

TSV ALTREYX WORKFLOW

SCREENSHOTS AND ANALYSIS

1. Avatar

The screenshot shows the Alteryx Designer x64 interface with a workflow titled 'TSV_File_Profiling_1.ymd'. The workflow consists of a 'Browse' tool connected to a 'Join' tool, which is then connected to a 'Parse' tool, and finally a 'Transform' tool. The 'Results - Browse (22) - Input' table shows 11 records for the movie 'Avatar'.

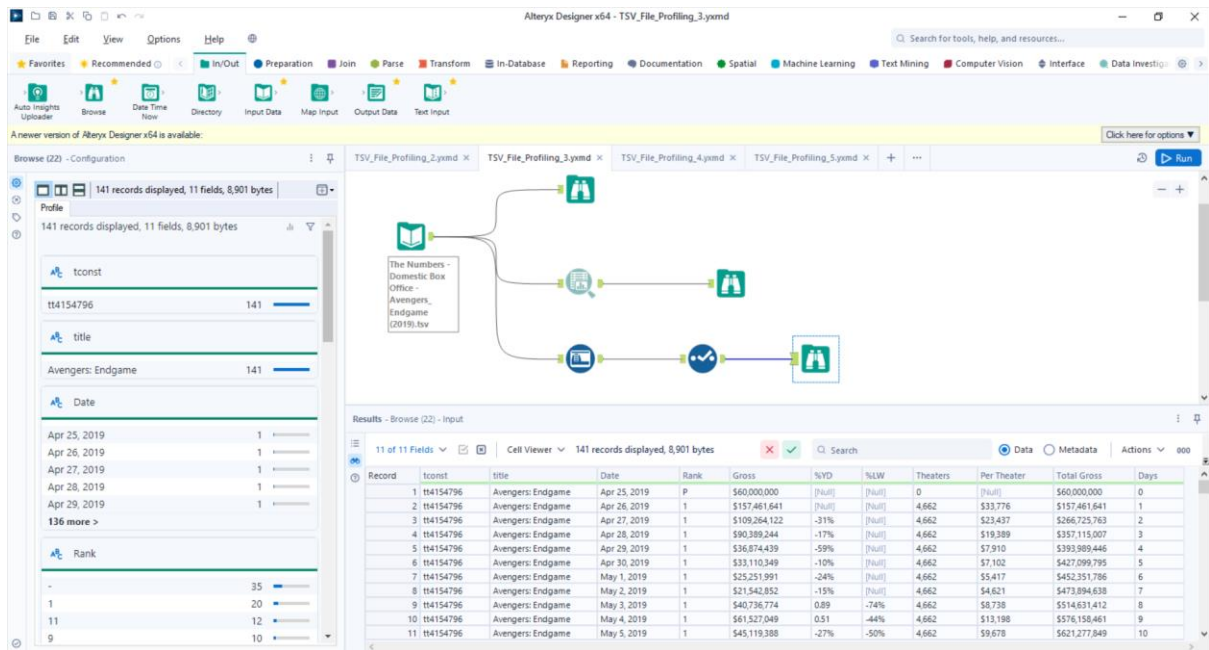
Record	tconst	title	Date	Rank	Gross	%YD	%LW	Theaters	Per Theater	Total Gross	Days
1	tt0499549	Avatar	Dec 18, 2009	1	\$36,752,099	[Null]	[Null]	3,452	\$7,750	\$36,752,099	1
2	tt0499549	Avatar	Dec 19, 2009	1	\$25,529,036	-5%	[Null]	3,452	\$7,395	\$52,281,135	2
3	tt0499549	Avatar	Dec 20, 2009	1	\$24,744,346	-3%	[Null]	3,452	\$7,168	\$77,025,481	3
4	tt0499549	Avatar	Dec 21, 2009	1	\$16,385,820	-34%	[Null]	3,452	\$4,747	\$93,411,301	4
5	tt0499549	Avatar	Dec 22, 2009	1	\$16,086,461	-2%	[Null]	3,452	\$4,660	\$109,497,762	5
6	tt0499549	Avatar	Dec 23, 2009	2	\$16,445,291	0.02	[Null]	3,452	\$4,764	\$125,943,053	6
7	tt0499549	Avatar	Dec 24, 2009	1	\$11,150,998	-32%	[Null]	3,452	\$3,230	\$137,094,051	7
8	tt0499549	Avatar	Dec 25, 2009	2	\$23,095,046	1.07	-14%	3,456	\$6,683	\$160,189,097	8
9	tt0499549	Avatar	Dec 26, 2009	1	\$28,274,406	0.22	0.11	3,456	\$8,181	\$188,463,503	9
10	tt0499549	Avatar	Dec 27, 2009	1	\$24,247,681	-14%	-2%	3,456	\$7,016	\$212,711,184	10
11	tt0499549	Avatar	Dec 28, 2009	1	\$19,418,139	-20%	0.19	3,456	\$5,619	\$232,129,323	11

