

## Proyecto Catalogo Animes

Se obtuvo el conjunto de Datos [Anime Dataset 2023](#) de la Plataforma Kaggle sobre las calificaciones y visualizaciones de animes de la plataforma MyAnimeList.com hasta el año 2023. El conjunto de datos incluía varias tablas de Animes y Usuarios, para este proyecto utilizare:

- anime-dataset-2023.csv: Contiene información del Catálogo de Animes y sus datos generales.
- anime-filtered.csv: Contiene la misma información, pero con columnas que ayudan a medir las visualizaciones del Anime como personas viéndolo, personas que lo dejaron, etc.

En un Excel se dejó únicamente la columna ID, nombre y las columnas Watching, Completed, On\_Hold y Dropped del Archivo anime-filtered.csv, para posterior unirlos en SQL.

Se importaron a SQL Managment Studio utilizando la herramienta de importación de SQL, por lo que muchas Columnas Incluyen valores NULL y/o se pasaron con algún formato de NVARCHAR o TEXT.

### Contenido de las Columnas

Tabla Animes (anime-dataset-2023.csv)

- **anime\_id: ID único de cada Anime.**
- **Name: Nombre del Anime.**
- English name: Nombre en Ingles del Anime.
- Other name: Nombre Nativo del Anime (Japones, Chino, Coreano).
- **Score: Rating dado al Anime.**
- Genres: Géneros a los que pertenece el Anime.
- Synopsis: Breve Descripción de la Historia
- **Type: Tipo de Contenido (TV series, movie, OVA, etc.).**
- **Episodes: Cantidad de Episodios.**
- **Aired: Fechas de Emisión.**
- **Premiered: Estación del Año en que se estrenó.**
- **Status: Estatus del Anime. (Finished Airing, Currently Airing, etc.).**
- Producers: Compañías productoras.
- Licensors: Plataformas / Empresas con la Licencia. (streaming platforms).
- Studios: Estudios de Animación que hicieron el Anime.
- **Source: Fuente del Material Original (manga, light novel, original).**
- **Duration: Duración cada Episodio.**
- **Rating: Clasificación.**
- **Rank: Ranking Basado en Popularidad u otros Criterios.**
- **Popularity: Popularidad del Anime.**
- **Favorites: Número de Personas que lo marcaron como favorito.**
- **Scored By: Número de Personas que calificaron al Anime.**
- **Members: Número de Personas que lo añadieron a sus Listas.**
- Image URL: URL a una Imagen.

Tabla Visualizaciones (anime-filtered.csv)

- Anime\_id: ID único del Anime.
- Name: Nombre del Anime
- WATCHING: Cantidad de personas que lo están viendo.
- COMPLETED: Cantidad de personas que lo terminaron de Ver.
- ON\_HOLD: Cantidad de Personas que lo dejaron en Pausa.

- DROPPED: Cantidad de personas que lo dejaron de Ver.

```
Select * from Catalogo_Animes;
Select * from Visualizaciones;
```

anime_id	Name	English_name	Other_name	Score	Genres	Synopsis	Type	Episodes	Aired
1	Cowboy Bebop	Cowboy Bebop	?????????	8.75	Action, Award Winning, Sci-Fi	Crime is timeless. By the year 2071, humanity has e...	TV	26	Apr 3, 1998 to Apr 24, 1999
5	Cowboy Bebop: Tengoku no Tobira	Cowboy Bebop: The Movie	????????? ????	8.38	Action, Sci-Fi	Another day, another bounty—such is the life of the of...	Movie	1	Sep 1, 2001
6	Trigun	Trigun	?????	8.22	Action, Adventure, Sci-Fi	Vash the Stampede is the man with a \$500,000,000...	TV	26	Apr 1, 1998 to Sep 30, 1998
7	Witch Hunter Robin	Witch Hunter Robin	Witch Hunter ROBIN (?????????????)	7.25	Action, Drama, Mystery, Supernatural	Robin Sena is a powerful craft user drafted into the ...	TV	26	Jul 3, 2002 to Dec 25, 2002
8	Bouken Ou Beet	Beet the Vandal Buster	???????	6.94	Adventure, Fantasy, Supernatural	It is the dark century and the people are suffering un...	TV	52	Sep 30, 2004 to Sep 29, 20
15	Eyeshield 21	UNKNOWN	???????21	7.92	Sports	Shy, reserved, and small-statured, Deimon High Sc...	TV	145	Apr 6, 2005 to Mar 19, 2006
16	Hachimitsu to Clover	Honey and Clover	???????????	8	Comedy, Drama, Romance	Yuuta Takemoto, a sophomore at an arts college, sh...	TV	24	Apr 15, 2005 to Sep 27, 200
17	Hungry Heart: Wild Striker	UNKNOWN	????????? Wild Striker	7.55	Comedy, Slice of Life, Sports	As the younger brother of Japanese soccer star Seis...	TV	52	Sep 11, 2002 to Sep 10, 20

anime_id	Name	Watching	Completed	On_Hold	Dropped
1	Cowboy Bebop	105808	718161	71513	26678
5	Cowboy Bebop: Tengoku no Tobira	4143	208333	1935	770
6	Trigun	29113	343492	25465	13925
7	Witch Hunter Robin	4300	46165	5121	5378
8	Bouken Ou Beet	642	7314	766	1108
15	Eyeshield 21	13907	78349	14228	11573
16	Hachimitsu to Clover	11909	81145	11901	11026
17	Hungry Heart: Wild Striker	817	13778	828	1168

Se cambio el Tipo de Dato de los ID's para trabajarlos como INT.

```
ALTER TABLE Catalogo_Animes ALTER COLUMN ANIME_ID INT;
ALTER TABLE Visualizaciones ALTER COLUMN ANIME_ID INT;
```

Se valido en las tablas originales que los Id's fueran únicos y no hubiera valores repetidos.

```
Limpieza Parte 1.s...4P5VPUB\Jorge (82))* x SQLQuery1.sql - LA...P5VPUB\Jorge (80))*
```

```
-- Validar si hay un ID repetido tabla Anime
WITH ID_REPETIDOS AS (
Select anime_id, count(anime_id) AS CANTIDAD from Catalogo_Animes
group by anime_id)
SELECT * FROM ID_REPETIDOS WHERE CANTIDAD >1;

-- Validar si hay un ID repetido tabla Visualziacion
WITH ID_REPETIDOS AS (
Select anime_id, count(anime_id) AS CANTIDAD from Visualizaciones
group by anime_id)
SELECT * FROM ID_REPETIDOS WHERE CANTIDAD >1;
```

anime_id	CANTIDAD
anime_id	CANTIDAD

Ciertas columnas que deberían tener datos Numéricos o de fecha para trabajarlos correctamente contienen la palabra UNKOWN, esto no permite que se puedan pasar al tipo de dato correcto, por lo que se cambiaron esos valores a 0.

