

## Data Profiling Document

### 1. Title\_episode

Overview:

Total number of records: 9,324,110

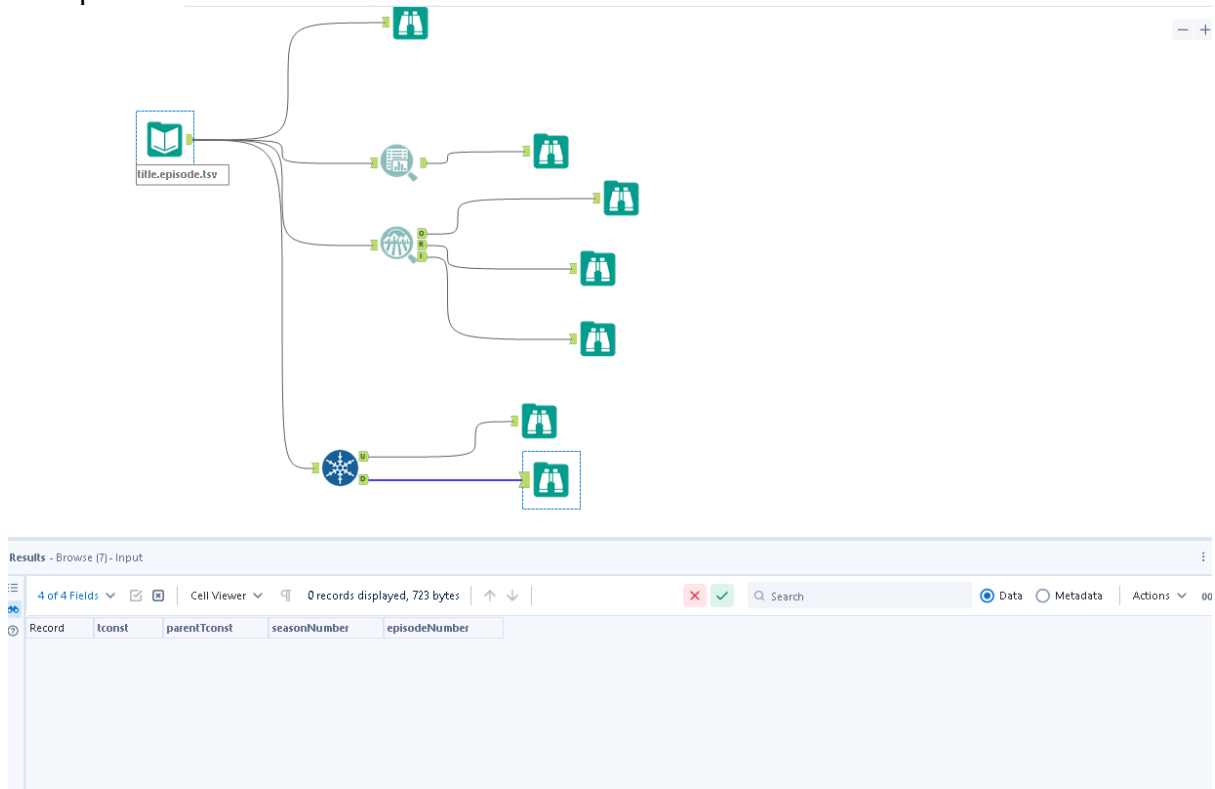
Total number of Columns:4

Duplicate records :0

Column Names	Source Data Type	Target Data Type
tconst	V_String	V_String
parentTconst	V_String	V_String
seasonNumber	V_String	Integer
episodeNumber	V_String	Integer

Data Quality Assessment:

#### 1. Duplicated Records



Detail : There are no duplicate record while selecting all the columns from the dataset.

## 2. Missing/Null values

Results - Browse (4) - Input

3 of 3 Fields | Cell Viewer | 4 records displayed, 2,032 bytes | Search

Record	FieldName	Name	Value
1	tconst	Nulls	0
2	parentTconst	Nulls	0
3	seasonNumber	Nulls	0
4	episodeNumber	Nulls	0

None of the fields have null /missing values

## 3. Leading and trailing whitespaces

There are no leading and trailing whitespaces in any of the columns in the entire dataset.  
Checked via basic profile tool

## 4. Contains '\N' values :

Columns affected : seasonNumber, episodeNumber

Column Names	Replace '\N' with	Reason
seasonNumber	-1	To maintain numeric format
episodeNumber	-1	To maintain numeric format

Summary Statistics:

1 record displayed, 2 fields, 1,985 bytes | Report | Profile

1 of 1 Fields | Records 1 to 1

Record 1

Report

String/Character Fields

Name	% Missing	Unique Values	Shortest Value	Longest Value	Min Value Count	Max Value Count	Remarks
parentTconst	0.0%	228,980	tt0041038	tt32857063	1	18,593	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
episodeNumber	0.0%	15,874	9	10312	1	1,901,449	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
seasonNumber	0.0%	329	1	1994	1	4,720,207	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
tconst	0.0%	9,324,110	tt0031458	tt10001006	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.