2. Avengers_Age of Ultron (2015)

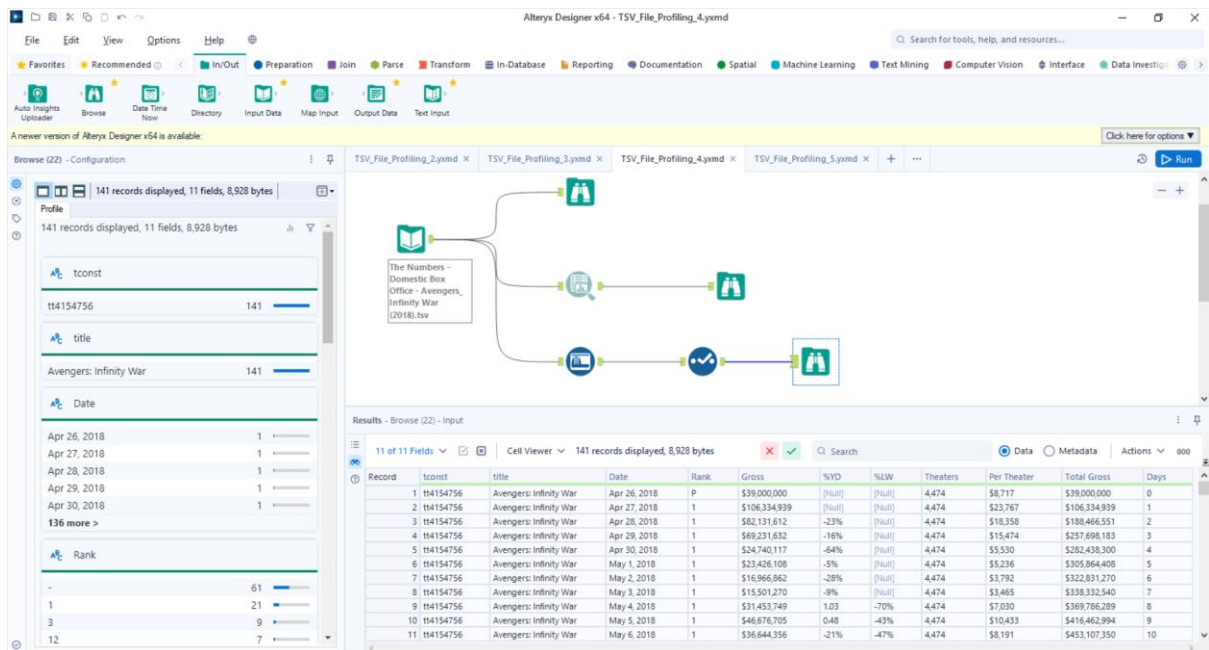
The screenshot shows the Alteryx Designer x64 interface with a workflow titled 'TSV_File_Profiling_2.ymd'. The workflow consists of a 'Browse' tool connected to a 'Join' tool, which is then connected to a 'Parse' tool, and finally a 'Transform' tool. The 'Results - Browse (22) - Input' table shows 11 records for the movie 'Avengers: Age of Ultron'.

Record	tconst	title	Date	Rank	Gross	%YD	%LW	Theaters	Per Theater	Total Gross	Days
1	tt2395427	Avengers: Age of Ultron	Apr 30, 2015	P	\$27,600,000	[Null]	[Null]	0	[Null]	\$27,600,000	[Null]
2	tt2395427	Avengers: Age of Ultron	May 1, 2015	1	\$84,424,532	[Null]	[Null]	4,276	\$19,744	\$84,424,532	1
3	tt2395427	Avengers: Age of Ultron	May 2, 2015	1	\$56,526,635	-33%	[Null]	4,276	\$13,220	\$140,951,167	2
4	tt2395427	Avengers: Age of Ultron	May 3, 2015	1	\$50,319,942	-11%	[Null]	4,276	\$11,768	\$191,271,109	3
5	tt2395427	Avengers: Age of Ultron	May 4, 2015	1	\$13,228,555	-74%	[Null]	4,276	\$3,094	\$204,499,664	4
6	tt2395427	Avengers: Age of Ultron	May 5, 2015	1	\$13,152,812	-1%	[Null]	4,276	\$3,071	\$217,652,476	5
7	tt2395427	Avengers: Age of Ultron	May 6, 2015	1	\$9,409,838	-28%	[Null]	4,276	\$2,201	\$227,062,314	6
8	tt2395427	Avengers: Age of Ultron	May 7, 2015	1	\$8,613,154	-8%	[Null]	4,276	\$2,014	\$235,655,468	7
9	tt2395427	Avengers: Age of Ultron	May 8, 2015	1	\$21,232,362	1.47	-75%	4,276	\$4,965	\$256,887,830	8
10	tt2395427	Avengers: Age of Ultron	May 9, 2015	1	\$33,835,656	0.59	-40%	4,276	\$7,913	\$290,723,486	9
11	tt2395427	Avengers: Age of Ultron	May 10, 2015	1	\$22,678,911	-33%	-55%	4,276	\$5,304	\$313,402,397	10

