

Disney Analysis

2022

Marcos Santillan
Marco Montoya



**Data
Analytics**

CODER HOUSE

Índice:

1. Descripción de la temática de los datos:	3
2. Para nuestro análisis responderemos a las siguientes preguntas de investigación:	3
3. Datasets.	4
4. Diagrama entidad relación:	4
5. Listado de Tablas:	5
1) Películas:.....	5
2) Director:.....	5
3) Ganancias disney:.....	5
4) Personajes:	5
6. Listado de columnas por tablas:	6
7. Transformación de datos:.....	7
8. Variables, Medidas y Columnas Calculadas:	10



DISNEY WORLD MOVIES ANÁLISIS

1. Descripción de la temática de los datos:

Este proyecto busca encontrar la relación entre la taquilla (ingreso) bruta, los ratings de la MPAA, personajes como protagonistas, el estudio y directores en la elaboración de las Disney películas. La suposición común es que las películas con clasificación G generan la mayor cantidad de ingresos porque a la mayor parte de los espectadores se les permite la entrada a estas películas, tanto niños como adultos y que los personajes más mundialmente conocidos como podría ser el caso de Pinocho son las películas de mayor ingreso.

¿Serán estas presunciones reales?

2. Para nuestro análisis responderemos a las siguientes preguntas de investigación:

Para nuestro análisis responderemos a las siguientes preguntas de investigación:

¿Existe una relación significativa entre los ingresos brutos y el rating de películas?

¿Cómo ha ido cambiado esta relación través de las décadas?

¿En los últimos años, las películas PG se han posicionado en la parte superior de las películas con mayor ingreso bruto?

¿Es esto únicamente porque hay más películas PG que G lanzadas por Disney? ahora que en el pasado?

