**TABLA DE CONTENIDO**

1.-Objetivo del proyecto.

2.-Esquema preliminar del trabajo a realizar.

3.-Lista de Herramientas a utilizar.

4.-Plan de Acción.

5.-Conclusiones y recomendaciones al cliente.

6.- Anexos

6.1 Notebook de Python- EDA

6.2 Backup de la base de datos

6.3 Notebook del API

**1.-Objetivo**

El proyecto individual tiene como finalidad realizar las tareas de un perfil de Data Engineer.

Las fuentes entregadas para el inicio de esta tarea son archivos CSV y Json, que contienen información de las películas y series que ofrecen las plataformas de Streaming.

El reto es realizar el análisis exploratorio de datos (EDA), las tareas de ETL y disponibilizar las consultas en una API, para realizar comparaciones entre plataformas, donde se detallan 4 consultas principales:

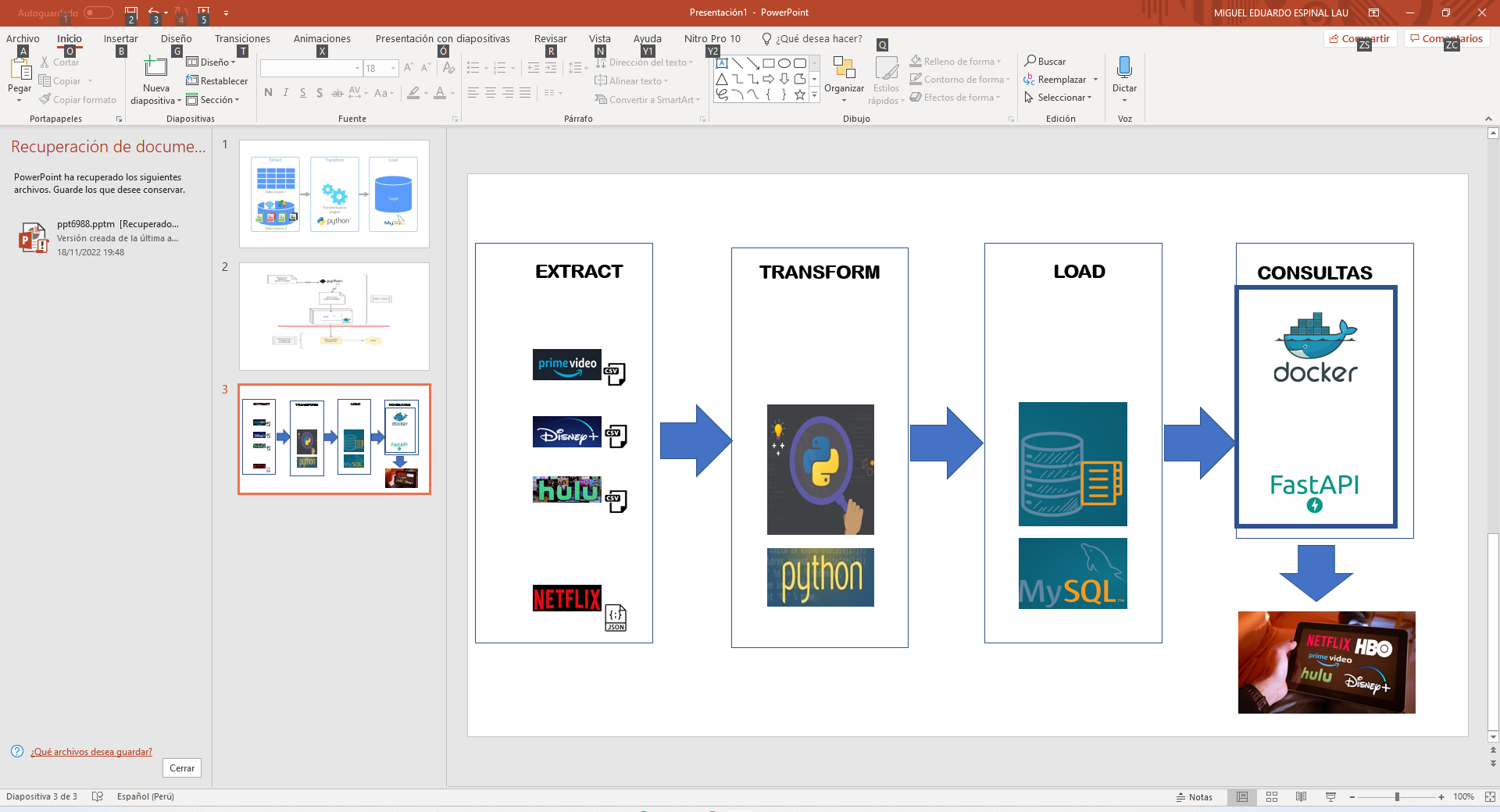
1.1.- Films con máxima duración por plataforma y por Año.

1.2. Cantidad de películas y series por plataforma

1.3.-Cantidad de películas por género y plataforma

1.4.- El actor que tiene más películas registradas en el servicio de streaming por plataforma y año.

