NORTHEASTERN UNIVERSITY

INFORMATION SYSTEMS FALL ’23

IMDB Database Project

DESIGNING ADVANCED DATA ARCHITECTURES FOR BUSINESS INTELLIGENCE

Guided By:

Prof. Naveen Kuragayala

Group 13:

Shreyasi Wakankar - 002771284

Muskan Srivastava - 002794929

Pramita Sandhyan - 002766881

Rhea Bajpai - 002702927

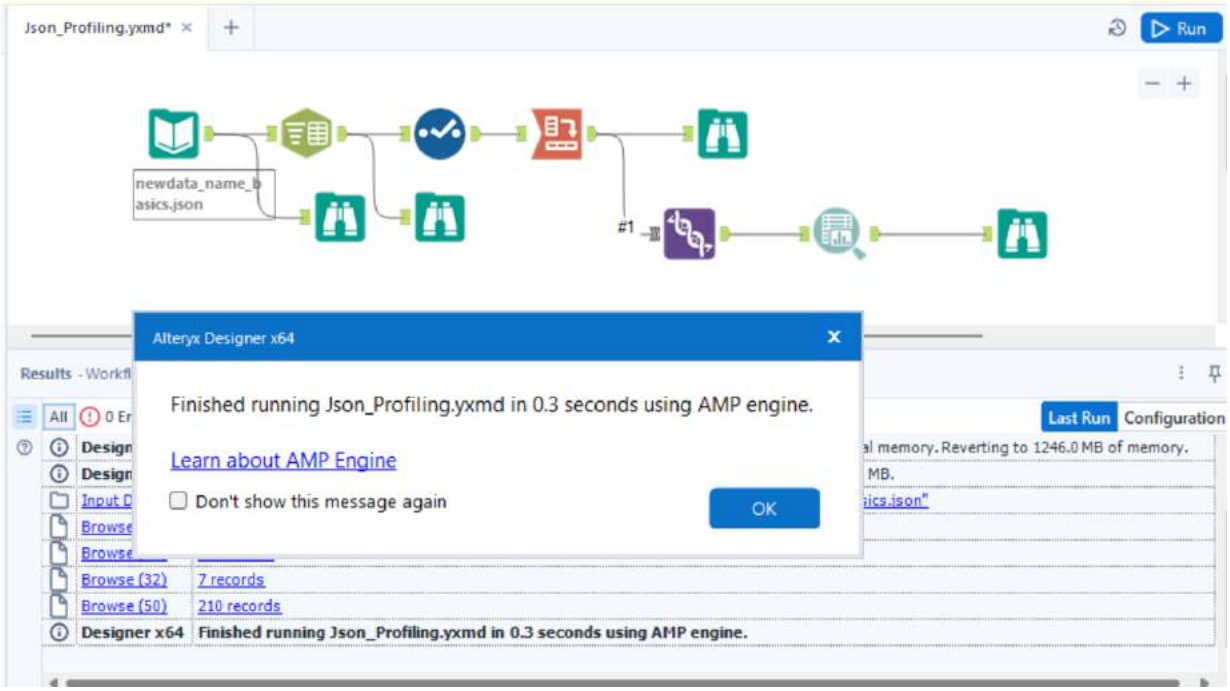
List of Tables:

r No.	Table Name	Table Type	Count
Source: MySQL Script			

1	stg_imdb_name_basics	Staging Table	14,54,602
2	stg_imdb_title_basics	Staging Table	6,07,423
3	stg_imdb_title_akas	Staging Table	26,12,798
4	stg_imdb_title_crew	Staging Table	6,07,423
5	stg_imdb_title_principal	Staging Table	42,83,620
6	stg_imdb_title_rating	Staging Table	2,77,171
Source: tsv File			
1	stg_movie	Staging Table	265
Source: JSON			
1	stg_newdata_name_baisc	Staging Table	7
2	stg_newdata_title_baisc	Staging Table	10
Dimensional Model Tables			
1	dim_person_profession	Dimension Table	2544773
2	dim_date	Dimension Table	10000
3	dim_person	Dimension Table	1454602
4	dim_person_knownfor	Dimension Table	3934774
5	dim_titleregion	Dimension Table	2612798
6	dim_crew	Dimension Table	4283620
7	dim_genre	Dimension Table	
8	im_movie_title_json_with_scd	Dimension Table	
9	dim_movie_title_scd	Dimension Table	10
10	dim_movie_title	Dimension Table	890688
11	fct_rating	Fact Table	277171
12	fct_IMDB	Fact Table	872504
13	fct_Movie_Earnings	Fact Table	265

Data Profiling in Alteryx:

1. **newdata_name_basics.json**



Observations:

3 of 3 Fields | 210 records displayed, 3,192 bytes | Search

Record	FieldName	Name	Value
20	1	Uniques	7
21	1	Unique Values	0
22	nconst	Name	nconst
23	nconst	Data Type	V_WString
24	nconst	Size	2048
25	nconst	Source	CrossTab:Header:Column_Name:nconst:Concat
26	nconst	Description	[Null]
27	nconst	OKs	7
28	nconst	Nulls	0
29	nconst	Non-Nulls	7
30	nconst	Blanks	0
31	nconst	Values with Leading Whitespace	0
32	nconst	Values with Trailing Whitespace	0
33	nconst	Values with Both Whitespace	0
34	nconst	Average Length	9.0
35	nconst	Longest Length	9

Profile

FieldName

Distinct Values	
1	21
DI_CreateDatetime	21
DI_JobId	21
DI_JobName	21
birthYear	21
deathYear	21
knownForTitles	21
nconst	21
primaryName	21
primaryProfession	21

FieldName ×

Summary

Type	Records	Data Type Size
V_WString	**210**	**1,073,741,823**

Ok	210	100.00%
Unique	10	4.76%
Null	0	0.00%
Not Ok	0	0.00%
Empty	0	0.00%

Length Statistics

Min	1
Max	17
Average	10.20
Shortest Value	1
Longest Value	primaryProfession
First Alphanumeric Value	1
Last Alphanumeric Value	primaryProfession
Blanks	0
Values with Leading Whitespace	0
Values with Trailing Whitespace	0

Frequent Values

Value by count

10 out of 10 values

Profile

Name ×

Summary

Type	Records	Data Type Size
V_WString	**210**	**1,073,741,823**

Ok	210	100.00%
Unique	21	10.00%
Null	0	0.00%
Not Ok	0	0.00%
Empty	0	0.00%

Length Statistics

Min	3
Max	31
Average	12.40
Shortest Value	OKs
Longest Value	Values with Trailing Whitespace
First Alphanumeric Value	Average Length
Last Alphanumeric Value	Values with Trailing Whitespace
Blanks	0
Values with Leading Whitespace	0
Values with Trailing Whitespace	0

Profile

Frequent Values

Value by count

21 out of 21 values

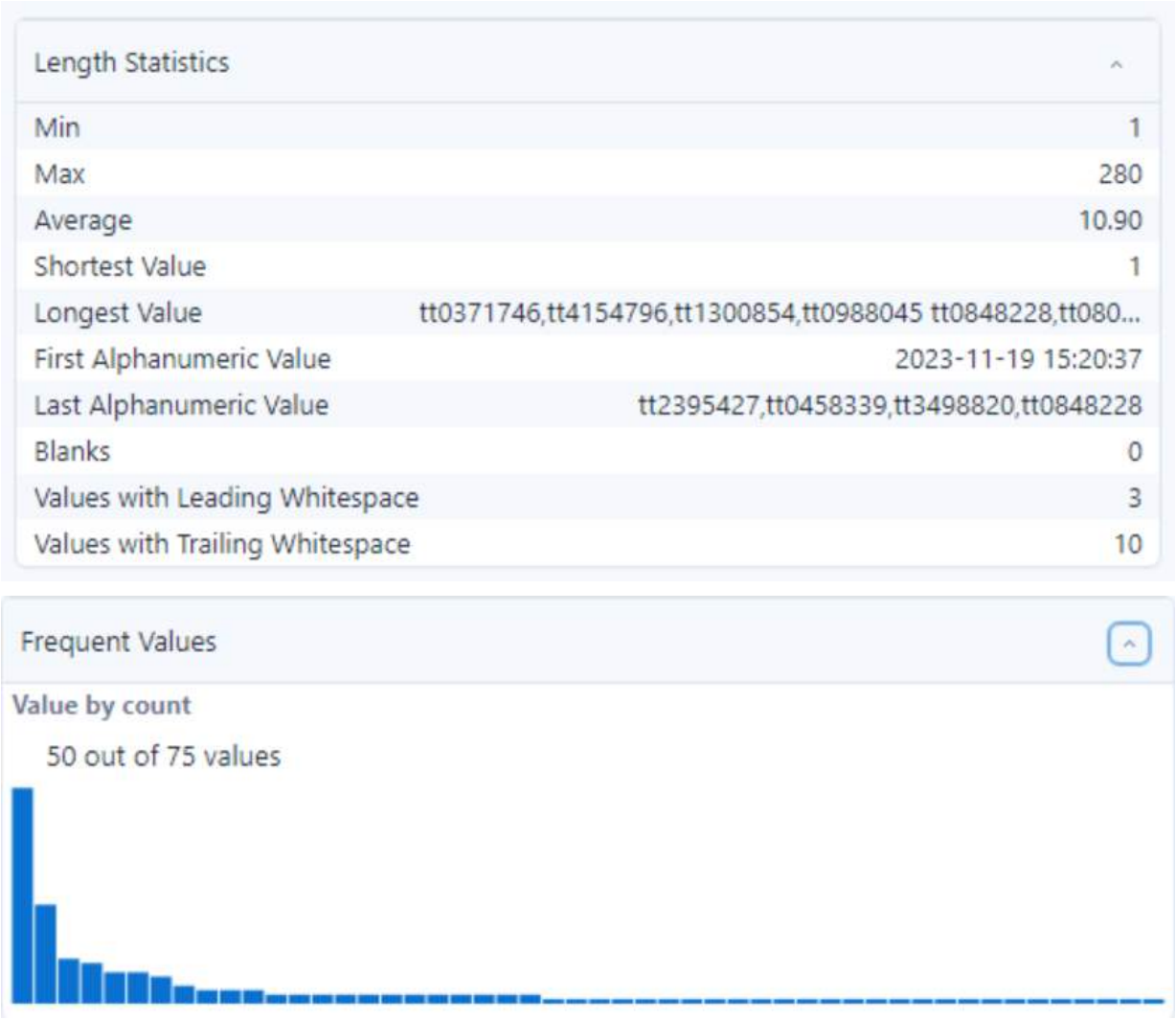
Profile

Value ×

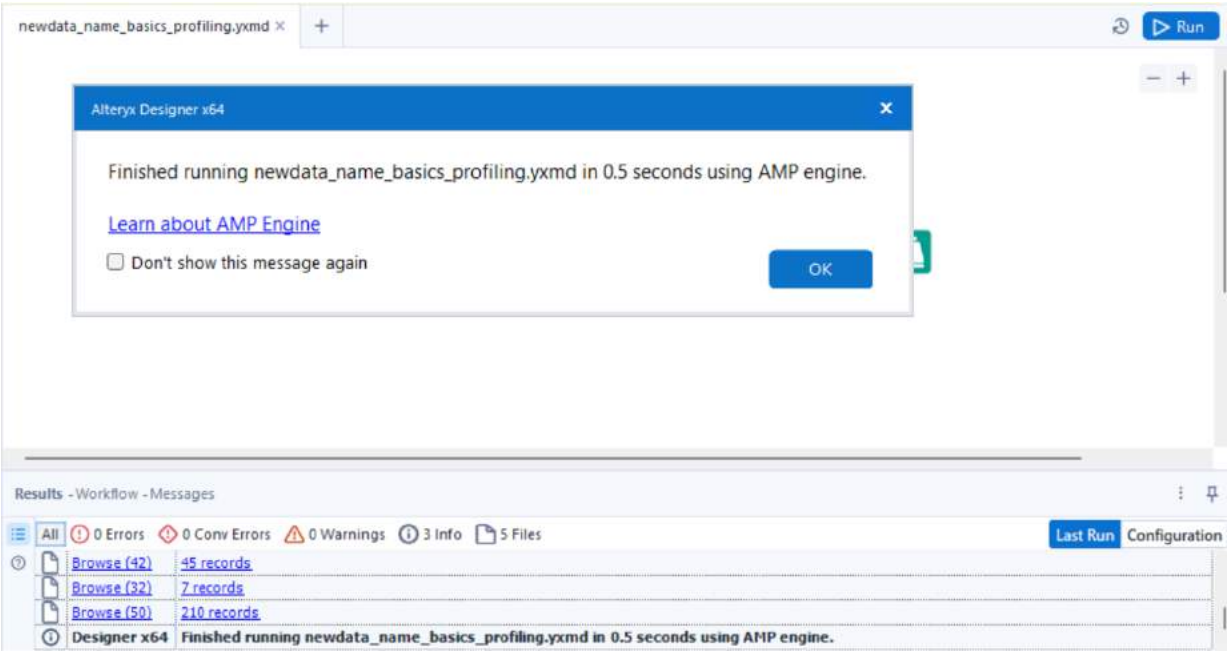
Summary

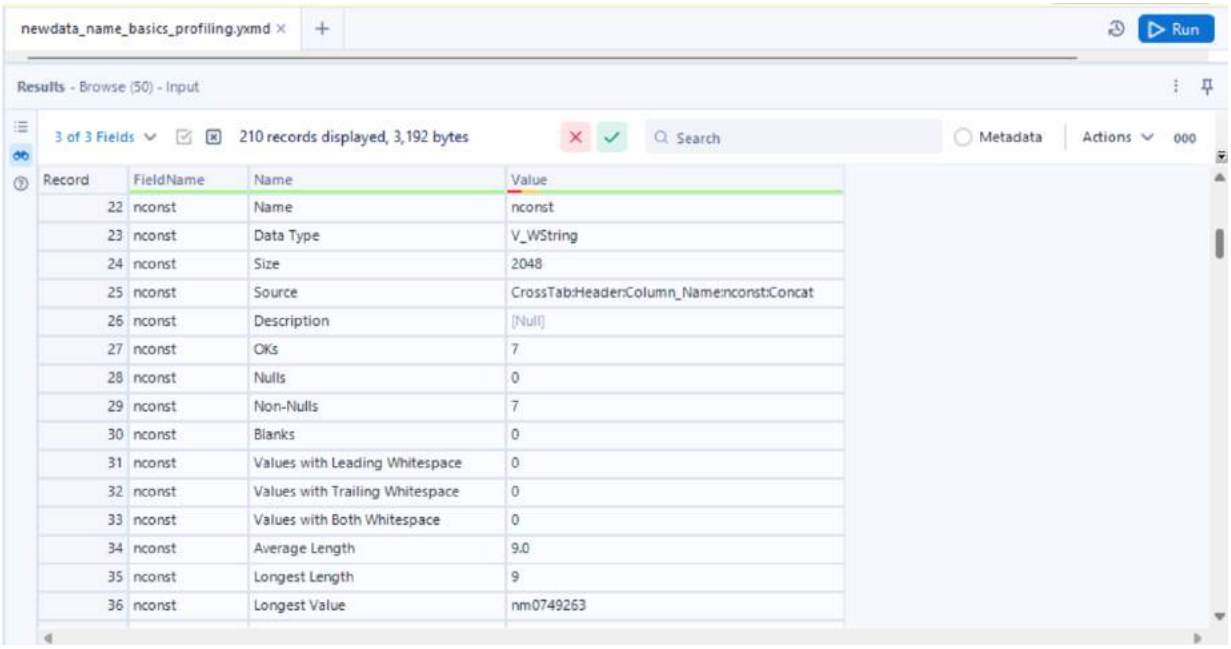
Type	Records	Data Type Size
V_WString	**210**	**1,073,741,823**

Ok	188	89.52%
Unique	75	35.71%
Null	12	5.71%
Not Ok	10	4.76%
Empty	0	0.00%

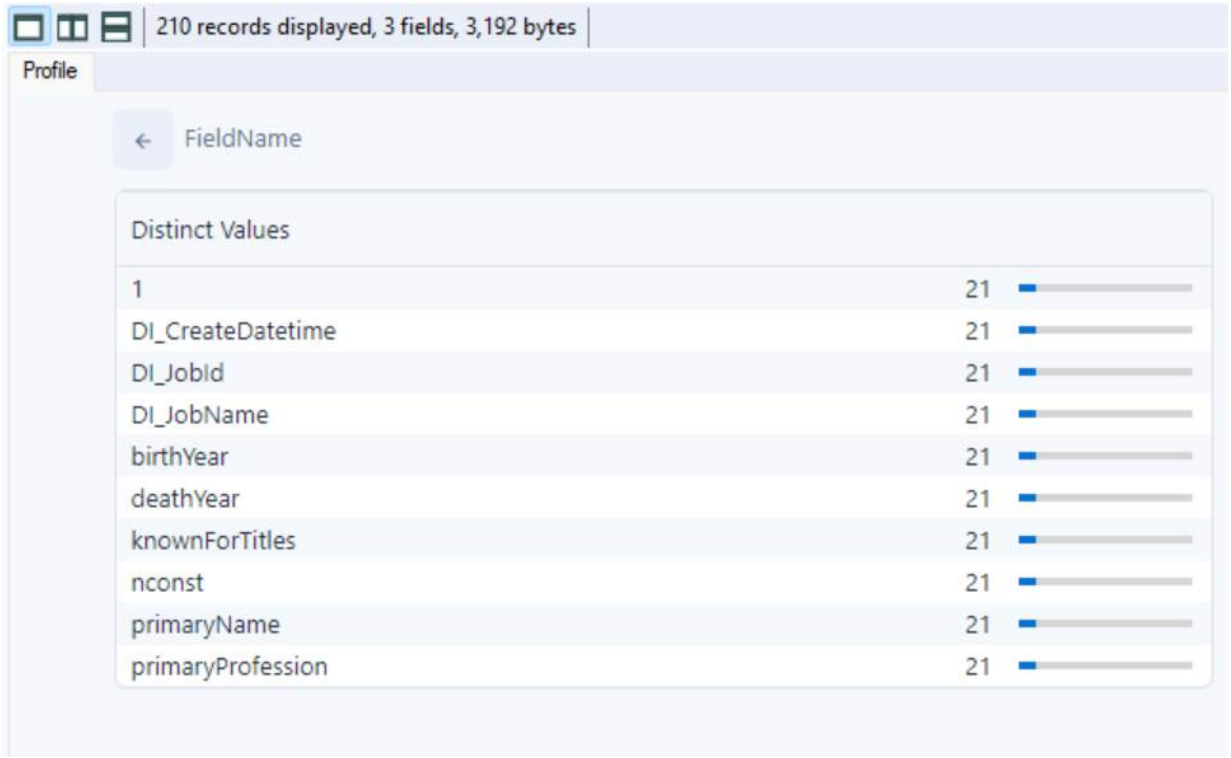


2. Newdata_title_basics.json:





Record	FieldName	Name	Value
22	nconst	Name	nconst
23	nconst	Data Type	V_WString
24	nconst	Size	2048
25	nconst	Source	CrossTab:Header:Column_Name:nconst:Concat
26	nconst	Description	[Null]
27	nconst	OKs	7
28	nconst	Nulls	0
29	nconst	Non-Nulls	7
30	nconst	Blanks	0
31	nconst	Values with Leading Whitespace	0
32	nconst	Values with Trailing Whitespace	0
33	nconst	Values with Both Whitespace	0
34	nconst	Average Length	9.0
35	nconst	Longest Length	9
36	nconst	Longest Value	nm0749263



FieldName ×

Summary

Type	Records	Data Type Size
V_WString	**210**	**1,073,741,823**

Ok	210	100.00%
Unique	10	4.76%
Null	0	0.00%
Not Ok	0	0.00%
Empty	0	0.00%

Length Statistics

Min	1
Max	17
Average	10.20
Shortest Value	1
Longest Value	primaryProfession
First Alphanumeric Value	1
Last Alphanumeric Value	primaryProfession
Blanks	0
Values with Leading Whitespace	0
Values with Trailing Whitespace	0

Frequent Values

Value by count

10 out of 10 values

Profile

Distinct Values	
Average Length	10
Blanks	10
Data Type	10
Description	10
Longest Length	10
Longest Value	10
Maximum	10
Minimum	10
Name	10
Non-Nulls	10
Nulls	10
OKs	10
Shortest (Non Blank) Length	10
Shortest Value	10
Size	10
Source	10

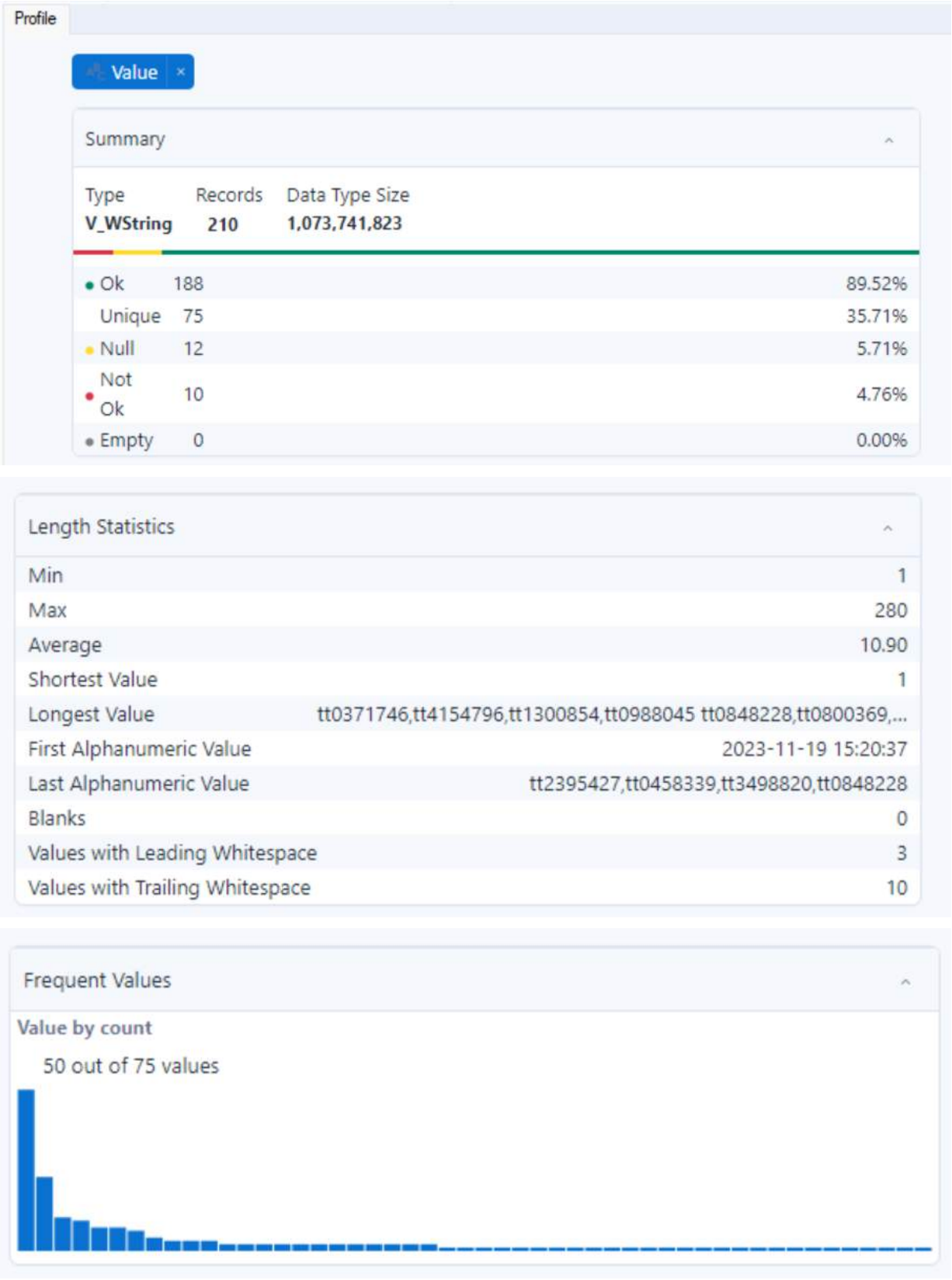
Summary

Type	Records	Data Type Size
V_WString	**210**	**1,073,741,823**

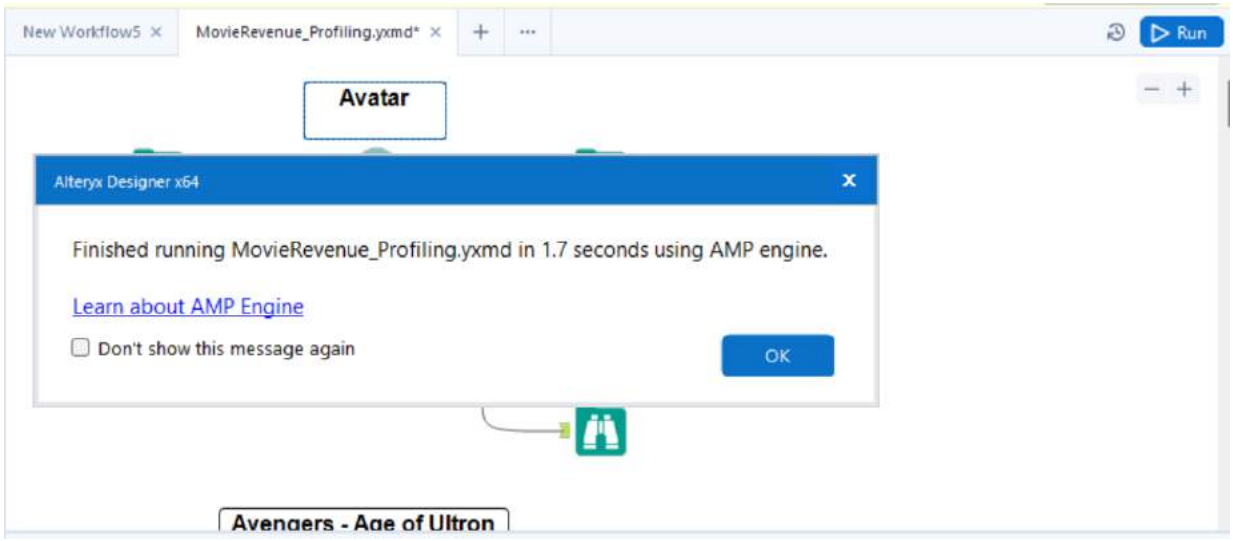
• Ok	210	100.00%
Unique	21	10.00%
• Null	0	0.00%
• Not Ok	0	0.00%
• Empty	0	0.00%