## 2. Title\_ratings

Overview:

Total number of records: 1605649

Total number of Columns:3

Duplicate records :0

Column Names	Source Data Type	Target Data Type
tconst	V String	V String
averageRating	V String	double
numVotes	V String	Integer

Data Quality Assesment:

### 1. Duplicate records

The diagram shows a data quality assessment workflow starting with a 'title\_ratings.tsv' source. It branches into three main paths: one for duplicate detection (using a 'Detect Duplicates' tool), one for missing value detection (using a 'Find Missing Values' tool), and one for schema validation (using a 'Validate Schema' tool). Each path leads to a 'Results' table.

**Results - Browse (18) - Input**

3 of 3 Fields | Cell Viewer | 1,605,649 records displayed, 16 MB | Search | Data | Metadata | Actions

Record	tconst	averageRating	numVotes
1	tt0000001	5.7	2186
2	tt0000002	5.5	306
3	tt0000003	6.4	2270
4	tt0000004	5.2	196
5	tt0000005	6.2	3012
6	tt0000006	5.0	221
7	tt0000007	5.3	934
8	tt0000008	5.4	2359
9	tt0000009	5.2	231

There are no duplicates in the entire dataset

### 2. Missing/Null values:

The diagram shows a data quality assessment workflow starting with a 'title\_ratings.tsv' source. It branches into three main paths: one for duplicate detection (using a 'Detect Duplicates' tool), one for missing value detection (using a 'Find Missing Values' tool), and one for schema validation (using a 'Validate Schema' tool). Each path leads to a 'Results' table.

**Results - Browse (4) - Input**

3 of 3 Fields | Cell Viewer | 3 records displayed, 2,025 bytes | Search | Data | Metadata | Actions

Record	FieldName	Name	Value
1	tconst	Nulls	0
2	averageRating	Nulls	0
3	numVotes	Nulls	0

There are no nulls in the entire dataset

### 3. Leading and trailing whitespaces

There are no leading /trailing whitespaces in the entire dataset .Checked all the columns via basic data profile

Summary Statistics:

Record	Report
--------	--------

1

#### String/Character Fields

Name	% Missing	Unique Values	Shortest Value	Longest Value	Min Value Count	Max Value Count	Remarks
averageRating	0.0%	91	5.7	10.0	355	59,073	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
numVotes	0.0%	23,926	9	2181557	1	62,979	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
tconst	0.0%	1,605,649	tt0000001	tt10001002	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.

### 3. Title-Principles

Overview:

Total number of records: 96,132,477

Total number of Columns:4

Duplicate records : 0

Source to Target Datatype:

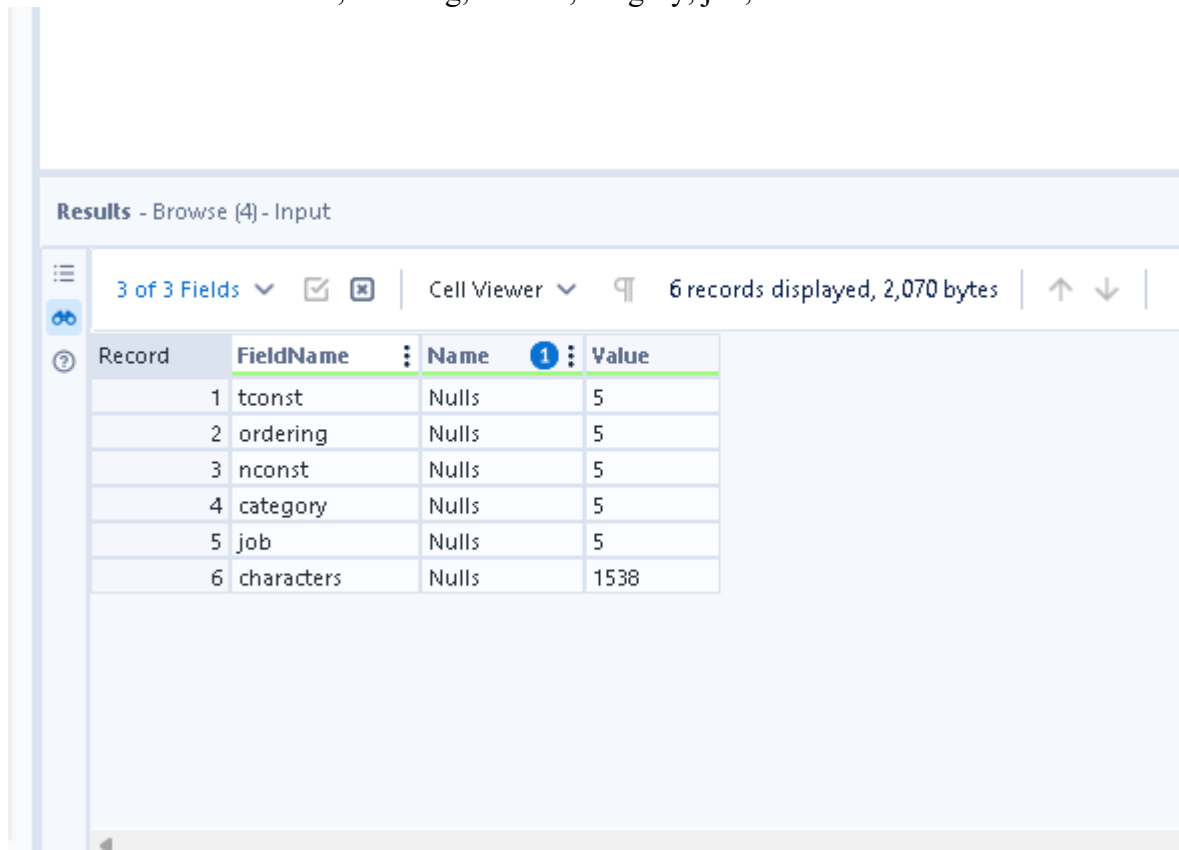
Column Names	Source Data Type	Target Data Type
tconst	V_String	V_String
ordering	V_String	Integer
nconst	V_String	V_String
job	V_String	V_String
characters	V_String	V_String
category	V_String	V_String