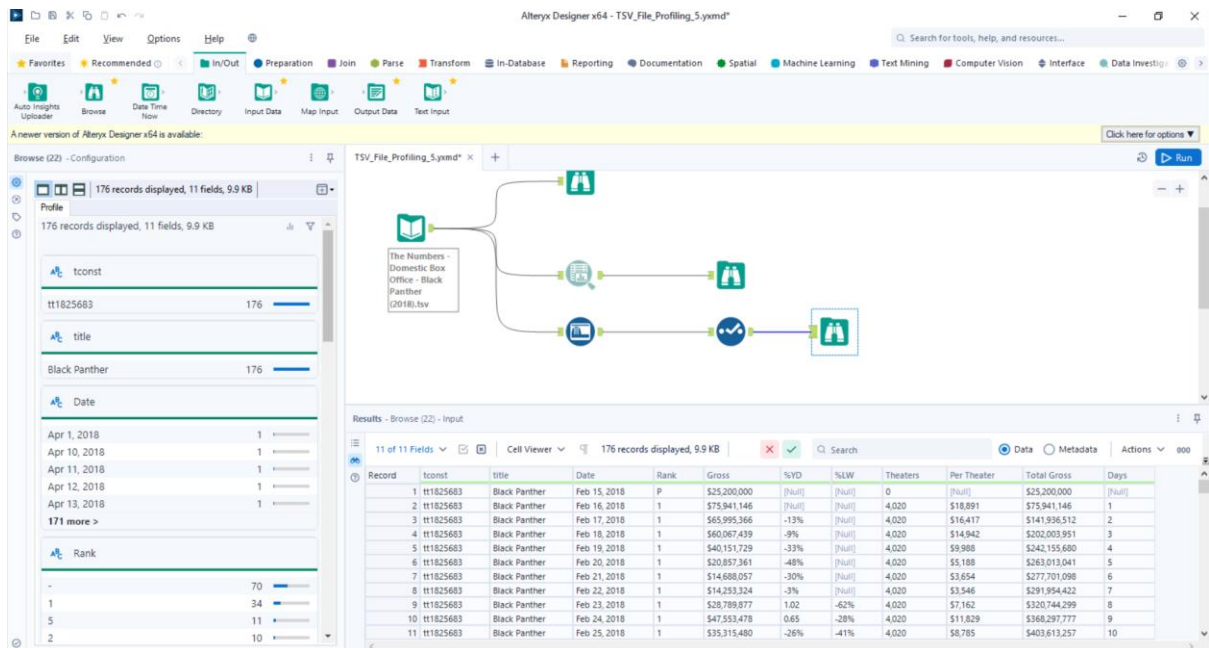
3. Avengers_Endgame (2019)



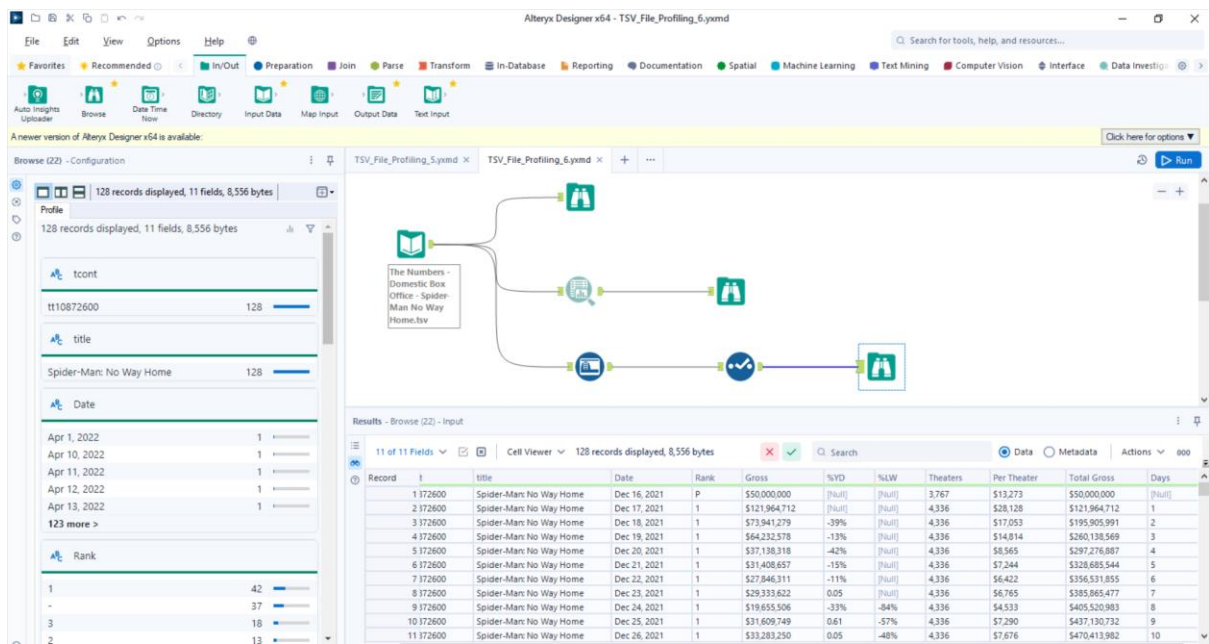
4. Avengers_Infinity War (2018)



