

Científico de datos en NEQUI: Prueba Técnica

Propósito

El propósito de esta prueba es verificar tus conocimientos en el campo del modelado analítico de datos (indiferente de los "framework" o lenguajes de programación que utilices) esperamos que sea la oportunidad de combinar lo que has aprendido a lo largo de tu experiencia profesional y mezclarlo con tu capacidad de abstracción y analítica.

En este test, tienes libertad de elegir el lenguaje de programación de tu preferencia.

Recursos [Dataset]

Recopilar y disponer de los datos correctos es una de las tareas más importantes para quienes construyen modelos analíticos. Para este ejercicio disponemos de un data set de prueba que contiene datos reales junto con datos sintéticos que conservan la distribución estadística.

Encontrarás las siguientes columnas:

_id: Identificador único del registro

merchant_id: Código único del Comercio o aliado.

subsidiary: Código unico de la sede o sucursal.

transaction_date: Fecha de contabilización de la transacción en el core Financiero.

account_number: numero único de cuenta

user_id: Código único del usuario dueño de la cuenta desde donde se registran las transacciones.

transaction_amount: Monto de la transaction en moneda ficticia.

transaction_type: Naturaleza de la transacción valores (DEBITO o CREDITO)

Instrucciones

Paso 1: Alcance del proyecto.

El objetivo de la prueba es idear una solución para identificar transacciones que evidencian un comportamiento de Mala Practica Transaccional, empleando un producto de datos.

Adicional, describir la solución y detallar cómo incorporar el producto de datos en un marco operativo.

Una Mala Practica Transaccional es un comportamiento donde se evidencia un uso de los

canales mal intencionado, para la prueba técnica nos centraremos en la práctica de Fraccionamiento Transaccional, esta mala práctica consiste en fraccionar una transacción en un número mayor de transacciones con menor monto que agrupadas suman el valor de la transacción original. Estas transacciones se caracterizan por estar en una misma ventana de tiempo que suele ser 24 horas y tienen como origen o destino la misma cuenta o cliente.

Con la solución de este reto, esperamos conocer las habilidades de modelado estadístico, matemático, heurístico y/o analítico en general que a lo largo de su vida académica y profesional han desarrollado.

Valoramos mucho la habilidad que tengan de detallar los hallazgos dados a partir de las diferentes etapas del análisis.

Paso 2: Explorar y evaluar los datos, el EDA.

- * Explorar los datos y realizar descriptivos de valor que permitan identificar los features que serán usando en el modelo analítico.

Paso 3: Definir el modelo analítico.

- * Trazar el flujo de datos y explicar bajo cuál criterio seleccionó el modelo final.
- * Diseñar la arquitectura y los recursos utilizados.
- * Indique claramente los motivos de la elección de las herramientas y tecnologías para el proyecto.
- * Proponga con qué frecuencia deben actualizarse los datos y por qué.

Criterios:

- * Para la entrega de la prueba publicar el proyecto final en un repositorio de GitHub y compartírnos el enlace.
- * Para el proceso de documentación escribe una guía “.readme” (Markdown) que tenga un paso a paso de entendimiento que pueda vincular los diferentes análisis construidos en los diferentes pasos y la explicación.
- * Reporta la(s) métrica(s) utilizada(s) para medir el performance del modelo.

Si llegaste hasta este punto, sabrás, al igual que nosotros que la complejidad de la prueba es alta, no te desanimes. Queremos conocer tu capacidad de respuesta ante escenarios de alta presión.