### Data Quality Assessment

1. Missing/Null values:

Verified using basic data profiling tool

Affected Columns: tconst, ordering, nconst ,category, job, characters



Results - Browse (4) - Input			
3 of 3 Fields ✓ ✕   Cell Viewer 🔍   6 records displayed, 2,070 bytes   ⬆ ⬇			
Record	FieldName	Name	Value
1	tconst	Nulls	5
2	ordering	Nulls	5
3	nconst	Nulls	5
4	category	Nulls	5
5	job	Nulls	5
6	characters	Nulls	1538

Plan of action:


Column Names	Replace with	Reason
tconst	‘unknown’	To maintain string format
ordering	‘-1 ‘	To main Integer format
nconst	‘unknown’	To maintain string format
job	‘unknown’	To maintain string format
characters	‘unknown’	To maintain string format

category	'unknown'	To maintain string format
----------	-----------	---------------------------

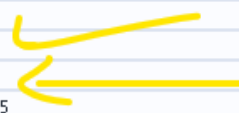
### 3. Leading and Trailing whitespaces:

Trailing and Both whitespaces:

85	job	Name	job
86	job	Data Type	V_String
87	job	Size	254
88	job	Source	File: C:\Users\prath\Downloads\IMDB Dataset\M...
89	job	Description	[Null]
90	job	OKs	96132471
91	job	Nulls	5
92	job	Non-Nulls	96132472
93	job	Blanks	0
94	job	Values with Leading Whitespace	0
95	job	Values with Trailing Whitespace	1
96	job	Values with Both Whitespace	1



106	characters	Name	characters
107	characters	Data Type	V_String
108	characters	Size	254
109	characters	Source	File: C:\Users\prath\Downloads\IMDB Dataset\M...
110	characters	Description	[Null]
111	characters	OKs	96130938
112	characters	Nulls	1538
113	characters	Non-Nulls	96130939
114	characters	Blanks	0
115	characters	Values with Leading Whitespace	0
116	characters	Values with Trailing Whitespace	1
117	characters	Values with Both Whitespace	1
118	characters	Average Length	8.5
119	characters	Longest Length	254



Affected columns: job and characters:

Plan of action :Remove the whitespaces in cleaning

### 4. Contains 'N' Values

Affected Columns

Column Names	Replace 'N' with	Reason
job	'unknown'	To maintain string format
characters	'unknown'	To maintain string format

Summary Statistics:

1 record displayed, 2 fields, 2,724 bytes

Report

Profile

1 of 1 Fields

Records 1 to 1

Record	Report
1	String/Character Fields

Name	% Missing	Unique Values	Shortest Value	Longest Value	Min Value Count	Max Value Count	Remarks
ordering	0.0%	76	1	10	1	10,929,831	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
category	0.0%	14	self	production_designer	5	22,797,878	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
job	0.0%	45,867	e	sketches "Allô", "ton ton, pourquoi tu tousses?", "La 2 CV de ma soeur", "Toto en pension", "L'hallebardier", "Le râ@gine sans selle", "Le plombier", "T'entends t'y, dis Paul?", "Les croissants", "Le vase de Soissons", "Moi, mon papa, il a un vÂ@lo", "Le b	1	78,106,531	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
nconst	0.0%	6,891,832	nm1588970	nm10501864	1	39,320	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
characters	0.0%	4,394,695	\N	["Self - MÃ@decin de famille et en soins palliatifs au Centre intÃ©grÃ© universitaire de santÃ© et de services sociaux (CIUSSS) du Nord-de-l'Ã®le-deMontrÃ©al, et professeure adjointe de clinique Ã la FacultÃ© de mÃ@decine de l'UniversitÃ© de MontrÃ©al"]	1	49,264,505	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
tconst	0.0%	10,929,832	tt00000001	tt10001000	1	75	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.

#### 4. Title\_crew

Overview:

Total number of records: 12,095,664

The screenshot displays the Alteryx Designer x64 interface. On the left, the 'Browse (2) - Configuration' pane shows the dataset 'title.crew.tsv' with 12,095,664 records and 3 fields. The 'Profile' section lists fields like 'tconst', 'directors', and 'writers' with their respective record counts. The 'Results - Browse (2) - Input' pane on the right shows a preview of the data with columns: Record, tconst, directors, and writers. The workflow canvas shows a 'title.crew.tsv' input tool connected to a 'Browse' tool.