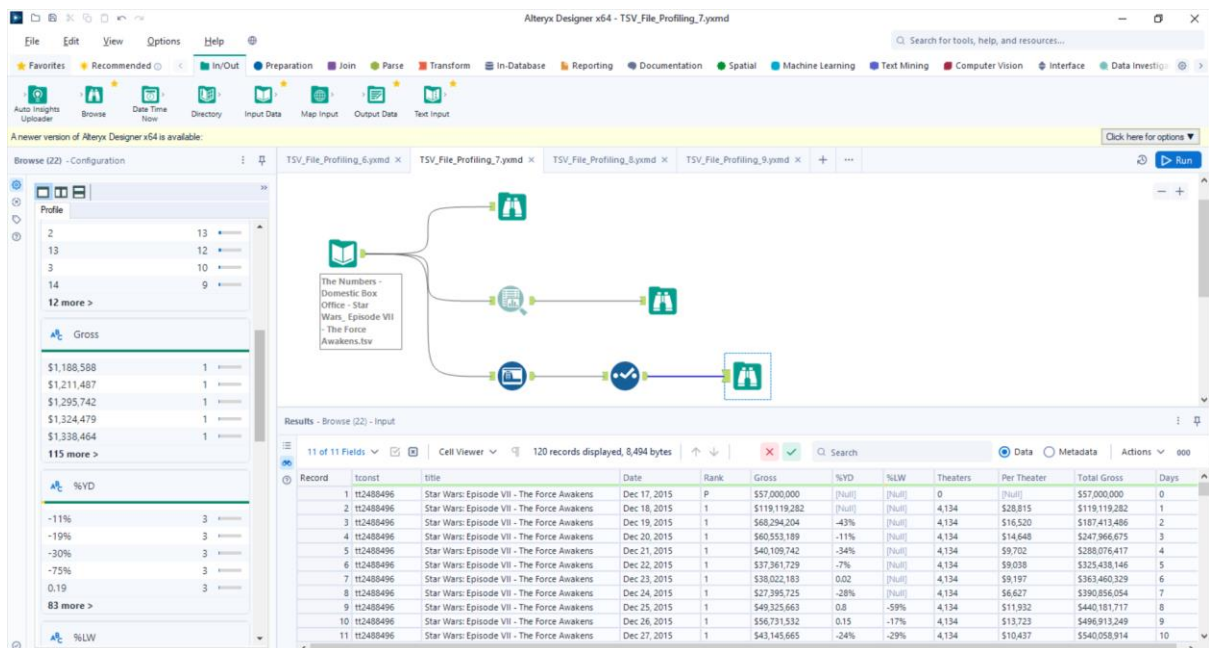
5. Black Panther (2018)