**2.- Esquema preliminar del trabajo a realizar**



3.- Lista de Herramientas a utilizar

3.1.- Python, Pandas, Sqlalchemy

3.2.- Mysql

3.3.-Docker desktop para Windows.

3.4. FastAPI y las siguientes herramientas:

anyio==3.6.2

charset-normalizer==2.1.1

click==8.1.3

colorama==0.4.6

docopt==0.6.2

fastapi==0.88.0

greenlet==2.0.1

h11==0.14.0

idna==3.4

mysql==0.0.3

mysqlclient==2.1.1

pipreqs==0.4.11

pydantic==1.10.2

PyMySQL==1.0.2

requests==2.28.1

sniffio==1.3.0

SQLAlchemy==1.4.44

starlette==0.22.0

typing\_extensions==4.4.0

urllib3==1.26.13

uvicorn==0.20.0

wincertstore==0.2

yarg==0.1.9

4.- Plan de Acción

4.1.- Trabajo de EDA a través de Python.

* Utilización de Pandas para Carga de datos y limpieza, se analiza tabla por tabla con las siguientes premisas:
  + - Eliminación de columnas no significativas para las consultas
    - Llenado de datos por ‘No dato’ en nulo para columnas relevantes, que serán informadas al cliente.
    - Separación de datos relevantes para las consultas y normalización
    - Creación de índices y columnas de clasificación de plataforma
    - Concatenación de tablas y extracción de maestros de actores y géneros.
* Utilización de Sqlalchemy para la carga a Mysql.

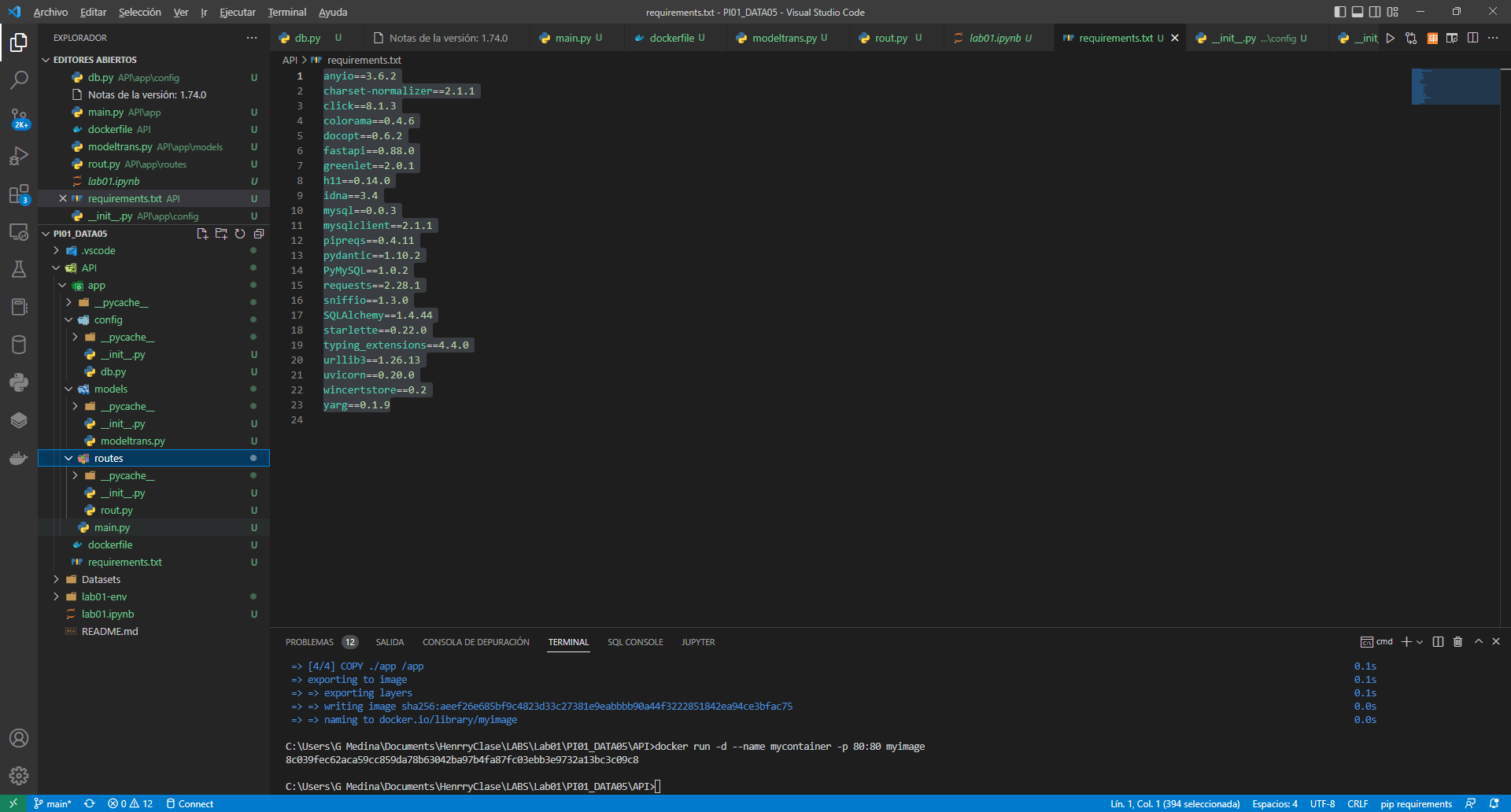
\*Nota. - detalle y documentación anexo 6.1

4.2.- Carga de tablas a Mysql, preparación de consultas. Revisión de la información, preparación y verificación de las consultas(modelado).