Total number of Columns:3

Duplicate records :0

Column Names	Source Data Type	Target Data Type
tconst	V_String	V_String
directors	V_String	V_String
writers	V_String	V_String

Data Quality Assesment:

1. Missing /Null values:

The screenshot shows the 'Results - Browse (7) - Input' pane. It displays a table with 3 records. The columns are 'Record', 'FieldName', 'Name', and 'Value'. The data shows that for the first three records, the 'Name' column contains 'Nulls' and the 'Value' column contains '0'.

Record	FieldName	Name	Value
1	tconst	Nulls	0
2	directors	Nulls	0
3	writers	Nulls	0

There are no nulls in the entire dataset



## 2. Leading/trailing whitespaces

There are no leading /trailing whitespaces in the dataset

## 3. Contains '\N' values:

directors		
\N	4,314,201	
nm1203430	11,458	
nm1409127	9,924	
nm0565214	5,325	
nm1402433	5,246	
995 more >		

writers		
\N	4,883,320	
nm0934807	10,917	
nm0417224	5,183	
nm0341311	4,230	
nm1879108,nm1876156	4,183	
995 more >		

Action – need to replace '\N' values in directors and writers with 'unknown' in cleaning

## 4. Comma-Separated Values

Need to flatten the table since multiple directors and writers are in same field in arrays which should be in different rows, so they would have to be exploded and a new field would have to be created called crew\_role which would have the directors and writers

Record	tconst	directors	writers
240	tt0000242	nm0617588	nm0617588
241	tt0000243	nm0349785	\N
242	tt0000244	nm0349785	\N
243	tt0000245	\N	\N
244	tt0000246	nm0617588	\N
245	tt0000247	nm2156608,nm0005690,nm0002504	nm0000636,nm0002504
246	tt0000248	nm0808310	\N
247	tt0000249	nm0808310	\N
248	tt0000250	nm0005717	\N
249	tt0000251	nm0177862	\N
250	tt0000252	nm0349785	\N
251	tt0000253	nm0095816	\N
252	tt0000254	nm0349785	\N
253	tt0000255	nm0349785	\N

## 5. Title basics:

Overview:

Total number of records: 12,095,664

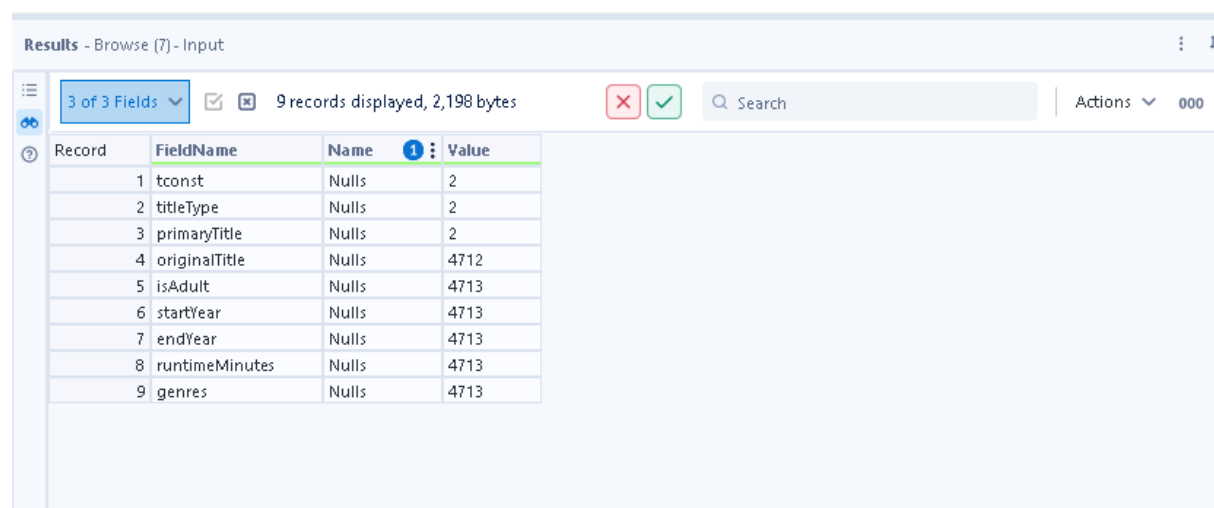
Total number of Columns:9

Duplicate records :0

Column Names	Source Data Type	Target Data Type
tconst	V_String	V_String
titleType	V_String	V_String
primaryTitle	V_String	V_String
originalTitle	V_String	V_String
isAdult	V_String	boolean
startYear	V_String	Integer
endYear	V_String	Integer
runtimeMinutes	V_String	integer
genres	V_String	V_String

Data Inconsistency:

1. Contains Nulls:



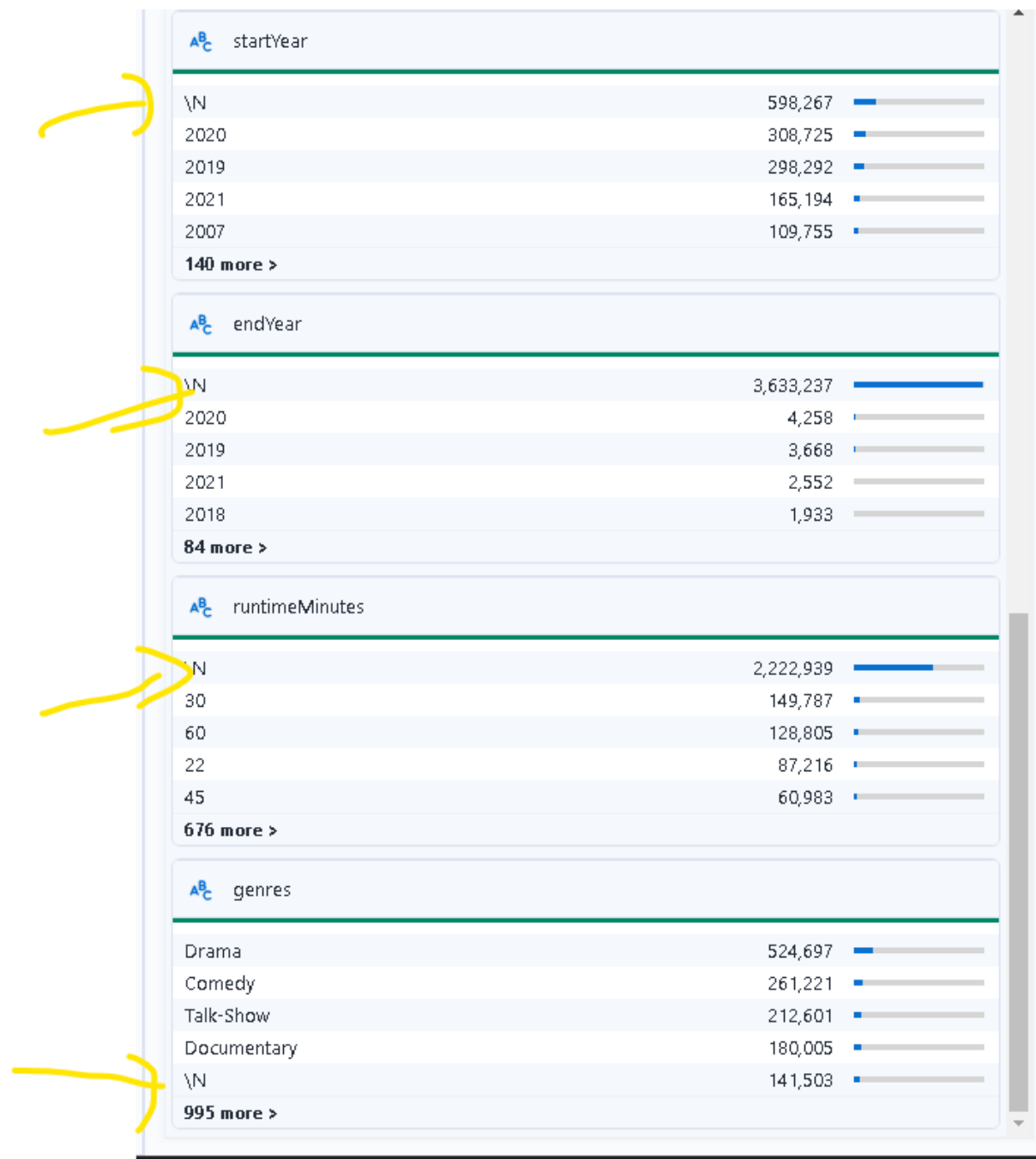
Record	FieldName	Name	Value
1	tconst	Nulls	2
2	titleType	Nulls	2
3	primaryTitle	Nulls	2
4	originalTitle	Nulls	4712
5	isAdult	Nulls	4713
6	startYear	Nulls	4713
7	endYear	Nulls	4713
8	runtimeMinutes	Nulls	4713
9	genres	Nulls	4713

Action

Column Name	Replace with	Reason
tconst	Unknown	To maintain datatype format
titletype	Unknown	To maintain datatype format
primaryTitle	Unknown	To maintain datatype format
originalTitle	Unknown	To maintain datatype format
isAdult	-1	To maintain datatype format
startYear	0000	To maintain YYYY format
endYear	9999	To maintain YYYY format
runtimeMinutes	-1	To maintain datatype format
genres	Unknown	To maintain datatype format

2.. Contains 'N' values:

Affected Columns: startYear ,endYear ,runtimeMinutes,genres



Action- Replace '\N' with following:

Column Name	Replace with	Reason
startYear	0000	To maintain YYYY format
endYear	9999	To maintain YYYY format
runtimeMinutes	-1	To maintain datatype format
genres	unknown	To maintain datatype format

2. Trailing whitespace:

Affected columns: primaryTitle , originalTitle

primaryTitle x

### Summary ^

Type	Records	Data Type Size
V_String	3,689,465	254

● Ok	3,689,449	100.00%
Unique	> 10,000	
● Null	2	0.00%
● Not Ok	14	0.00%
● Empty	0	0.00%

### Length Statistics ^

Min	1
Max	254
Average	19.30
Shortest Value	M
Longest Value	Bob Hope Looks at Women's Lib with Connie Steve...
First Alphanumeric Value	!OUTING!
Last Alphanumeric Value	é□□æ□³ Tekken 4
Blanks	0
Values with Leading Whitespace	0
Values with Trailing Whitespace	14

