



UNIVERSIDAD DE LAS AMÉRICAS

**FACULTAD DE INGENIERÍA Y CIENCIAS
APLICADAS**

ICBS0003-2441

ANÁLISIS Y VISUALIZACIÓN DE DATOS

PASOS PARA EJECUTAR EL PROYECTO

**CARLOS MEJIA, CALEB NARANJO, FABIÁN
CRUZ**

Enlace para acceder al repositorio en GitHub, en donde se encuentra el proyecto:

https://github.com/CralosMejia/UDLAICBS0003202310_PROYECTOINTEGRADOR_GRUPO6

Para ejecutar el proyecto, abrir el archivo "py_startup.py" y ejecutarlo.

Si no desea generar datos aleatorios, puede comentar las líneas en el archivo generateData.py.

Estructura del proyecto

El proyecto realizado contiene cinco carpetas, las cuales se explicarán a continuación:

✓ **.idea**

Aquí se encuentran los archivos xml.

✓ **Documentación**

En esta carpeta se encuentra el manual de instalación de infraestructura utilizada en este proyecto y los pasos para ejecutar el proyecto.

✓ **ETL**

Dentro se encuentran las siguientes carpetas:

- **Data:** Aquí se encuentra el csv de las áreas académicas y los archivos:
generate data: genera datos aleatorios.
seed: Carga en la base de datos del source, llama a los métodos y carga en la base de datos del source.
- **Extract:** Aquí se encuentran los archivos de extracción.
- **Load:** Aquí se encuentran los archivos de carga.
- **Transform:** Aquí se encuentran los archivos de transformación.
- **Util:** Aquí se encuentran los archivos necesarios para la ejecución del proyecto como:
connection: contiene la conexión a la base de datos,
properties: contiene la configuración de properties,
sql: contiene funciones de apoyo para el proceso de ETL.
Además del archivo py_startup.py que es el archivo principal con el cuál se ejecutará el proyecto.

✓ **POWER BI**

Aquí se encuentra el archivo realizado en Power BI, el cual contiene el informe realizado respondiendo a las preguntas planteadas.

✓ **SQL SCRIPTS**

Aquí se encuentran los scripts utilizados para la creación de las bases de datos en MySQL.