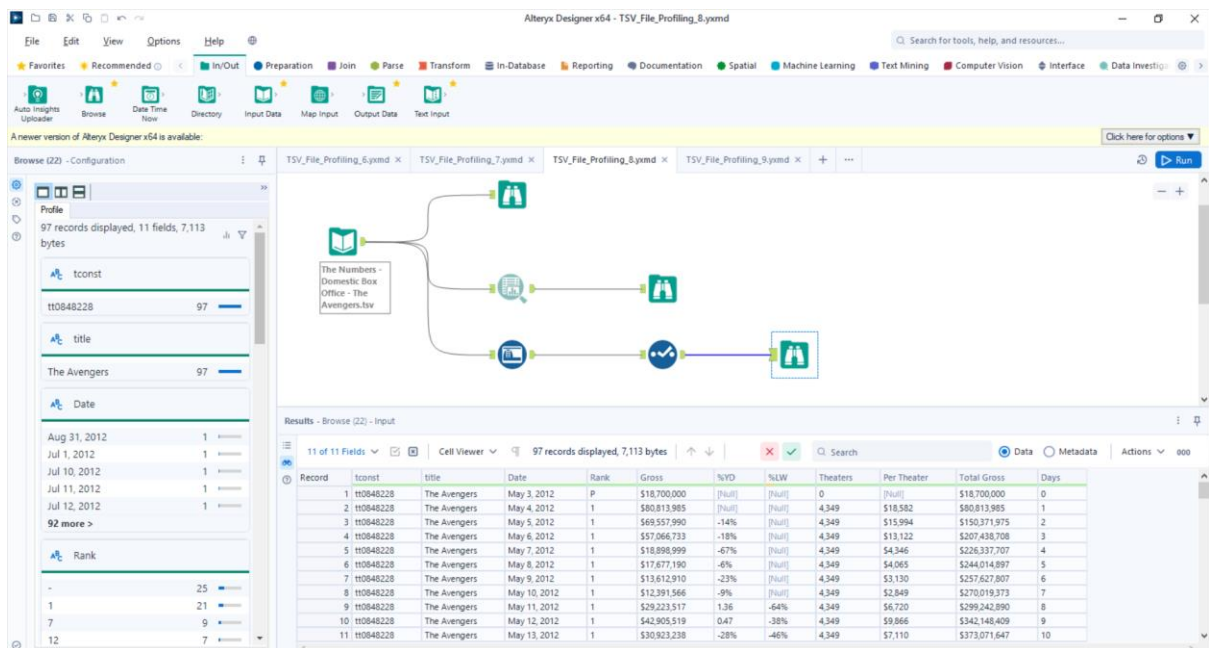
6. Spider-Man No Way Home



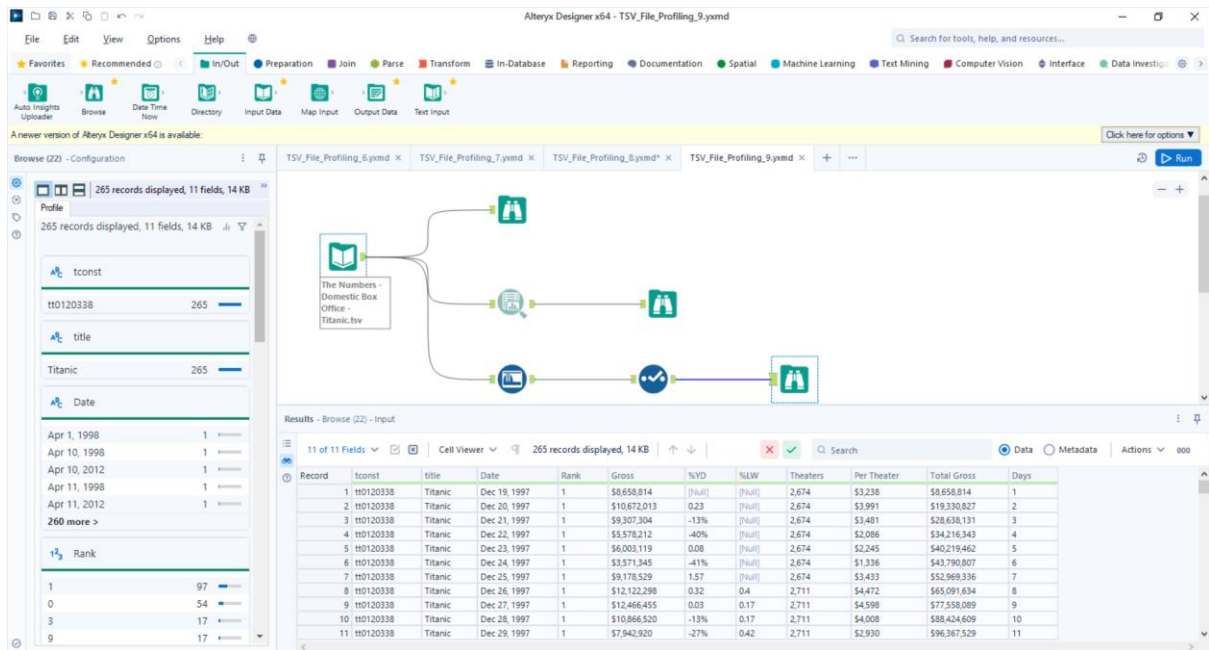
7. Star Wars_Episode VII - The Force Awakens



