

PROYECTO PROYECCIONES ACOSUX

INTRODUCCION:

En este proyecto, desarrollaré un proceso de Extracción, Transformación y Carga (ETL) utilizando Python para el análisis de datos. Dividiré el proyecto en dos etapas: desarrollo y producción.

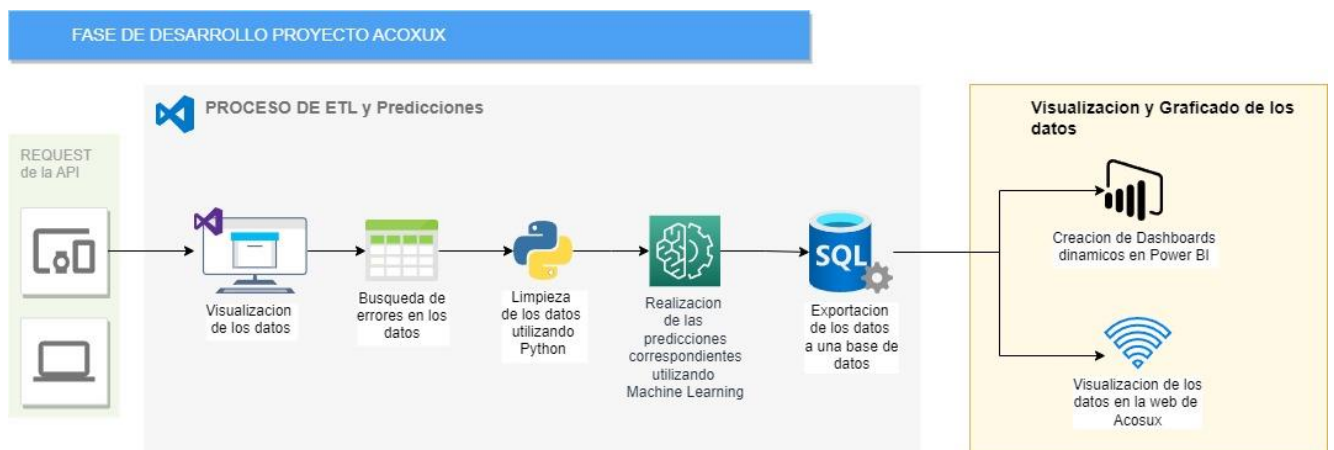
En la etapa de desarrollo, ejecutaré el proceso localmente para realizar pruebas exhaustivas y garantizar el buen funcionamiento. Utilizaré Python y bibliotecas especializadas para extraer datos de una API, realizar la limpieza, transformación y generar predicciones con Machine Learning. Exportaré los resultados a una base de datos local y utilizaré programas como Power BI para visualizar y analizar los datos.

En la etapa de producción, automatizaré los procesos utilizando servicios en la nube como Google Cloud o AWS. Configuraré tareas programadas para que el proceso de ETL se ejecute automáticamente en respuesta a eventos, como la carga de nuevos datos en la API. Esto brindará escalabilidad y confiabilidad al proyecto.

Con estas dos etapas, garantizaré resultados precisos y confiables, aprovechando la eficiencia y escalabilidad que ofrecen las soluciones en la nube.

Fase de Desarrollo:

Diagrama de Flujo:



Introducción:

El objetivo de este proyecto es desarrollar un proceso de Extracción, Transformación y Carga (ETL) utilizando Python en un entorno local. Utilizaremos Python para extraer datos de una API, realizar la limpieza y transformación de estos mediante técnicas de ETL, y posteriormente realizar predicciones utilizando algoritmos de Machine Learning. Los datos procesados serán exportados a una base de datos local y podrán ser analizados y visualizados en programas como Power BI o utilizados para incorporarlos en la web de Acosux.

Objetivos del Proyecto:

Extraer datos de una API utilizando Python en un entorno local.

Aplicar técnicas de limpieza y transformación de datos mediante ETL utilizando Python.

Utilizar algoritmos de Machine Learning para generar predicciones basadas en los datos procesados.

Exportar los datos limpios y las predicciones a una base de datos local.

Integrar los datos y visualizaciones en programas de análisis de datos como Power BI.

Incorporar los datos procesados y las visualizaciones en la web de Acosux.

Metodología:

Extracción de datos: Utilizaremos Python y bibliotecas como requests o httpplib para interactuar con la API y obtener los datos necesarios en un entorno local.

Transformación de datos (ETL): Aplicaremos técnicas de limpieza y transformación utilizando bibliotecas como pandas en un entorno local para asegurar la calidad y coherencia de los datos.

Predicciones con Machine Learning: Utilizaremos algoritmos de Machine Learning, como regresión o clasificación, para generar predicciones basadas en los datos procesados en un entorno local.

Exportación a la base de datos local: Utilizaremos bibliotecas adecuadas, como sqlite3, para exportar los datos limpios y las predicciones a una base de datos local en el entorno local, asegurando un almacenamiento seguro y eficiente.

Integración con programas de análisis de datos como Power BI: Utilizaremos las capacidades de integración de programas de análisis de datos como Power BI para importar y visualizar los datos y las predicciones generadas en el entorno local, creando informes y paneles interactivos para un análisis más profundo.

Incorporación en la web de Acosux: Utilizaremos las herramientas y tecnologías adecuadas para incorporar los datos procesados y las visualizaciones en la web de Acosux en el entorno local, permitiendo a los usuarios acceder y explorar la información de manera fácil y amigable.

Recursos Necesarios:

Entorno de desarrollo Python con las bibliotecas necesarias (pandas, numpy, scikit-learn, etc.).