Length Statistics

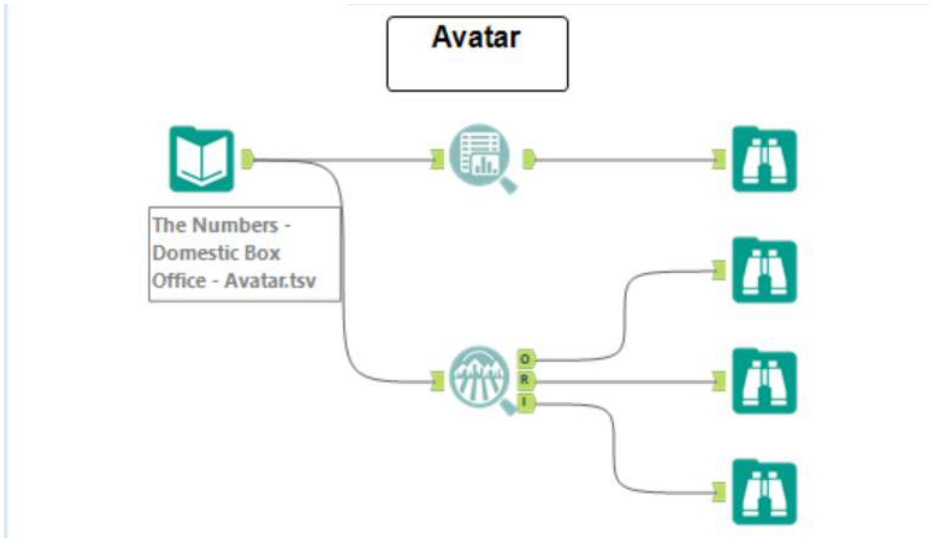
Min	3
Max	31
Average	12.40
Shortest Value	OKs
Longest Value	Values with Trailing Whitespace
First Alphanumeric Value	Average Length
Last Alphanumeric Value	Values with Trailing Whitespace
Blanks	0
Values with Leading Whitespace	0
Values with Trailing Whitespace	0



3. Movie Revenue Profiling:


New Workflow5
MovieRevenue_Profiling.yxmd*
Run
Avatar
Alteryx Designer x64
Finished running MovieRevenue_Profiling.yxmd in 1.7 seconds using AMP engine.
Learn about AMP Engine
Don't show this message again
OK
Avengers - Age of Ultron

3.1 Avatar


Avatar
The Numbers - Domestic Box Office - Avatar.tsv
O
R
I

Results - Browse (3) - Input

3 of 3 Fields | 231 records displayed, 4,789 bytes | Search

Record	FieldName	Name	Value
1	tconst	Name	tconst
2	tconst	Data Type	V_String
3	tconst	Size	254
4	tconst	Source	File: D:\dadabi\Final Project\IMDB\IMDB\TheNu...
5	tconst	Description	[Null]
6	tconst	OKs	318
7	tconst	Nulls	0
8	tconst	Non-Nulls	318
9	tconst	Blanks	0
10	tconst	Values with Leading Whitespace	0
11	tconst	Values with Trailing Whitespace	0
12	tconst	Values with Both Whitespace	0
13	tconst	Average Length	9.0
14	tconst	Longest Length	9
15	tconst	Longest Value	tt0499549
16	tconst	Shortest (Non Blank) Length	9

Profile

FieldName

Summary

Type	Records	Data Type Size
V_WString	**231**	**1,073,741,823**

Ok	231	100.00%
Unique	11	4.76%
Null	0	0.00%
Not Ok	0	0.00%
Empty	0	0.00%

Profile

Name ×

Summary

Type	Records	Data Type Size
V_WString	**231**	**1,073,741,823**

Ok	231	100.00%
Unique	21	9.09%
Null	0	0.00%
Not Ok	0	0.00%
Empty	0	0.00%

Profile

Value ×

Summary

Type	Records	Data Type Size
V_WString	**231**	**1,073,741,823**

Ok	209	90.48%
Unique	81	35.06%
Null	15	6.49%
Not Ok	7	3.03%
Empty	0	0.00%

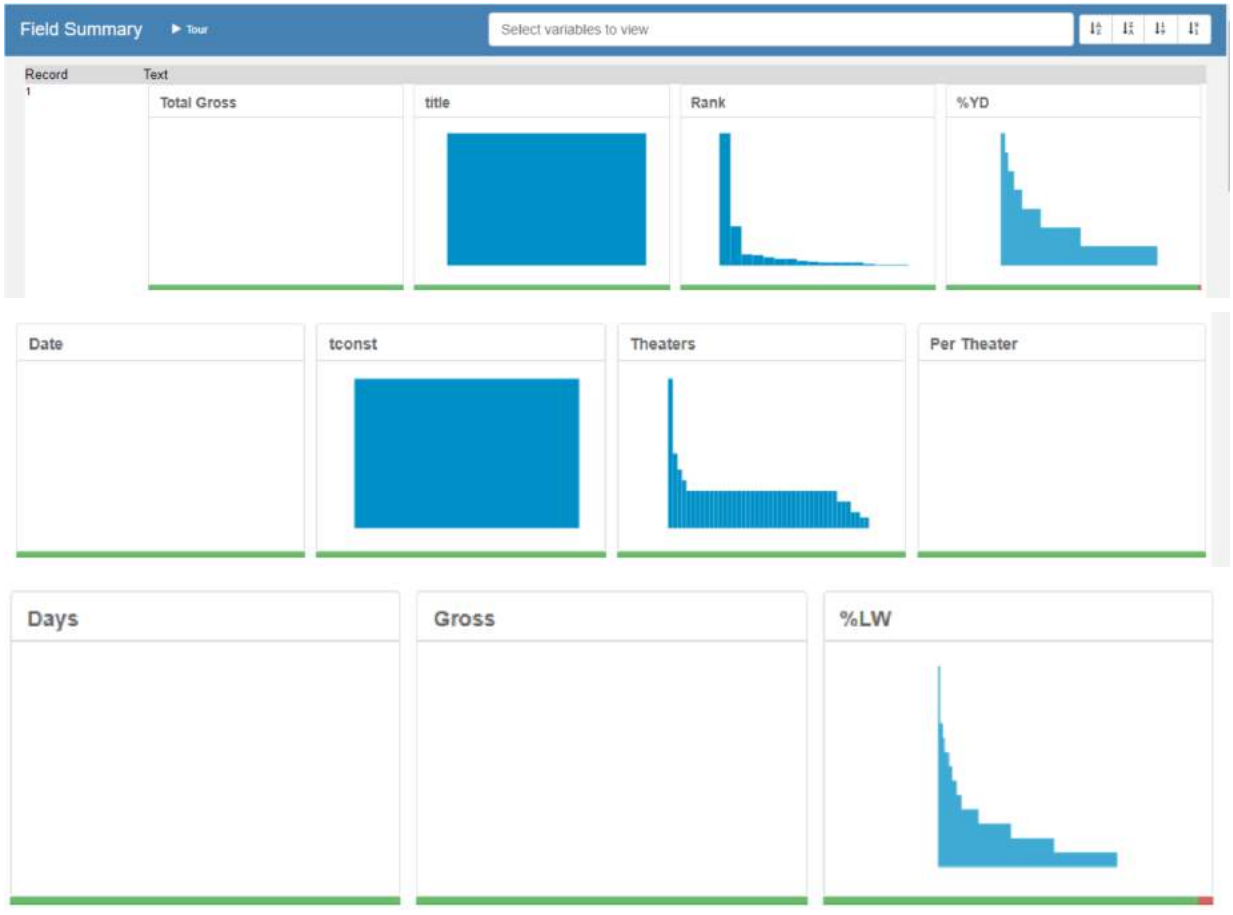
Report Profile

1 of 1 Fields | Records 1 to 1

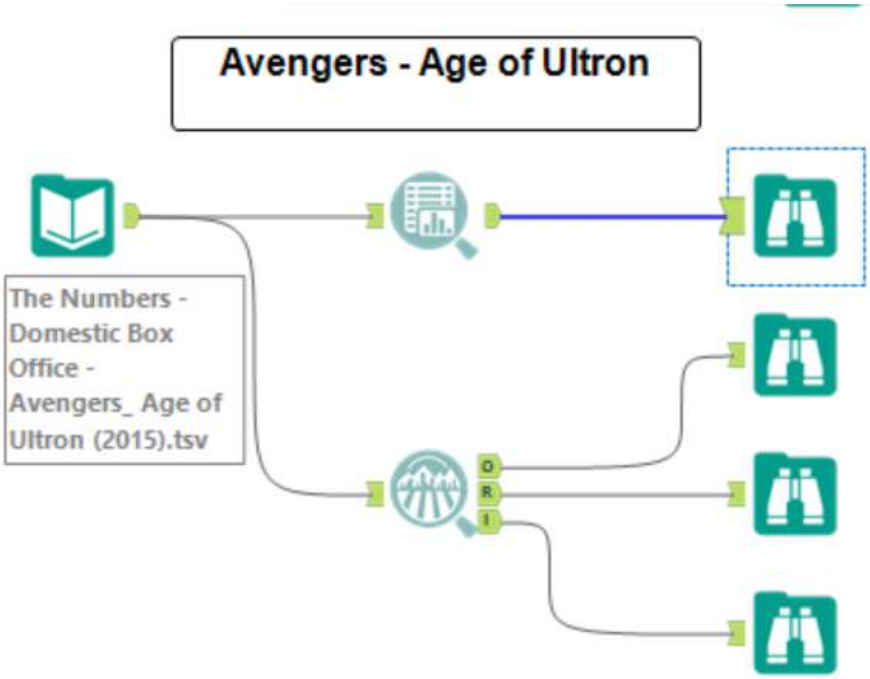
1

String/Character Fields

Name	% Missing	Unique Values	Shortest Value	Longest Value	Min Value Count	Max Value Count	Remarks
Total Gross	0.0%	318	$26,752,099	$109,497,762	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Rank	0.0%	17	1	10	1	176	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
%YD	0.6%	155	2	-34%	1	7	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Date	0.0%	318	Jan 1, 2010	Dec 18, 2009	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Theaters	0.0%	44	9	3,452	2	28	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Per Theater	0.0%	293	$60	$7,750	1	3	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Days	0.0%	318	1	100	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining
Gross	0.0%	317	$571	$26,752,099	1	2	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
%LW	4.4%	116	-2%	-14%	1	14	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
title	0.0%	1	Avatar	Avatar	318	318	
tconst	0.0%	1	tt0499549	tt0499549	318	318	



3.2 Avengers - Age of Ultron



3 of 3 Fields | 231 records displayed, 5,633 bytes | Search

Record	FieldName	Name	Value
1	tconst	Name	tconst
2	tconst	Data Type	V_String
3	tconst	Size	254
4	tconst	Source	File: D:\dadabi\Final Project\IMDB\IMDB\TheNu...
5	tconst	Description	[Null]
6	tconst	OKs	79
7	tconst	Nulls	0
8	tconst	Non-Nulls	79
9	tconst	Blanks	0
10	tconst	Values with Leading Whitespace	0
11	tconst	Values with Trailing Whitespace	0
12	tconst	Values with Both Whitespace	0
13	tconst	Average Length	9.0
14	tconst	Longest Length	9
15	tconst	Longest Value	tt2395427

Profile

FieldName ×

Summary

Type	Records	Data Type Size
V_WString	**231**	**1,073,741,823**

Ok	231	100.00%
Unique	11	4.76%
Null	0	0.00%
Not Ok	0	0.00%
Empty	0	0.00%

Profile

Name ×

Summary

Type	Records	Data Type Size
V_WString	**231**	**1,073,741,823**

Ok	231	100.00%
Unique	21	9.09%
Null	0	0.00%
Not Ok	0	0.00%
Empty	0	0.00%

Profile

Value

Summary

Type	Records	Data Type Size
V_WString	231	1,073,741,823

Ok	209	90.48%
Unique	80	34.63%
Null	11	4.76%
Not Ok	11	4.76%
Empty	0	0.00%

Report Profile

1 of 1 Fields | Records 1 to 1

1

String/Character Fields

Name	% Missing	Unique Values	Shortest Value	Longest Value	Min Value Count	Max Value Count	Remarks
Total Gross	0.0%	79	$27,600,000	$140,951,167	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Rank	0.0%	15	P	10	1	14	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
%YD	2.5%	55	-1%	-33%	1	5	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Date	0.0%	79	May 1, 2015	Apr 30, 2015	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Theaters	0.0%	13	0	4,276	1	21	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Per Theater	1.3%	78	$944	$19,744	1	2	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Days	1.3%	79	1	10	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Gross	0.0%	79	$69,841	$27,600,000	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
%LW	10.1%	39	-1%	-75%	1	8	This field has over 10% missing values. Consider imputing these values. Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
title	0.0%	1	Avengers: Age of Ultron	Avengers: Age of Ultron	79	79	
tconst	0.0%	1	tt2395427	tt2395427	79	79	

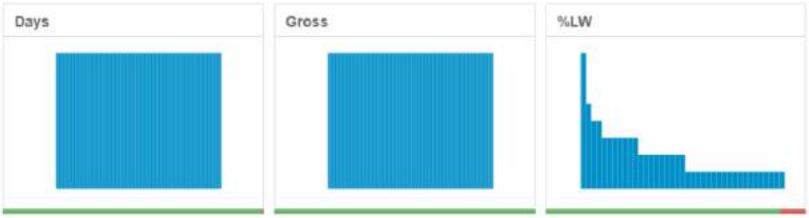
Field Summary ▶ Tour

Select variables to view

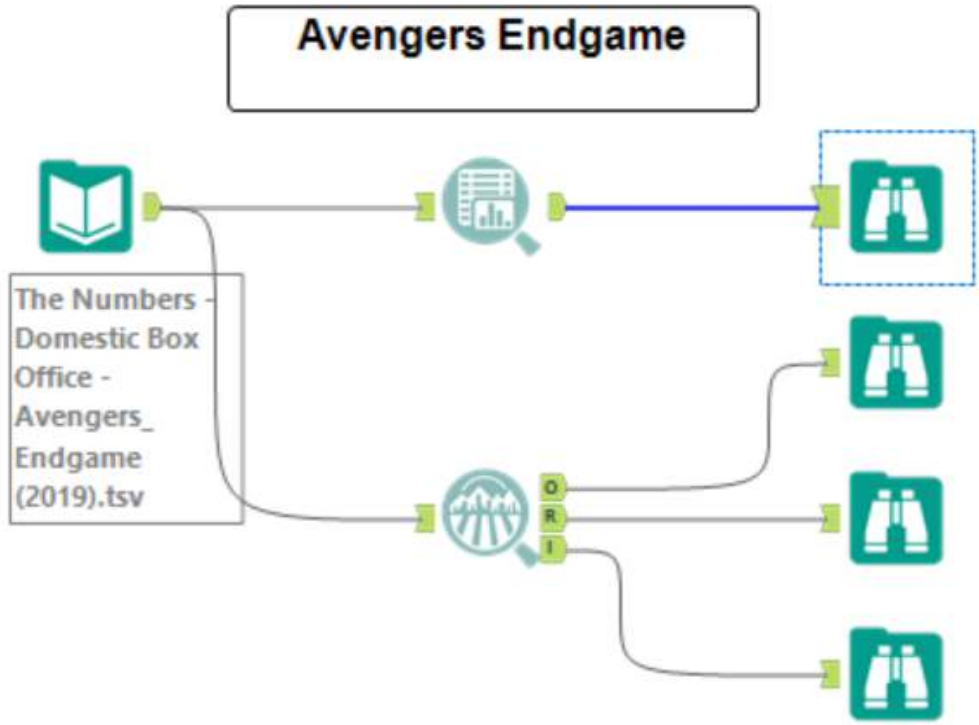
Record

1

Text



3.3 Avengers Endgame



3 of 3 Fields | 231 records displayed, 4,366 bytes | Search

Record	FieldName	Name	Value
1	tconst	Name	tconst
2	tconst	Data Type	V_String
3	tconst	Size	254
4	tconst	Source	File: D:\dadabi\Final Project\IMDB\IMDB\TheNu...
5	tconst	Description	[Null]
6	tconst	OKs	141
7	tconst	Nulls	0
8	tconst	Non-Nulls	141
9	tconst	Blanks	0
10	tconst	Values with Leading Whitespace	0
11	tconst	Values with Trailing Whitespace	0
12	tconst	Values with Both Whitespace	0
13	tconst	Average Length	9.0
14	tconst	Longest Length	9
15	tconst	Longest Value	tt4154796
16	tconst	Shortest (Non Blank) Length	9