Excepto la columna de SCORE, se le dio valor de 1 para que no afecte.

Las columnas con fecha DATE se pasaron como NULL.

La idea es trabajar en una tabla nueva compuesta por ciertas columnas de ambas tablas y sobre esa comenzar a limpiar la información.

Se validó el largo de algunas de las columnas que deben ir como VARCHAR para ponerles un valor adecuado al momento de crear la tabla.

```
--Validar Largo Columnas para cambiar a NVARCHAR
SELECT max(LEN(NAME)) Name, max(LEN(English_name)) English_name, max(LEN(Genres)) Genres,
max(LEN(Episodes)) Episodes, max(LEN(Producers)) Producers,
max(LEN(Licensors)) Licensors, max(LEN(Studios)) Studios, max(LEN(Source)) Source,
max(LEN(Rating)) Rating, max(LEN(image_url)) image_url
FROM Catalogo_Animes
```

9 %

Results Messages

Name	English_name	Genres	Episodes	Producers	Licensors	Studios	Source	Rating	image_url
126	120	73	4	375	69	126	12	30	64

## Creación y llenado de la Tabla.

```
CREATE TABLE animes
(
  anime_id INT,
  NAME VARCHAR(150),
  score FLOAT,
  genero NVARCHAR(100),
  type NVARCHAR(20),
  episodes INT,
  fecha_estreno NVARCHAR(50),
  temporada NVARCHAR(50),
  status NVARCHAR(20),
  producers NVARCHAR(400),
  licensors NVARCHAR(100),
  studio NVARCHAR(150),
  source NVARCHAR(200),
  duration NVARCHAR(50),
  raiting NVARCHAR(50),
  ranking INT,
  popularity INT,
  favorites INT,
  scored_by INT,
  members INT,
  watching INT,
  completed INT,
  on_hold INT,
  dropped INT
);
```

Se lleno la información haciendo un LEFT JOIN para traer la información de todo el catalogo de Animes, tenga información o no de en la tabla de visualizaciones.

```
-- Insert tabla Nueva
Insert into Animes
SELECT a.anime_id, a.name, cast(score as float), Genres, Type, Episodes, Aired, Premiered, Status, Producers, Licensors, Studios,
Source, Duration, Rating, Rank, Popularity, Favorites, Scored_By, Members
, WATCHING, COMPLETED, ON_HOLD, DROPPED FROM
Catalogo_Animes a LEFT JOIN Visualizaciones b ON a.anime_id = b.anime_id;
```

## Limpieza de las Tablas.

### Columna NAME

Se valida Contenido Tabla de Catalogo\_Animes de la columna NAME para ver valores duplicados o que se pudieran Agrupar. Se encontraron 6 nombres repetidos, pero al validarlos todos son diferentes registros, ya que tienen diferente TYPE O Genero, el único que considere duplicado fue el ID 55680, el cual se eliminó.

```
-- Validar Duplicados Nombre
With duplicados as (
Select name, count(*) cant
from anims group by name)
select * from duplicados where cant >1;

Select * from anims where
name in ('Awakening', '1', 'Azur Lane', 'Kanojo mo Kanojo', 'Souiseiki', 'Utopia')
order by 2;
```

Results Messages

name	cant
1	2
Awakening	2
Azur Lane	2
Kanojo mo Kanojo	2
Souiseiki	2
Utopia	2

anime_id	NAME	SCORE	GENERO	TYPE	EPISODES	Fecha_Estreno	Temporada	status	producers	licensors	Studio	Source	duration	rating	ranking	popularity	favorites	scored_5
29978	1	3.35	Comedy	ONA	1	Feb 7, 2013	UNKNOWN	Finished Airing	UNKNOWN	UNKNOWN	UNKNOWN	Original	37 sec	G - All Ages	12664	5022	5	11440
54453	1	NULL	Horror	Movie	NULL	2007	UNKNOWN	Finished Airing	UNKNOWN	UNKNOWN	UNKNOWN	Original	19 min	PG-13 - Teens 13 or older	NULL	NULL	NULL	NULL
54618	Awakening	NULL	UNKNOWN	Music	NULL	May 11, 2020	UNKNOWN	Finished Airing	UNKNOWN	UNKNOWN	UNKNOWN	Mixed media	3 min	UNKNOWN	NULL	NULL	NULL	NULL
55688	Awakening	NULL	UNKNOWN	Music	NULL	Oct 30, 2002	UNKNOWN	Finished Airing	UNKNOWN	UNKNOWN	UNKNOWN	Original	4 min	G - All Ages	NULL	NULL	NULL	NULL
55351	Azur Lane	NULL	Action, Slice of Life	Special	NULL	Apr 17, 2020 to ?	UNKNOWN	Currently Airing	UNKNOWN	UNKNOWN	UNKNOWN	Game	30 sec	PG-13 - Teens 13 or older	NULL	NULL	NULL	NULL
38328	Azur Lane	6.26	Action, Sci-Fi	TV	12	Oct 3, 2019 to Mar 20, 2020	fall 2019	Finished Airing	Frontier Works, Sotsu, Nitroplus, TOHO animation, L.	Funimation	Bibury Animation Studios	Game	23 min per ep	R+ - Mild Nudity	7502	1343	960	50398
43969	Kanojo mo Kanojo	6.54	Comedy, Romance	TV	12	Jul 3, 2021 to Sep 18, 2021	summer 2021	Finished Airing	Mainichi Broadcasting System, Kodansha, Half H.P.	UNKNOWN	Tezuka Productions	Manga	24 min per ep	PG-13 - Teens 13 or older	6132	654	2292	161567
55680	Kanojo mo Kanojo	NULL	UNKNOWN	UNKNOWN	NULL	Not available	UNKNOWN	Finished Airing	UNKNOWN	UNKNOWN	UNKNOWN	Unknown	Unknown	UNKNOWN	NULL	NULL	NULL	NULL
55610	Souiseiki	NULL	Fantasy	OVA	NULL	Not available	UNKNOWN	Finished Airing	UNKNOWN	UNKNOWN	UNKNOWN	Original	Unknown	G - All Ages	NULL	NULL	NULL	NULL
3473	Souiseiki	5.38	UNKNOWN	Movie	1	Oct 1, 1968	UNKNOWN	Finished Airing	Mushi Production	UNKNOWN	UNKNOWN	Unknown	4 min	PG - Children	11141	10929	1	965
21103	Utopia	NULL	UNKNOWN	Music	NULL	Feb 26, 2014	UNKNOWN	Finished Airing	UNKNOWN	UNKNOWN	P.A. Works	Original	1 min	PG-13 - Teens 13 or older	NULL	NULL	NULL	NULL
55582	Utopia	NULL	UNKNOWN	Music	NULL	Aug 12, 2021	UNKNOWN	Finished Airing	UNKNOWN	UNKNOWN	UNKNOWN	Original	3 min	G - All Ages	NULL	NULL	NULL	NULL