8. Domestic Box Office - The Avengers



9. Domestic Box Office – Titanic



For all of the above TSV files here are the inferences:

tconst, title, Date, Days- Everything is in proper format

Rank- Few file's contain 'P' and '-' at some places, replace with values like 0 or -1 or 99

Gross- Remove \$ and comma

%YD- Remove %(percentage), -(minus) if we need this column to calculate or leave it as string , some files have 'n/c' in the field

%LW- Remove %(percentage), -(minus) if we need this column to calculate or leave it as string

Theaters- Remove comma

Per Theater- Remove \$ and comma

Total Gross- Remove \$ and comma for calculation

JSON ALTREYX WORKFLOW

SCREENSHOTS AND ANALYSIS

- Used Blob Input
- Convert to text using Blob Converter
- Parse out the JSON
- Transfer individual column we get into separate columns/split using Text to Column
- Filter data by taking required field and using Select
- Using CrossTab for transpose

1. newdata_name basics

The screenshot shows the Alteryx Designer interface with a workflow titled "JSON_File_Profiling_1.ymd". The workflow includes a "Select" tool with a configuration table on the left. The configuration table lists fields and their types, with "deathYear" set to "V_String" and "Unknown" set to "Unknown". The results table at the bottom shows 7 records displayed, with columns including "Record", "Records", "birthYear", "deathYear", "Dx_CreateOutTime", "Dx_jobId", "Dx_jobName", "knownForTitles", and "ncount".

Field	Type	Size	Format	Description
Records	V_String	107		Test To Column
birthYear	V_String	214		
deathYear	V_String	214		
Dx_CreateOutTime	V_String	214		
Dx_jobId	V_String	214		
Dx_jobName	V_String	214		
knownForTitles	V_String	214		
ncount	V_String	214		
primaryName	V_String	214		
primaryProfession	V_String	214		
Unknown	Unknown	0		Dynamic or Unk.

Record	Records	birthYear	deathYear	Dx_CreateOutTime	Dx_jobId	Dx_jobName	knownForTitles	ncount
1	2	1967	UN	[null]	[null]	[null]	[null]	nm0740263
2	4	1984	UN	2023-11-19 15:20:37	srMSLk	stg_name_basics	nm0424060	nm0424060
3	0	1965	UN	[null]	[null]	[null]	nm0000375	nm0000375
4	6	1996	UN	[null]	[null]	[null]	nm0436118	nm0436118
5	1	1981	UN	[null]	[null]	[null]	nm0262635	nm0262635
6	5	1976	UN	[null]	[null]	[null]	nm1212722	nm1212722
7	3	1983	UN	[null]	[null]	[null]	nm1165110	nm1165110

- The deathYear column has NULL (N) that is not adding value and is also not necessary for any visualization, so we can ignore the column or replace with 0

2. newdata_title basics

The screenshot shows the Alteryx Designer interface with a workflow titled "JSON_File_Profiling_2.ymd". The workflow includes a "Select" tool with a configuration table on the left. The configuration table lists fields and their types, with "endYear" set to "V_String" and "Unknown" set to "Unknown". The results table at the bottom shows 10 records displayed, with columns including "Record", "Records", "endYear", "genre", "adult", "originalTitle", "primaryTitle", "numberOfEpisodes", "startYear", "ncount", and "titleType".

Field	Type	Size	Format	Description
Records	V_String	107		Test To Column
endYear	V_String	214		
genre	V_String	214		
adult	V_String	214		
originalTitle	V_String	214		
primaryTitle	V_String	214		
numberOfEpisodes	V_String	214		
startYear	V_String	214		
ncount	V_String	214		
titleType	V_String	214		
Unknown	Unknown	0		Dynamic or Unk.

Record	Records	endYear	genre	adult	originalTitle	primaryTitle	numberOfEpisodes	startYear	ncount	titleType
1	2	UN	Action-Adventure-Sci-Fi	0	The Avengers	The Avengers (2012)	143	2012	nm0402228	movie
2	4	UN	Action-Adventure-Sci-Fi	0	Black Panther	Black Panther (2018)	134	2018	nm1025683	movie
3	0	UN	Drama-Romance	0	Titanic	Titanic (1997)	194	1997	nm0120338	movie
4	6	UN	Action-Adventure-Sci-Fi	0	Star Wars: Episode VII - The Force Awakens	Star Wars: Episode VII - The Force Awakens (2015)	138	2015	nm4854986	movie
5	1	UN	Action-Adventure-Fantasy	0	Avatar	Avatar (2009)	162	2009	nm0499549	movie
6	7	UN	Action-Adventure-Sci-Fi	0	Avengers: Infinity War	Avengers: Infinity War (2018)	149	2018	nm1143736	movie
7	9	UN	Drama-Romance	0	Titanic	Titanic (1997) remake	194	1997	nm0120338	movie
8	8	UN	Action-Adventure-Drama	0	Avengers: Endgame	Avengers: Endgame (2019)	181	2019	nm1143736	movie
9	5	UN	Action-Adventure-Sci-Fi	0	Avengers: Age of Ultron	Avengers: Age of Ultron (2015)	141	2015	nm2395427	movie
10	3	UN	Action-Adventure-Fantasy	0	Spider-Man: No Way Home	Spider-Man: No Way Home (2021)	148	2021	nm10872800	movie