Profile

FieldName ×

Summary

Type	Records	Data Type Size
V_WString	**231**	**1,073,741,823**

Ok	231	100.00%
Unique	11	4.76%
Null	0	0.00%
Not Ok	0	0.00%
Empty	0	0.00%

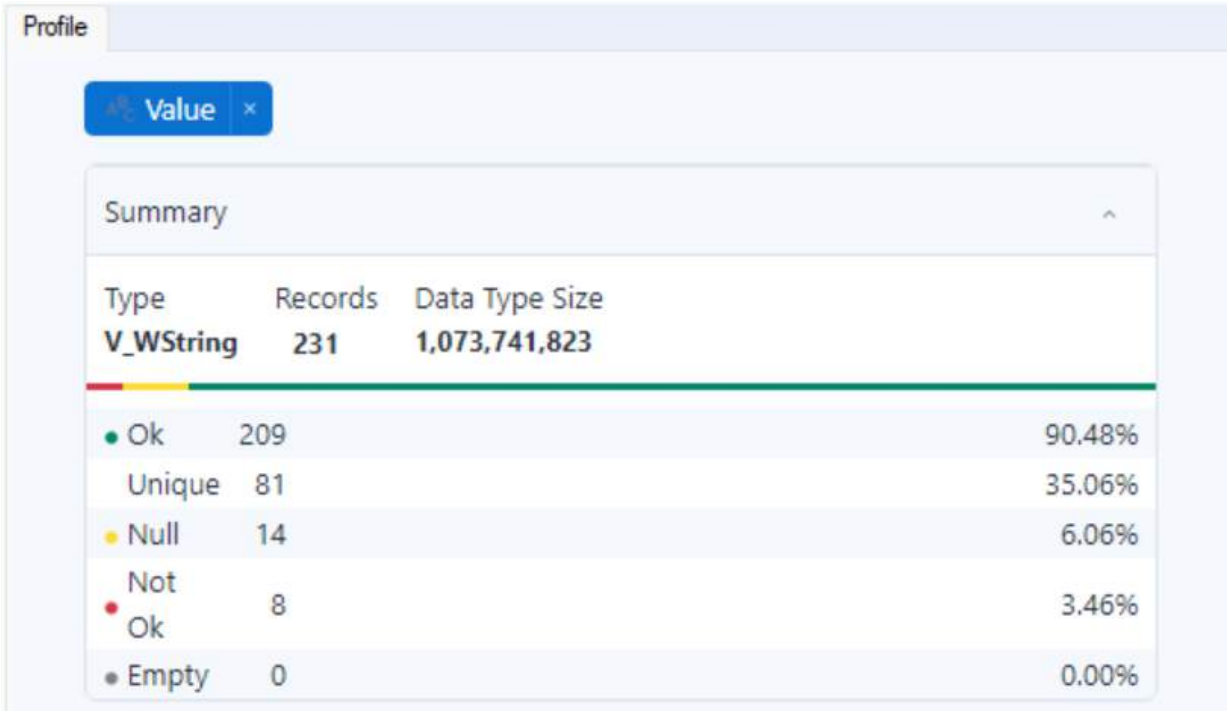
Profile

Name ×

Summary

Type	Records	Data Type Size
V_WString	**231**	**1,073,741,823**

Ok	231	100.00%
Unique	21	9.09%
Null	0	0.00%
Not Ok	0	0.00%
Empty	0	0.00%



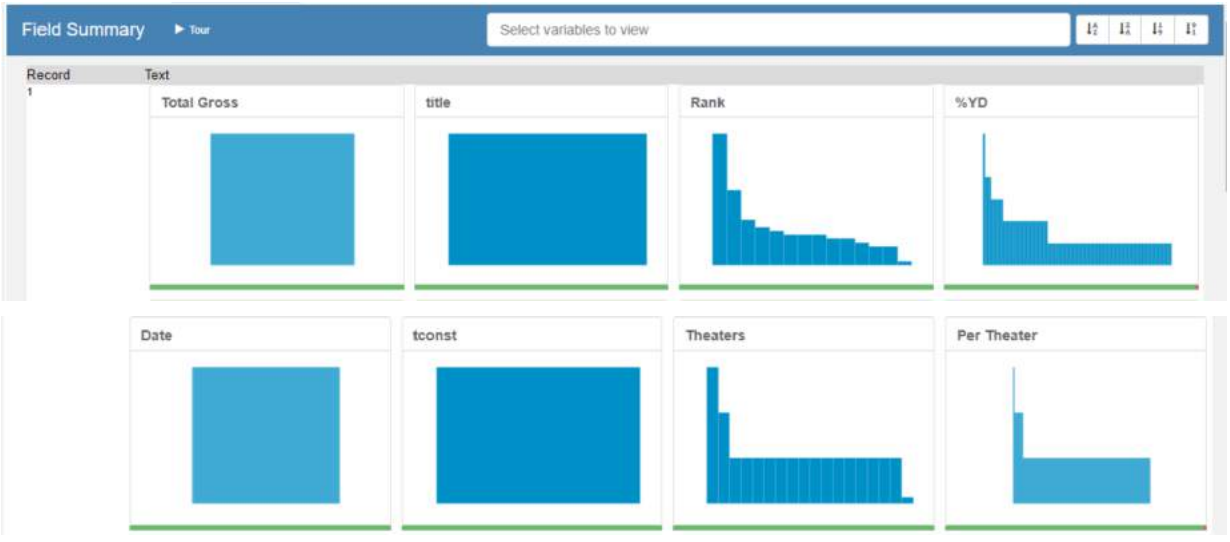
Report Profile

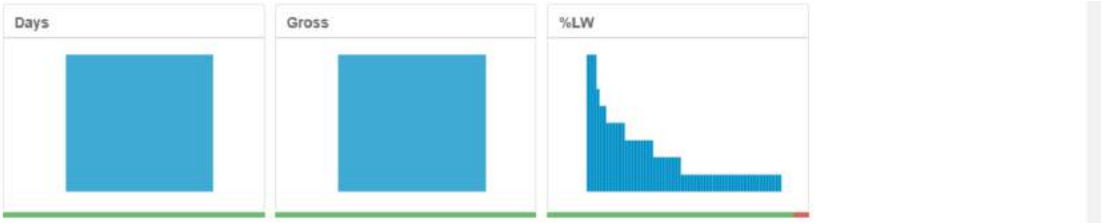
1 of 1 Fields | Records 1 to 1

1

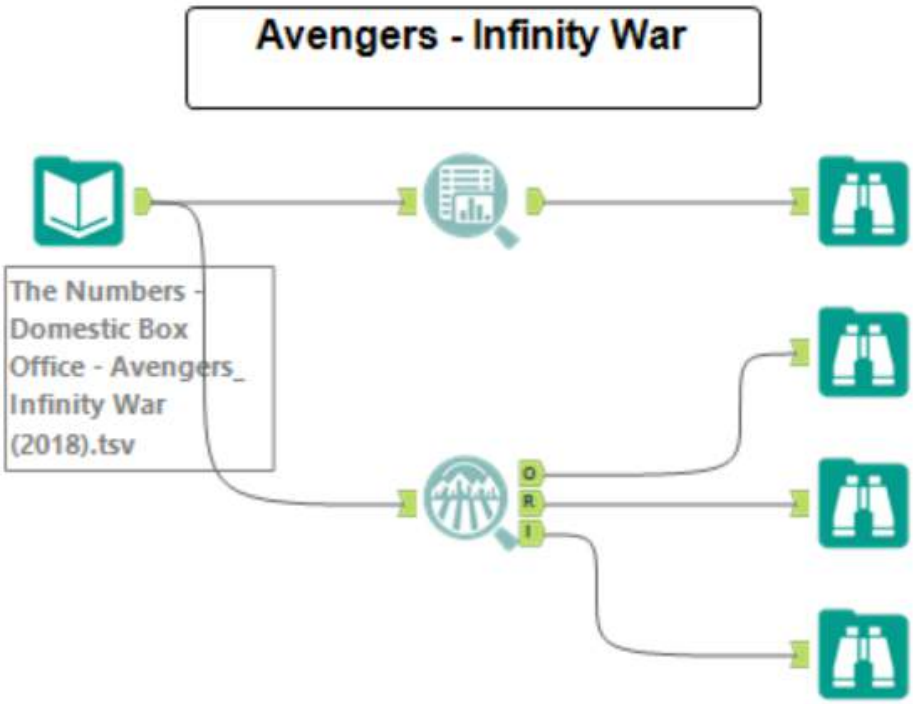
String/Character Fields

Name	% Missing	Unique Values	Shortest Value	Longest Value	Min Value Count	Max Value Count	Remarks
Total Gross	0.0%	141	$60,000,000	$157,461,641	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Rank	0.0%	14	P	12	1	35	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
%YD	1.4%	93	0.7	-31%	1	6	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Date	0.0%	141	May 1, 2019	Apr 25, 2019	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Theaters	0.0%	18	0	4,662	1	21	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Per Theater	0.7%	131	$91	$33,776	1	3	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Days	0.0%	141	0	100	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Gross	0.0%	141	$940	$157,461,641	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
%LW	5.7%	62	-5%	-74%	1	8	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
title	0.0%	1	Avengers: Endgame	Avengers: Endgame	141	141	
tconst	0.0%	1	tt4154796	tt4154796	141	141	





3.4 Avengers - Infinity War



3 of 3 Fields | Cell Viewer | 231 records displayed, 4,360 bytes | Search | Data

Record	FieldName	Name	Value
1	tconst	Name	tconst
2	tconst	Data Type	V_String
3	tconst	Size	254
4	tconst	Source	File: D:\dadabi\Final Project\IMDB\IMDB\TheNu...
5	tconst	Description	[Null]
6	tconst	OKs	141
7	tconst	Nulls	0
8	tconst	Non-Nulls	141
9	tconst	Blanks	0
10	tconst	Values with Leading Whitespace	0
11	tconst	Values with Trailing Whitespace	0
12	tconst	Values with Both Whitespace	0
13	tconst	Average Length	9.0
14	tconst	Longest Length	9
15	tconst	Longest Value	tt4154756
16	tconst	Shortest (Non Blank) Length	9

Profile

FieldName ×

Summary

Type	Records	Data Type Size
V_WString	**231**	**1,073,741,823**

• Ok	231	100.00%
Unique	11	4.76%
• Null	0	0.00%
• Not Ok	0	0.00%
• Empty	0	0.00%

Profile

Name ×

Summary

Type	Records	Data Type Size
V_WString	**231**	**1,073,741,823**

• Ok	231	100.00%
Unique	21	9.09%
• Null	0	0.00%
• Not Ok	0	0.00%
• Empty	0	0.00%

Profile

Value

Summary

Type	Records	Data Type Size
V_WString	231	1,073,741,823

Ok	209	90.48%
Unique	81	35.06%
Null	14	6.06%
Not Ok	8	3.46%
Empty	0	0.00%

Report Profile

1 of 1 Fields | Records 1 to 1

1

String/Character Fields

Name	% Missing	Unique Values	Shortest Value	Longest Value	Min Value Count	Max Value Count	Remarks
Total Gross	0.0%	141	$39,000,000	$106,334,939	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Rank	0.0%	14	P	12	1	61	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
%YD	1.4%	100	1	-23%	1	4	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Date	0.0%	141	May 1, 2018	Apr 26, 2018	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Theaters	0.0%	18	92	4,474	7	22	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Per Theater	0.0%	132	$68	$23,767	1	3	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Days	0.0%	141	0	100	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Gross	0.0%	141	$9,993	$106,334,939	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
%LW	5.7%	58	-6%	-70%	1	8	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
title	0.0%	1	Avengers: Infinity War	Avengers: Infinity War	141	141	
tconst	0.0%	1	tt4154756	tt4154756	141	141	

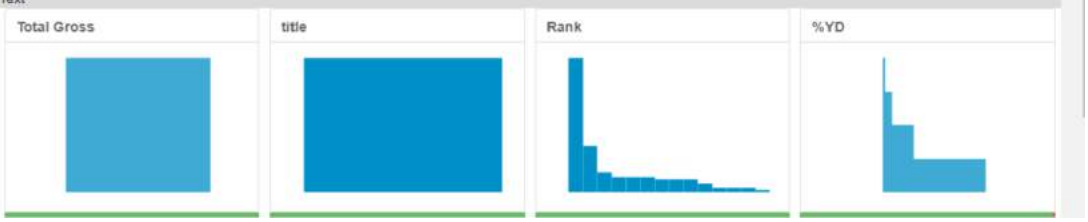
Field Summary ▶ Tour

Select variables to view

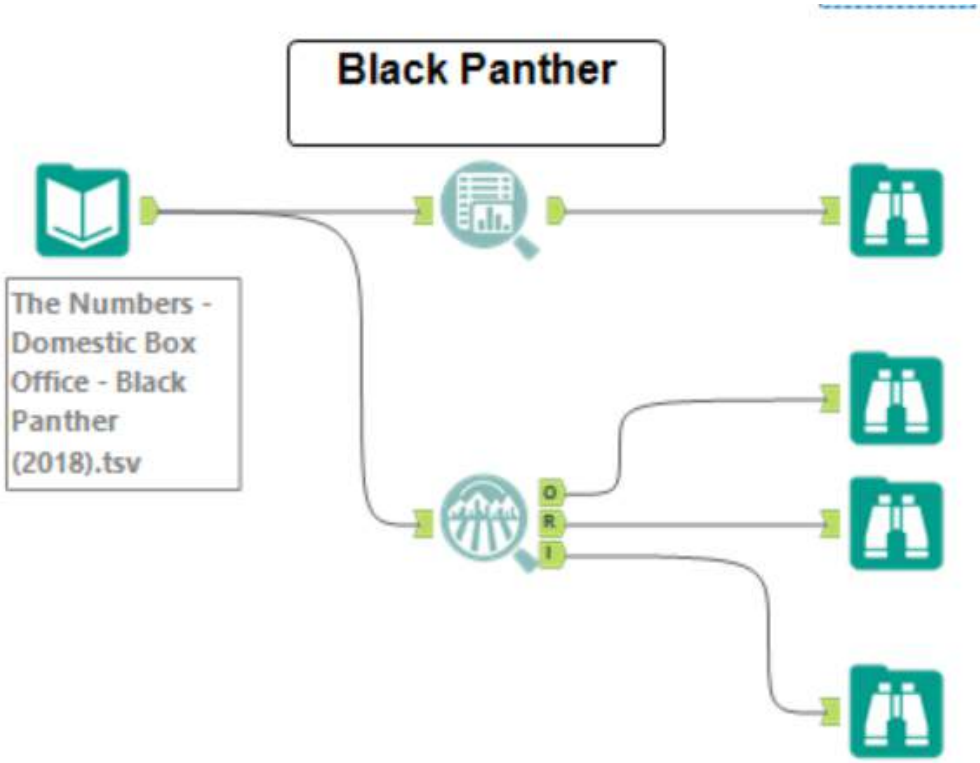
Record

1

Text



3.5 Black Panther



3 of 3 Fields | 231 records displayed, 4,679 bytes | Search | Data

Record	FieldName	Name	Value
1	tconst	Name	tconst
2	tconst	Data Type	V_String
3	tconst	Size	254
4	tconst	Source	File: D:\dadabi\Final Project\IMDB\IMDB\TheNu...
5	tconst	Description	[Null]
6	tconst	OKs	176
7	tconst	Nulls	0
8	tconst	Non-Nulls	176
9	tconst	Blanks	0
10	tconst	Values with Leading Whitespace	0
11	tconst	Values with Trailing Whitespace	0
12	tconst	Values with Both Whitespace	0
13	tconst	Average Length	9.0
14	tconst	Longest Length	9
15	tconst	Longest Value	tt1825683
16	tconst	Shortest (Non Blank) Length	9

Profile

FieldName

Summary

Type	Records	Data Type Size
V_WString	231	1,073,741,823

Ok	231	100.00%
Unique	11	4.76%
Null	0	0.00%
Not Ok	0	0.00%
Empty	0	0.00%

Profile

Name

Summary

Type	Records	Data Type Size
V_WString	231	1,073,741,823

Ok	231	100.00%
Unique	21	9.09%
Null	0	0.00%
Not Ok	0	0.00%
Empty	0	0.00%

Profile

Value

Summary

Type	Records	Data Type Size
V_WString	231	1,073,741,823

Ok	207	89.61%
Unique	85	36.80%
Null	16	6.93%
Not Ok	8	3.46%
Empty	0	0.00%

Report Profile

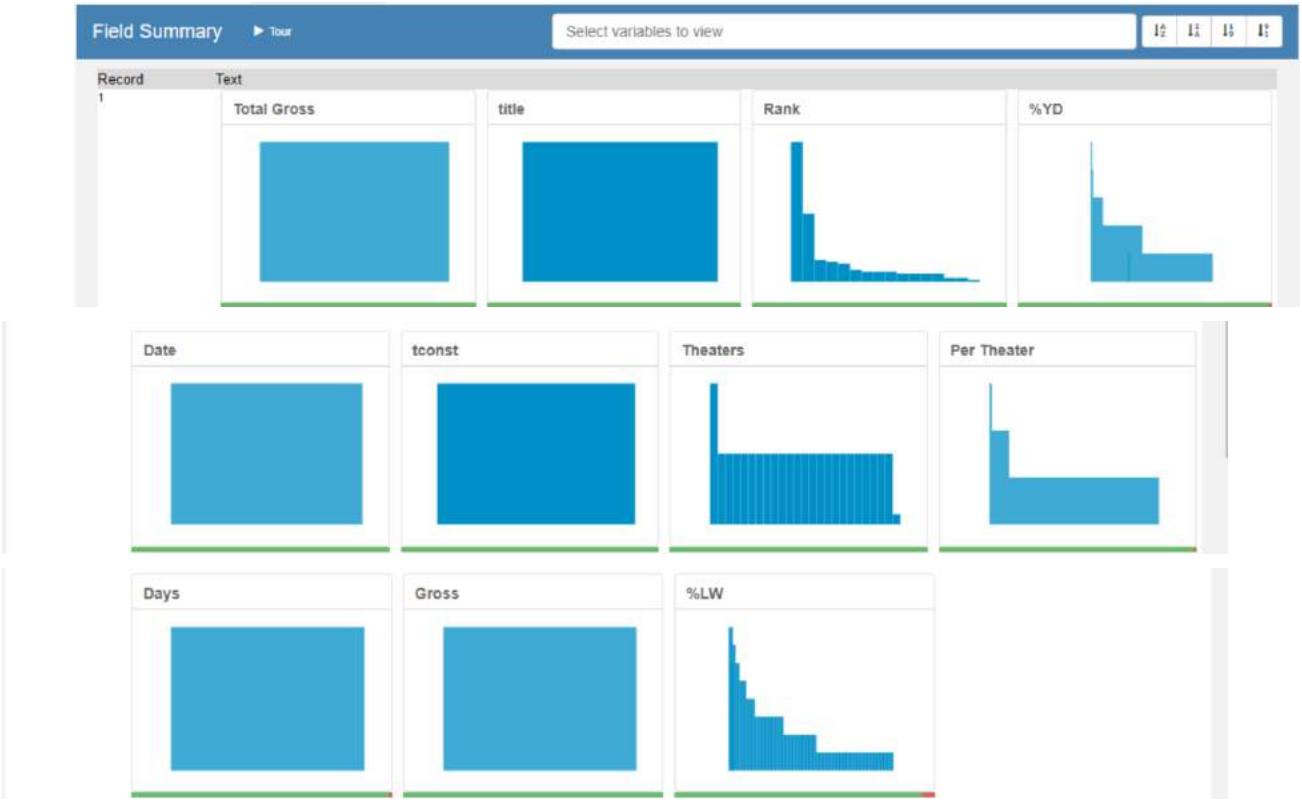
1 of 1 Fields | Records 1 to 1

1

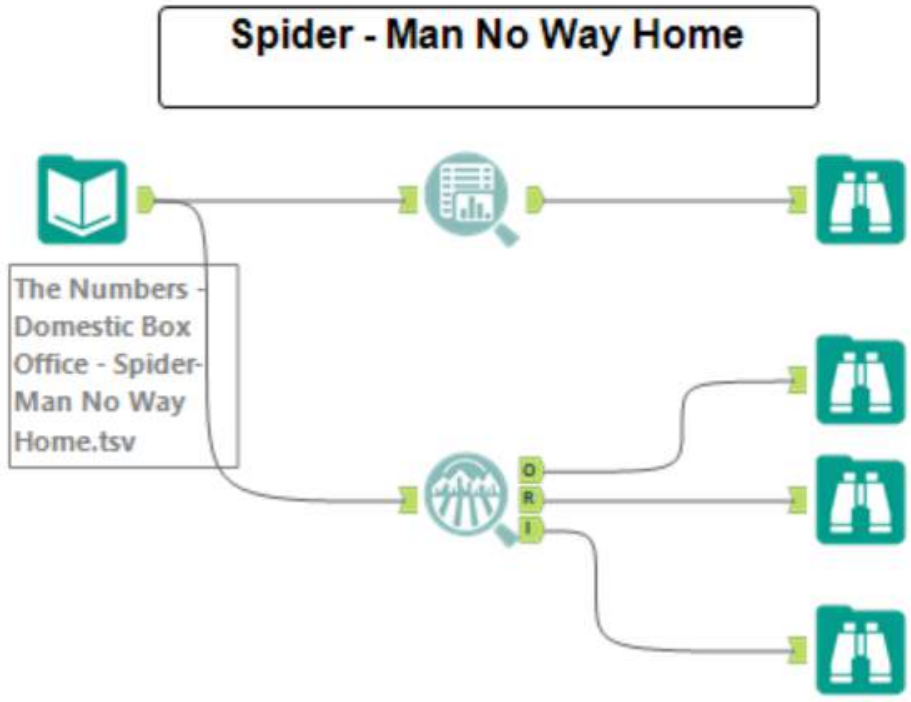
String/Character Fields

Name	% Missing	Unique Values	Shortest Value	Longest Value	Min Value Count	Max Value Count	Remarks
Total Gross	0.0%	176	$25,200,000	$141,936,512	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Rank	0.0%	16	P	10	1	70	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
%YD	1.1%	114		-13%	1	5	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Date	0.0%	176	Mar 1, 2018	Feb 15, 2018	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Theaters	0.0%	25	0	4,020	1	14	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Per Theater	0.6%	156	$79	$18,891	1	3	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Days	0.6%	176	1	100	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.

Gross	0.0%	176	$780	$25,200,000	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
%LW	4.5%	76		-62%	1	8	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
title	0.0%	1	Black Panther	Black Panther	176	176	
tconst	0.0%	1	tt1825683	tt1825683	176	176	



3.6 Spider Man - No Way Home



3 of 3 Fields | 231 records displayed, 4,233 bytes | Search

Record	FieldName	Name	Value
1	tcont	Name	tcont
2	tcont	Data Type	V_String
3	tcont	Size	254
4	tcont	Source	File: D:\dadabi\Final Project\IMDB\IMDB\TheNu...
5	tcont	Description	[Null]
6	tcont	OKs	128
7	tcont	Nulls	0
8	tcont	Non-Nulls	128
9	tcont	Blanks	0
10	tcont	Values with Leading Whitespace	0
11	tcont	Values with Trailing Whitespace	0
12	tcont	Values with Both Whitespace	0
13	tcont	Average Length	10.0
14	tcont	Longest Length	10
15	tcont	Longest Value	tt10872600
16	tcont	Shortest (Non Blank) Length	10

Profile

FieldName ×

Summary

Type	Records	Data Type Size
V_WString	**231**	**1,073,741,823**

Ok	231	100.00%
Unique	11	4.76%
Null	0	0.00%
Not Ok	0	0.00%
Empty	0	0.00%

Profile

Name ×

Summary

Type	Records	Data Type Size
V_WString	**231**	**1,073,741,823**

Ok	231	100.00%
Unique	21	9.09%
Null	0	0.00%
Not Ok	0	0.00%
Empty	0	0.00%

Profile

Value ×

Summary

Type	Records	Data Type Size
V_WString	**231**	**1,073,741,823**

Ok	209	90.48%
Unique	83	35.93%
Null	14	6.06%
Not Ok	8	3.46%
Empty	0	0.00%

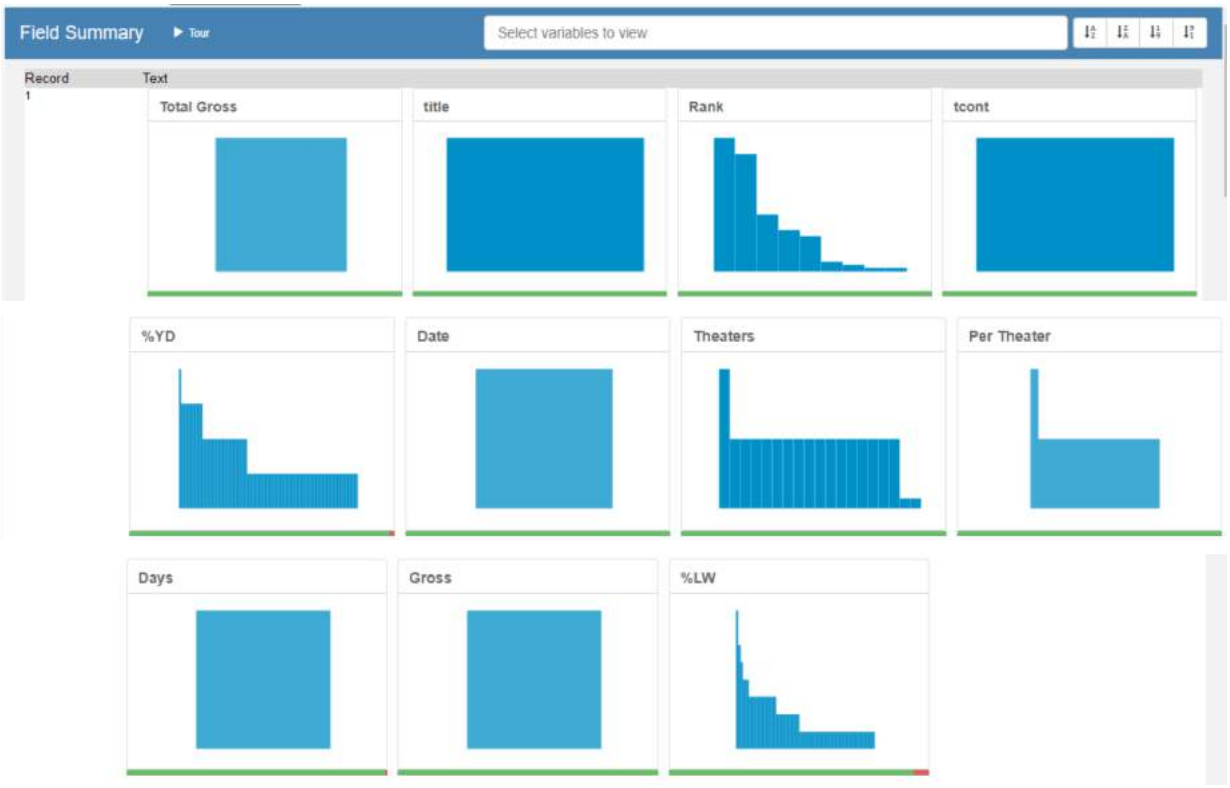
Report Profile

1 of 1 Fields | Records 1 to 1

1

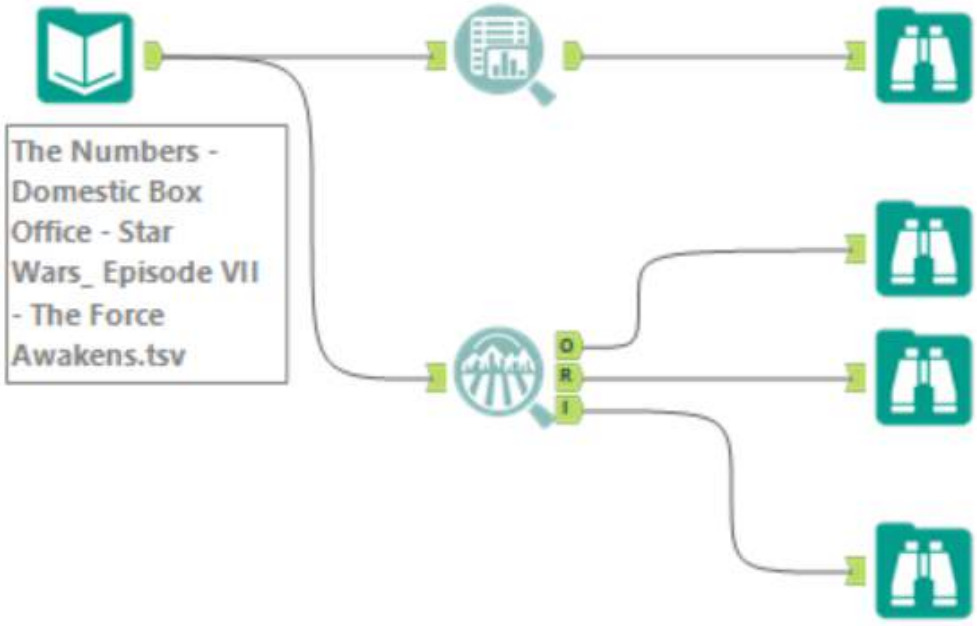
String/Character Fields

Name	% Missing	Unique Values	Shortest Value	Longest Value	Min Value Count	Max Value Count	Remarks
Total Gross	0.0%	128	$50,000,000	$121,964,712	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Rank	0.0%	9	P	P	1	42	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
%YD	1.6%	84	1	-39%	1	4	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Date	0.0%	128	Jan 1, 2022	Dec 16, 2021	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Theaters	0.0%	19	387	3,767	1	14	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Per Theater	0.0%	121	$95	$13,273	1	2	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Days	0.8%	128	1	100	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Gross	0.0%	128	$98,951	$121,964,712	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
%LW	6.2%	66	-6%	-84%	1	8	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
title	0.0%	1	Spider-Man: No Way Home	Spider-Man: No Way Home	128	128	
tcont	0.0%	1	tt10872600	tt10872600	128	128	