### Columna TYPE

No se encontraron valores extraños o diferentes dentro de la Columna Type, todo esta uniforme y no se hacen más cambios.

Los valores UNKOWN se dejan así.

```
Select type, count(distinct type) cant
from anims group by type;
```

Results Messages

type	cant
Movie	1
Music	1
ONA	1
OVA	1
Special	1
TV	1
UNKNOWN	1

## Columna STATUS

La columna no presenta valores irregulares, repetidos o valores NULL, todo está bien.

```
-- Validar Duplicados status
Select * from animes;

Select status, count(distinct status) cant
from animes group by status;
```

9 %

	status	cant
1	Currently Airing	1
2	Finished Airing	1
3	Not yet aired	1

## Columna SOURCE

La columna no presenta valores irregulares o valores NULL, todo está bien.

```
-- Validar Duplicados Source
Select * from animes;

Select Source, count(distinct source) cant
from animes group by Source
order by 1;
```

9 %

	Source	cant
1	4-koma manga	1
2	Book	1
3	Card game	1
4	Game	1
5	Light novel	1
6	Manga	1
7	Mixed media	1
8	Music	1

## Columna DURATION

Se observa que hay valores que se pueden agrupar, ya que tienen la misma duración, pero algunos se les agrega la palabra “per episode”, se quitará esta palabra para agrupar más los valores.

```
-- Validar Duplicados duration
```

```
Select * from animes;
```

```
With duplicados as (  
Select distinct duration, count(*) cant  
from animes group by duration)  
select * from duplicados where cant >1  
order by 1;
```

234 %

Results Messages

duration	cant
76 12 sec	4
77 13 min	100
78 13 min per ep	439
79 14 min	119
80 14 min per ep	172
81 14 sec	3
82 15 min	178
83 15 min per ep	357
84 15 sec	33
85 15 sec per ep	28
86 16 min	127
87 16 min per ep	139
88 16 sec	5
89 16 sec per ep	4

Se utilizo el UPDATE para corregir las filas, se obtuvo el texto Antes del Delimitador “per” usando la función SUBSTRING y CHAR\_INDEX.

```
with Duration as(  
Select anime_id, name, duration,  
SUBSTRING(duration,1, CHARINDEX('per',duration)-1) as Duracion_Correcta  
from animes |  
where duration LIKE '%per%')  
UPDATE duration set duration = Duracion_Correcta;
```

14 %

```
Select * from animes;
```

```
With duplicados as (  
Select distinct duration, count(*) cant  
from animes group by duration)  
select * from duplicados where cant >1  
order by 1;
```

234 %

Results Messages

duration	cant
66 12 sec	5
67 13 min	539
68 14 min	291
69 14 sec	4
70 15 min	535
71 15 sec	61
72 16 min	266
73 16 sec	9
74 17 min	190
75 17 sec	2
76 18 min	205
77 18 sec	2
78 19 min	157
79 19 sec	4
80 2 hr	38
81 2 hr 1 min	8
82 2 hr 10 min	9
83 2 hr 11 min	2
84 2 hr 12 min	3
85 2 hr 14 min	2
86 2 hr 15 min	7
87 2 hr 16 min	6
-- --	--

Con la columna DURATION quise crear una columna con el Tiempo en minutos en formato INT, y no en el formato de texto de “1 hora y 20 minutos” por ejemplo que se maneja en el archivo.

Para sacar los valores que tienen horas se utilizó la siguiente consulta.

Usando las Funciones CAST para transformar el Obtenido a Entero, TRIM para quitarle espacios que pudieran quedar, y SUBSTRING – CHARINDEX para obtener los valores separados.

Para las películas que tienen Horas, multiplique el valor que saque primero por 60 para que fueran minutos.

```
--Columna duracion en Minutos
With T_Minutos as (
Select anime_id, name,duration,cast(TRIM(SUBSTRING(duration,1,1)) as INT)*60 as Horas,
cast(TRIM(SUBSTRING(duration,CHARINDEX('hr',duration)+3,2)) as INT) as Minutos
from animes WHERE duration NOT IN ('1 hr') and
DURATION LIKE '%hr%')
Select *, horas + Minutos as Tiempo_Minutos from T_minutos ;
```