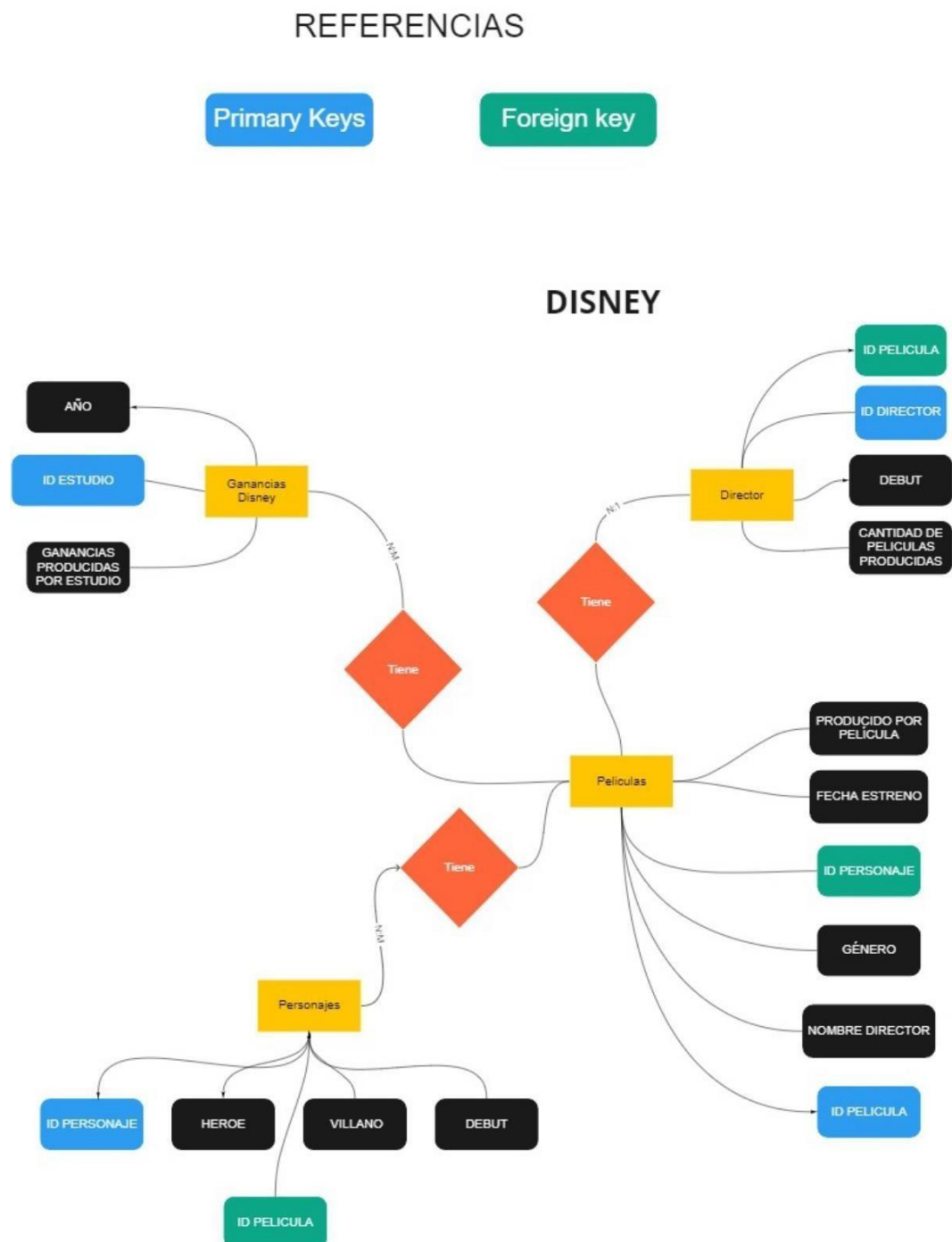


3. Datasets.

Se adjunta el archivo Excel con las bases de datos Organizadas y estructuradas utilizadas en el trabajo.

“Base-de-datos-del-proyecto-final-base-de-datos-Disney”

4. Diagrama entidad relación:



5. Listado de Tablas:

En este apartado, se hará mención de cada una de las tablas identificando sus claves primarias y foráneas:

1) Películas:

Esta es la tabla principal donde bajo la PK ID Película se relaciona con las FK de todas las demás tablas, pudiendo relacionar la información del tipo que personajes realizo más películas así como que director, que película es la que más ganancia genero a un estudio, ganancias a través de los años abarcando fechas de estreno desde 1937 a 2016.

- PK: ID PELICULA.
- FK: ID PERSONAJE.
- FK: ID DIRECTOR.
- FK: ID ESTUDIO.

2) Director:

Esta tabla contiene el nombre y apellido de cada director con su respectivo ID como clave PK y como clave FK ID Película para poder vincular ambas relaciones.

- PK: ID DIRECTOR.