Profile		
originalTitle		
Summary		
Type	Records	Data Type Size
V_String	3,689,465	254
Ok	3,688,591	99.98%
Unique	> 10,000	
Null	863	0.02%
Not Ok	11	0.00%
Empty	0	0.00%
Length Statistics		
Min		1
Max		254
Average		19.30
Shortest Value		M
Longest Value	Bob Hope Looks at Women's Lib with Connie Steve...	
First Alphanumeric Value		!OUTING!
Last Alphanumeric Value	Ävalker Danke Reklami	
Blanks		0
Values with Leading Whitespace		0
Values with Trailing Whitespace		11
Frequent Values		

Action: Remove the trailing whitespaces using trim function in cleaning

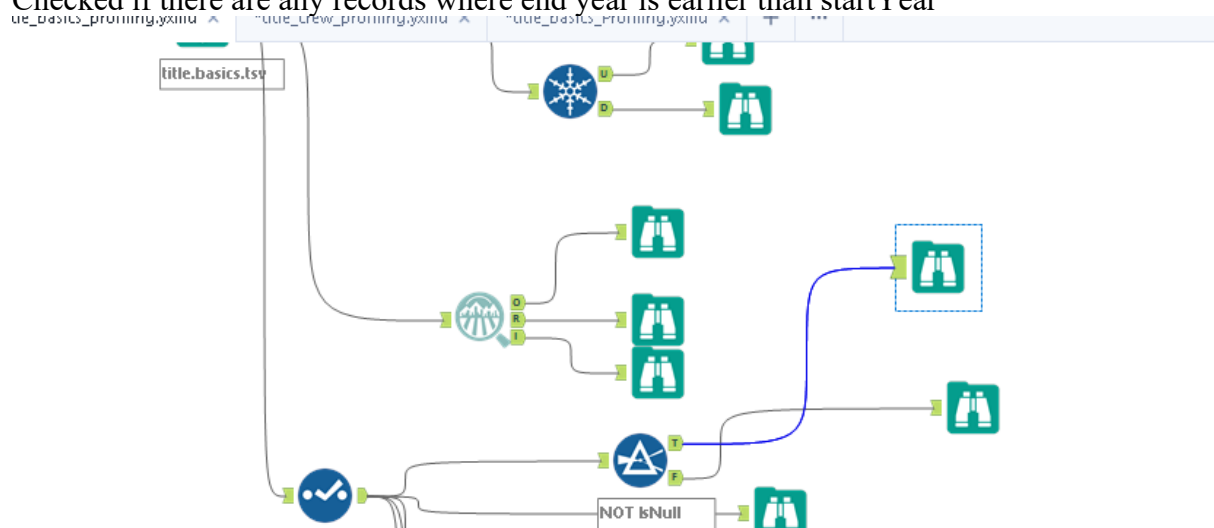
### 3. Comma Separated values

Genre field would have to be flattened since its having multiple values in a single field ie explore

runtimeMinutes	genres
1	Documentary,Short
5	Animation,Short
5	Animation,Comedy,Romance
12	Animation,Short
1	Short
1	Short
1	Short,Sport
1	Documentary,Short
45	Romance
1	Documentary,Short
1	Documentary,Short
1	Documentary,Short
1	Documentary,Short
1	Comedy,Short
2	Animation,Comedy,Short
1	Documentary,Short
1	Documentary,Short
1	Short,Sport
\N	Comedy,Short
1	Documentary,Short,Sport
1	Documentary,Short
1	Documentary,Short
1	News Short

### Additional verification

Checked if there are any records where end year is earlier than startYear



There are none in the dataset

## 6. Title\_Regions

Overall Records: 244

Total Number of columns: 2

Missing values overall :0

Duplicates :0

Column Name	Source Target Data type
REGION_NAME	String
REGION_CODE	String

Results - Browse [9] - Input							
16 of 22 Fields		Cell Viewer	2 records displayed, 2,296 bytes		Search		
						<input checked="" type="radio"/> Data	<input type="radio"/> Metadata
						Actions	
						000	
Record	Name	Field Category	Percent Missing	Unique Values	Remarks	Shortest_Value	Longest_Value
1	REGION_NAME	String	0	244	Some values of this field have a small number of...	Chad	Korea Democratic People
2	REGION_CODE	String	0	244	Some values of this field have a small number of...	af	af

Conclusion- No issues identified in the dataset as its directly extracted from source

## 7. Title\_Language\_Codes

Overall Records: 183

Total Number of columns: 2

Missing values overall :0

Duplicates :0

Column Name	Source Target Data type
LANGUAGE_NAME	String
LANGUAGE_CODE	String

Results - Browse [9] - Input							
16 of 22 Fields		Cell Viewer	2 records displayed, 2,277 bytes		Search		
						<input checked="" type="radio"/> Data	<input type="radio"/> Metadata
						Actions	
						000	
Record	Name	Field Category	Percent Missing	Unique Values	Remarks	Shortest_Value	Longest_Value
1	LANGUAGE_CODE	String	0	183	Some values of this field have a small number of...	ab	ab
2	LANGUAGE_NAME	String	0	183	Some values of this field have a small number of...	Ewe	Church iꞤꞤSlavonic Old iꞤ