anime_id	name	duration	Horas	Minutos	Tiempo_Minutos
5	Cowboy Bebop: Tengoku no Tobira	1 hr 55 min	60	55	115
31	Neon Genesis Evangelion: Death & Rebirth	1 hr 44 min	60	44	104
32	Neon Genesis Evangelion: The End of Evangelion	1 hr 26 min	60	26	86
43	Koukaku Kidoutai	1 hr 22 min	60	22	82
46	Rurouni Kenshin: Meiji Kenkaku Romantan - Ishinshis...	1 hr 31 min	60	31	91
47	Akira	2 hr 4 min	120	4	124
54	Appleseed (Movie)	1 hr 45 min	60	45	105
67	Kidou Senshi Gundam: Gyslainu no Char	1 hr 59 min	60	59	119
88	Kidou Senshi Gundam F91	2 hr	120	0	120
155	X/1999	1 hr 37 min	60	37	97
164	Mononoke Hime	2 hr 13 min	120	13	133
187	Initial D Third Stage	1 hr 44 min	60	44	104
199	Sen to Chihiro no Kamikakushi	2 hr 4 min	120	4	124
212	Project A-Ko	1 hr 23 min	60	23	83
219	Kidou Senkan Nadesico: The Prince of Darkness	1 hr 19 min	60	19	79
264	Hajime no Ippo: Champion Road	1 hr 30 min	60	30	90
265	Hajime no Ippo: Mashiba vs. Kimura	1 hr 1 min	60	1	61
304	Aoi Megami-sama! Movie	1 hr 46 min	60	46	106
313	Ace wo Nerae! (1979)	1 hr 30 min	60	30	90
315	Xiao Qian	1 hr 24 min	60	24	84
317	Final Fantasy VII: Advent Children	1 hr 41 min	60	41	101
371	Cardcaptor Sakura Movie 1	1 hr 21 min	60	21	81

Posterior se agregará la columna a la tabla y con un Update se actualizarán primero esos valores para las que sean de Horas y algunos Minutos.

```
--Añadir Columna a la Tabla
Alter Table Animes ADD T_Minutos INT;

With T_Minutos as (
Select anime_id, name,duration,cast(TRIM(SUBSTRING(duration,1,1)) as INT)*60 as Horas,
cast(TRIM(SUBSTRING(duration,CHARINDEX('hr',duration)+3,2)) as INT) as Minutos, T_Minutos
from animes WHERE duration NOT IN ('1 hr') and
DURATION LIKE '%hr%')
update t minutos set t_minutos = horas + Minutos
--Select *, horas + Minutos as Tiempo_Minutos
from T_minutos ;
```

(2055 rows affected)

Completion time: 2024-10-20T16:22:33.6808129-06:00

Los registros que únicamente tienen segundos se transforman y actualizan de la siguiente forma utilizando las mismas funciones.

```
-- Separando los Segundos a Valores Numericos quitandoles el PER
Select anime_id, name, duration,
cast(TRIM(SUBSTRING(duration,1,CHARINDEX('sec',duration)-1)) as float)/60 as Segundos
from Animes
where duration like '%sec%';

-- Update Valores
WITH Segundos as (
Select anime_id, name, duration,
cast(TRIM(SUBSTRING(duration,1,CHARINDEX('sec',duration)-1)) as float)/60 as Segundos,
T_Minutos
from Animes
where duration like '%sec%')
UPDATE Segundos set T_Minutos = Segundos;
```

anime_id	name	duration	Segundos
2616	Sora Iro no Tane	30 sec	0.5
4966	Jigazou	12 sec	0.2
5812	Doubutsu Sumo Taikai	52 sec	0.8666666666666667
6064	Higurashi no Naku Koro ni Kai Specials	33 sec	0.55
6234	True Tears: Mini Chara 4-koma Oekijou	30 sec	0.5
6907	Mikan-Seijin	22 sec	0.3666666666666667
7041	Seto no Hanayome OVA Specials	40 sec	0.6666666666666667
7261	Mameshiba	30 sec	0.5
7466	Hashire!	32 sec	0.5333333333333333
7643	Domo-kun	39 sec	0.65
8231	Princess Lover! Magical Knight Maria-chan	30 sec	0.5
8486	Kihei Senki Legacies	52 sec	0.8666666666666667
9561	Shokupan Mimi	30 sec	0.5
9777	Umacha	15 sec	0.25

(605 rows affected)

Completion time: 2024-10-20T16:31:29.8290656-06:00

Por ultimo los Minutos Normales se sacan y actualizan con la misma consulta y mismas funciones.

```
-- VALDIAR MINUTOS
Select anime_id, name, duration,
cast(TRIM(SUBSTRING(duration,1,CHARINDEX('Min',duration)-1)) as INT) as Minutos
from Animes
where duration like '%min%' and T_Minutos is null;

-- Update Valores
```

anime_id	name	duration	Minutos
111	Corrector Yui	25 min	25
112	Chou Henshin Cosplayers	17 min	17
113	Uchou no Stelvia	23 min	23
114	Sakigake! Cromartie Koukou	12 min	12
115	Ei Hazard: The Alternative World	25 min	25
116	Ei Hazard: The Wanderers	23 min	23
117	Ei Hazard: The Magnificent World	34 min	34
118	Ei Hazard 2: The Magnificent World	30 min	30
119	Final Approach	14 min	14
120	Fruits Basket	24 min	24
121	Fullmetal Alchemist	24 min	24



```

-- Update Valores
WITH Minutos as (
Select anime_id, name, duration,
cast(TRIM(SUBSTRING(duration,1,CHARINDEX('Min',
, T_Minutos
from Animes
where duration like '%min%' and T_Minutos is not null) as int) as T_Minutos
UPDATE Minutos set T_Minutos = Minutos;

```

(21498 rows affected)

Completion time: 2024-10-20T16:35:03.3116673-06:00

## Columna RATING

Columna Rating no tiene duplicados y/o valores que se puedan agrupar más, todos se presentan 1 sola vez.

```

-- Validar Duplicados Raiting
Select * from animes;

Select distinct Raiting, count(distinct raiting) cant
from animes group by Raiting
order by 1;

```

99 %

Results Messages

	Raiting	cant
1	G - All Ages	1
2	PG - Children	1
3	PG-13 - Teens 13 or older	1
4	R - 17+ (violence & profanity)	1
5	R+ - Mild Nudity	1
6	Rx - Hentai	1
7	UNKNOWN	1

Columna FECHAS/FECHA\_ESTRENO/FECHA\_FIN.

