

Prueba Ingreso Analista de Datos / Científico de datos junior.

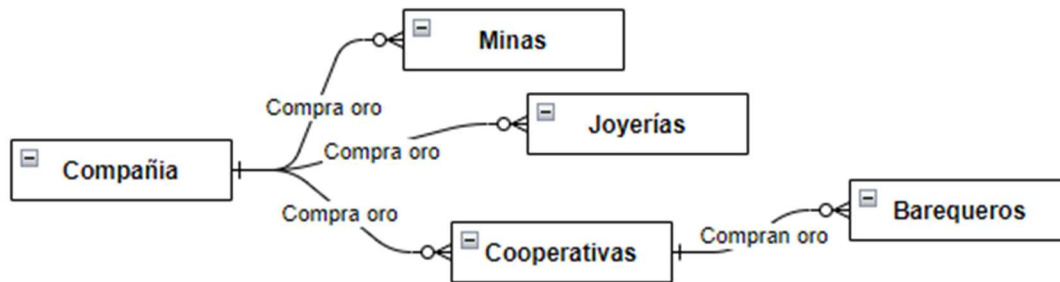
Contexto

Usted forma parte de una compañía dedicada a la comercialización de oro. Sin embargo, debido a su alto valor, existen riesgos asociados con su adquisición, siendo el más importante la procedencia o relación con actividades ilegales.

El oro que entra a la compañía puede llegar de las siguientes maneras:

- **Joyería en desuso:** Aleaciones de oro con otros elementos como plata y cobre. Este material siempre tiene una segunda procedencia, es decir, el dueño anterior antes del joyero.
- **Minería:** Oro en roca triturada, generalmente con muy baja concentración y acompañado de otros elementos como arsénico, zinc, plomo, calcio, etc. Cuando la compañía firma un convenio con una mina, envía ingenieros a tomar muestras del suelo de la mina.
- **Minería formal a pequeña escala:** Oro recolectado en pequeñas cantidades por mineros formalizados, generalmente con un grado mayor de concentración. Debido a la escasez del material, no se suele aceptar más de 70 gramos de oro al día o 420 gramos al mes por cada minero registrado.

Para cada categoría hay un proveedor específico: los **comercializadores venden joyería**, las **minas realizan la minería**, y las **cooperativas mineras venden oro explotado a pequeña escala**. Cada cooperativa minera posee la información de los pequeños mineros registrados, conocidos como *barequeros*.



Problema para resolver

Los proveedores se suelen ver inmersos en problemas legales debido al alto riesgo de comerciar oro. De manera que, si la compañía tiene lazos con un proveedor riesgoso, su reputación frente a la autoridad regulatoria se viene abajo. Ocasionando consecuencias legales y económicas. Su **objetivo** como analista de datos es encontrar **alertas tempranas** que permitan mitigar este riesgo.

Insumos

Usted cuenta con las siguientes tablas:

Tablas de condiciones iniciales:

- **Tabla de minas:** Cuenta con las minas (tres) con las cuales la compañía tiene convenio, por cada mina hay un promedio de **concentración de elementos** en el suelo de esta, el teléfono, la ubicación y otros datos.
- **Tabla de joyeros:** Posee los datos de ubicación, representante legal, etc. (Diez joyeros en total)
- **Tabla de cooperativas:** Posee los datos de cada cooperativa, su ubicación, teléfono etc, son un total de tres cooperativas.
- **Tabla de barequeros:** Posee los datos de cada barequero, su ubicación, cedula teléfono y la cooperativa a la cual pertenece.

Tablas de eventos (oro que entró a la compañía pero que no ha sido comprado):

- **Oro proveniente de minas (último mes):** oro que ingresó a la compañía por concepto de minería, contiene los datos del volumen, la cantidad oro, la concentración de elementos etc.
- **Oro proveniente de joyería (último mes):** oro que ingresó a la compañía por concepto de joyería en desuso, contiene los datos del joyero y de la segunda procedencia.
- **Oro proveniente de cooperativas (último mes):** oro que ingresó a la compañía por concepto de minería a pequeña escala, contiene los datos de la cooperativa y del pequeño minero que lo entregó.

Información relevante

Las tablas de eventos poseen transacciones de oro hechas entre el proveedor y la compañía, sin embargo, la compañía se reserva el derecho de devolver material si ve que el oro entregado puede haber estado involucrado en delitos (hurto, lavado de dinero, minería ilegal, etc). Lo cual genera una alerta roja. En las tablas de eventos usted encontrará anomalías, generando columnas donde se alerte de un posible hecho delictivo.

Señales de fraude

Minas: Diferencias en la concentración de elementos respecto a las muestras iniciales le indicarán que el oro enviado no proviene de donde se acordó.

Joyerías: Datos faltantes en la procedencia del oro, datos repetidos, o individuos que vendieron demasiado a los joyeros le mostrarán oro de dudosa procedencia.

Cooperativas: Oro proveniente de barequeros no afiliados, oro de barequeros cuyos datos coincidan entre sí, o de barequeros que hayan entregado más de 35 gramos durante el mes. Este oro no se debe recibir ya que puede estar involucrado en actos de minería ilegal.

Resultado esperado:

1. Con sus conocimientos como analista de datos (Python, Power BI, Excel, etc.) usted creará las reglas para señalar el oro fraudulento.
2. Una vez tenga el oro fraudulento, usted generará una tabla donde clasifique el oro que entró a la compañía por cada rubro (Minería, Joyería y Cooperativa) y cuanto de este oro tiene una procedencia fraudulenta.
3. Por cada rubro haga un grafico donde muestre cada proveedor y el porcentaje de oro fraudulento que envió a la compañía, Utilice su creatividad y herramientas de visualización para escoger el grafico que mejor se ajuste. (Plotly, Seaborn, Power BI, Excel etc).
4. Tendrá 15 minutos para exponer la metodología y los resultados a los cuales llegó.

Criterios para evaluar:

- Análisis de patrones
- Presentación concisa de la información
- Escalabilidad de la solución

Tiempo para la prueba: de 6 a 8 hrs como máximo.