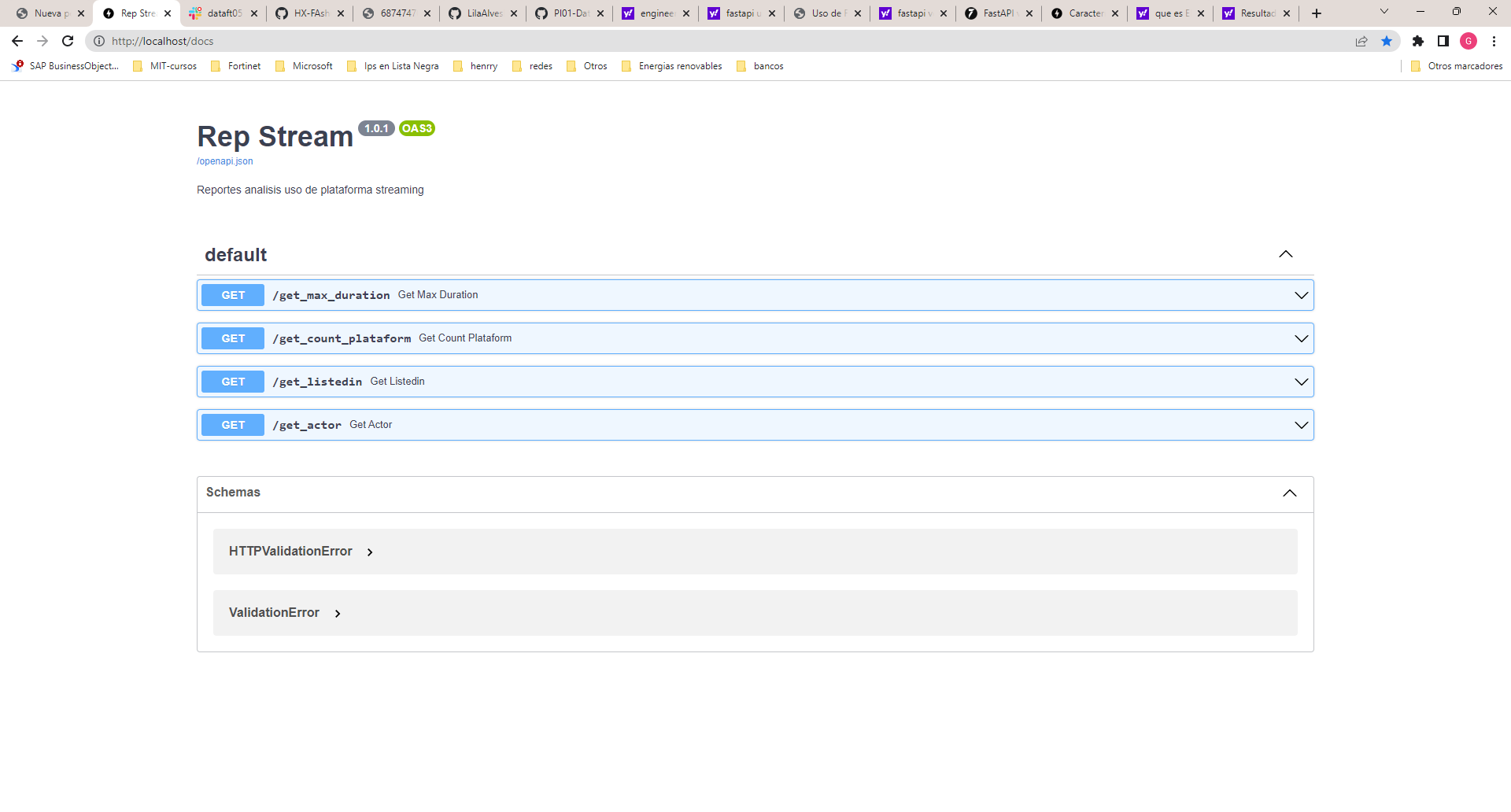
4.3.-Instalacion de Docker desktop para Windows.

4.4. Instalación de FastAPI donde se organiza:

App que contiene las librerías Config, Models, Router



En router se detallan la lógica de las consultas para que son visualizadas a través del localhost



5.- Conclusiones y recomendaciones

5.1 En la base de Hulu en la columna Rating se encontró información duraciones de películas/serie. No se procedió a realizar la modificación, ya que se validó los resultados pruebas sin modificaciones.

5.2 Analizar y completar el llenado de otros campos, como por ejemplo Rating, que es una buena medida para análisis de plataformas.

6.- Revisar documentación anexa.