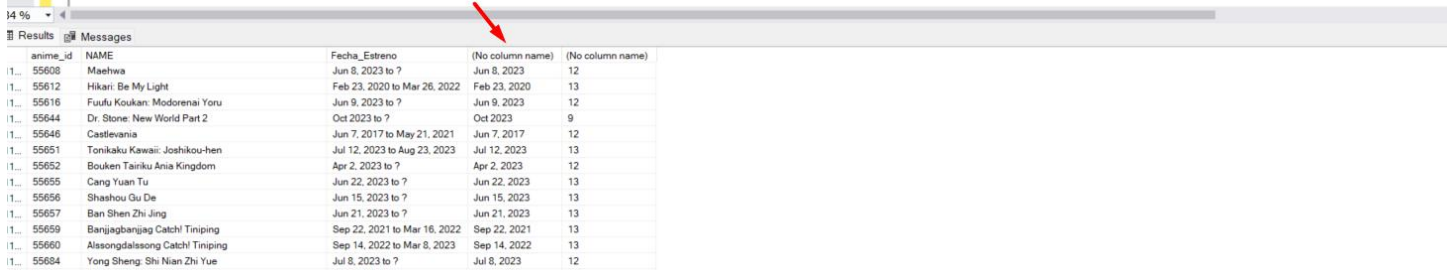
Validación Columna Fechas, esta columna se va a separar en 2 partes, la primera columna será para la fecha de estreno, y esta misma columna se duplico para cuando el registro tenga fecha Final.

Se creo la columna y se pobló.

```
-- Creacion y llenado de la Columna
Alter Table Animes ADD Fecha_Fin NVARCHAR(30)
update animes set Fecha_Fin = Fecha_Estreno;
```

Para dejar únicamente la fecha de estreno se utilizó esta consulta extrayendo todo lo que se encuentre antes del delimitador “to” que indica que seguía otra fecha, con el UPDATE se actualizarán los valores para dejar lista la columna. Todo lo que no tuviera delimitador “to” se queda igual e indica que solamente hay una fecha en esa columna.

```
SELECT anime_id, NAME, Fecha_Estreno,
TRIM(SUBSTRING(Fecha_Estreno,1,CHARINDEX('TO',Fecha_Estreno)-1)) as Fecha_Inicio,
CHARINDEX('TO',Fecha_Estreno)-1
FROM Animes
WHERE Fecha_Estreno LIKE '% to %';
```



anime_id	NAME	Fecha_Estreno	(No column name)	(No column name)
55608	Maehwa	Jun 8, 2023 to ?	Jun 8, 2023	12
55612	Hikari: Be My Light	Feb 23, 2020 to Mar 26, 2022	Feb 23, 2020	13
55616	Fuufu Koukan: Modorenai Yoru	Jun 9, 2023 to ?	Jun 9, 2023	12
55644	Dr. Stone: New World Part 2	Oct 2023 to ?	Oct 2023	9
55646	Castlevania	Jun 7, 2017 to May 21, 2021	Jun 7, 2017	12
55651	Tonikaku Kawaii: Joshikou-hen	Jul 12, 2023 to Aug 23, 2023	Jul 12, 2023	13
55652	Bouken Tairiku Ania Kingdom	Apr 2, 2023 to ?	Apr 2, 2023	12
55655	Cang Yuan Tu	Jun 22, 2023 to ?	Jun 22, 2023	13
55656	Shashou Gu De	Jun 15, 2023 to ?	Jun 15, 2023	13
55657	Ban Shen Zhi Jing	Jun 21, 2023 to ?	Jun 21, 2023	13
55659	Banjagbanjag Catch! Tiniping	Sep 22, 2021 to Mar 16, 2022	Sep 22, 2021	13
55660	Alsongdalssong Catch! Tiniping	Sep 14, 2022 to Mar 8, 2023	Sep 14, 2022	13
55684	Yong Sheng Shi Nian Zhi Yue	Jul 8, 2023 to ?	Jul 8, 2023	12

```
WITH FI AS
(
SELECT anime_id, NAME, Fecha_Estreno,
TRIM(SUBSTRING(Fecha_Estreno,1,CHARINDEX('TO',Fecha_Estreno)-1)) as Fecha_Inicio
FROM Animes
WHERE Fecha_Estreno LIKE '% to %'
)
update FI set Fecha_Estreno = Fecha_Inicio;
```

(11655 rows affected)

Completion time: 2024-10-20T16:57:29.5856585-06:00

Se extraen los valores de la misma manera para la nueva columna de Fecha Fin, trayendo la fecha que se encuentra después del delimitador “to” y posterior se realiza un UPDATE.

```
SELECT anime_id, NAME, fecha_fin,  
TRIM(SUBSTRING(fecha_fin,CHARINDEX('TO', fecha_fin)+2,LEN(fecha_fin))) as Fecha_Inicio,  
CHARINDEX('TO', fecha_fin)+2  
FROM Animes  
WHERE fecha_fin LIKE '% to %';
```