No issues identified in the dataset as its directly extracted from source

## 8. Name\_Basics

Total Records – 14,903,875

Total attributes – 6

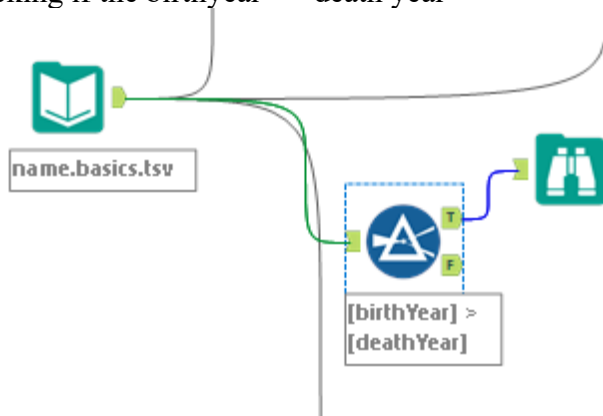
Duplicates – 0

Column Name	Source Data type	Target Data Type
Nconst	V_String	V_String
primaryName	V_String	V_String
BirthYear	V_String	Integer
DeathYear	V_String	Integer
primaryProfession	V_String	V_String
knownForTitles	V_String	V_String

Key Observations –

1. There are no duplicates in the dataset.

2. Checking if the birthyear  $\geq$  death year



The \N values were handled by replacing with 0000

3. There are nulls in the birth Year which are represented by “\N” indicating that these details were not captured. We will replace all the “\N” by nulls first and then replace all these nulls with 0000 to indicate incorrect year format.

Results - Browse (17) - Input						
6 of 6 Fields						
Cell Viewer						
15,739 records displayed, 756 KB						
Data						
Record	nconst	primaryName	birthYear	deathYear	primaryProfession	knownForTitles
1	nm0001424	Udo Kier	\N	2025	actor,writer,producer	tt30333632,tt27847051,tt7456722,tt10832274
2	nm0002334	Douglas C. Hart	\N	2023	camera_department,visual_effeds	tt0097770,tt0100404,tt0089853,tt0108550
3	nm0002578	Thomas J. Mack	\N	2021	assistant_director,producer,camera_department	tt0149261,tt0120749,tt0100685,tt0106977
4	nm0002729	Matthew J. Clark	\N	2020	assistant_director,location_managementeditor	tt0119215,tt0089886,tt0207109,tt0100661
5	nm0003085	Alan Roberts	\N	2016	editor,producer,director	tt0147479,tt0226429,tt0076944,tt0070507
6	nm0003229	Roger Cardinal	\N	2017	director,writer,producer	tt0097818,tt0244500,tt0396514,tt0244403
7	nm0003381	Miguel Ángel Rivas	\N	2003	assistant_director,writer,director	tt0079544,tt0077263,tt0081289,tt0255379
8	nm0003528	Raúl Dávalos	\N	2017	editor,editorial_department,producer	tt0285531,tt0106387,tt0104029,tt0090022
9	nm0003805	Geoffrey Brown	\N	2015	producer,actor,director	tt0238217,tt0163179,tt0087612,tt0083868
10	nm0004153	Carol Goldwasser	\N	2024	casting_director,casting_department,miscellaneous	tt1830491,tt0493093,tt7631348,tt2222135
11	nm0005998	K. Chakravarthy	\N	2002	composer,actor,music_department	tt0246313,tt0245980,tt0245731,tt0257362
12	nm0006421	Robert Grayson	\N	2024	assistant_director,miscellaneous	tt1706620,tt0120738,tt0119116,tt0475276
13	nm0006693	Malcolm Davies	\N	2019	sound_department,camera_department,miscellan...	tt2317225,tt3622110,tt0084750,tt0491807
14	nm0007607	Ignasi Abadal	\N	2010	actor	tt0247586,tt0064904,tt2235501,tt0297174
15	nm0007870	Claire Abbey	\N	2011	actress	tt0107972,tt0087837
16	nm0007994	Joan Abbott	\N	1968	actress,soundtrack	tt0026527,tt013603,tt4983554,tt2113671



4. There are nulls in the death Year which are represented by “\N” indicating that these people are alive yet. We will replace all the “\N” by nulls first and then replace all these nulls with 9999 to indicate incorrect year format.

5. There are \N in the primary profession and the knownForTitles will be replaced with “Unknown”.

Results - Browse (2) - Input

6 of 6 Fields | Cell Viewer | 2,968,917 records displayed, 71 MB | Search | Data | Metadata | Actions | 000

Record	nconst	primaryName	birthYear	deathYear	primaryProfession	knownForTitles
1	nm0002652	Gary Rosen	\N	\N	\N	tt0358197,tt0268808
2	nm0003936	Elvind Fredagsvik Ådyangen	1979	\N	\N	\N
3	nm0004346	Richard A. Roth	\N	\N	\N	tt0110724
4	nm0007243	Antonio 'El de Chipiona'	\N	\N	\N	tt0122520
5	nm0007257	Izzy 'The Religious Nut'	\N	\N	\N	tt0111522
6	nm0007260	Sharon 10X	\N	\N	\N	tt0054416
7	nm0007265	Benjamin 2X	1932	2005	\N	tt0108089,tt0094416
8	nm0007271	Lucky 7	\N	\N	\N	tt0192069,tt0421356
9	nm0007286	A-Man	\N	\N	\N	tt0127499
10	nm0007353	Aalami	\N	\N	\N	tt0031874
11	nm0007411	Marjorie Aamodt	\N	\N	\N	tt0117228
12	nm0007454	Eddie Aaron	\N	\N	\N	tt0279745
13	nm0007516	Karl Johan AarsÅther	1920	1971	\N	tt0049808
14	nm0007530	Svein Roar Aas	\N	\N	\N	tt0277500
15	nm0007570	Elena Ababy	\N	\N	\N	\N