- FK: ID PELICULA.

3) Ganancias disney:

Esta tabla contiene el año y ganancias de cada Estudio realizador de películas con su respectivo ID para poder vincularlo a su respectiva película.

- PK: ID ESTUDIO.

4) Personajes:

Esta tabla contiene la fecha de estreno de cada película con su personaje protagonista con su respectivo ID como clave PK y la clave FK ID Película para poder vincular ambas relaciones.

- PK: ID PERSONAJE.
- FK: ID PELICULA.



6. **Listado de columnas por tablas:**

A continuación, se hará mención de las columnas que posee cada tabla junto con su tipo de campo y clave:

TABLA 1: PELÍCULAS

TIPO DE CLAVE	CAMPO	TIPO DE CAMPO
-	Fecha_Estreno	datetime
FK	ID PERSONAJE	nvarchar(100)
-	Género	text(100)
FK	ID DIRECTOR	text(100)
PK-index	ID PELÍCULA	nvarchar(100)
FK	ID ESTUDIO	nvarchar(100)
	Rating	text(100)
-	GANANCIAS	int
-	inflation_adjusted_gross	int

TABLA 2 : DIRECTOR

TIPO DE CLAVE	CAMPO	TIPO DE CAMPO
PK-index	ID DIRECTOR	text(100)
FK	ID PELÍCULA	nvarchar(100)

