. Tabla “pokemon” con los siguientes campos:
 - a. Nombre
 - b. Id
 - c. Base_experience
 - d. Especie
 - e. Height
2. Tabla “ataques” (moves) con los siguientes campos:

- a. Pokemon_id
 - b. Nombre del movimiento
3. Tabla “habilidades” con los siguientes campos:
- a. Pokemon_id
 - b. Nombre de habilidad
 - c. Esta oculta o no

Nota: Su proceso debe de ser flexible para la incorporación de nuevos Pokemons.

Parte 3 - Preguntas adicionales

Se te pide contestar los puntos detallados a continuación:

1. Que buenas prácticas / consideraciones tomas en los siguientes puntos:
 - i. Seguridad y privacidad de la información al manejar datos sensibles
 - ii. cómo implementas gobernanza de datos en tus proyectos
 - iii. que aplicas para optimizar el tiempo de procesamiento y rendimiento en las consultas a las bases de datos / SQL
 - iv. Qué utilizarías para manejar datos faltantes en un conjunto de datos
2. ¿Cuál es el flujo de trabajo de ETL que has implementado? Describir o realizar diagrama
3. Se te indica que apoyarás al equipo de Inteligencia artificial, por lo que te consultan también:
 - i. ¿Qué consideraciones tomarías cómo ingeniero de datos para armar los conjuntos de datos en los proyectos de IA?
 - ii. ¿Cómo definirías la calidad de un conjunto de datos para un proyecto de inteligencia artificial?
 - iii. ¿Qué herramientas y tecnologías consideras esenciales para el trabajo con datos en proyectos de IA?