



## Entrega 1

### Tema: Abstracción de Datos y Tareas

Fecha de entrega: 27 de agosto a las 20:00 hrs

## Contexto

Una [Película Biográfica](#) (*Biopic*) es una película donde se representa la vida o los años más importantes de una persona o conjunto de personas reales. Las películas biográficas usualmente se enfocan en personas queridas e históricamente importantes.

## Dataset

El *dataset* a utilizar en este ejercicio es [biopics.csv](#) y contiene información de varias películas biográficas<sup>1</sup>. Las columnas que lo componen son las siguientes:

Columna	Explicación
title	Título de la película
site	Enlace a página de IMDB
country	País de origen
year_released	Año de estreno
box_office	Ganancias monetarias en USD
director	Director de la película
subject	Nombre de la persona cuya biografía está siendo representada
type_of_subject	Por lo qué es conocida esa persona
race_known	Indica si la etnicidad del personaje es determinable
subject_race	Etnicidad del personaje
person_of_color	Indica si personaje es de color
subject_sex	Sexo biológico del personaje
lead_actor_actress	Actor o actriz que representa al sujeto

## 1. Abstracción de tareas

### 1.1. Objetivo

A continuación se te presentará un problema de visualización, el cual puede ser resuelto utilizando el *dataset* de películas biográficas. Basándote en la explicación del problema, identifica el o los **por qué** según el *framework* de Tamara Munzner revisado en el curso.

<sup>1</sup>Dataset original extraído del repositorio de datos de [FiveThirtyEight](#)

Un grupo de amigxs tienen una obsesión por todo tipo de películas, y discuten constantemente sobre ellas en su tiempo libre. Últimamente, se han reunido a ver varias películas biográficas y a hablar sobre ellas.

Un tema en particular que ha aparecido durante sus conversaciones es sobre las personas protagonistas de dichas películas y su sexo biológico: algunos integrantes del grupo creen que la cantidad de películas con protagonistas de sexo masculino es mucho mayor a aquellas con protagonistas de sexo femenino de forma consistente, mientras que otros creen que a medida que pasan los años, esa desigualdad ha ido disminuyendo de forma gradual. Es por esto que les gustaría crear una visualización que les ayude a responder esta inquietud.

Basándote en el párrafo anterior, tendrás que especificar tareas que contengan una **acción** y **objetivo** correspondientes al problema. Debes identificar al menos una tarea que ataque el problema, y justificar brevemente la elección de términos. También deberás indicar qué **datos** del *dataset* entregado, o derivado de este, se utilizarían en cada par identificado.

## 2. Abstracción de datos

### 2.1. Objetivo

Para cada columna del *dataset* entregado al inicio del enunciado, y para cada dato derivado que hayas detectado en la parte anterior, deberás clasificar cada atributo según los tipos con los que se identifica y justificar porqué aplica o no tal clasificación. Las siguientes son las distintas clasificaciones mencionadas en el curso para atributos:

- ¿Es un atributo categórico u ordenado?
- ¿Es ordinal o cuantitativo?
- ¿Es secuencial o divergente?
- ¿Es cíclico o no cíclico?
- ¿Es un atributo de llave o valor?

Por ejemplo, para la columna `box_office`, su clasificación sería:

- **Ordenado**, ya que la cantidad de dinero ganada sí cuenta con un orden intrínseco y numérico.
- **Cuantitativo**. La cantidad de dinero sí cuenta con operaciones aritméticas exactas, la diferencia entre valores permite apreciar cuanto menos ganancia de dinero existe.
- **Secuencial**. *Box office* en particular solo puede tomar valores positivos, y no hay un potencial valor “cero” que divida su rango.
- **No cíclico**, ya que sus valores podrían aumentar indefinidamente.
- Es atributo de **valor**, ya que dos películas pueden compartir el mismo *box office*, lo que no lo vuelve una columna válida para identificar una película.

Finalmente responde, **justificando** brevemente tu respuesta:

- ¿De qué tipo de *dataset* se trata **biopics.csv**?
- ¿Es un *dataset* estático o dinámico?
- ¿Qué tipos de datos singulares hay presentes en el *dataset*? (Atributo, ítem, enlace, posición o grilla)

### 3. Entregables

La entrega se realizará mediante la plataforma [Canvas](#) en la [evaluación correspondiente](#). Para ésta entrega se deberá subir un archivo de extensión PDF con todas las respuestas solicitadas anteriormente.

### 4. Dudas

Cualquier duda que tengas sobre esta evaluación, prefiere publicarla en el [foro](#) correspondiente a esta evaluación. También siéntete libre de responder dudas de compañerxs si crees que manejas la respuesta.

### 5. Política de atraso

Las entregas tienen como fin preparar al estudiante para la instancia evaluativa que lo sigue (Hito). Es por esto que **no son evaluadas** y tienen un carácter completamente **opcional**. Sin embargo, la no entrega de alguna o entrega atrasada (por cualquier cantidad de tiempo) significará que el estudiante no recibirá *feedback* de su trabajo por parte del equipo docente. Este *feedback* tiene como finalidad corregir errores en la entrega para tener un mejor desempeño en el hito, por lo que se recomienda fuertemente la realización de este informe.