TABLA 3 : GANANCIAS DISNEY

TIPO DE CLAVE	CAMPO	
PK	ID_ESTUDIO	nvarchar(100)
-	Año	int
-	Ganancias_por estudio EN MILLONES U\$S	int



7. Transformación de datos:

- a) En BI accedemos a la sección transformar datos. Abrimos el Power Query. Tildamos la opción “Usar la Primera Fila como Encabezado”, como paso 2, detectamos el tipo de dato, accedemos a la sección “Transformar” y tildamos “Detección Tipo de Datos” previamente presionando la tecla Shift para seleccionar todas las columnas. Verificamos que la detección coincida con los tipos de datos.
- b) En sección “**Vista**” tildamos “calidad de columnas” para ver en qué estado se encuentra nuestra base de datos.
- c) **Agregamos columna condicional en tabla Películas:** Diferenciamos las películas en categorías según su nivel de ingreso. Siendo la categoría 2, las de mayor ganancia.
- d) **Combinar Consultas:** Desde Inicio, opción “combinar consultas”: Combinamos las tablas “**Películas**”; “**Ganancias**”; “**Director**”; “**Personajes**” a través de la columna ID Estudio, Director y Personajes Respectivamente de manera de poder agregar a nuestra tabla las ganancias de cada estudio en millones asociada a su película, el código ID de cada director y de Personaje.

▲ PASOS APLICADOS	
Origen	✳
Navegación	✳
Encabezados promovidos	✳
Tipo cambiado	
Columna condicional agregada	✳
Columnas quitadas	
Consultas combinadas	✳
Consultas combinadas1	✳
Consultas combinadas2	✳
Se expandió Director	✳
Se expandió Ganancias Disney	✳
✕ Se expandió Personajes	✳



- e) **Modo Editor Avanzado en sección Vista:** Visualizamos los cambios que venimos realizando en modo backend.

Películas

```
let
    Origen = Excel.Workbook(File.Contents("C:\Users\W10-PC\Documents\Data Analyst\Coderhouse; Data Analisis\Datasets\Base-de-datos-del-proyecto-final-base-de-datos-disney.xlsx"), null, true),
    Peliculas_Sheet = Origen[Item="Peliculas",Kind="Sheet"][Data],
    #Encabezados promovidos = Table.PromoteHeaders(Peliculas_Sheet, [PromoteAllScalars=true]),
    #Tipo cambiado = Table.TransformColumnTypes(#Encabezados promovidos,{{"ID_PELÍCULAS", type text}, {"Nombre_Personaje", type text}, {"Nombre_director", type text}, {"ID_ESTUDIO", type text}, {"Fecha_estreno", type date}},
    #Columna condicional agregada = Table.AddColumn(#Tipo cambiado, "Personalizado", each if [GANANCIAS] <= 100000 then "Categoría 1" else "Categoría 2"),
    #Columnas quitadas = Table.RemoveColumns(#Columna condicional agregada,{"Column11", "Column10"}),
    #Consultas combinadas = Table.NestedJoin(#Columnas quitadas, {"Nombre_director"}, Director, {"Nombre_director"}, "Director", JoinKind.LeftOuter),
    #Consultas combinadas1 = Table.NestedJoin(#Consultas combinadas, {"ID_ESTUDIO"}, #Ganancias Disney, {"ID_ESTUDIO"}, "Ganancias Disney", JoinKind.LeftOuter),
    #Consultas combinadas2 = Table.NestedJoin(#Consultas combinadas1, {"Nombre_Personaje"}, Personajes, {"Nombre_Personaje"}, "Personajes", JoinKind.LeftOuter),
    #Se expandió Director = Table.ExpandTableColumn(#Consultas combinadas2, "Director", {"ID_Director"}, {"Director.ID_Director"}),
    #Se expandió Ganancias Disney = Table.ExpandTableColumn(#Se expandió Director, "Ganancias Disney", {"Ganancias_por estudio en millones US$"}, {"Ganancias Disney.Ganancias_por estudio en millones US$"}),
    #Se expandió Personajes = Table.ExpandTableColumn(#Se expandió Ganancias Disney, "Personajes", {"ID_Personaje"}, {"Personajes.ID_Personaje"})
in
    #Se expandió Personajes
```

- f) **Creación de Parámetro:** Al trabajar en equipo creamos este parámetro para que el origen de nuestra Base de Datos se encuentre en una URL y poder utilizarla de manera compartida.