anime_id	NAME	fecha_fin	Fecha_Inicio	(No column name)
18.. 2935	Ryokan Shirasagi	Jun 27, 2003 to Apr 30, 2004	Apr 30, 2004	16
18.. 2936	Private Emotion	Mar 25, 2003 to Jul 25, 2003	Jul 25, 2003	16
18.. 2938	Idol Densetsu Eiko	Apr 3, 1989 to Mar 26, 1990	Mar 26, 1990	15
18.. 2942	Sketchbook: Full Color's	Oct 2, 2007 to Dec 25, 2007	Dec 25, 2007	15
18.. 2943	Inbaku Gakuen	Jun 25, 2002 to May 27, 2003	May 27, 2003	16
18.. 2944	Seisai	Jan 10, 2003 to Apr 11, 2003	Apr 11, 2003	16
18.. 2945	Hitozuma Ryoujoku Sankinbi	Feb 25, 2005 to May 25, 2005	May 25, 2005	16
18.. 2946	Angel Core: Tenshi-tachi no Sumika	Jan 25, 2003 to Apr 25, 2003	Apr 25, 2003	16
18.. 2954	Maria-sama ga Miteru 3rd Specials	Dec 29, 2006 to Jul 25, 2007	Jul 25, 2007	16
18.. 2955	Hana no Joshi Announcer: Newscaster Etsuko	Jun 25, 2004 to Sep 24, 2004	Sep 24, 2004	16
18.. 2956	Shin-ban Megami Tentei Virus File	Feb 27, 2004 to Aug 27, 2004	Aug 27, 2004	16
18.. 2957	Okusama wa Mahoutsukai	Dec 8, 2006 to Mar 9, 2007	Mar 9, 2007	15
18.. 2958	Hitozuma Cosplay Kisea	Mar 25, 2004 to Aug 25, 2004	Aug 25, 2004	16
18.. 2959	Utsukushiki Sei no Dendoushi Reirei	Jul 8, 1994 to Sep 9, 1994	Sep 9, 1994	15
18.. 2960	Kininaru Roommate	Feb 25, 2006 to Sep 25, 2006	Sep 25, 2006	16
18.. 2963	Minami-ke	Oct 8, 2007 to Dec 31, 2007	Dec 31, 2007	15
18.. 2964	Blue Drop: Tenshi-tachi no Gikyoku	Oct 3, 2007 to Dec 26, 2007	Dec 26, 2007	15
18.. 2965	Suteki Tantei Labyrinth	Oct 3, 2007 to Mar 26, 2008	Mar 26, 2008	15
18.. 2966	Okami to Koushinryou	Jan 9, 2008 to Mar 26, 2008	Mar 26, 2008	15
18.. 2967	Saishuu Shiken Kujira	Aug 25, 2007 to Nov 10, 2007	Nov 10, 2007	16
18.. 2970	Night Wizard	Oct 3, 2007 to Dec 26, 2007	Dec 26, 2007	15
18.. 2971	Kyou kara Mao! R	Oct 26, 2007 to Mar 28, 2008	Mar 28, 2008	16
18.. 2977	Yuujuu Sensen Adventure Kid	Jul 21, 1992 to Oct 21, 1993	Oct 21, 1993	16
18.. 2978	Shin Getter Robo	Apr 9, 2004 to Sep 10, 2004	Sep 10, 2004	15
18.. 2979	Aiko	Mar 15, 1996 to Jul 26, 1996	Jul 26, 1996	16
18.. 2980	Kuusou Kagaku Sekai Gulliver Boy	Jan 8, 1995 to Dec 24, 1995	Dec 24, 1995	15

```
WITH FF AS  
(  
SELECT anime_id, NAME, fecha_fin,  
TRIM(SUBSTRING(fecha_fin,CHARINDEX('TO', fecha_fin)+2,LEN(fecha_fin))) as Fecha_Final  
FROM Animes  
WHERE fecha_fin LIKE '% to %'  
)  
update Ff set fecha_Fin = Fecha_Final;
```

(11655 rows affected)

Completion time: 2024-10-20T17:02:14.1977573-06:00

## Limpieza Columna FECHA\_EXTRENO / FECHA\_FINAL

Para los Que tienen fecha final con el carácter '?' que significa que son animes aun en emisión, se pondrá el Valor NULL.

Revisando la columna Fecha\_Estreno hay valores fechas que tienen diferente formato (YYYY, MMM-YY), o el valor "Not Aviable"

Cambios:

- Formato YYYY: Se pondrá el 1 de Enero de su respectivo Año
- Formato MMM-YY y MMM YYYY: Se pondrá el día 1 de su respectivo mes y año.
- Valor "Not Aviable": Pasaran a NULL.

```

-- Select ANIME_ID, name, Fecha_Estreno, Fecha_Fin from animes
where Fecha_Estreno not like '%,%';

```

	ANIME_ID	name	Fecha_Estreno	Fecha_Fin
4721	55134	Detective Conan: TDK Video Head Cleaner	1996	1996
4722	55143	Hidari	Not available	Not available
4723	55149	Komada Jouryuusho e Youkoso	nov-23	nov-23
4724	55150	Yarinaoshi Reijou wa Ryutei Heika wo Kouryakuch...	Not available	Not available
4725	55151	Da Zhu Zai Nian Fan	2023	NULL
4726	55152	Jodie	2021	2021
4727	55153	Yuzuki-san Chi no Yonkyoudai.	Oct 2023	NULL
4728	55166	Yami Shibai 11	Jul 2023	NULL
4729	55193	Midnight: Eraser Wars Spin-off	2017	2017
4730	55194	Ooi Tonbo	2024	NULL
4731	55237	Jashin-chan Dropkick: Seikimatsu-hen	Not available	Not available
4732	55243	A Girl Meets A Boy And A Robot	Not available	Not available
4733	55253	Wuxia	2017	2017
4734	55260	Chunfen	2022	2022
4735	55261	Daoying	2022	2022
4736	55263	Yume Mitsuru yoru	Not available	Not available
4737	55265	Tensei Kizoku, Kantei Skill de Nariagaru	2024	NULL
4738	55266	Wang Shen Zhi Ye	2022	2022

Corregir Fechas sin información.

```

-- UPDATE ANIMES SET FECHA_FIN = NULL, Fecha_Estreno=null WHERE
(Fecha_Fin = 'Not available' or Fecha_Estreno = 'Not available')

```

(914 rows affected)

Consulta para trabajar las que de formato YYYY

Utilice 2 columnas de Largo para poder identificar las que sean con ese formato, ya que hasta este punto las columnas son datos NVARCHAR. Solamente se concatena el Año a la cadena 'Jan 1. Y posterior se actualiza.

```

-- Se Validan las fechas que tienen unicamente el Año
with Largo as (
Select ANIME_ID, name, Fecha_Estreno, Fecha_Fin, len(Fecha_Estreno) as Largo_FI,
len(Fecha_Fin) as Largo_Ff,
CONCAT('Jan 1,', Fecha_Estreno) as Fecha_Correcta
from animes
where Fecha_Estreno not like '%,%')
Select * from largo where Largo_FI = 4 and Largo_Ff = Largo_FI

Select * from animes;

```

	ANIME_ID	name	Fecha_Estreno	Fecha_Fin	Largo_FI	Largo_Ff	Fecha_Correcta
34	55261	Daoying	2022	2022	4	4	Jan 1, 2022
35	55266	Wang Shen Zhi Ye	2022	2022	4	4	Jan 1, 2022
36	55353	Cemedine Bond to Q17-gou 75	2017	2017	4	4	Jan 1, 2017
37	55357	Bocchi the Rock! Movie	2024	2024	4	4	Jan 1, 2024
38	55373	Cemedine Bond to Q17-gou	1979	1979	4	4	Jan 1, 1979
39	55396	Lushang	2022	2022	4	4	Jan 1, 2022
40	55462	Lusi He	2022	2022	4	4	Jan 1, 2022
41	55464	Nee, Koko ni Iru	2016	2016	4	4	Jan 1, 2016
42	55532	Anata ga Miemashita	2017	2017	4	4	Jan 1, 2017
43	55533	Uwanosora kara Kaettekuru	2010	2010	4	4	Jan 1, 2010
44	55535	Coyote no Kubikazari	1993	1993	4	4	Jan 1, 1993
45	55540	The Anus	2004	2004	4	4	Jan 1, 2004
46	55550	On Time Off Time	2020	2020	4	4	Jan 1, 2020
47	55552	Dedari	2007	2007	4	4	Jan 1, 2007