3.7 Star Wars Episode VII - The Force Awakens

Star Wars Episode VII - The Force Awakens



3 of 3 Fields | 231 records displayed, 4,237 bytes | Search | Data

Record	FieldName	Name	Value
1	tconst	Name	tconst
2	tconst	Data Type	V_String
3	tconst	Size	254
4	tconst	Source	File: D:\dadabi\Final Project\IMDB\IMDB\TheNu...
5	tconst	Description	[Null]
6	tconst	OKs	120
7	tconst	Nulls	0
8	tconst	Non-Nulls	120
9	tconst	Blanks	0
10	tconst	Values with Leading Whitespace	0
11	tconst	Values with Trailing Whitespace	0
12	tconst	Values with Both Whitespace	0
13	tconst	Average Length	9.0
14	tconst	Longest Length	9
15	tconst	Longest Value	tt2488496
16	tconst	Shortest (Non Blank) Length	9

Profile

FieldName ×

Summary

Type	Records	Data Type Size
V_WString	**231**	**1,073,741,823**

Ok	231	100.00%
Unique	11	4.76%
Null	0	0.00%
Not Ok	0	0.00%
Empty	0	0.00%

Profile

Name ×

Summary

Type	Records	Data Type Size
V_WString	**231**	**1,073,741,823**

Ok	231	100.00%
Unique	21	9.09%
Null	0	0.00%
Not Ok	0	0.00%
Empty	0	0.00%

Profile

Value

Summary

Type	Records	Data Type Size
V_WString	231	1,073,741,823

Ok	209	90.48%
Unique	81	35.06%
Null	14	6.06%
Not Ok	8	3.46%
Empty	0	0.00%

1 of 1 Fields | Records 1 to 1

1

String/Character Fields

Name	% Missing	Unique Values	Shortest Value	Longest Value	Min Value Count	Max Value Count	Remarks
Total Gross	0.0%	120	$57,000,000	$119,119,282	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Rank	0.0%	17	P	11	1	25	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
%YD	1.7%	89	-7%	-43%	1	3	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Date	0.0%	120	Jan 1, 2016	Dec 17, 2015	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Theaters	0.0%	16	0	4,134	1	28	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Per Theater	0.8%	113	$72	$28,815	1	2	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Days	0.0%	120	0	100	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Gross	0.0%	120	$56,853	$119,119,282	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
%LW	6.7%	62	-5%	-59%	1	8	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
title	0.0%	1	Star Wars: Episode VII - The Force Awakens	Star Wars: Episode VII - The Force Awakens	120	120	
tconst	0.0%	1	tt2488496	tt2488496	120	120	

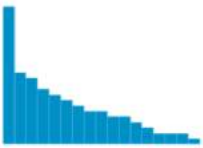
Field Summary | Tour | Select variables to view

Record | Text

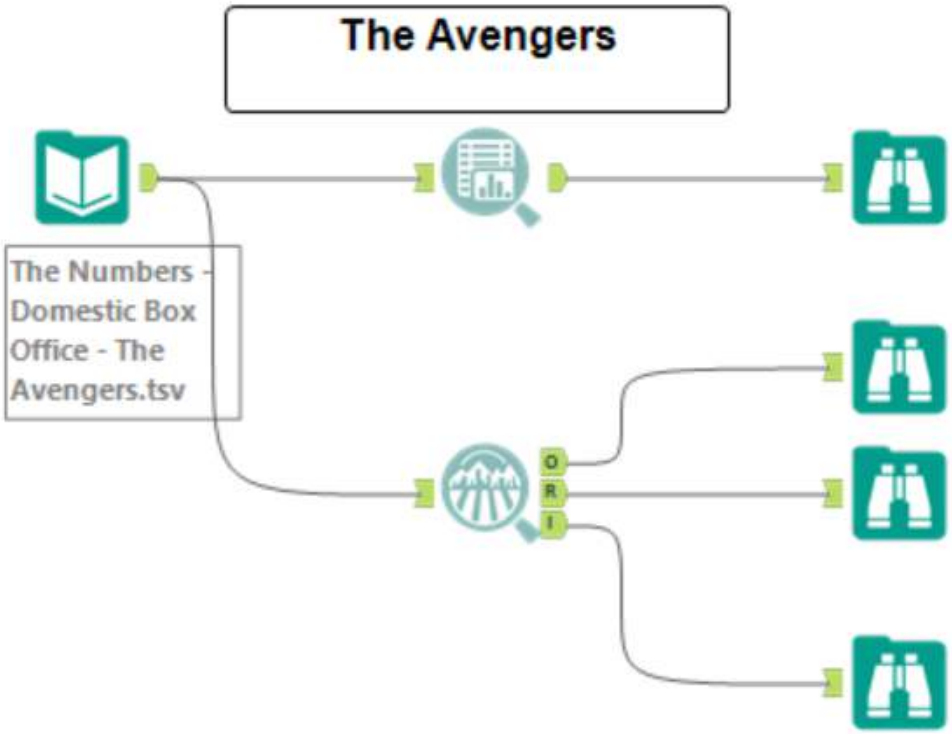
1

Total Gross | title | Rank | %YD






3.8 The Avengers



3 of 3 Fields | 231 records displayed, 4,812 bytes | Search

Record	FieldName	Name	Value
1	tconst	Name	tconst
2	tconst	Data Type	V_String
3	tconst	Size	254
4	tconst	Source	File: D:\dadabi\Final Project\IMDB\IMDB\TheNu...
5	tconst	Description	[Null]
6	tconst	OKs	97
7	tconst	Nulls	0
8	tconst	Non-Nulls	97
9	tconst	Blanks	0
10	tconst	Values with Leading Whitespace	0
11	tconst	Values with Trailing Whitespace	0
12	tconst	Values with Both Whitespace	0
13	tconst	Average Length	9.0
14	tconst	Longest Length	9
15	tconst	Longest Value	tt0848228
16	tconst	Shortest (Non Blank) Length	9

Profile

FieldName ×

Summary

Type	Records	Data Type Size
V_WString	**231**	**1,073,741,823**

Ok	231	100.00%
Unique	11	4.76%
Null	0	0.00%
Not Ok	0	0.00%
Empty	0	0.00%

Profile

Name ×

Summary

Type	Records	Data Type Size
V_WString	**231**	**1,073,741,823**

● Ok	231		100.00%
Unique	21		9.09%
● Null	0		0.00%
● Not Ok	0		0.00%
● Empty	0		0.00%

Profile

Value ×

Summary

Type	Records	Data Type Size
V_WString	**231**	**1,073,741,823**

● Ok	209		90.48%
Unique	77		33.33%
● Null	13		5.63%
● Not Ok	9		3.90%
● Empty	0		0.00%

Report Profile

1 of 1 Fields | Records 1 to 1

1

String/Character Fields

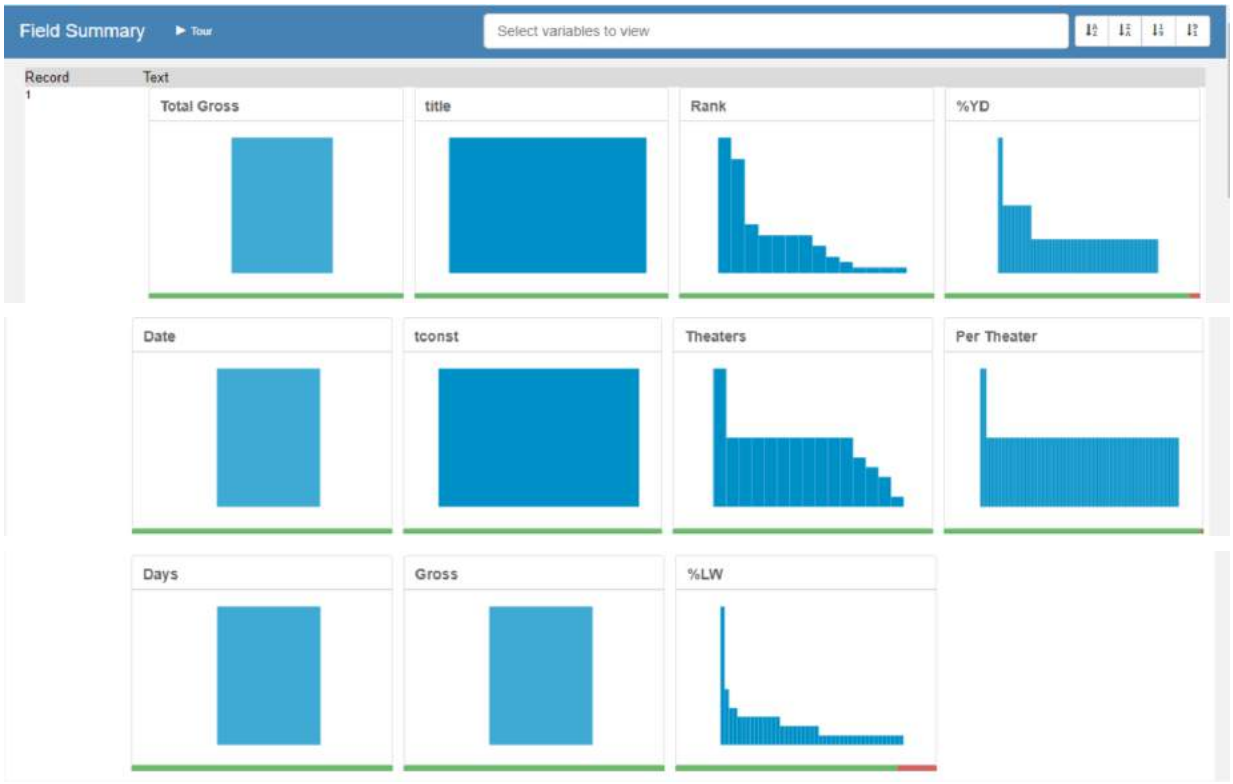
Name	% Missing	Unique Values	Shortest Value	Longest Value	Min Value Count	Max Value Count	Remarks
Total Gross	0.0%	97	$18,700,000	$150,371,975	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Rank	0.0%	14	P	11	1	25	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
%YD	4.1%	77	-6%	-14%	1	4	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Date	0.0%	97	May 3, 2012	May 10, 2012	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Theaters	0.0%	15	0	4,349	1	14	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Per Theater	1.0%	94	$56	$18,582	1	2	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Days	0.0%	97	0	120	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Gross	0.0%	97	$98,364	$18,700,000	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
%LW	15.5%	43	n/c	-64%	1	15	This field has over 10% missing values. Consider imputing these values. Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
title	0.0%	1	The Avengers	The Avengers	97	97	
tconst	0.0%	1	tt0848228	tt0848228	97	97	

Field Summary ▶ Tour

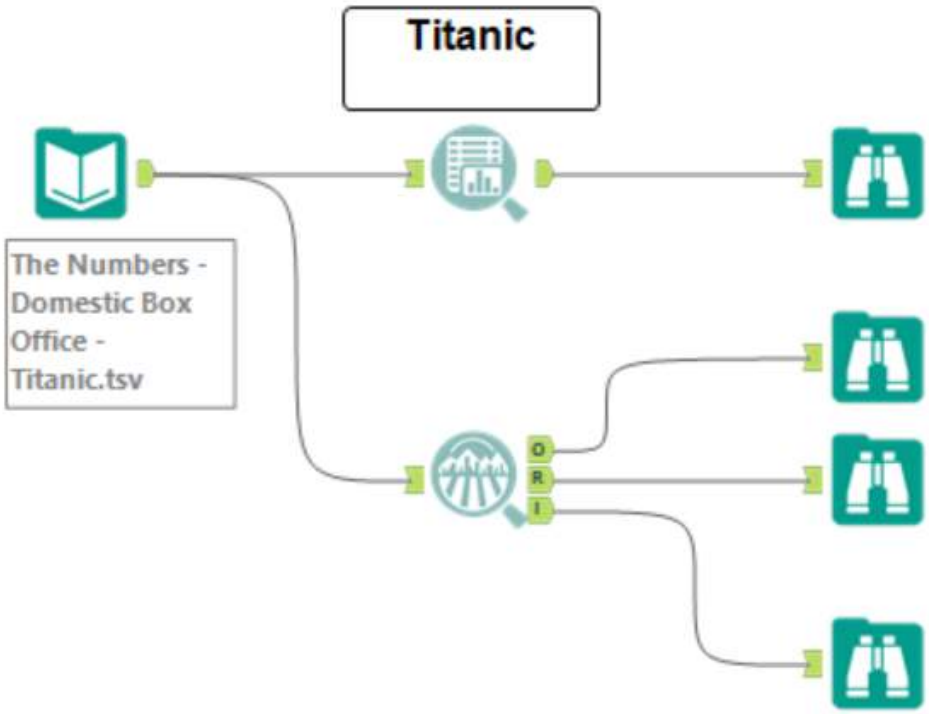
Select variables to view

Record Text

1



3.9 Titanic



3 of 3 Fields | 231 records displayed, 4,570 bytes | Search

Record	FieldName	Name	Value
1	tconst	Name	tconst
2	tconst	Data Type	V_String
3	tconst	Size	254
4	tconst	Source	File: D:\dadabi\Final Project\IMDB\IMDB\TheNu...
5	tconst	Description	[Null]
6	tconst	OKs	265
7	tconst	Nulls	0
8	tconst	Non-Nulls	265
9	tconst	Blanks	0
10	tconst	Values with Leading Whitespace	0
11	tconst	Values with Trailing Whitespace	0
12	tconst	Values with Both Whitespace	0
13	tconst	Average Length	9.0
14	tconst	Longest Length	9
15	tconst	Longest Value	tt0120338
16	tconst	Shortest (Non Blank) Length	9

Profile

FieldName ×

Summary

Type	Records	Data Type Size
V_WString	**231**	**1,073,741,823**

• Ok	231		100.00%
Unique	11		4.76%
• Null	0		0.00%
• Not Ok	0		0.00%
• Empty	0		0.00%

Profile

Name ×

Summary

Type	Records	Data Type Size
V_WString	**231**	**1,073,741,823**

• Ok	231		100.00%
Unique	21		9.09%
• Null	0		0.00%
• Not Ok	0		0.00%
• Empty	0		0.00%

Profile

Value ×

Summary

Type	Records	Data Type Size
V_WString	**231**	**1,073,741,823**

• Ok	209		90.48%
Unique	84		36.36%
• Null	15		6.49%
• Not Ok	7		3.03%
• Empty	0		0.00%

Report | Profile

1 of 1 Fields | Records 1 to 1

1

String/Character Fields

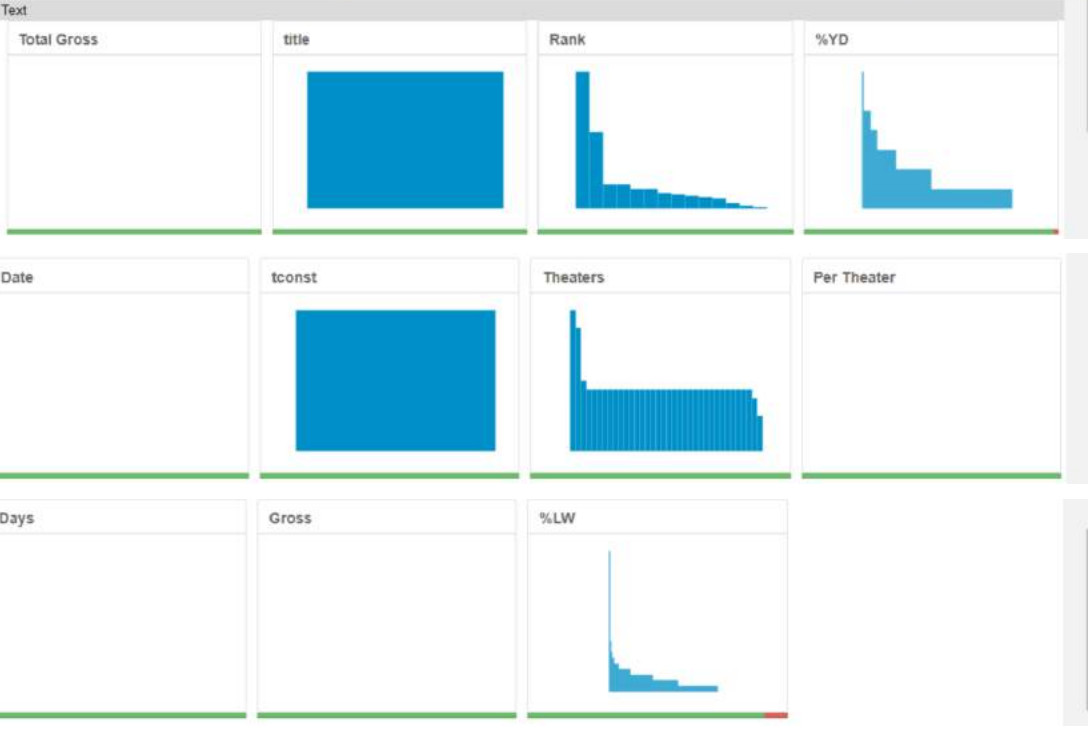
Name	% Missing	Unique Values	Shortest Value	Longest Value	Min Value Count	Max Value Count	Remarks
Total Gross	0.0%	265	$8,658,814	$104,380,438	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Rank	0.0%	14	1	11	1	97	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
%YD	1.5%	143	-7%	0.23	1	7	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Date	0.0%	265	Jan 1, 1998	Dec 19, 1997	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Theaters	0.0%	36	70	2,674	4	16	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Per Theater	0.0%	228	$88	$3,238	1	3	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Days	0.0%	265	1	5,221	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Gross	0.0%	264	$8,128	$10,672,013	1	2	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
%LW	9.4%	102	0.4	0.17	1	25	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
title	0.0%	1	Titanic	Titanic	265	265	
tconst	0.0%	1	tt0120338	tt0120338	265	265	

Field Summary

▶ Tour

Select variables to view

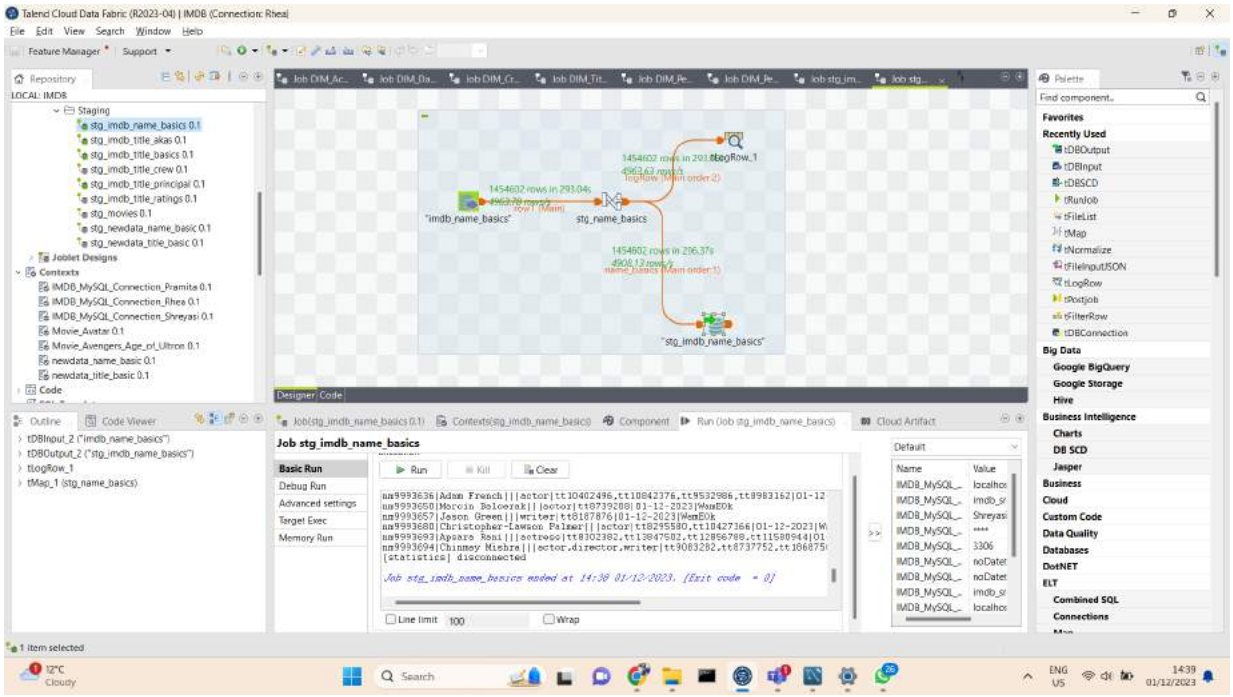
Record	Text
1	

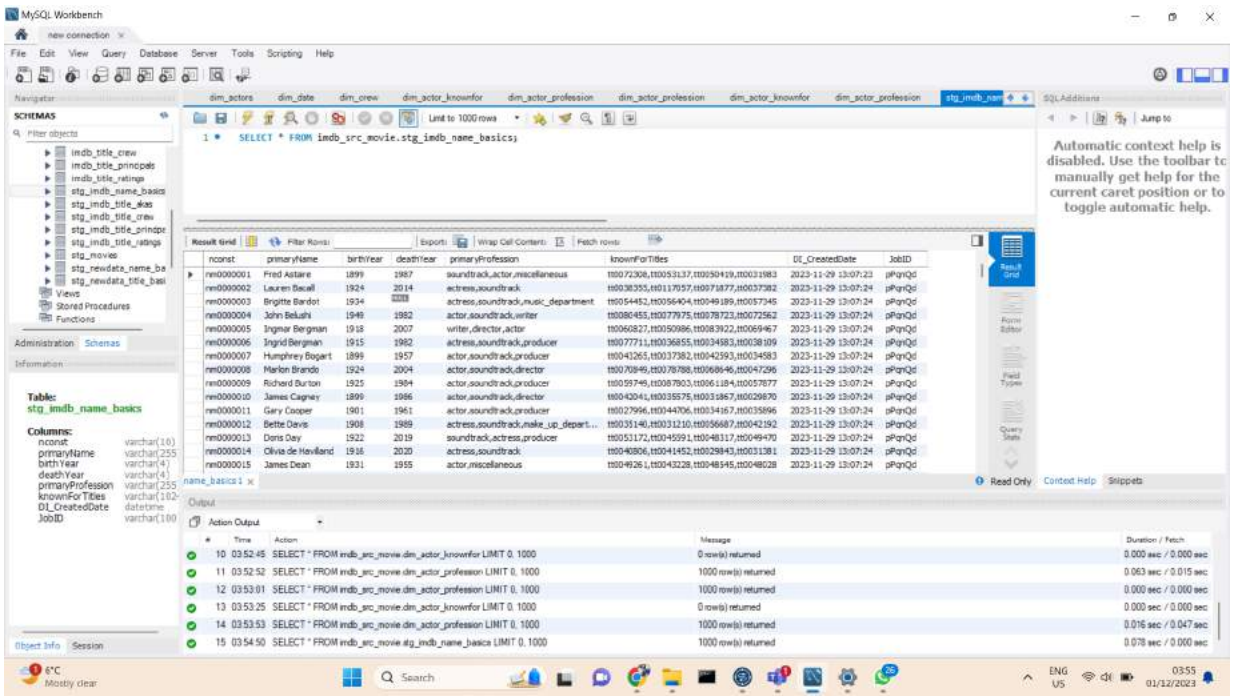


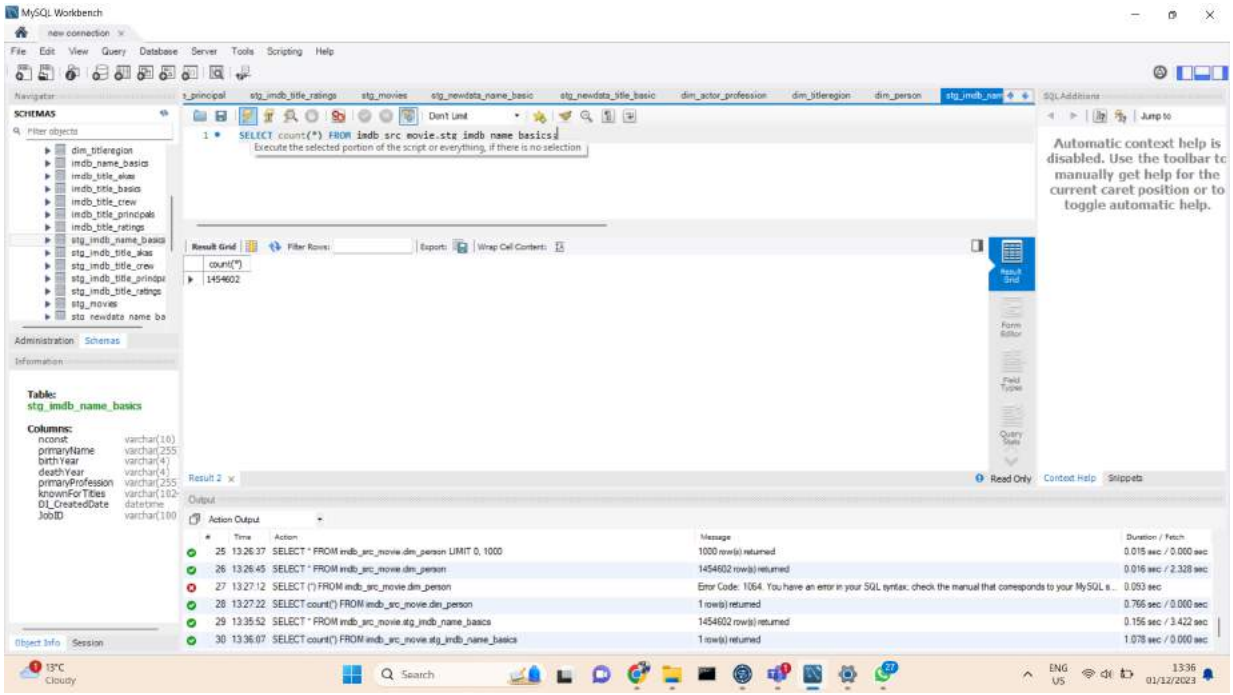
Jobs and Screenshots of Tables:

Stage Tables:

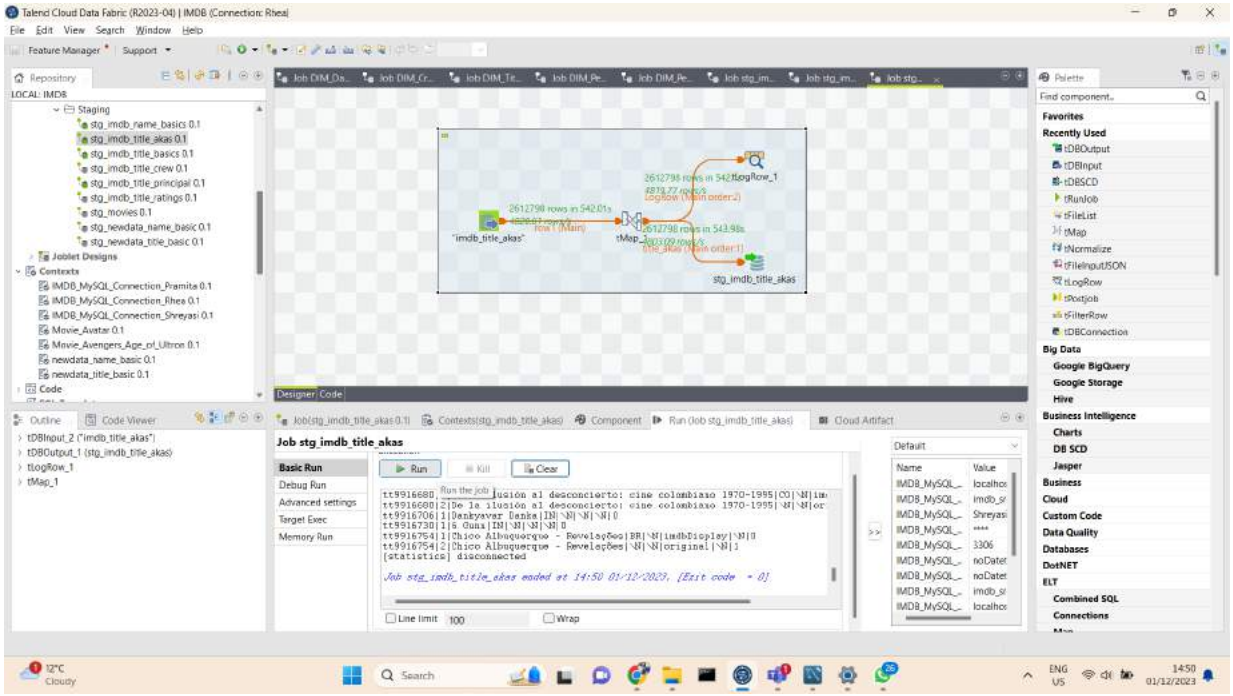
stg_imdb_name_basics:

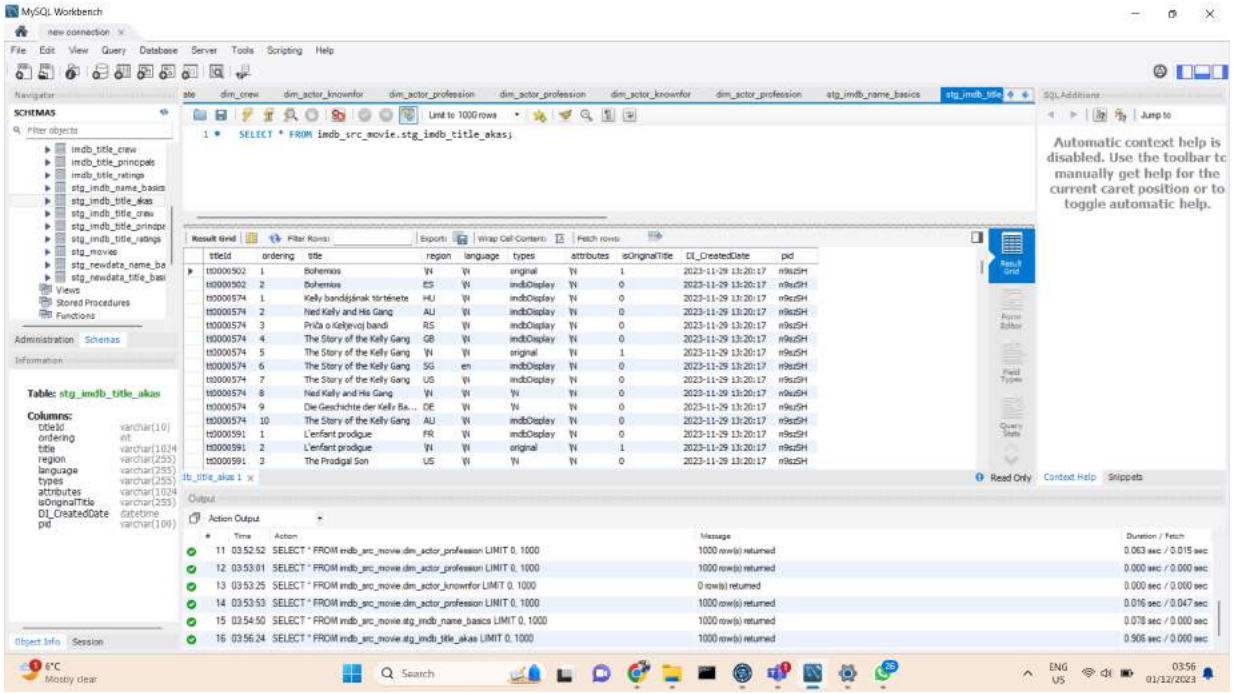


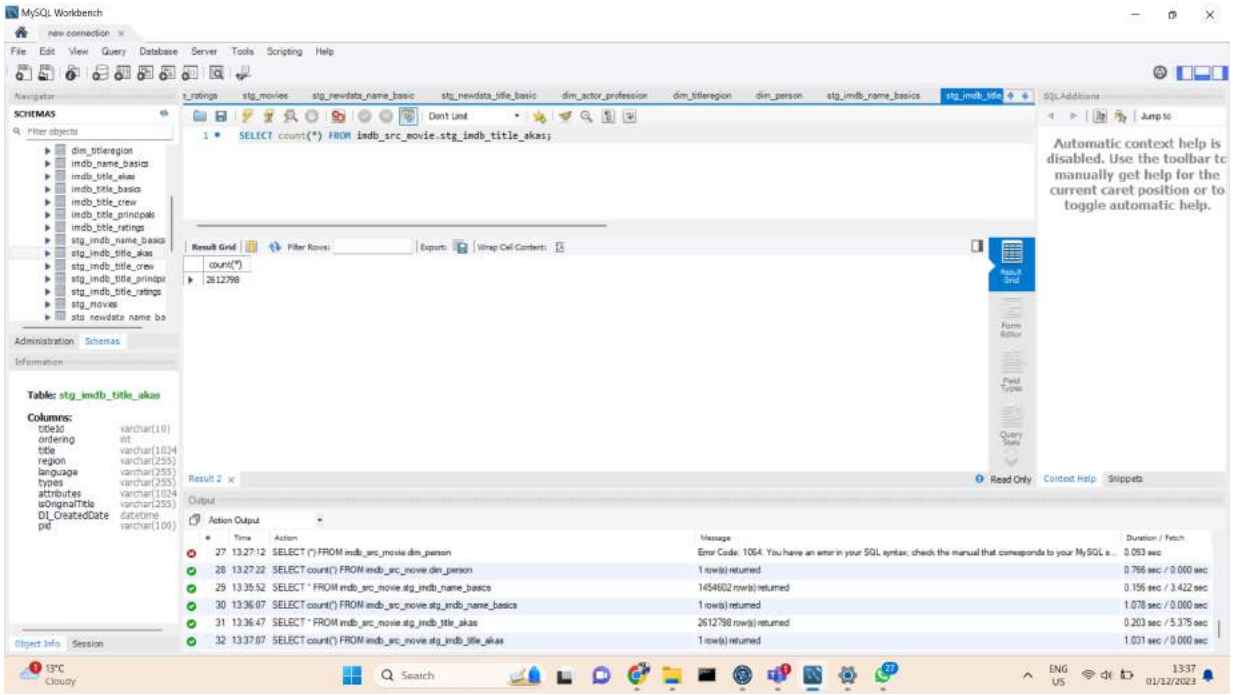




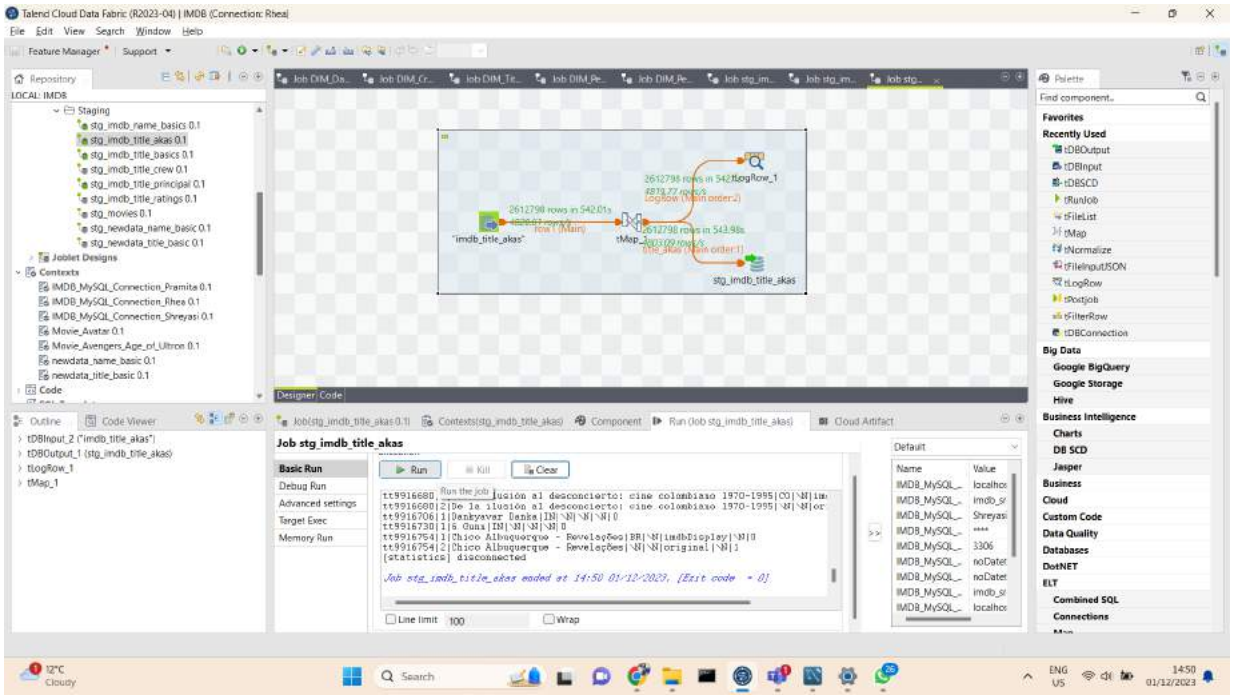
stg_imdb_title_akas

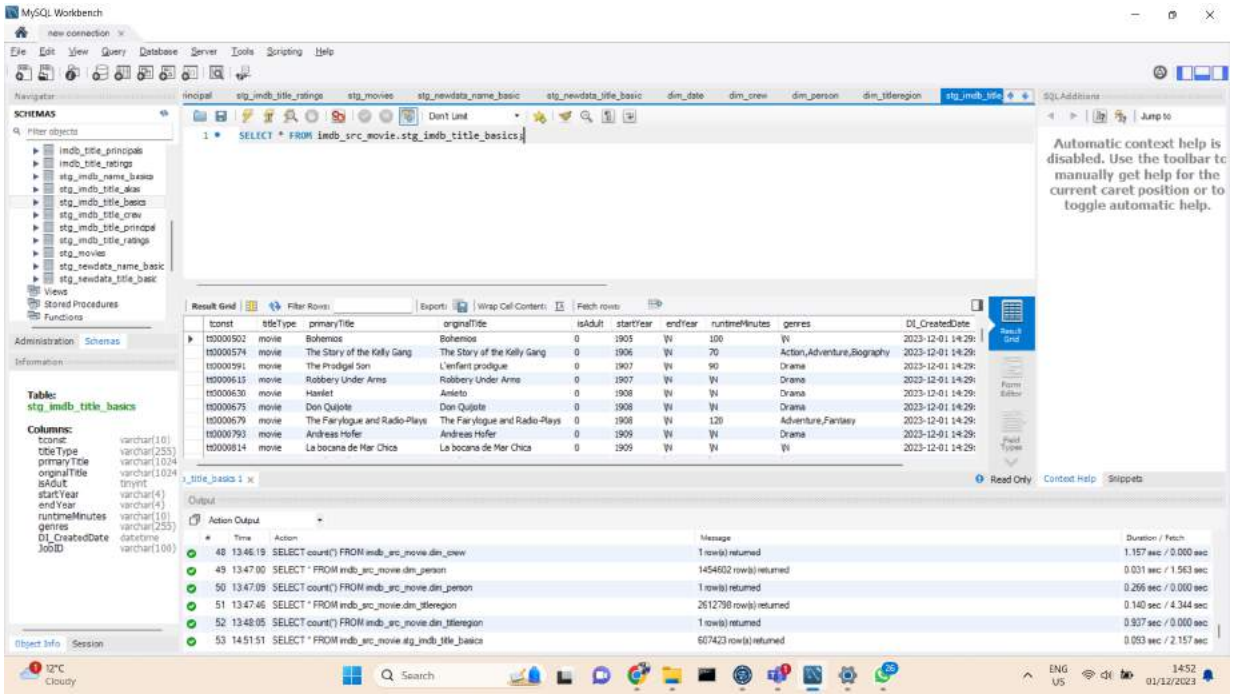


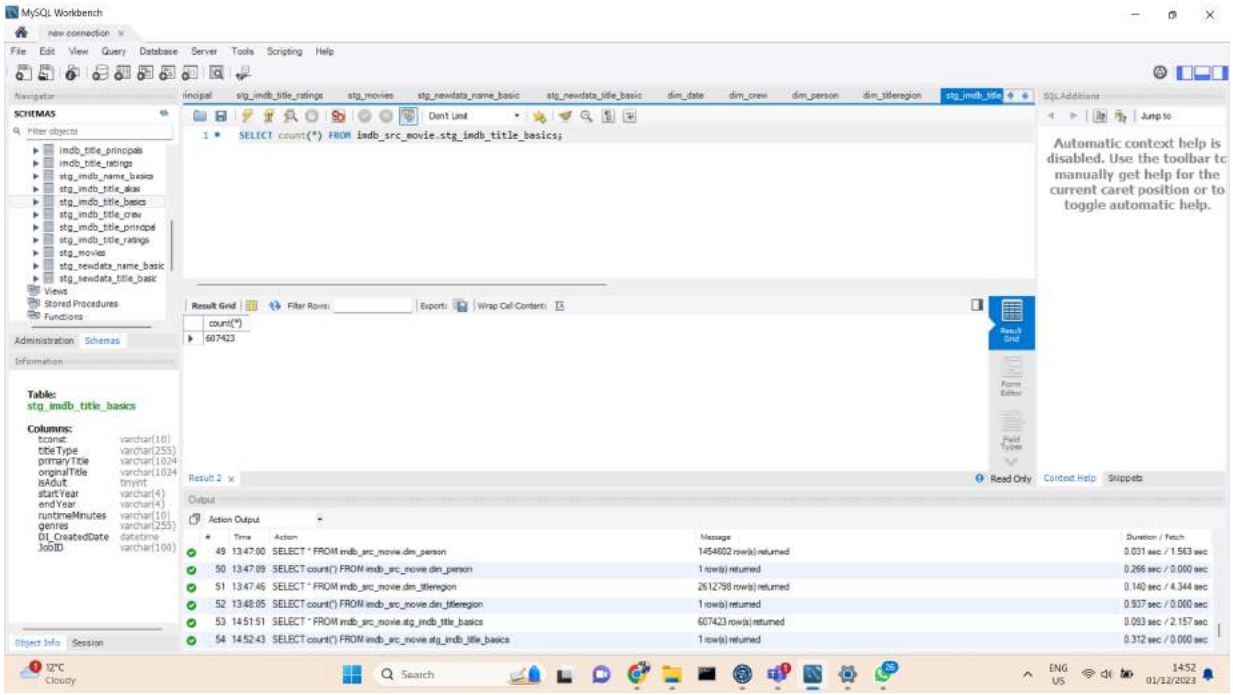

MySQL Workbench
SELECT * FROM imdb_src_movie.stg_imdb_title_akas;
Table: stg_imdb_title_akas
Columns:
titleId varchar(10)
ordering int
title varchar(1024)
region varchar(255)
language varchar(255)
types varchar(255)
attributes varchar(1024)
isOriginalTitle varchar(255)
DI_CreatedDate datetime
pid varchar(100)
Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.


MySQL Workbench
SELECT count(*) FROM imdb_src_movie.stg_imdb_title_akas;
count(*)
2612798
Table: stg_imdb_title_akas
Columns:
titleId varchar(10)
ordering int
title varchar(1024)
region varchar(255)
language varchar(255)
types varchar(255)
attributes varchar(1024)
isOriginalTitle varchar(255)
DI_CreatedDate datetime
pid varchar(100)
Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.

