

## Método

Esta prueba técnica intenta evaluar tus habilidades técnicas. A partir del momento de recibida la prueba el candidato cuenta con 5 días para entregar su resultado. Para la entrega, siga las siguientes instrucciones:

1. Analiza el documento y date el tiempo de pensar su solución. 2. Toma decisiones de diseño detallado y hazlas realidad en código fuente. Itera en este paso hasta que la solución te satisfaga;
2. La tarea ha sido diseñada para permitirte ser creativo en el diseño e implementación de la solución. No se tienen restricciones técnicas o de lenguaje de programación de ningún tipo;
3. Realiza un screencast de máximo 5 minutos en el cual expongas lo siguiente:
  - a. Las decisiones de diseño que tomaste. Haz esto mostrando el código fuente y si lo quieres o necesitas los esbozos gráficos de tus decisiones.
  - b. La prueba corriendo. Haz esto mediante la ejecución de los diferentes casos de prueba.
4. Sube el código a un repositorio público en algún servidor de git (github, bitbucket etc).
5. Revisaremos el video (solo revisaremos hasta el minuto 5 para hacer nuestra evaluación) y el código para determinar si continuamos con el proceso de selección o si se detiene dado que consideramos que, técnicamente, hay aspectos fundamentales que se deben mejorar en la aproximación al diseño de software.

## Enunciado

El candidato debe implementar una tarea de streaming de lectura, transformación y almacenamiento del resultado. Los datos sobre los cuales se apalanca la prueba se encuentran en el siguiente [enlace](#).

La tarea debe modelar una tubería (pipeline) de datos que ejecuta los siguientes pasos:

- Paso 1 (ejecución en paralelo)
  - Verificar el esquema de datos;
  - Verificar datos perdidos/faltantes;
  - Verificar datos erróneos/anómalos.
- Paso 2 (ejecución secuencial)
  - Normalizar una columna (cualquiera de valores continuos);
  - Filtrar una columna por cierto valor (cualquiera de valores categóricos);
  - Agrupar ciertas columnas (cualesquiera que correspondan a fechas).

Nota: Se debe tener la capacidad de recibir un conjunto de parámetros para determinar las variables usadas en este paso.

- Paso 3 (ejecución secuencial)
  - Transformar una variable y agregarla al conjunto de datos. (Aplique la función  $x^3 + \exp(y)$  sobre cualquier tupla de variables continuas);

- Agregación - Conteo de registros únicos (sobre cualquier columna de valores categóricos).
- Paso 4 (modelo)  
Presenta el ciclo completo de desarrollo de un modelo predictivo basado en el dataset entregado. Es decir, un notebook o conjunto de scripts donde se observe:
  - Extracción de datos;
  - Preparación de datos;
  - Transformación de datos;
  - Entrenamiento del modelo;
  - Predicción/inferencia del modelo;
  - Cualquier recomendación al equipo de analítica de datos.

Nota: Se espera un análisis del desempeño (performance) del modelo así como una explicación del proceso de pensamiento durante el proceso.