



# ESCUELA POLITÉCNICA NACIONAL

## ESCUELA DE FORMACIÓN DE TECNÓLOGOS



### ANÁLISIS DE DATOS

ASIGNATURA:

Análisis de Datos

PROFESOR:

Ing. Lorena Chulde / Ing. Juan Pablo Zaldumbide

PERÍODO ACADÉMICO:

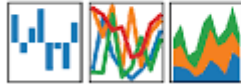
2023-B

### EXAMEN BIMESTRAL



pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



	ItemName	MinWeight	MaxWeight
0	Caribbean	450	450
1	Blue	510	450
2	Green	580	530
3	Red	570	640
4	Neutered	880	850
5	ShortNeutered_1	1850	1570
6	ShortNeutered_2	2280	2100
7	Gray	1380	1380

2023 - B

En los archivos proporcionados se encuentra datos de Netflix y la aceptación que cada película/serie tiene, 18337 registros se encuentran distribuidos en 4 archivos. Los campos del dataset son:

Title: Nombre de la película.

Available Globally?: Explica si se encuentra disponible a nivel mundial.

Release Date: Fecha de estreno en la plataforma

Hours Viewed: Horas vistas

Number of Ratings: Número de calificaciones "pulgar arriba"

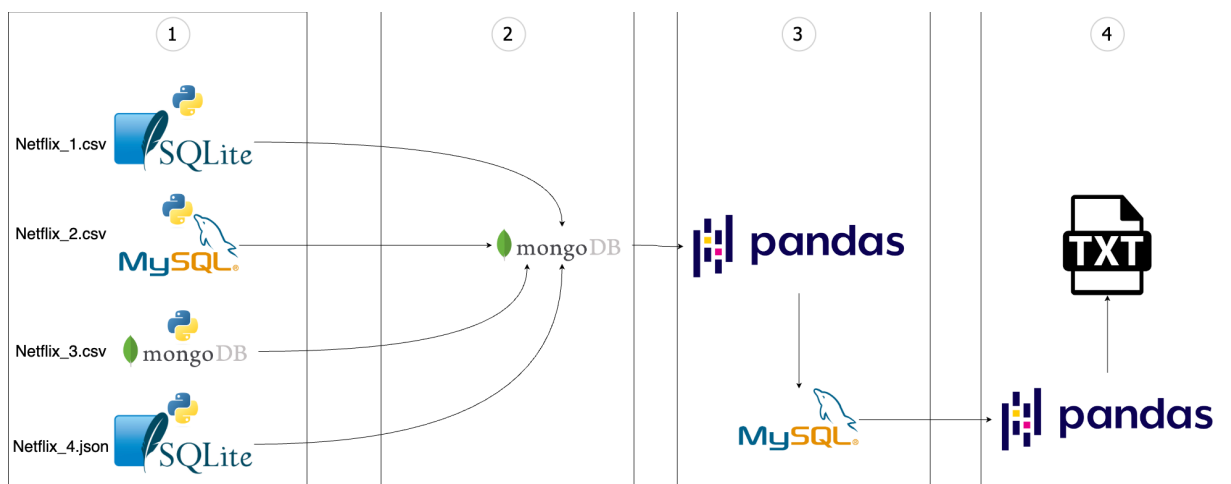
Rating: Clasificación general

Genre: Género

Key Words: Palabras clave

Description: Descripción

Dada la siguiente arquitectura:



**Se solicita:**

### **Etapas 1 - Importación de datos**

Debe importar los archivos suministrados a cada una de las bases de datos mencionadas en el esquema. Puede usar cualquier método que usted crea conveniente. (25% del puntaje total)

### **Etapas 2 - Consolidación de datos**

Una vez que tiene las diferentes fuentes de datos debe exportar en cualquier formato y consolidar las bases de datos en mongoDB utilizando cualquier método (25% del puntaje total)

### **Etapas 3 - Limpieza de datos**

En esta etapa debe importar los datos desde MongoDB y limpiarlos con pandas o cualquier herramienta que considere, deberá fijarse en campos vacíos y reemplazar por el promedio en el caso de que sea numérico, si es texto puede poner un texto estándar "info no disponible". Una vez limpia la data, se deberá exportar a MySQL (25% del puntaje total)

**Etapas 4 - Presentación de resultados**

La respuesta a cada una de las siguientes preguntas se guardará en un archivo de texto diferente:

a.txt - ¿Cuál es porcentaje de películas/series que están disponibles a nivel mundial?

b.txt - ¿Cuáles son las 10 películas/series más antiguas y las 10 más recientes?

c.txt - Tomando en cuenta que el género es una lista de Python, ¿Cuál es el género que más se repite?

d.txt - Según el rating ¿Las 10 mejores películas/series y las 10 peores películas/series?

e.txt - La keyword que más se repite.

Cada consulta corresponde al 5% del puntaje total.

**Entregables:**

Subir en el aula virtual en un solo archivo comprimido zip, rar, tar, etc:

- 1.- Código fuente <apellido.py> o <apellido.ipynb>
- 2.- pdf con capturas de pantalla de la ejecución de cada ejercicio.
- 3.- Archivos .txt generados