URL MARCOS

```
"URL MARCOS" meta [IsParameterQuery=true, Type="Text", IsParameterQueryRequired=true]
```

Consultas [5]

Director

Ganancias Disney

Películas

Personajes

URL MARCOS (URL MAR...

Valor actual

C:\Users\W10-PC\Documents\Data Analyst\Coderhouse; D

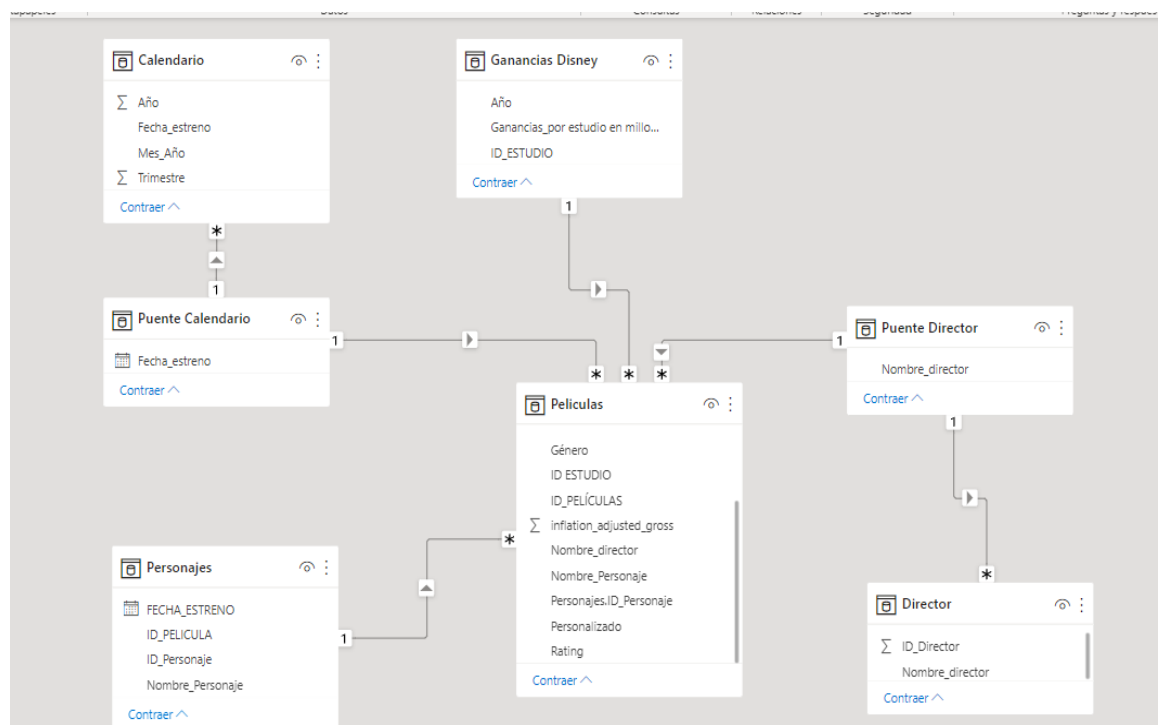
Administrar parámetro



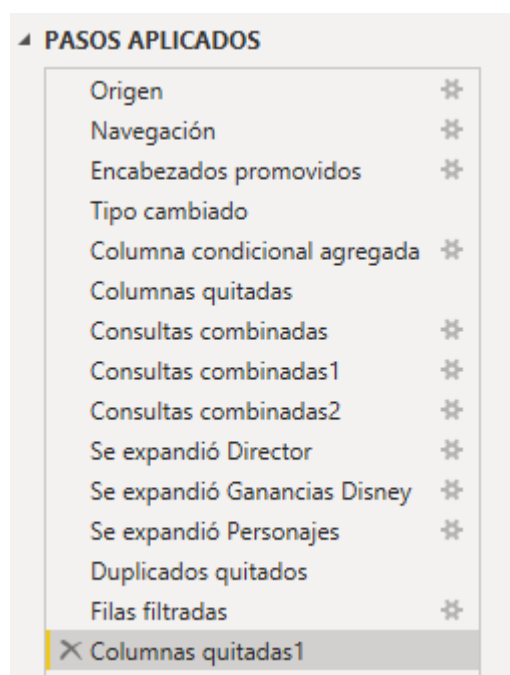
- g) **Tabla Calendario:** Creamos esta tabla que nos permite gestionar los indicadores y la segmentación de manera eficiente. Para evitar que se nos muestren días que no contengan datos y simplificar el análisis. Duplicamos Tabla Películas, y borramos todas las columnas de la tabla duplicada menos la de fecha de ingreso. **Renombramos** la nueva tabla como calendario. Creamos columnas Año, Mes Año y Trimestre.

	Fecha_estreno	Año	Mes_Año	Trimestre
1	21/12/1937	1937	12/1937	4
2	21/12/1937	1937	12/1937	4
3	21/12/1937	1937	12/1937	4
4	21/12/1937	1937	12/1937	4
5	21/12/1937	1937	12/1937	4
6	21/12/1937	1937	12/1937	4
7	21/12/1937	1937	12/1937	4
8	21/12/1937	1937	12/1937	4
9	21/12/1937	1937	12/1937	4
10	21/12/1937	1937	12/1937	4
11	21/12/1937	1937	12/1937	4
12	21/12/1937	1937	12/1937	4
13	21/12/1937	1937	12/1937	4
14	21/12/1937	1937	12/1937	4

- h) **Creamos tabla Puente:** Para crear las relaciones de tablas Director y Calendario con la tabla Película para evitar la duplicidad que genera una relación de muchos a muchos (*-*). Empezamos duplicando la tabla calendario en transformación de datos, quitamos todas las columnas menos fecha de estreno, eliminamos duplicados de esta columna, y ahora si relacionamos esta tabla puente con la tabla calendario y a su vez con nuestra tabla principal Película. Mismo proceso para tabla Director