- The endYear column has NULL (N) that is not adding value and is also not necessary for any visualization, so we can ignore the column or replace with 0

SQL ALTREYX WORKFLOW

SCREENSHOTS AND ANALYSIS

1. imdb name basics

Profile

Count	Min_birthYear
1	3,787
2	1,598
3	954
4	608
5	491
567 more >	
Min_birthYear	
\N	6,383
1971	86
1975	85
1970	84
1969	79
201 more >	
primaryProfession	
actor	1
actor,author	1

Results - Browse (5) - Input

Record	Count	Min_birthYear	primaryProfession
1	28	1991	make_up_department,actress,writer
2	2	1961	cinematographer,producer,animation_department
3	4	1981	producer,editorial_department,music_department
4	12	1963	art_department,camera_department,art_director
5	1	\N	make_up_department,art_director,script_department
6	1	\N	visual_effects,miscellaneous,executive
7	1	\N	art_director,camera_department,sound_department
8	10	1890	miscellaneous,cinematographer,editorial_department
9	10	1986	miscellaneous,art_department,editor
10	8	1983	visual_effects,editor,production_manager
11	2	\N	actress,transportation_department,producer

- **nconst, primaryName, birthYear, primaryProfession, knownForTitles, DI_Create_DT** - Everything is in proper format

2. imdb title akas

Profile

CountDistinct_region	CountDistinct_language
1	1,149,522
2	372,533
3	100,253
4	30,507
5	13,246
40 more >	
CountDistinct_language	
1	1,571,956
2	102,128
3	12,812
4	2,507
5	572
9 more >	
titleId	
tt0076759	92
tt0104431	92
tt0077749	74

Results - Browse (5) - Input

Record	CountDistinct_region	CountDistinct_language	titleId	title
1	2	1	tt597586	'88
2	1	1	tt0194484	Pelindra Milionário
3	2	1	tt13662380	El final de la inici
4	1	1	tt0115683	Mallucos em Apuros
5	1	1	tt480956	嵐さま、ごきげんよう
6	3	1	tt11311824	Hollywood Desperado: America's Latin Prince
7	1	1	tt0098360	This Man He's Wild
8	1	1	tt0165935	Eight Brave Brothers
9	1	1	tt4263482	Vrijitoarea
10	1	1	tt0146429	Mannequin
11	2	1	tt0237358	La juez Lobo

- **titleId, ordering, title, DI_Create_DT** – Everything is in proper format
- **region, language, types, attributes, isOriginalTitle** – Replace \N with 0/-1/99

3. imdb_title_basics

Profile: 607,423 records displayed, 5 fields, 22 MB

Results - Browse (5) - Input

Record	tconst	primaryTitle	originalTitle	Min_runtimeMinutes	genres
1	tt1498864	Queen & Country	Queen & Country	\N	Action,Thriller
2	tt10549128	Used Mercs	Used Mercs	\N	Action
3	tt1887954	Yigian Chou	Yigian Chou	\N	\N
4	tt5506740	Demon Warrior	Fu mo zhen shi	69	Drama
5	tt3404219	Playing Dirty	Playing Dirty	\N	Thriller
6	tt0309090	Confessions of a Woman	Yeojaga gobaekehal dae	90	Drama
7	tt5377526	Dnevnik Karlosa Espinoly	Dnevnik Karlosa Espinoly	74	Family
8	tt3417570	Seppun Jijiro	Seppun Jijiro	\N	Drama
9	tt0214462	Alik Babu	Alik Babu	\N	Drama
10	tt0470685	The Bathroom	Banyo	87	Comedy,Crime
11	tt3582032	We Get Around	We Get Around	\N	Drama

- **tconst, titleType, primaryTitle, originalTitle, isAdult, DI_Create_DT**- Everything is in proper format
- **startYear, endYear, runtimeMinutes, genres**- Replace \N values with 0 or -1

