



Challenge Data Engineer

¡Muchas **gracias** por querer ser parte de la familia de Reba! En el documento a continuación vas a encontrar un desafío que nos gustaría que hagas para poder evaluar tus aptitudes técnicas. Tené en cuenta que podés tomarte libertades si querés mostrarnos algo más allá de lo pedido.

Si tienes alguna duda no dudes en preguntar, y no te preocupes ningún punto es excluyente.

Gracias nuevamente por tomarte tu tiempo para realizarlo.

VISIÓN GENERAL

En este Challenge se tendrá en cuenta:

- Manejo de Librerías.
- Manejo de Excepciones.
- Preparación y Manipulación de datos.
- Comentarios.
- Logs.
- Modularización.
- Tratamiento de datos duplicados y/o faltantes.

OBJETIVOS

1. Desarrollar un script en Python utilizando la plataforma Jupyter Notebook que realice lo siguiente:
 - a. Descargar el archivo de la siguiente fuente:
https://www.indec.gob.ar/ftp/cuadros/economia/sh_ipc_aperturas.xls



- b. Realizar a partir de la lectura de la pestaña 'Índices aperturas' del archivo un análisis exploratorio de los datos.
 - c. A partir de lo anterior realizar 5 observaciones que le parezcan relevantes.
- 2. Desarrollar uno o más scripts que implementen un DAG para la plataforma Apache Airflow, pensados para ejecutarse periódicamente para realizar lo siguiente:
 - a. Descargar el archivo de la fuente anteriormente citada.
 - b. Volcar el contenido de la pestaña "Índices aperturas" en una base de datos relacional como por ejemplo SqlServer.
 - c. Copiar el archivo completo en un bucket del servicio Google Cloud Store de GCP.

(en caso de no tener experiencia por Apache Airflow, realizar los puntos en un script)

- 3. A partir de la tabla generada en el punto anterior desarrollar un Stored procedure en Transact-SQL que tome como parámetro de entrada una región y que devuelva en una tabla temporal los 5 rubros que tuvieron para dicha región una mayor variación interanual junto con su medida.