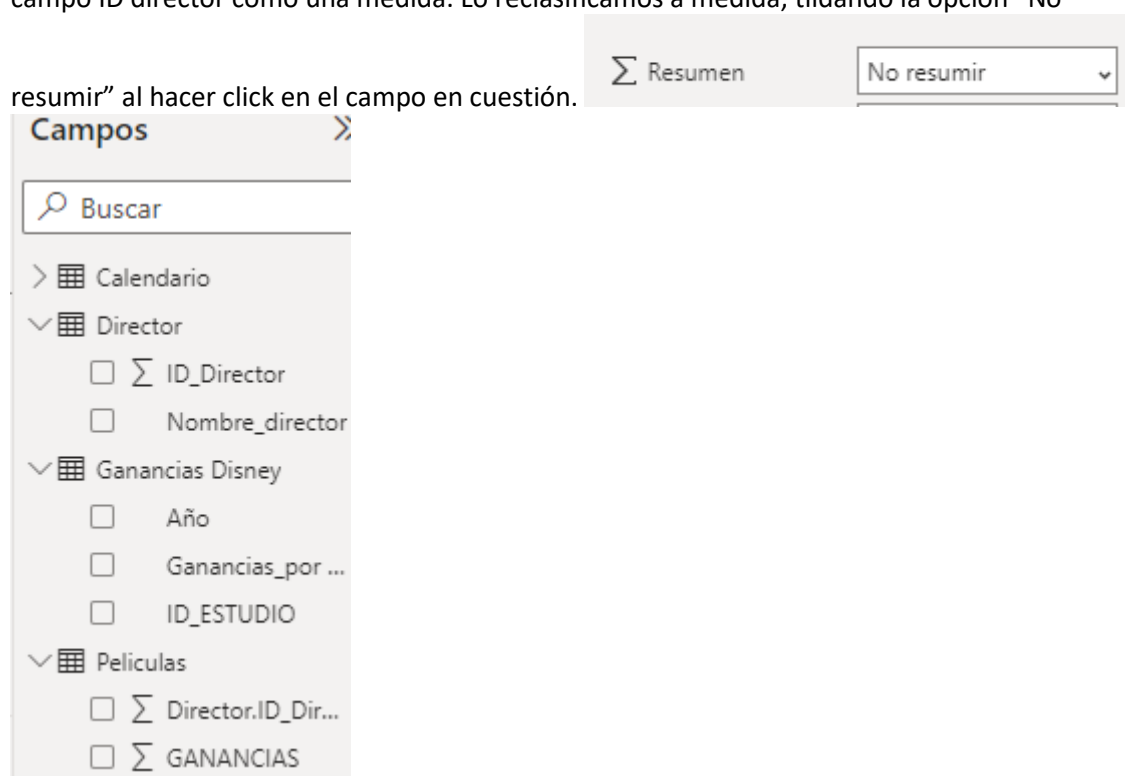
e) **Eliminamos duplicados** en columna ID_Películas en tabla Películas. Quitamos columna Ganancias Ajustados x Inflación,



8. Variables, Medidas y Columnas Calculadas:

- a) **Reclasificación campos de medida a dimensión:** Power Bi clasifico de manera automática el campo ID director como una medida. Lo reclasificamos a medida, tildando la opción “No

resumir” al hacer click en el campo en cuestión.



Mismo proceso, pero al revés para los campos Ganancias que estaban como dimensión los pasamos a medida.



b) Generación Nuevas Columnas Calculadas:

1. Creamos en Tabla Principal Películas la columna Antigüedad_Películas para en base a su fecha estreno saber hace cuantos años se estrenaron.

```
Antigüedad_Películas = DATEDIFF(Películas[Fecha_estreno], TODAY(), YEAR)
```

2. Creamos en Tabla Personajes la columna Antigüedad_Personajes para en base a su fecha estreno saber hace cuantos años debutaron.

```
Antigüedad_Personaje = DATEDIFF(Personajes[FECHA_ESTRENO], TODAY(), YEAR)
```

c) Generación Nuevas Medidas Calculadas:

1. Creamos una tabla nueva "AMedidas" donde albergaremos todas las medidas creadas.

2. Creamos la medida Cantidad_Películas.

```
Cantidad_Pelicula = COUNT(Películas[ID_PELÍCULAS])
```

Creamos una variable que almacene el valor de contar las películas.

3. Creamos la medida Cantidad_Personajes.

```
Cantidad_Personajes = COUNT(Personajes[ID_Personaje])
```

Creamos una variable que almacene el valor de contar la cantidad de personajes.

4. Creamos la medida Cantidad_Directores.

```
Cantidad_Directores = COUNT('Puente Director'[Nombre_director])
```

Creamos una variable que almacene el valor de contar la cantidad de directores.

5. Creamos la medida Max_Ganancia_Pelicula.

```
1 Max_Ganancia_Pelicula = MAX(Películas[Ganancias_Películas_MillonesUSD_AjustadaPorInflacion])
```

Creamos una variable que almacene el valor de la máxima ganancia por película

6. Creamos la medida Fecha_PrimerPelícula.

```
Fecha_PrimerPelícula = MIN(Calendario[Fecha_estreno])
```

Creamos una variable que almacene el valor de la fecha de estreno de las películas.