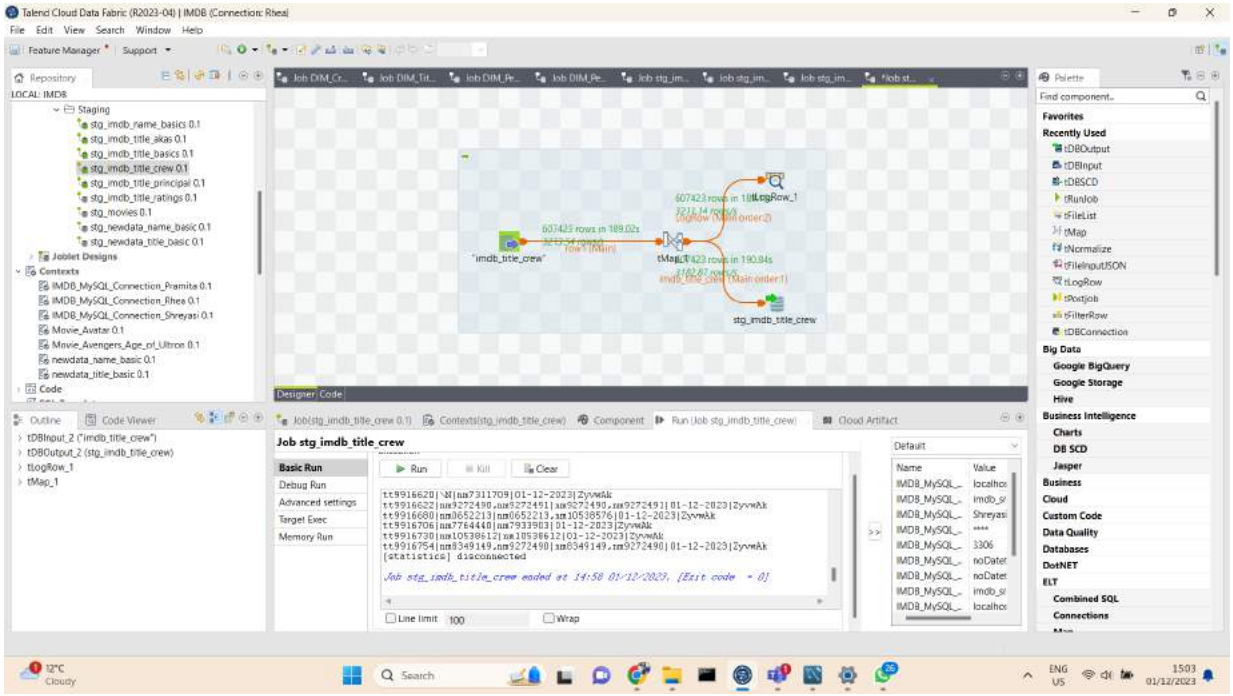
stg_imdb_title_basics

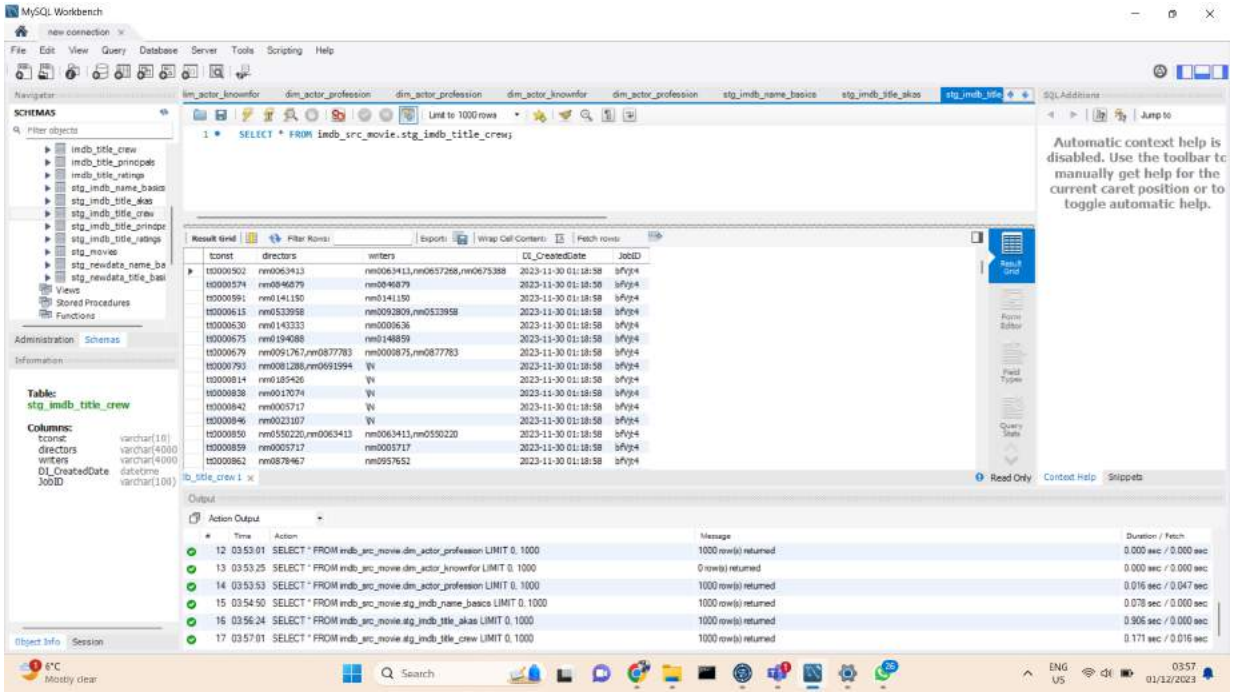


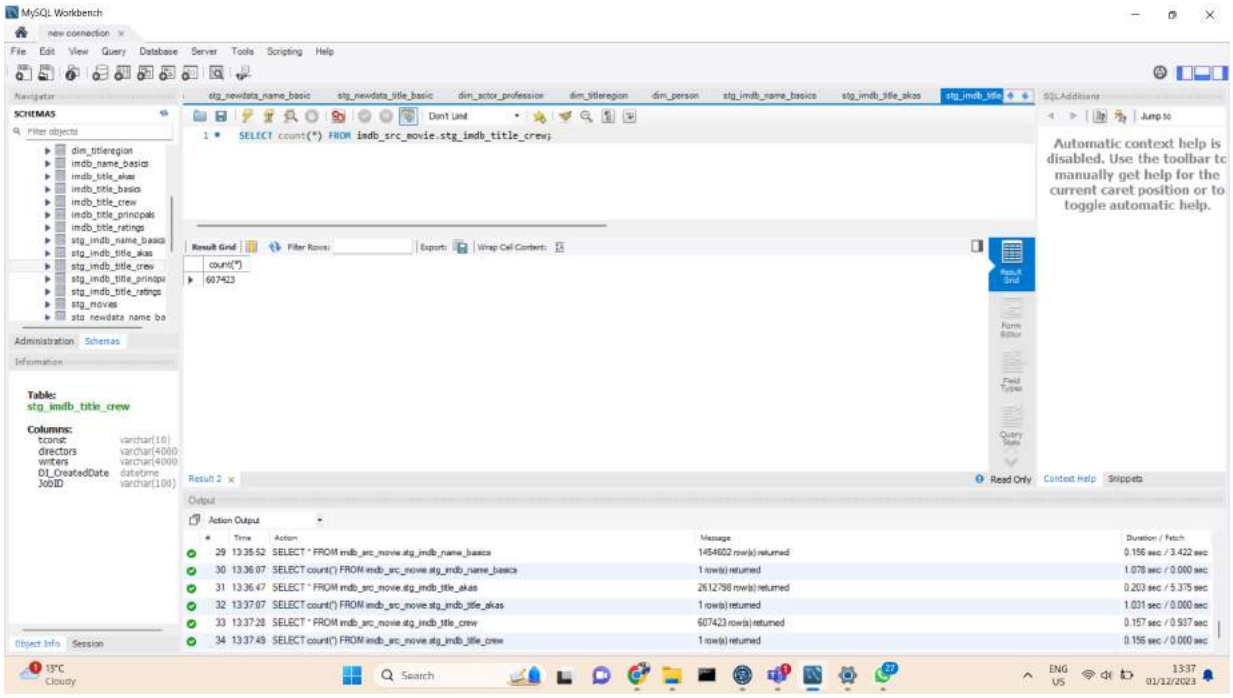




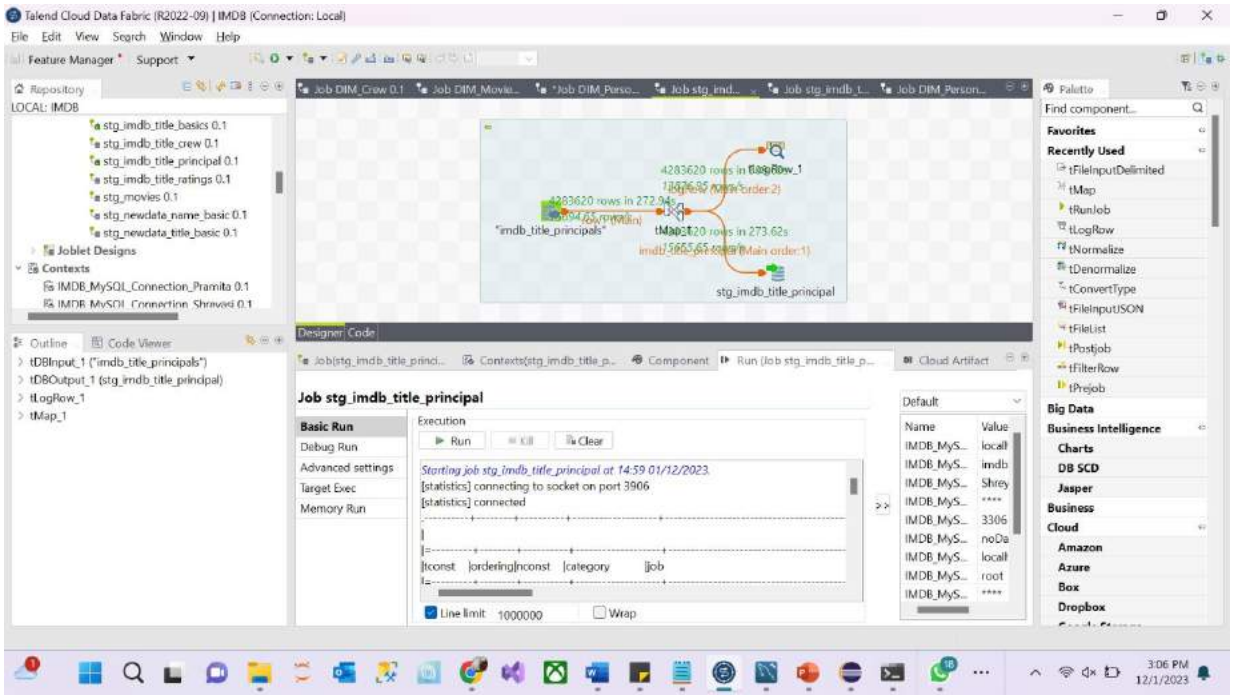
stg_imdb_title_crew

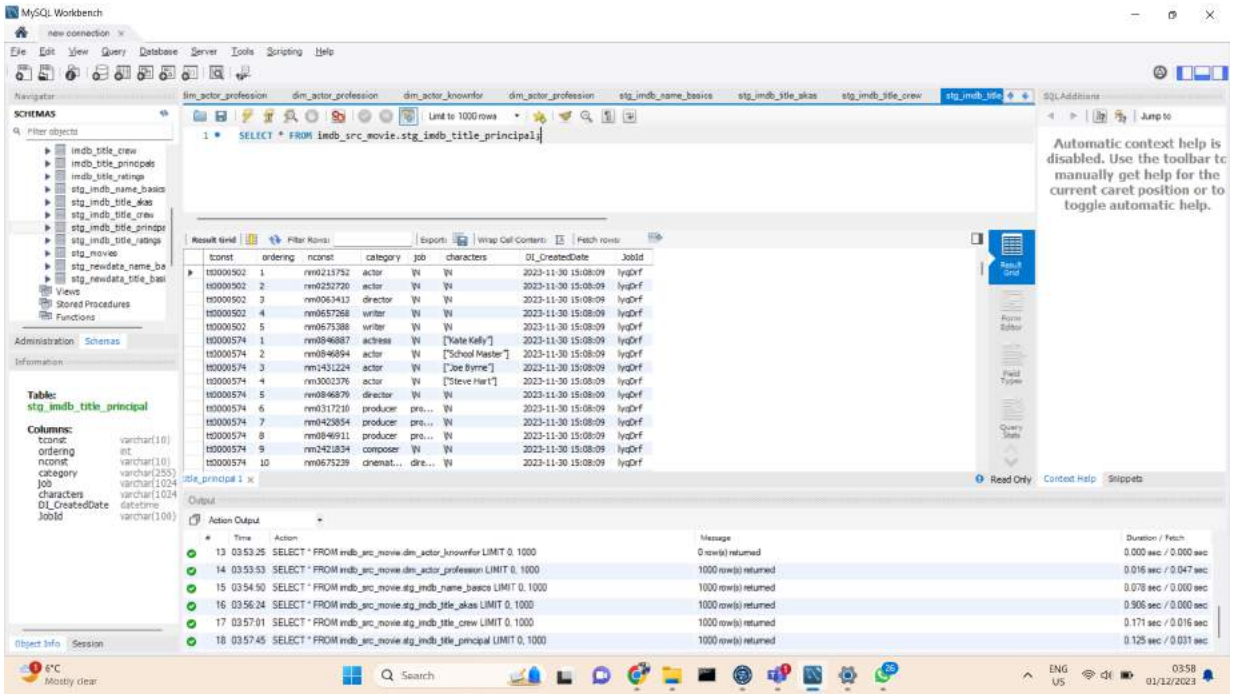


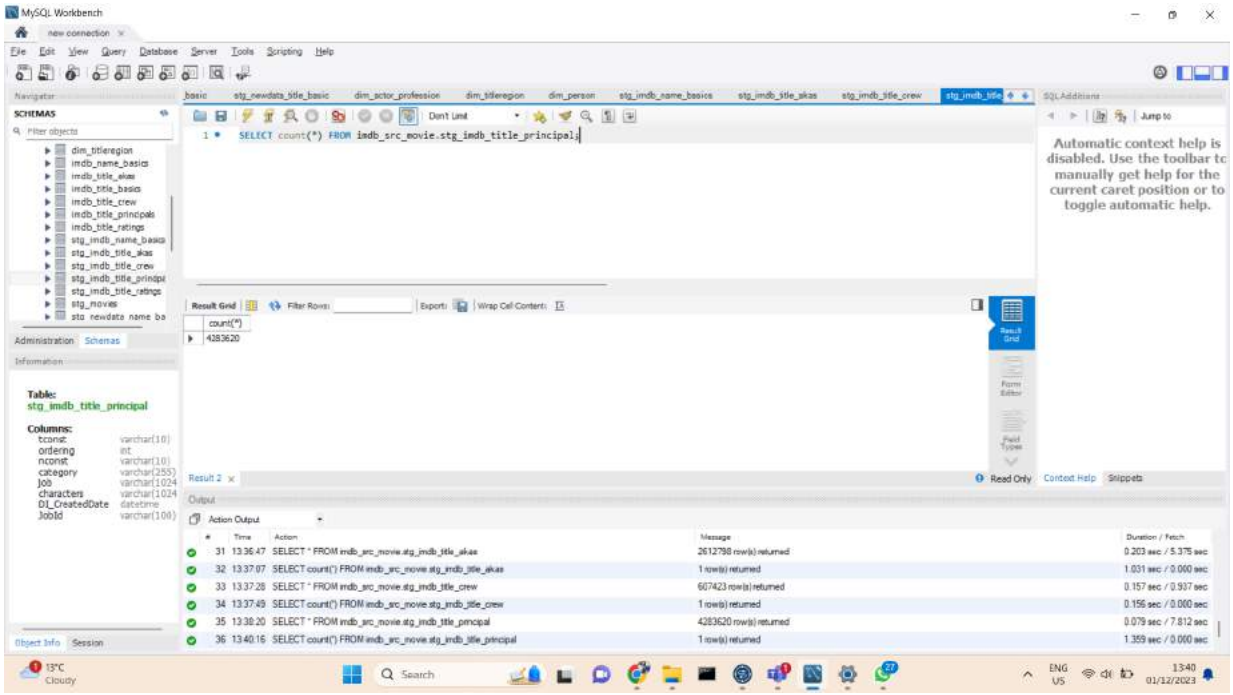

MySQL Workbench
SELECT * FROM imdb_src_movie.stg_imdb_title_crew;
Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.
Table: stg_imdb_title_crew
Columns:
tconst varchar(10)
directors varchar(4000)
writers varchar(4000)
DI_CreatedDate datetime
JobID varchar(100)
Action Output
12 03:53:01 SELECT * FROM imdb_src_movie.dim_actor_profession LIMIT 0, 1000 1000 row(s) returned 0.000 sec / 0.000 sec
13 03:53:25 SELECT * FROM imdb_src_movie.dim_actor_knownfor LIMIT 0, 1000 0 row(s) returned 0.000 sec / 0.000 sec
14 03:53:53 SELECT * FROM imdb_src_movie.dim_actor_profession LIMIT 0, 1000 1000 row(s) returned 0.016 sec / 0.047 sec
15 03:54:50 SELECT * FROM imdb_src_movie.stg_imdb_name_basics LIMIT 0, 1000 1000 row(s) returned 0.078 sec / 0.000 sec
16 03:56:24 SELECT * FROM imdb_src_movie.stg_imdb_title_akas LIMIT 0, 1000 1000 row(s) returned 0.906 sec / 0.000 sec
17 03:57:01 SELECT * FROM imdb_src_movie.stg_imdb_title_crew LIMIT 0, 1000 1000 row(s) returned 0.171 sec / 0.016 sec


MySQL Workbench
SELECT count(*) FROM imdb_src_movie.stg_imdb_title_crew;
count(*)
607423
Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.
Table: stg_imdb_title_crew
Columns:
tconst varchar(10)
directors varchar(4000)
writers varchar(4000)
DI_CreatedDate datetime
JobID varchar(100)
Action Output
29 13:35:52 SELECT * FROM imdb_src_movie.stg_imdb_name_basics 1454602 row(s) returned 0.156 sec / 3.422 sec
30 13:36:07 SELECT count(*) FROM imdb_src_movie.stg_imdb_name_basics 1 row(s) returned 1.078 sec / 0.000 sec
31 13:36:47 SELECT * FROM imdb_src_movie.stg_imdb_title_akas 2612798 row(s) returned 0.203 sec / 5.375 sec
32 13:37:07 SELECT count(*) FROM imdb_src_movie.stg_imdb_title_akas 1 row(s) returned 1.031 sec / 0.000 sec
33 13:37:28 SELECT * FROM imdb_src_movie.stg_imdb_title_crew 607423 row(s) returned 0.157 sec / 0.937 sec
34 13:37:49 SELECT count(*) FROM imdb_src_movie.stg_imdb_title_crew 1 row(s) returned 0.156 sec / 0.000 sec

