



monokera

Prueba técnica

Data Engineer



Contexto

Somos Monokera y hacemos parte del grupo Credicorp, como una de las primeras empresas de tecnología que busca transformar la industria aseguradora en Latam. Nuestra oferta de valor se basa en una **plataforma abierta de gestión de seguros digitales** que permite **integrar rápidamente y sin fricciones** a las aseguradoras, reaseguradoras y servicios de asistencia con canales de venta, para que ofrezcan seguros adecuados a sus clientes. El reto para 2024 es tener productos de seguros digitales en canales de Perú, Chile y Colombia.



Escenario y definición de la Prueba Técnica

Dado que monokera se basa, como negocio, en la digitalización de procesos, consecución y experiencia digital, busca crear una infraestructura/flujos robustos y dinámicos, que permitan construir un espacio seguro para la generación de información valiosa a partir de los datos y cuyo objetivo es disponibilizar a las aseguradoras, reaseguradoras, canales de venta, etc acceso a la información generada de manera simple y para la solución a interrogantes de las reglas de negocio.

Reto General: Construir un pipeline que tome los datos del archivo proporcionado para el reto, se realice un proceso de ETL y finalmente se construya un reporte.

Retos específicos

- Análisis y limpieza de la información del archivo CSV.
- Guarde la información resultante del paso anterior, en una base datos relacional que tenga las siguientes tablas:
 - Policy: Tabla que contiene los campos asociados a la información básica al momento de adquirir una póliza.
 - Insured: Tabla que contiene la información de la persona que adquirió la póliza y en caso de siniestro, hará el reclamo correspondiente.
 - Premium: Tabla que contiene los campos "premium_amount, deductible_amount, coverage_limit" y se relaciona con los valores asociados a la prima de la póliza.
 - Payments: Tabla que contiene los campos "payment_status, payment_date, payment_amount, payment_method" y se relaciona con la obligación monetaria del tomador de la póliza.
 - Claims: Tabla que contiene información de los reclamos realizados por los asegurados cuando ocurre un siniestro.
 - Agents: Tabla que contiene información de los canales de ventas(o canal) para las pólizas.
- Agregue el modelo relacional a las descripciones generales del reto.
- Construya el reporte (cuya fuente es la DB relacional del punto anterior) descrito a continuación:
 - Reporte que te permite mostrar únicamente las pólizas con fecha de finalización superior al 01-07-2023 (Ten presente que las fechas deben tener el formato, día-mes-año)
- Suponiendo en la necesidad de actualización del reporte de manera diaria, orquestar el flujo usando Airflow. En este ítem, describe y aclara cuál es la mejor manera de realizarlo según criterios propios.



Notas adicionales y **Entregas esperadas**

- Para los datos faltantes, aplique el método o métodos que considere adecuados y explíquelo en las descripciones generales del reto.
- Use la base de datos relacional que más le guste y se adecue al ejercicio, de nuestra parte sugerimos PostgreSQL.
- Use notebooks, scripts de Python o las herramientas que crea convenientes, sin dejar pasar por alto agregar los detalles de sus elecciones en las generalidades del reto.

Entrega

- Repositorio público en Github que permita mostrar lo realizado.
- Documentación con la información que consideres adecuada y los pasos para la ejecución del proyecto.
- Información, fuentes de ayuda, etc extras que le permitieron desarrollar el reto.



Tu conocimiento es potente

Esta prueba está estimada para su **entrega:**

Tres días después de ser enviada.

Luego de este paso, se programará reunión de sustentación de la entrega realizada.

