

CODERFLEX

Data Engineering **Consigna de Primer** **Pre entregable**

CODERHOUSE

Script que extraiga datos de una API pública y crear la tabla en Redshift para posterior carga de sus datos.

Para finalizar tu primer pre entregable, te proponemos entregar un Script que extraiga datos de una API pública y crear la tabla en Redshift para posterior carga de sus datos.

Objetivos

- Generar un script (formato .py o .ipynb) que funcione como prototipo (MVP) de un ETL para el proyecto final
- El script debería extraer datos desde una API en formato JSON para ser manipulado como diccionario utilizando el lenguaje Python
- Generar una tabla para ser almacenada en una base de datos a partir de información una API.

Requisitos

Este trabajo cuenta con dos instancias que se deben **mostrar dentro de una misma presentación**:

1

[Código/Script](#)

2

[Tabla en Amazon Redshift](#)

1. Código/Script

El Código deberá entregarse en formatos (.py o .ipynb) disponibles en un repositorio Github o en Google drive

El objetivo es tener un código inicial que será usado en el proyecto final como un MVP de ETL . El script debería extraer datos en JSON desde una API pública o gratuita para luego convertir estos datos en diccionario Python y posterior manipulación. Considera los siguientes puntos:

- 1. La API seleccionada debería contener datos en cambio constante:** Es decir, elegir datos que estén cambiando día con día, como pueden ser reportes de clima, datos de finanzas, casos de criminalidad, etc.
- 2. Revisa que la información de la API seleccionada pueda ser consultada por lo menos una vez al día:** Es importante que tengas la capacidad de actualizar los datos constantemente y con un rate de por lo menos una vez al día, sin que se limite su acceso.
- 3. El Código o script debería extraer varios datos de diferentes categorías/variables:** Por ejemplo, si se eligen acciones de bolsa, se pueden elegir diferentes acciones, o subdividir por ETFs, en caso de elegir, el clima, dividir por regiones, etc.
- 4. Respecto a la estructura del script/código:**
 - Revisa que no tenga bugs
 - Revisa que el código no sea redundante
 - Cerciorate de que no haya datos sensibles que se muestren en el mismo, como pueden ser credenciales.
 - Revisa que tu código sea eficiente, es decir, que ejecute lo que se propone de la forma más sencilla, clara y rápida posible.

Recuerda que esto será la base para tu proyecto final.

2. Tabla en Amazon Redshift

Tabla creada en Redshift y con información obtenida desde la API elegida

A su vez, la entrega involucra la creación de una versión inicial de la tabla donde los datos serán cargados posteriormente. Considera los siguientes puntos para su elaboración:

1. **La tabla debería contener información correspondiente a la que se extrae de la API.**
2. **Incluir una columna temporal para el control de ingesta de datos**

Recuerda que esto será la base para tu proyecto final.

Script que extraiga datos de una API pública y crear la tabla en Redshift para posterior carga de sus datos.

Recomendaciones

- Utiliza la guía de actividades para poder tener mayor referencia de lo que se puede hacer para cumplir con este pre entregable.
- Consultar la documentación del módulo [requests](#) en Python.
- Se sugiere ampliamente la creatividad del estudiante. Usar Apis que les resulten interesantes y están relacionadas a su contexto. En este [enlace](#) podrás encontrar información gratuita de APIs

Ejemplos

Para guiarte, te compartimos el siguiente ejemplo:

- [Ejemplo](#)

Criterios de evaluación

Para la evaluación de tu Proyecto Final, tendremos en cuenta los siguientes [criterios de evaluación](#).