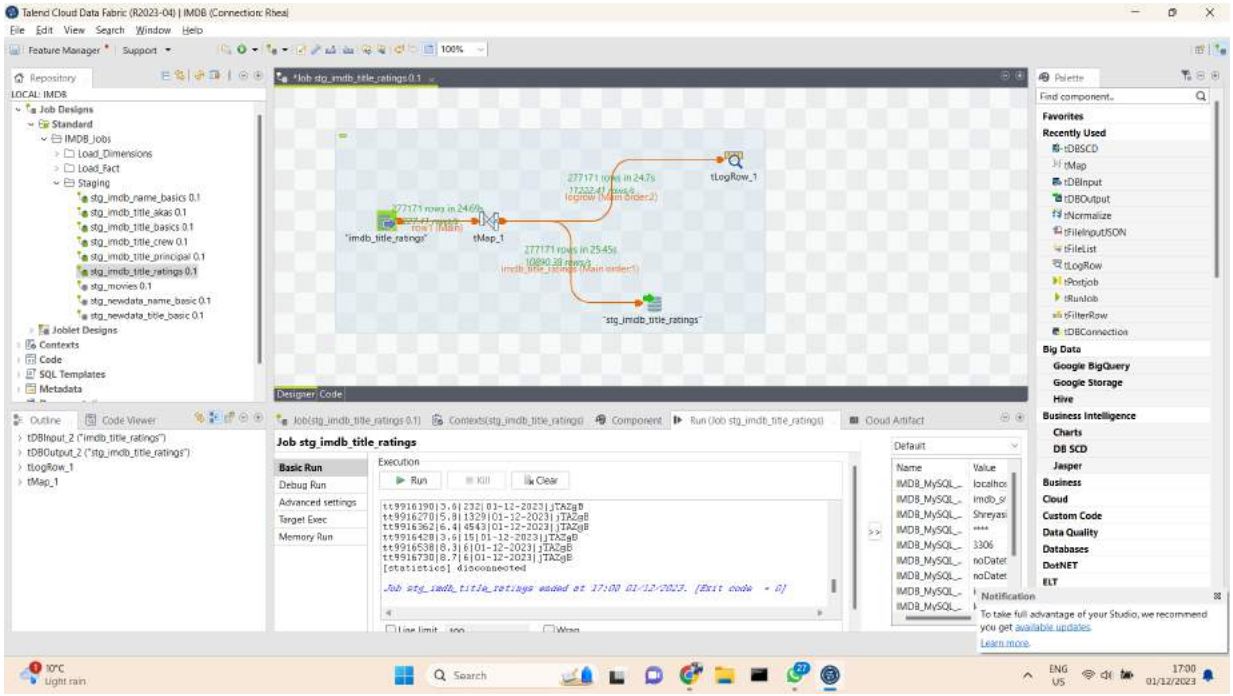
stg_imdb_title_principal

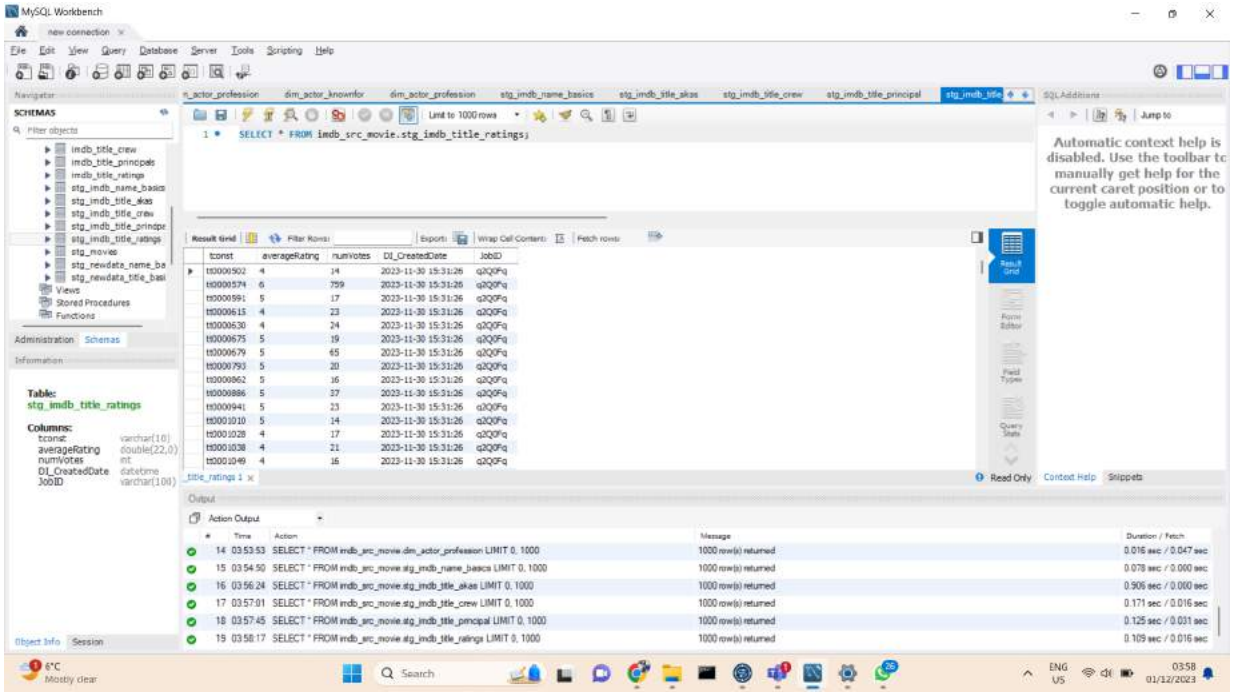


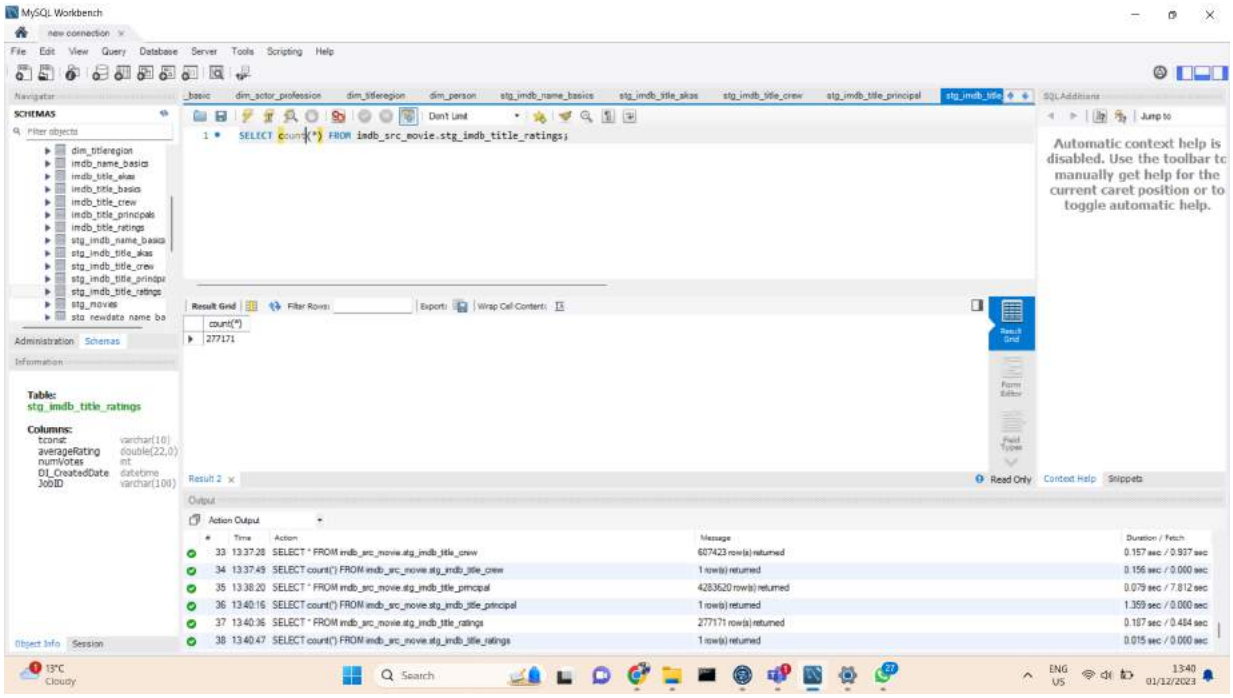




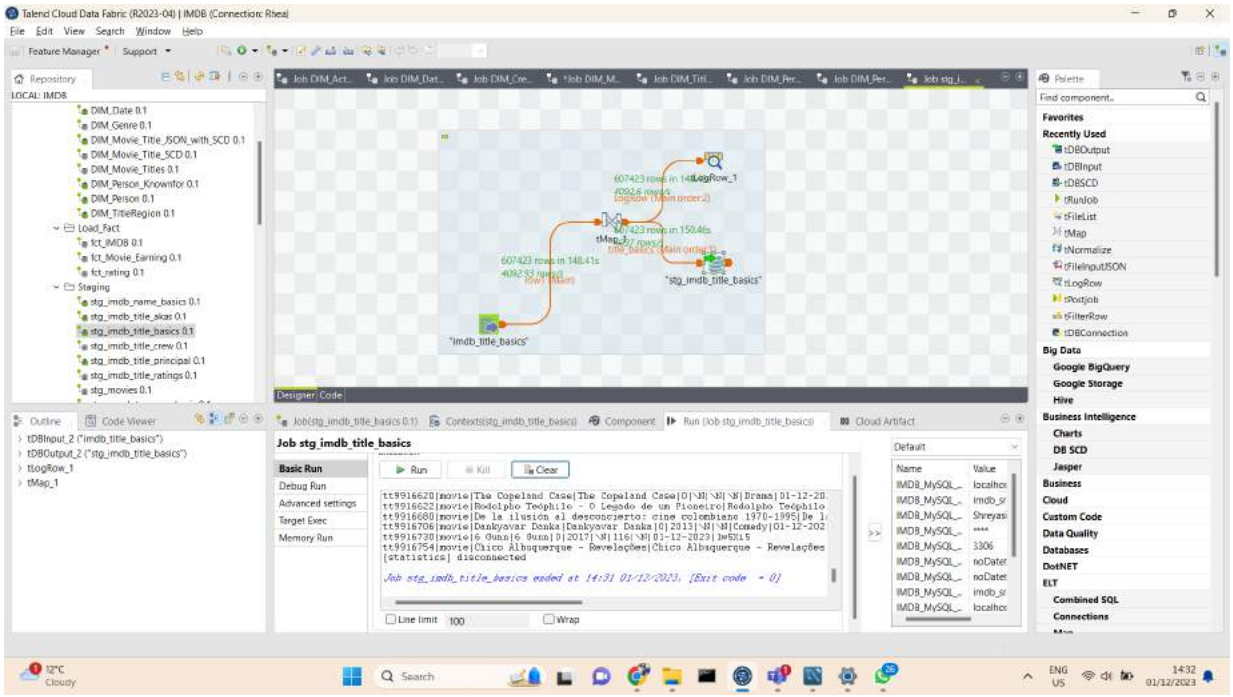
stg_imdb_title_ratings

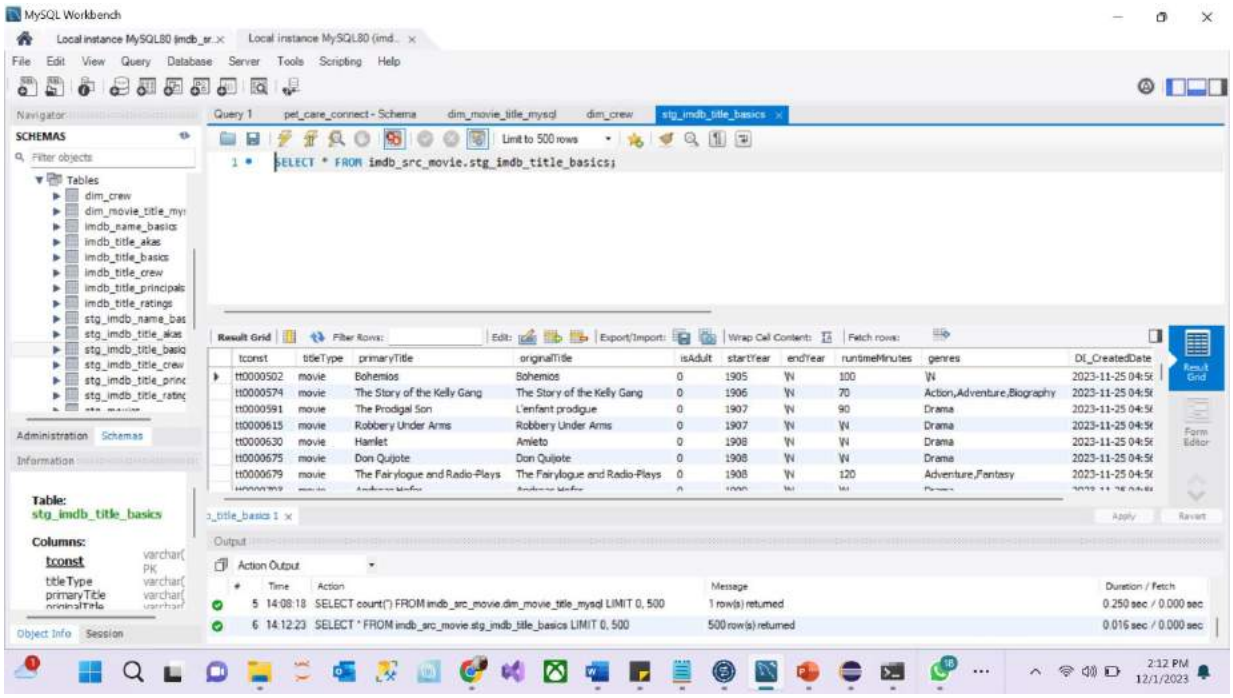


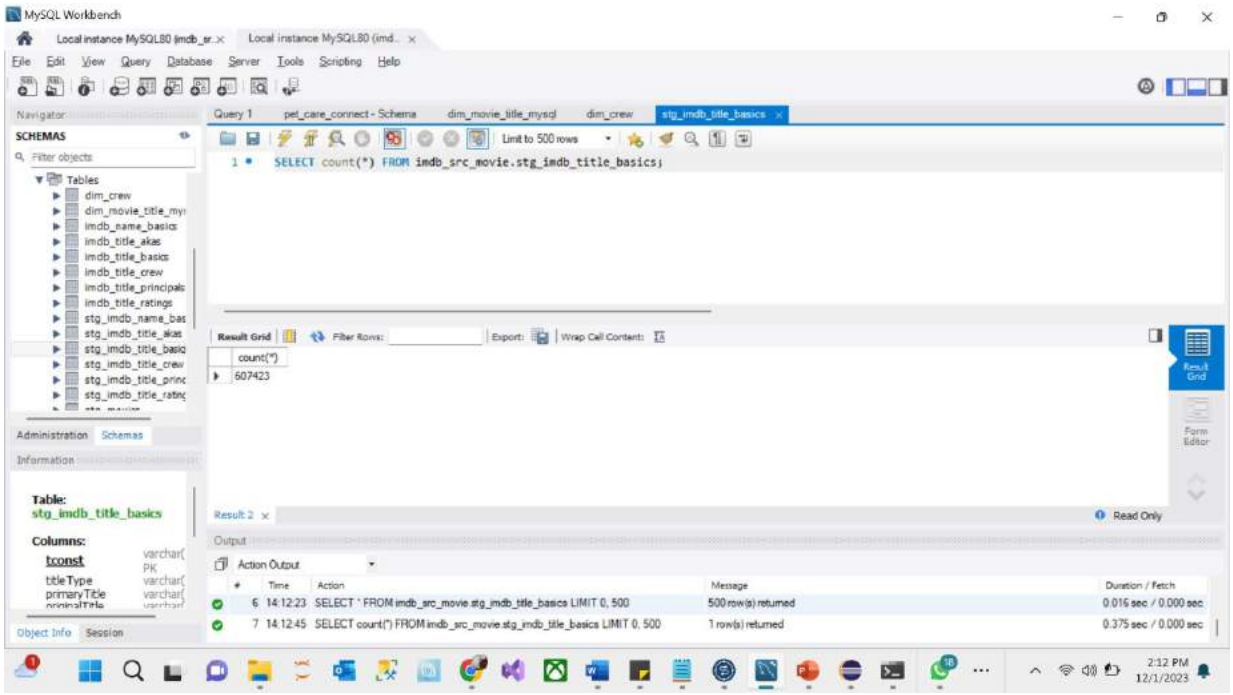

MySQL Workbench
SELECT * FROM imdb_src_movie.stg_imdb_title_ratings;
Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.
Table: stg_imdb_title_ratings
Columns:
tconst varchar(10)
averageRating double(22,0)
numVotes int
DI_CreatedDate datetime
JobID varchar(100)


MySQL Workbench
SELECT count(*) FROM imdb_src_movie.stg_imdb_title_ratings;
count(*)
277171
Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.
Table: stg_imdb_title_ratings
Columns:
tconst varchar(10)
averageRating double(22,0)
numVotes int
DI_CreatedDate datetime
JobID varchar(100)

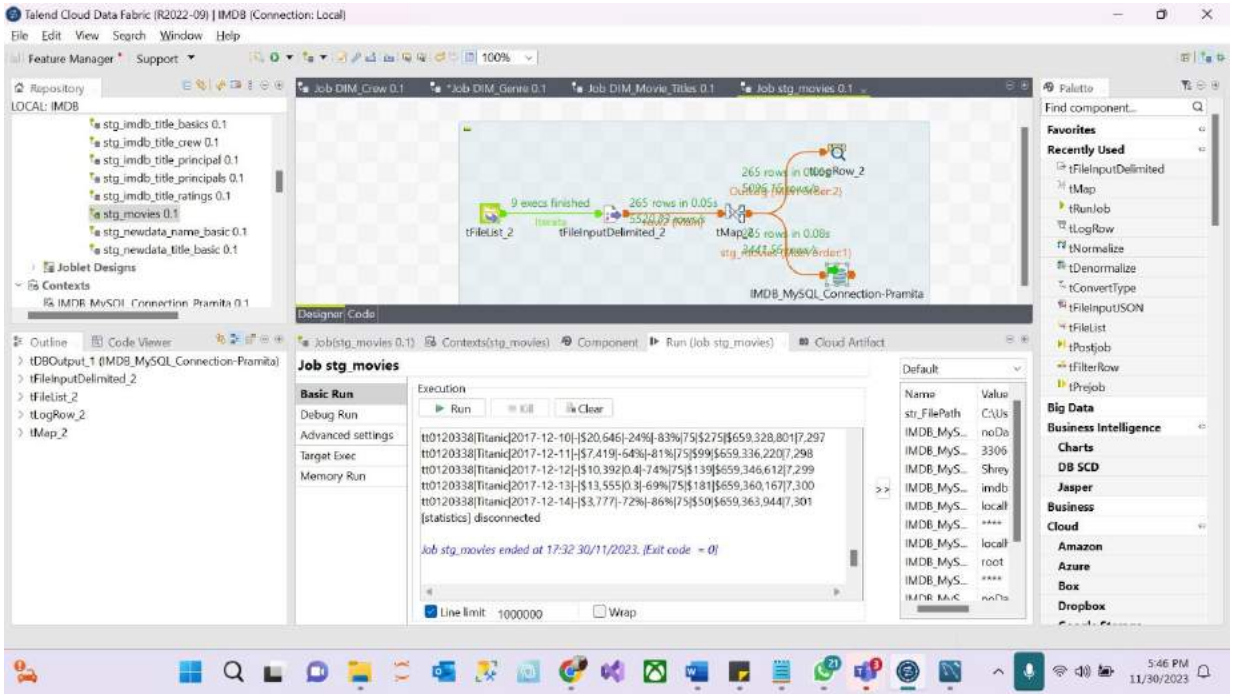
stg_title_basics

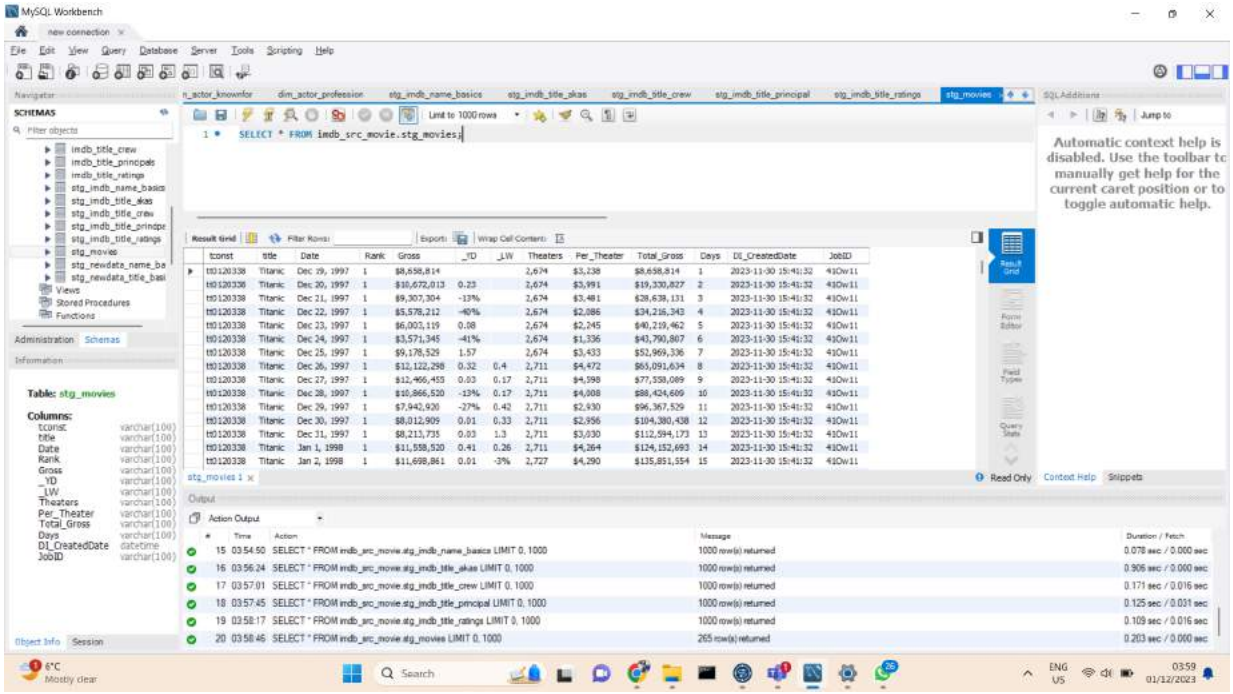


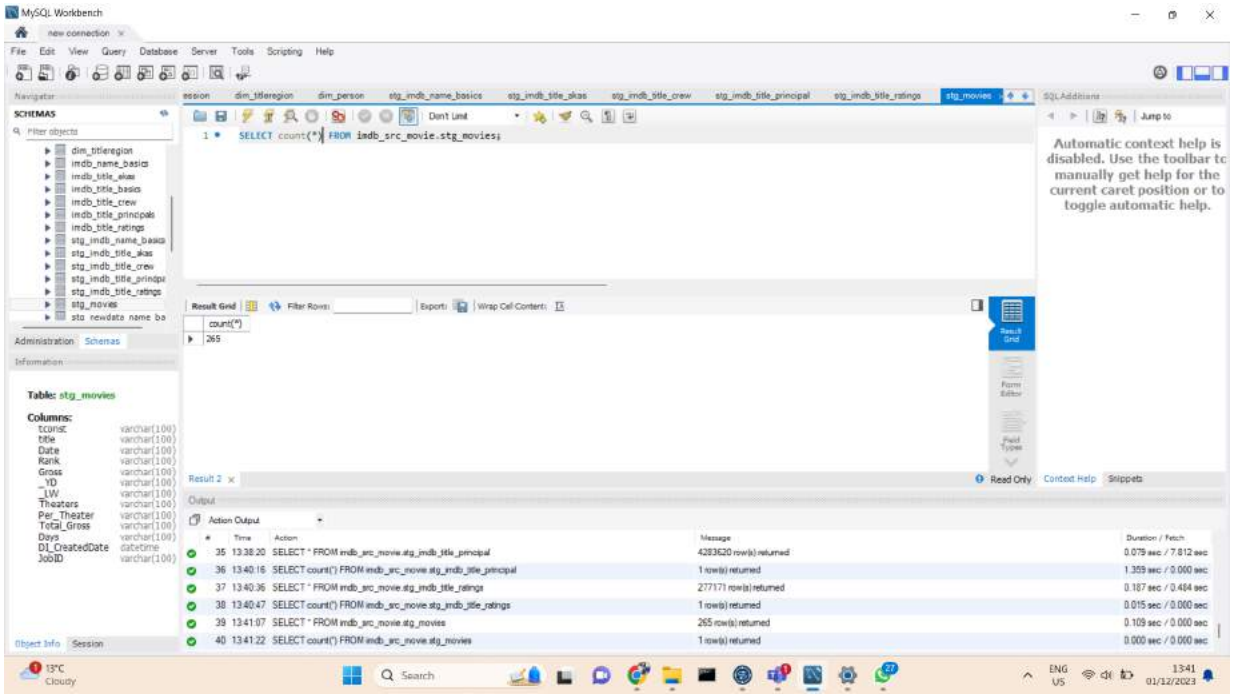



MySQL Workbench
SELECT count(*) FROM imdb_src_movie.stg_imdb_title_basics;
count(*)
607423
Table:
stg_imdb_title_basics
Columns:
tconst
titleType
primaryTitle
Action Output
6 14:12:23 SELECT * FROM imdb_src_movie.stg_imdb_title_basics LIMIT 0, 500 500 row(s) returned 0.016 sec / 0.000 sec
7 14:12:45 SELECT count(*) FROM imdb_src_movie.stg_imdb_title_basics LIMIT 0, 500 1 row(s) returned 0.375 sec / 0.000 sec

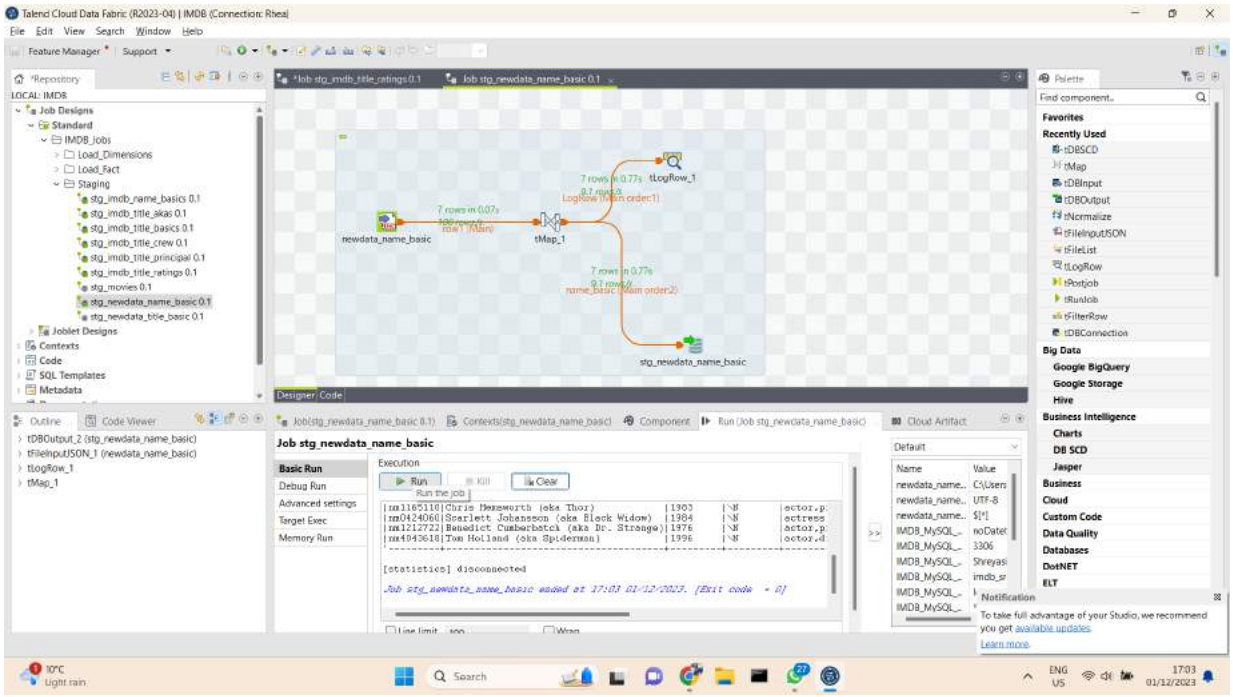
stg_movies

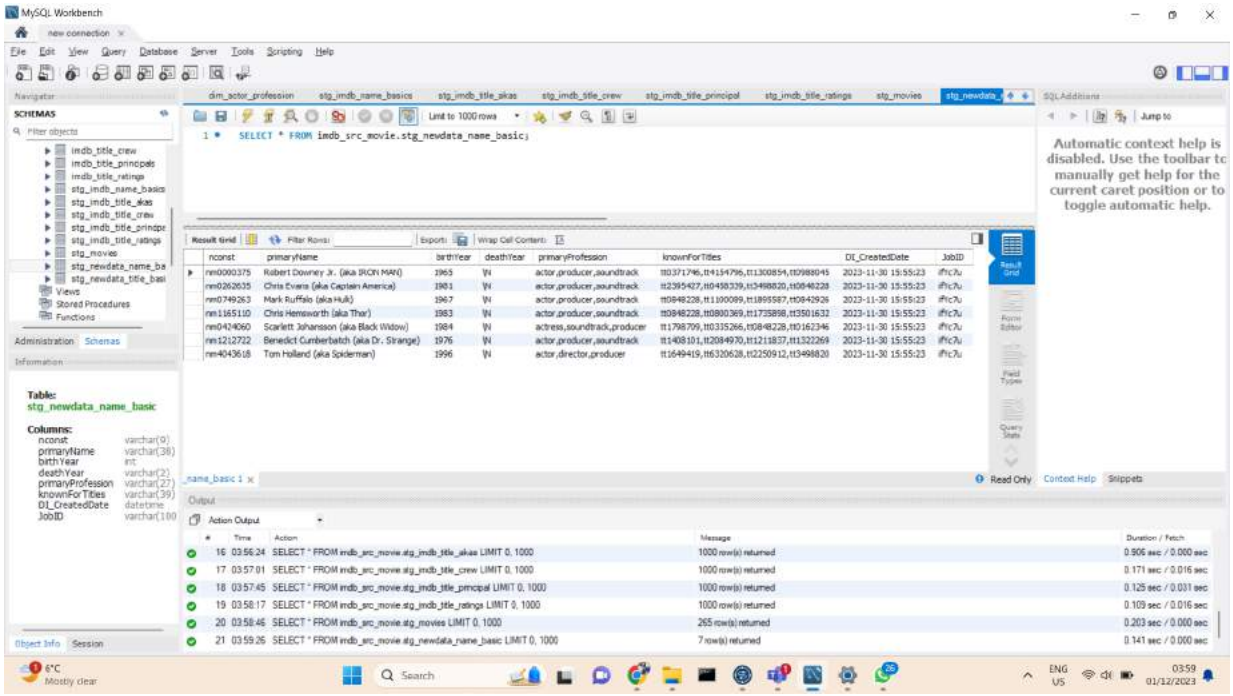

Talend Cloud Data Fabric (R2022-09) | IMDB (Connection: Local)
Job stg_movies
Basic Run
Debug Run
Advanced settings
Target Exec
Memory Run
Run
Clear
tt0120338|Titanic|2017-12-10|-|$20,646|-24%|-83%|75|$275|$659,328,801|7,297
tt0120338|Titanic|2017-12-11|-|$7,419|-64%|-81%|75|$99|$659,336,220|7,298
tt0120338|Titanic|2017-12-12|-|$10,392|0.4|-74%|75|$139|$659,346,612|7,299
tt0120338|Titanic|2017-12-13|-|$13,555|0.3|-69%|75|$181|$659,360,167|7,300
tt0120338|Titanic|2017-12-14|-|$3,777|-72%|-86%|75|$50|$659,363,944|7,301
[statistics] disconnected
Job stg_movies ended at 17:32 30/11/2023. [Exit code = 0]
Line limit 1000000
Wrap
IMDB_MySQL_Connection-Pramita