We Cannot drop the rows that have no primary name as they are associated with a nconst and therefore these \N values will be replaced with default values

Column Names	Replace ‘\N’ with	Reason
primaryName	Unknown	To maintain string format
birthYear	0000	To maintain numeric format
deathYear	9999	To indicate person is alive
primaryProfession	Unknown	To maintain string format
knownForTitles	Unknown	To maintain string format

## 9. Title.Akas

Total records – 54,124,127

Total Attributes – 8

Duplicates -

Basic view of data–


Attributes	Source Data type	Target Data Type
titleID	V_WString	V_String
Ordering	V_String	Integer
Title	V_String	V_String
region	V_String	V_String
language	V_String	V_String
types	V_String	Array
attributes	V_WString	Array
isOriginalTitle	V_WString	Integer

Results - Browse (2) - Input

8 of 8 Fields 54,124,127 records displayed, 760 MB

Record	titleId	ordering	title	region	language	types	attributes	isOriginalTitle
1	tt00000001	1	Carmencita	\N	\N	original	\N	1
2	tt00000001	2	Carmencita	DE	\N	\N	literal title	0
3	tt00000001	3	Carmencita	US	\N	imdbDisplay	\N	0
4	tt00000001	4	Carmencita - spanyol tánc	HU	\N	imdbDisplay	\N	0
5	tt00000001	5	Карменцита	GR	\N	imdbDisplay	\N	0
6	tt00000001	6	Карменсита	RU	\N	imdbDisplay	\N	0
7	tt00000001	7	Карменсита	UA	\N	imdbDisplay	\N	0
8	tt00000001	8	カルメンチータ	JP	ja	imdbDisplay	\N	0
9	tt00000002	1	Le clown et ses chiens	\N	\N	original	\N	1
10	tt00000002	2	A bohóc és kutyái	HU	\N	imdbDisplay	\N	0
11	tt00000002	3	Clovnul si cainii sai	RO	\N	imdbDisplay	\N	0
12	tt00000002	4	Der Clown und seine Hunde	DE	\N	\N	literal title	0
13	tt00000002	5	Le clown et ses chiens	FR	\N	imdbDisplay	\N	0
14	tt00000002	6	The Clown and His Dogs	US	\N	\N	literal English title	0
15	tt00000002	7	Клоун и его собаки	RU	\N	imdbDisplay	\N	0
16	tt00000002	8	道化師と犬	JP	ja	imdbDisplay	\N	0
17	tt00000003	1	Pauvre Pierrot	\N	\N	original	\N	1
18	tt00000003	10	Бідний П'єро	UA	\N	imdbDisplay	\N	0
19	tt00000003	11	哀れなピエロ	JP	ja	imdbDisplay	\N	0
20	tt00000003	2	Poor Pierrot	GB	\N	imdbDisplay	\N	0
21	tt00000003	3	Poor Pierrot	US	\N	imdbDisplay	\N	0
22	tt00000003	4	¡Pobre Pierrot!	ES	\N	imdbDisplay	\N	0
23	tt00000003	5	Armer Pierrot	DE	\N	\N	literal title	0
24	tt00000003	6	Pauvre Pierrot	FR	\N	imdbDisplay	\N	0

All the values in the data are distinct.



Results - Browse (15) - Input

8 of 8 Fields 54,124,123 records displayed, 760 MB

Record	titleId	ordering	title	region	language	types	attributes	isOriginalTitle
1	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]
2	tt00000001	1	Carmencita	\N	\N	original	\N	1
3	tt00000001	2	Carmencita	DE	\N	\N	literal title	0
4	tt00000001	3	Carmencita	US	\N	imdbDisplay	\N	0
5	tt00000001	4	Carmencita - spanyol tánc	HU	\N	imdbDisplay	\N	0
6	tt00000001	5	Карменцита	GR	\N	imdbDisplay	\N	0
7	tt00000001	6	Карменсита	RU	\N	imdbDisplay	\N	0
8	tt00000001	7	Карменсита	UA	\N	imdbDisplay	\N	0
9	tt00000001	8	カルメンチータ	JP	ja	imdbDisplay	\N	0
10	tt00000002	1	Le clown et ses chiens	\N	\N	original	\N	1
11	tt00000002	2	A bohóc és kutyái	HU	\N	imdbDisplay	\N	0
12	tt00000002	3	Clovnul si cainii sai	RO	\N	imdbDisplay	\N	0
13	tt00000002	4	Der Clown und seine Hunde	DE	\N	\N	literal title	0

The corrected data types

All the nulls and the \N values will be replaced with appropriate default values.

Column Names	Replace 'N' with	Reason
ordering	-1	To maintain numeric format
title	Unknown	To maintain string format
Region	Unknown	To maintain string format
Language	Unknown	To maintain string format
types	Unknown	To maintain string format
Attributes	Unknown	To maintain string format
isOriginalTitle	-1	To maintain numeric format