Se actualizarán las fechas Fin y Estreno con el valor de la fecha correcta.

```
with Largo as (
Select ANIME_ID,name, Fecha_Estreno, Fecha_Fin, len(fecha_estreno) as Largo_FI,
len(Fecha_Fin) as Largo_Ff,
CONCAT('Jan 1,',Fecha_Estreno) as Fecha_Correcta
from animes
where Fecha_Estreno not like '%,%')
UPDATE Largo SET Fecha_Estreno = Fecha_Correcta, Fecha_Fin = Fecha_Correcta
where Largo_FI = 4 and Largo_Ff = Largo_FI
```

(2317 rows affected)

Completion time: 2024-10-20T17:24:56.0660323-06:00

Las que tienen el formato MMM-YY utilice un CASE que valida los 2 últimos dígitos de la columna, y compara esos valores como enteros.

Si es un número Menor que 25 que significa que su año debe ser del 2000 al 2025, se le concatenara el Mes y la Cadena 20, para formar el Año completo.

De lo contrario se le concatena un 19, indicado que puede ser de 1926 hasta 1999, y su respectivo mes para armar la columna.

```
with Largo as (
Select ANIME_ID,name, Fecha_Estreno, Fecha_Fin, len(fecha_estreno) as Largo_FI,
len(Fecha_Fin) as Largo_Ff,
CASE WHEN
CAST(RIGHT(Fecha_Estreno,2) AS INT) < 25
THEN CONCAT(LEFT(Fecha_Estreno, 3),' 1, 20', RIGHT(Fecha_Estreno,2))
ELSE
CONCAT(LEFT(Fecha_Estreno, 3),' 1, 19', RIGHT(Fecha_Estreno,2))
END AS FECHA_CORRECTA
--CONCAT('Jan 1,',1) as Fecha_Correcta
from animes
where Fecha_Estreno not like '%,%')
Select * from largo where Largo_FI = 6
```

ANIME_ID	name	Fecha_Estreno	Fecha_Fin	Largo_FI	Largo_Ff	FECHA_CORRECTA	
76	52139	Hungry Days x Bump of Chicken: Kinen Satsuei	nov-19	nov-19	6	6	nov 1, 2019
77	52161	La Golondrina	feb-72	feb-72	6	6	feb 1, 1972
78	52325	Yume wo Kanaeyo	feb-06	feb-06	6	6	feb 1, 2006
79	52327	Fufu!	feb-84	feb-84	6	6	feb 1, 1984
80	52328	Toshokan Rocket	oct-13	oct-13	6	6	oct 1, 2013
81	52332	Aoi Douwa	oct-01	oct-01	6	6	oct 1, 2001
82	52395	Hengao Taisou	jun-19	jun-19	6	6	jun 1, 2019
83	52397	Noman	jul-68	jul-68	6	6	jul 1, 1968
84	52453	Koumori Bat wa Good na Shinshi	oct-19	oct-19	6	6	oct 1, 2019
85	52605	Shounen Sunday CM: Konjiki no Gash!!	oct-01	oct-01	6	6	oct 1, 2001
86	52620	Shounen Sunday CM: Megumi no Daigo	nov-96	nov-96	6	6	nov 1, 1996
87	52621	Shounen Sunday CM: Yakitate!! Japan	mar-02	mar-02	6	6	mar 1, 2002
88	52622	Shounen Sunday CM: Fantasia	jul-01	jul-01	6	6	jul 1, 2001
89	52642	Shounen Sunday CM: Shijou Saikyou no Deshi Keni...	feb-04	feb-04	6	6	feb 1, 2004



```
with Largo as (
Select ANIME_ID,name, Fecha_Estreno, Fecha_Fin, len(fecha_estreno) as Largo_FI,
len(Fecha_Fin) as Largo_Ff,
CASE WHEN
CAST(RIGHT(Fecha_Estreno,2) AS INT) < 25
THEN CONCAT(LEFT(FECHA_ESTRENO, 3),' 1, 20', RIGHT(Fecha_Estreno,2))
ELSE
CONCAT(LEFT(FECHA_ESTRENO, 3),' 1, 19', RIGHT(Fecha_Estreno,2))
END AS FECHA_CORRECTA
--CONCAT('Jan 1,',1) as Fecha_Correcta
from animes
where Fecha_Estreno not like '%,%')
UPDATE LARGO set Fecha_Estreno = FECHA_CORRECTA, Fecha_Fin = FECHA_CORRECTA
from largo where Largo_FI = 6

(203 rows affected)

Completion time: 2024-10-20T17:39:39.3550689-06:00
```

Formato MMM YYYY, únicamente se le concatena el día 1 al mes y año.

```
-- Siguiete Formato de Fecha.
with Largo as (
Select ANIME_ID,name, Fecha_Estreno, Fecha_Fin, len(fecha_estreno) as Largo_FI,
len(Fecha_Fin) as Largo_Ff,
CONCAT(LEFT(fecha_estreno,3),' 1,',RIGHT(fecha_estreno,4)) as Fecha_Correcta
from animes
where Fecha_Estreno not like '%,%')
Select * from largo where Largo_FI = 8 and Largo_Ff is null
```