4. imdb_title_crew

Profile: 26 records displayed, 3 fields, 753 bytes

Results - Browse (5) - Input

Record	CountDistinct_directors	CountDistinct_writers	DI_Create_DT
1	3599	15791	2022-04-20 07:20:37
2	17374	15699	2022-04-20 07:20:50
3	15626	14223	2022-04-20 07:20:51
4	2035	7569	2022-04-20 07:20:36
5	21255	19067	2022-04-20 07:20:49
6	11934	17153	2022-04-20 07:20:40
7	14897	16105	2022-04-20 07:20:43
8	15389	14891	2022-04-20 07:20:55
9	17999	16336	2022-04-20 07:20:52
10	18517	17328	2022-04-20 07:20:47
11	16101	14935	2022-04-20 07:20:58
12	18265	16833	2022-04-20 07:20:59
13	4634	13980	2022-04-20 07:20:38

- **tconst,DI_Create_DT** – Everything is in proper format
- **directors,writers** - Replace \N values with 0 or -1

5. imdb_title_principals

Alteryx Designer x64 - SQL_Profiling5-imdb_title_principals_table.yxmd

Search for tools, help, and resources...

File Edit View Options Help

Favorites Recommended In/Out Preparation Join Parse Transform In-Database Reporting Documentation Spatial Machine Learning Text Mining Computer Vision Interface Data Investigation

Browse Input Data Output Data Text Input Data Cleansing Filter Formula Sample Select Sort Join Union Text To Columns Summarize Comment

A newer version of Alteryx Designer x64 is available: [Click here for options](#)

Browse (5) - Configuration 4,283,620 records displayed, 7 fields, 118 MB

Profile 4,283,620 records displayed, 7 fields, 118 MB

tconst

tt0001116	10
tt0001175	10
tt0001630	10
tt0001790	10
tt0002101	10
995 more >	

ordering

1	591,300
2	545,574
3	508,997
4	474,836
5	442,350
5 more >	

nconst

nm0006137	1,009
nm0756966	1,001
nm1113677	768

SQL_Profiling5-imdb_title_principals_table.yxmd

SQL_Profiling5-imdb_title_ratings_table.yxmd

Results - Browse (5) - Input 7 of 7 Fields Cell Viewer 4,283,620 records displayed, 118 MB

Record #1, Field tconst (V_String)

tt1580426

Record	tconst	ordering	nconst	category	job	characters	DI_Create_DT
1	tt1580426	5	nm0078762	director	IN	["Self"]	2022-04-20 07:23:29
2	tt0618181	10	nm0423690	self	IN	["Barbara"]	2022-04-20 07:23:19
3	tt1569363	6	nm0542029	director	IN	["Colonel Galko"]	2022-04-20 07:22:04
4	tt1568814	4	nm0679167	actress	IN		2022-04-20 07:23:14
5	tt0055866	1	nm059412	actor	IN		2022-04-20 07:22:45
6	tt042522	3	nm0370144	actor	IN		2022-04-20 07:22:50
7	tt1237034	6	nm0759683	producer	IN		
8	tt0293289	7	nm1475378	writer	IN		
9	tt0352178	8	nm0109590	production_designer	IN		

- **tconst, ordering, nconst, category, DI_Create_DT** - Everything is in proper format
- **job** - Replace \N with 0
- **characters** - Replace [""] and \N with 0

6. imdb_title_ratings

Alteryx Designer x64 - SQL_Profiling6-imdb_title_ratings_table.yxmd

Search for tools, help, and resources...

File Edit View Options Help

Favorites Recommended In/Out Preparation Join Parse Transform In-Database Reporting Documentation Spatial Machine Learning Text Mining Computer Vision Interface Data Investigation

Browse Input Data Output Data Text Input Data Cleansing Filter Formula Sample Select Sort Join Union Text To Columns Summarize Comment

A newer version of Alteryx Designer x64 is available: [Click here for options](#)

Browse (5) - Configuration 277,171 records displayed, 2 fields, 2.7 MB

Profile 277,171 records displayed, 2 fields, 2.7 MB

Max_numVotes

8	7,048
9	7,017
10	6,604
7	6,465
11	6,195
995 more >	

tconst

tt0000574	1
tt0000615	1
tt0000630	1
tt0000675	1
tt0000793	1
995 more >	

SQL_Profiling5-imdb_title_principals_table.yxmd

SQL_Profiling6-imdb_title_ratings_table.yxmd

Results - Browse (5) - Input 2 of 2 Fields Cell Viewer 277,171 records displayed, 2.7 MB

Field Information

Field #3

Record	Max_numVotes	tconst
1	942	tt0097410
2	419	tt0380758
3	198	tt0403598
4	97	tt0077575
5	749	tt0006189
6	42	tt0012814
...

- **tconst, averageRatings, DI_Create_DT** - Everything is in proper format
- **numVotes** - remove comma