Documentación de la API y acceso a los datos necesarios.

Base de datos local (puede ser SQLite u otro sistema de gestión de bases de datos).

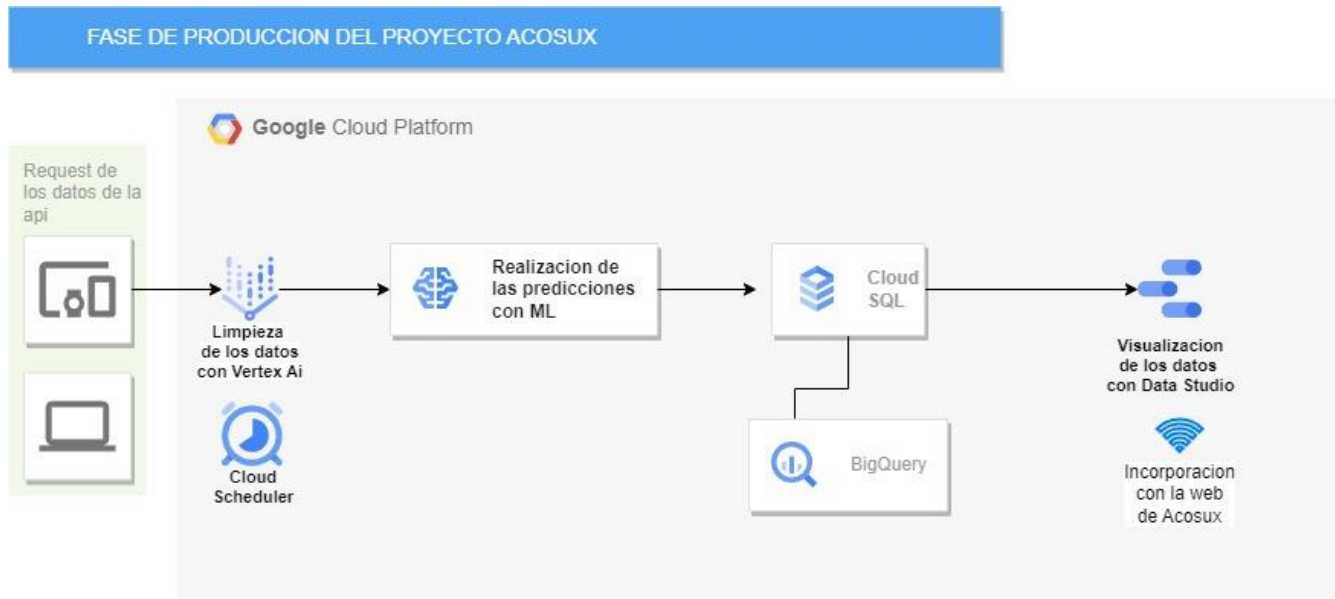
Programas de análisis de datos como Power BI instalados en el entorno local.

Acceso y conocimientos de desarrollo web para la incorporación en la web de Acosux.

Cronograma Tentativo:

Fase de Producción:

Diagrama de Flujo:



En esta fase, el objetivo principal es automatizar los procesos para que se ejecuten automáticamente de forma periódica, asegurando que se realice todo el proceso nuevamente al cargarse nuevos datos. Para lograr esto, utilizaré tecnologías en la nube como Google Cloud Platform (GCP) o Amazon Web Services (AWS).

En GCP, aprovecharé las siguientes herramientas y servicios para la automatización:

Cloud Scheduler: Utilizaré Cloud Scheduler para programar y desencadenar la ejecución periódica del proceso. Configuraré el cronograma adecuado para que el proceso se ejecute automáticamente a intervalos regulares.

Vertex AI: Emplearé Vertex AI para automatizar la descarga de datos de la API de manera periódica. Utilizaré las capacidades de Vertex AI para gestionar y orquestar la extracción de datos de la API de manera eficiente y automática.

Vertex AI (ETL): Usaré Vertex AI también para realizar el proceso de ETL y la limpieza de los datos. Aprovecharé las funcionalidades y herramientas proporcionadas por Vertex AI para llevar a cabo la transformación y limpieza de los datos de manera automatizada.

AutoML / BigQuery ML: Para las predicciones de Machine Learning, utilizaré opciones como AutoML y BigQuery ML de GCP. Estos servicios me permitirán entrenar y desplegar modelos de Machine Learning de manera automatizada, aprovechando las capacidades de aprendizaje automático sin necesidad de conocimientos avanzados en ML.

BigQuery: Moveré los datos limpios y procesados a una base de datos, como BigQuery, que proporciona un almacenamiento escalable y eficiente para su posterior análisis y vinculación con otras fuentes de datos.

Data Studio / Looker: Exportaré los datos desde BigQuery y utilizaré herramientas de visualización y gráficos como Data Studio o Looker para crear paneles interactivos y visualizaciones personalizadas. Estas herramientas me permitirán analizar y presentar los datos de manera efectiva.

Además, tengo la intención de vincular los resultados y visualizaciones obtenidas a la página web de Acosux para que los datos sean accesibles y puedan ser compartidos con otros usuarios.

En resumen, en esta fase de automatización del proceso, utilizaré tecnologías en la nube como GCP y servicios como Cloud Scheduler, Vertex AI, AutoML/BigQuery ML, BigQuery y herramientas de visualización como Data Studio/Looker. Estos servicios y herramientas me permitirán automatizar la descarga de datos, realizar el proceso de ETL, realizar predicciones de Machine Learning, almacenar los datos en una base de datos, crear visualizaciones y vincular los resultados a la página web de Acosux.