ANIME_ID	name	Fecha_Estreno	Fecha_Fin	Largo_FI	Largo_Ff	Fecha_Correcta	
167	54595	Kage no Jitsuryokusha ni Nantakute! 2n...	Oct 2023	NULL	8	NULL	Oct 1,2023
168	54688	Naruto (Shinsaku Anime)	Sep 2023	NULL	8	NULL	Sep 1,2023
169	54716	Kibou no Chikara: Otona Precure '23	Oct 2023	NULL	8	NULL	Oct 1,2023
170	54727	Rhapsody	Jul 2023	NULL	8	NULL	Jul 1,2023
171	54743	Dead Mount Death Play Part 2	Oct 2023	NULL	8	NULL	Oct 1,2023
172	54758	The IDOLM@STER Shiny Colors	Apr 2024	NULL	8	NULL	Apr 1,2024
173	54794	Metallic Rouge	Jan 2024	NULL	8	NULL	Jan 1,2024
174	54798	Kamierabi	Oct 2023	NULL	8	NULL	Oct 1,2023
175	54803	Captain Tsubasa: Junior Youth-hen	Oct 2023	NULL	8	NULL	Oct 1,2023
176	54835	Kono Sekai wa Fukanzan Suguru	Apr 2024	NULL	8	NULL	Apr 1,2024
177	54858	Hypnosis Mic: Division Rap Battle - Rhy...	Oct 2023	NULL	8	NULL	Oct 1,2023

```
with Largo as (
Select ANIME_ID,name, Fecha_Estreno, Fecha_Fin, len(fecha_estreno) as Largo_FI,
len(Fecha_Fin) as Largo_Ff,
CONCAT(LEFT(fecha_estreno,3),' 1,',RIGHT(fecha_estreno,4)) as Fecha_Correcta
from animes
where Fecha_Estreno not like '%,%')
UPDATE largo SET Fecha_Estreno= Fecha_Correcta, Fecha_Fin = Fecha_Correcta
WHERE Largo_FI = 8 and Largo_Ff is null

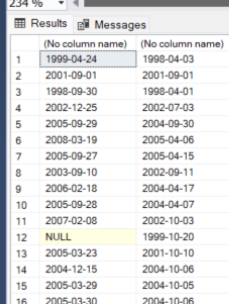
(190 rows affected)

Completion time: 2024-10-20T17:45:02.9645170-06:00
```

Se aplicaron los mismos pasos para limpiar la nueva columna de Fecha Fin, ya que presentaba los mismos escenarios.

Ya limpias permite cambiar su formato a DATE.

```
Select cast(fecha_fin as date), cast(fecha_estreno as date) from animes
```



	(No column name)	(No column name)
1	1999-04-24	1998-04-03
2	2001-09-01	2001-09-01
3	1998-09-30	1998-04-01
4	2002-12-25	2002-07-03
5	2005-09-29	2004-09-30
6	2008-03-19	2005-04-06
7	2005-09-27	2005-04-15
8	2003-09-10	2002-09-11
9	2006-02-18	2004-04-17
10	2005-09-28	2004-04-07
11	2007-02-08	2002-10-03
12	NULL	1999-10-20
13	2005-03-23	2001-10-10
14	2004-12-15	2004-10-06
15	2005-03-29	2004-10-05
16	2005-03-30	2004-10-06

Se cambia el tipo de dato a DATE.

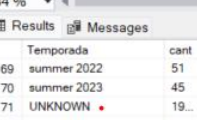
```
alter table animes alter column fecha_fin date;  
alter table animes alter column fecha_estreno date;
```

## Columna TEMPORADA

No presenta valores duplicados, o que se puedan agrupar, sin embargo, algunas filas se encuentran con valor UNKNOWN o de acuerdo con su fecha de estreno no corresponde con la estación que deberían.

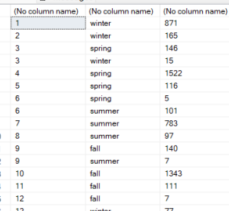
Se actualizará desde 0 la columna.

```
With duplicados as (  
Select distinct Temporada, count(*) cant  
from animes group by Temporada)  
select * from duplicados where cant >1  
order by 1;
```



	Temporada	cant
169	summer 2022	51
170	summer 2023	45
171	UNKNOWN	19...

```
SELECT MONTH(FECHA_ESTRENO), SUBSTRING(TEMPORADA,1,CHARINDEX(' ', TEMPORADA)), COUNT(*)  
FROM ANIMES  
WHERE TEMPORADA != 'UNKNOWN'  
GROUP BY MONTH(FECHA_ESTRENO), SUBSTRING(TEMPORADA,1,CHARINDEX(' ', TEMPORADA))  
ORDER BY 1;
```



	(No column name)	(No column name)	(No column name)
1	1	winter	871
2	2	winter	165
3	3	spring	146
4	3	winter	15
5	4	spring	1522
6	5	spring	116
7	6	spring	5
8	6	summer	101
9	7	summer	783
10	8	summer	97
11	9	fall	140
12	9	summer	7
13	10	fall	1343
14	11	fall	111
15	12	fall	7
16	12	winter	77

Se corrigen los Valores y la tabla queda Ok.

```
UPDATE ANIMES SET Temporada = (CASE
  WHEN MONTH(FECHA_ESTRENO) IN (12,1,2) THEN CONCAT('Winter ', YEAR(fecha_estreno))
  WHEN MONTH(FECHA_ESTRENO) IN (3,4,5) THEN CONCAT('Spring ', YEAR(fecha_estreno))
  WHEN MONTH(FECHA_ESTRENO) IN (6,7,8) THEN CONCAT('Summer ', YEAR(fecha_estreno))
  WHEN MONTH(FECHA_ESTRENO) IN (9,10,11) THEN CONCAT('Fall ', YEAR(fecha_estreno))
end)

(24904 rows affected)

Completion time: 2024-10-20T19:21:52.2925676-06:00
```

Y la columna quedo de la siguiente manera, los valores NULL quedaron porque no se tiene una fecha de Inicio para esos registros.

```
SELECT MONTH(FECHA_ESTRENO), SUBSTRING(TEMPORADA,1,CHARINDEX(' ', TEMPORADA)), COUNT(*)
FROM ANIMES
--WHERE TEMPORADA != 'UNKNOWN'
GROUP BY MONTH(FECHA_ESTRENO), SUBSTRING(TEMPORADA,1,CHARINDEX(' ', TEMPORADA))
ORDER BY 1;
```

Results Messages

	(No column name)	(No column name)	(No column name)
1	NULL	NULL	914
2	1	Winter	5201
3	2	Winter	1245
4	3	Spring	1744
5	4	Spring	2866
6	5	Spring	1035
7	6	Summer	1267
8	7	Summer	2445
9	8	Summer	1439
10	9	Fall	1259
11	10	Fall	2698
12	11	Fall	1112
13	12	Winter	1679