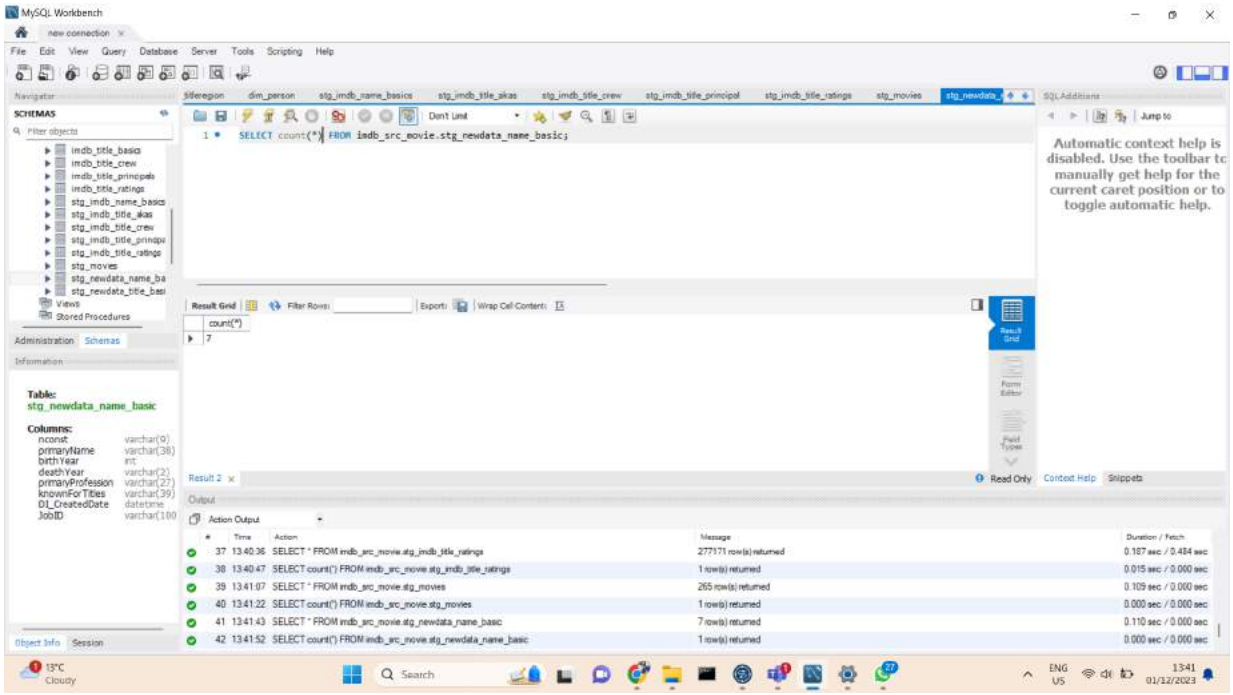

MySQL Workbench
SELECT * FROM imdb_src_movie.stg_movies;
Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.
Table: stg_movies
Columns:
tconst varchar(100)
title varchar(100)
Date varchar(100)
Rank varchar(100)
Gross varchar(100)
_YD varchar(100)
_LW varchar(100)
Theaters varchar(100)
Per_Theater varchar(100)
Total_Gross varchar(100)
Days varchar(100)
DI_CreatedDate datetime
JobID varchar(100)


MySQL Workbench
SELECT count(*) FROM imdb_src_movie.stg_movies;
count(*)
265
Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.
Table: stg_movies
Columns:
tconst varchar(100)
title varchar(100)
Date varchar(100)
Rank varchar(100)
Gross varchar(100)
_YD varchar(100)
_LW varchar(100)
Theaters varchar(100)
Per_Theater varchar(100)
Total_Gross varchar(100)
Days varchar(100)
DI_CreatedDate datetime
JobID varchar(100)

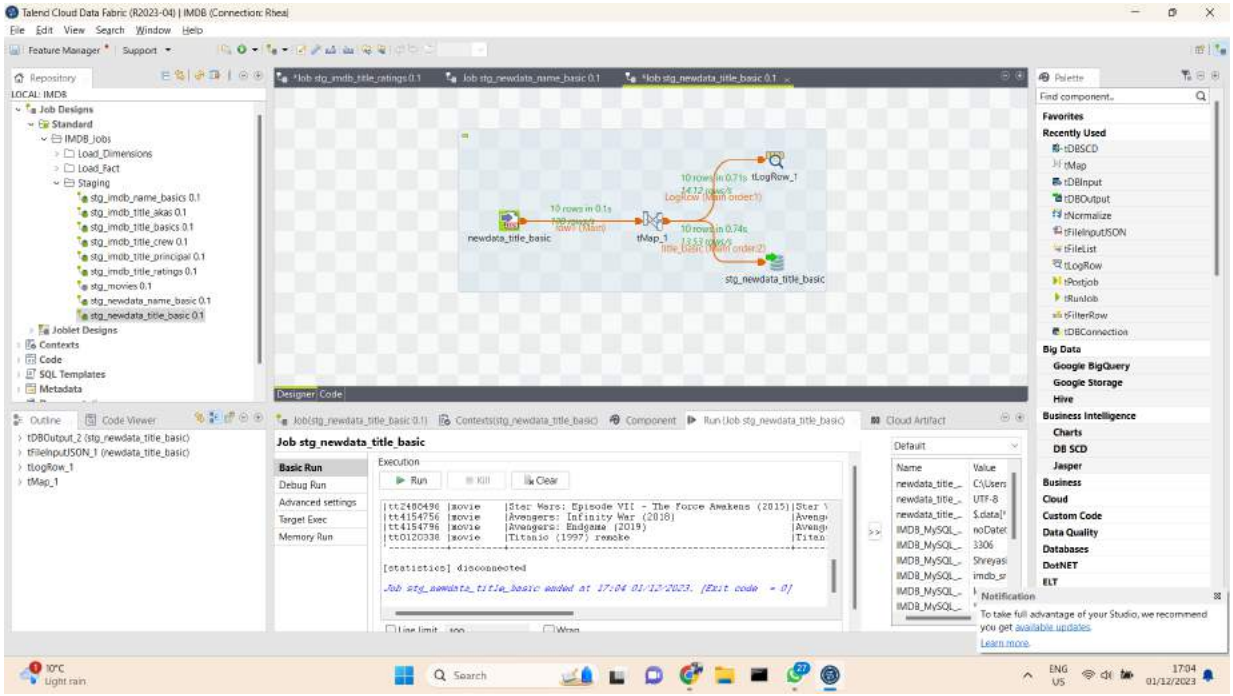
stg_newdata_name_basic

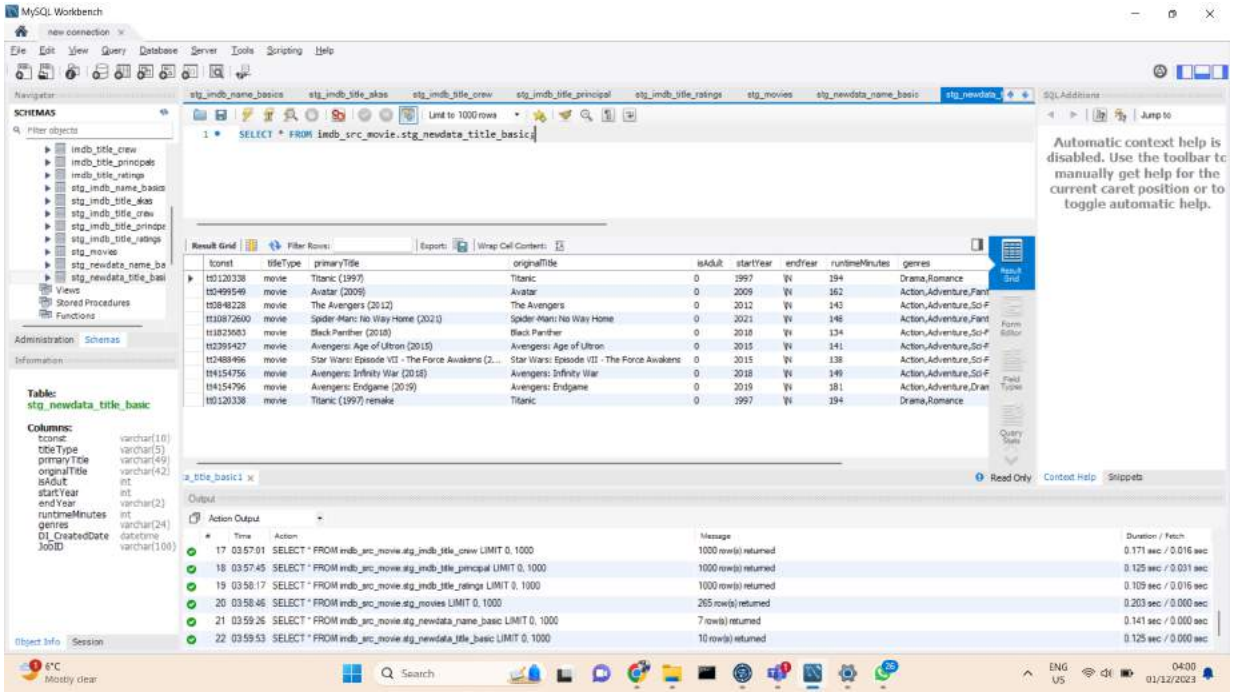


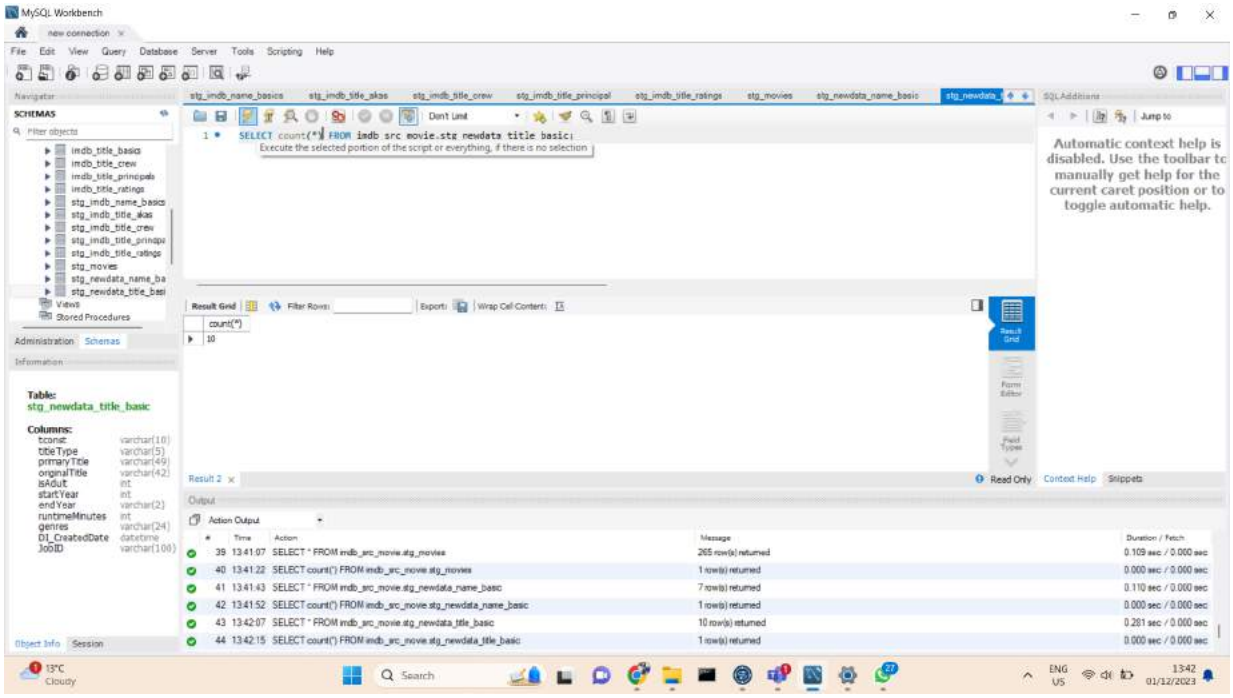




stg_newdata_title_basic

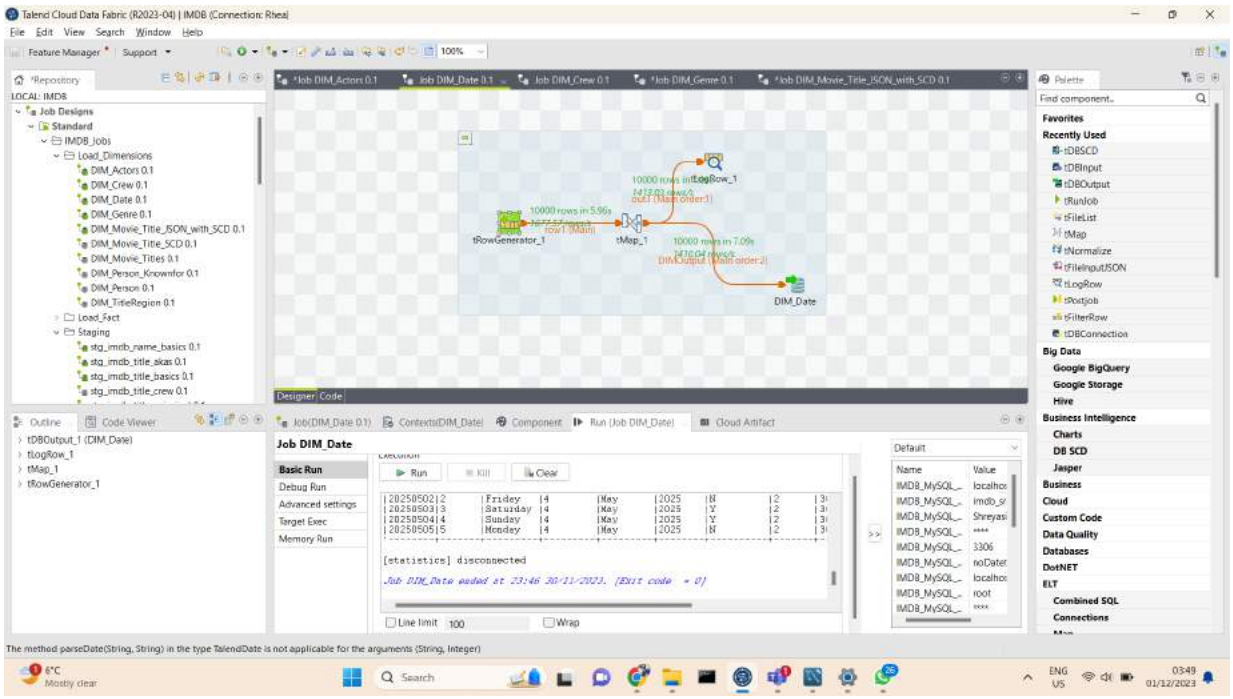


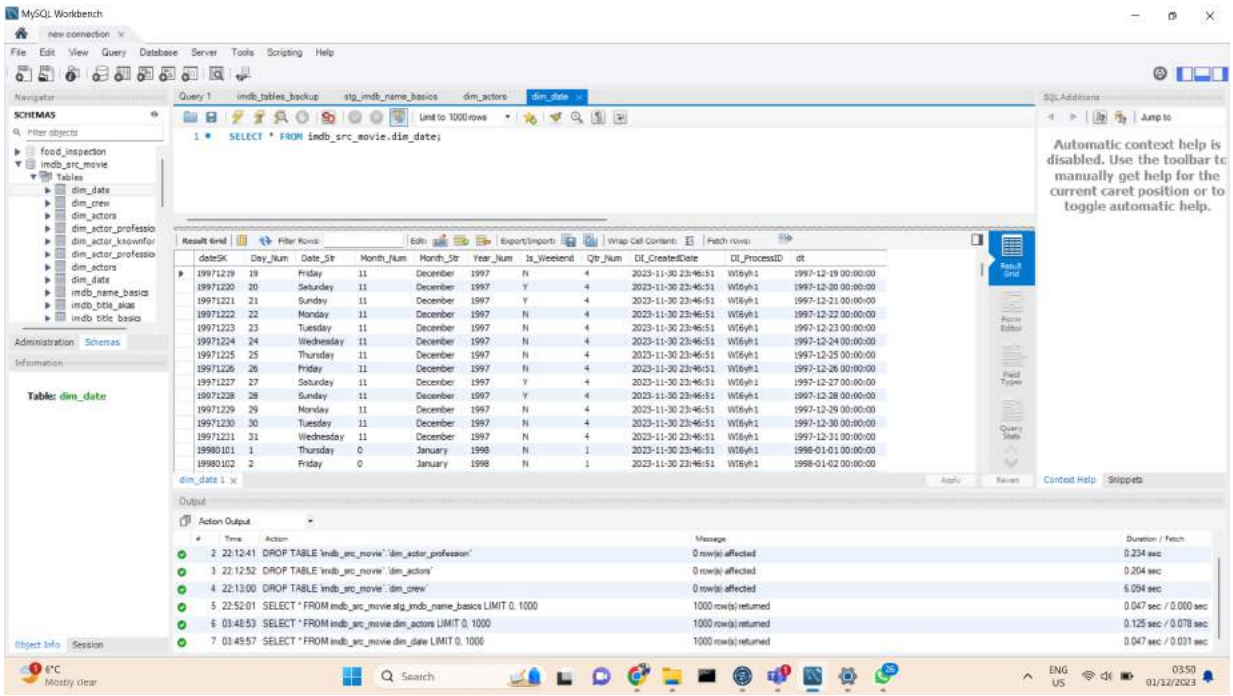

MySQL Workbench
SELECT * FROM imdb_src_movie.stg_newdata_title_basic;
Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.
Table: stg_newdata_title_basic
Columns:
tconst varchar(10)
titleType varchar(5)
primaryTitle varchar(49)
originalTitle varchar(42)
isAdult int
startYear int
endYear varchar(2)
runtimeMinutes int
genres varchar(24)
DI_CreatedDate datetime
JobID varchar(100)

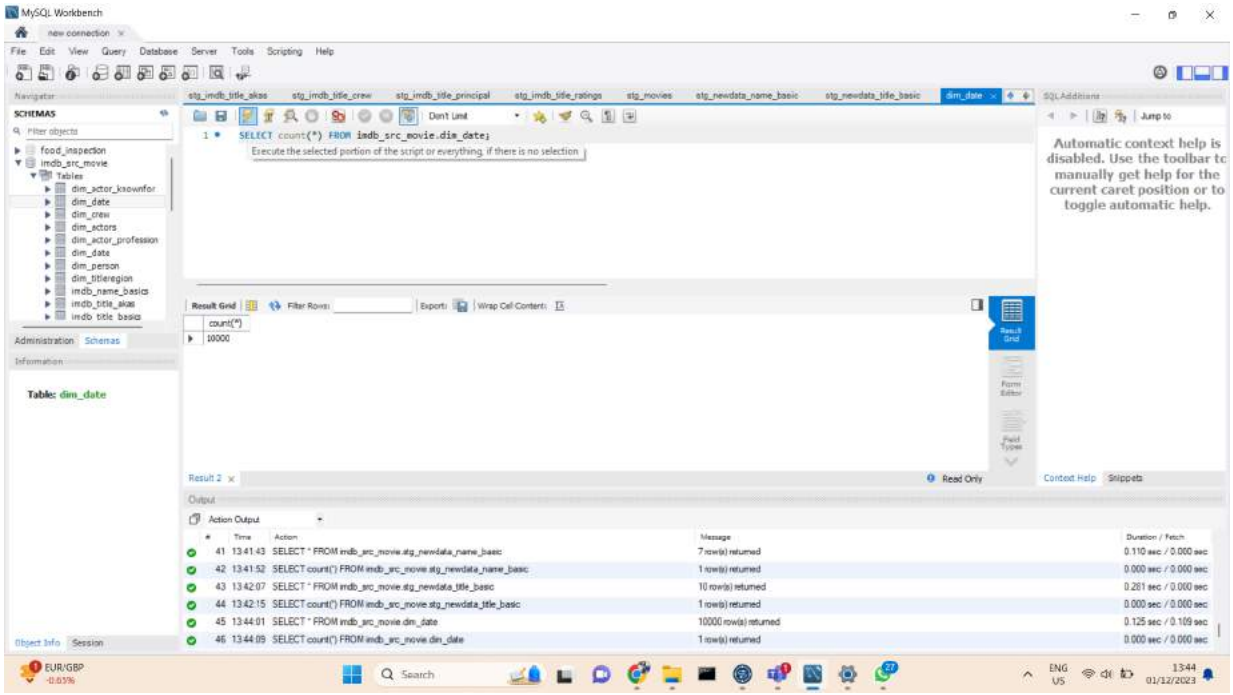

MySQL Workbench
SELECT count(*) FROM imdb_src_movie.stg_newdata_title_basic;
Execute the selected portion of the script or everything, if there is no selection
Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.
Table: stg_newdata_title_basic
Columns:
tconst varchar(10)
titleType varchar(5)
primaryTitle varchar(49)
originalTitle varchar(42)
isAdult int
startYear int
endYear varchar(2)
runtimeMinutes int
genres varchar(24)
DI_CreatedDate datetime
JobID varchar(100)

Dim Tables:

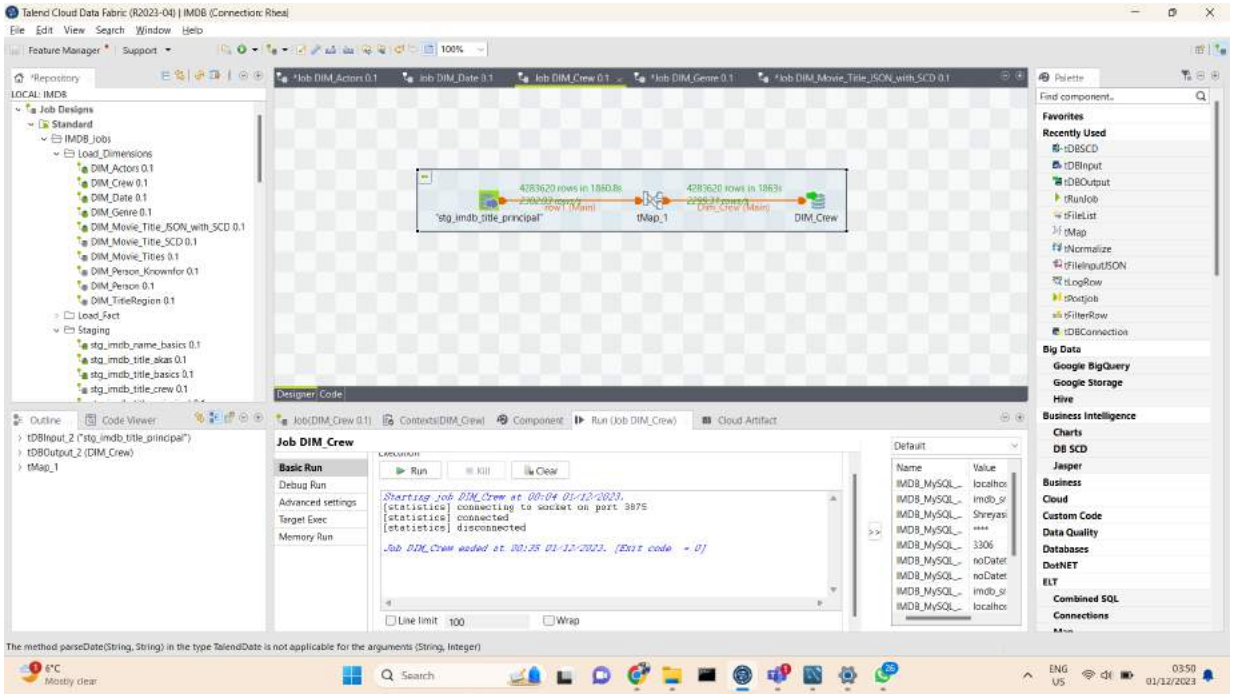
DIM Date:

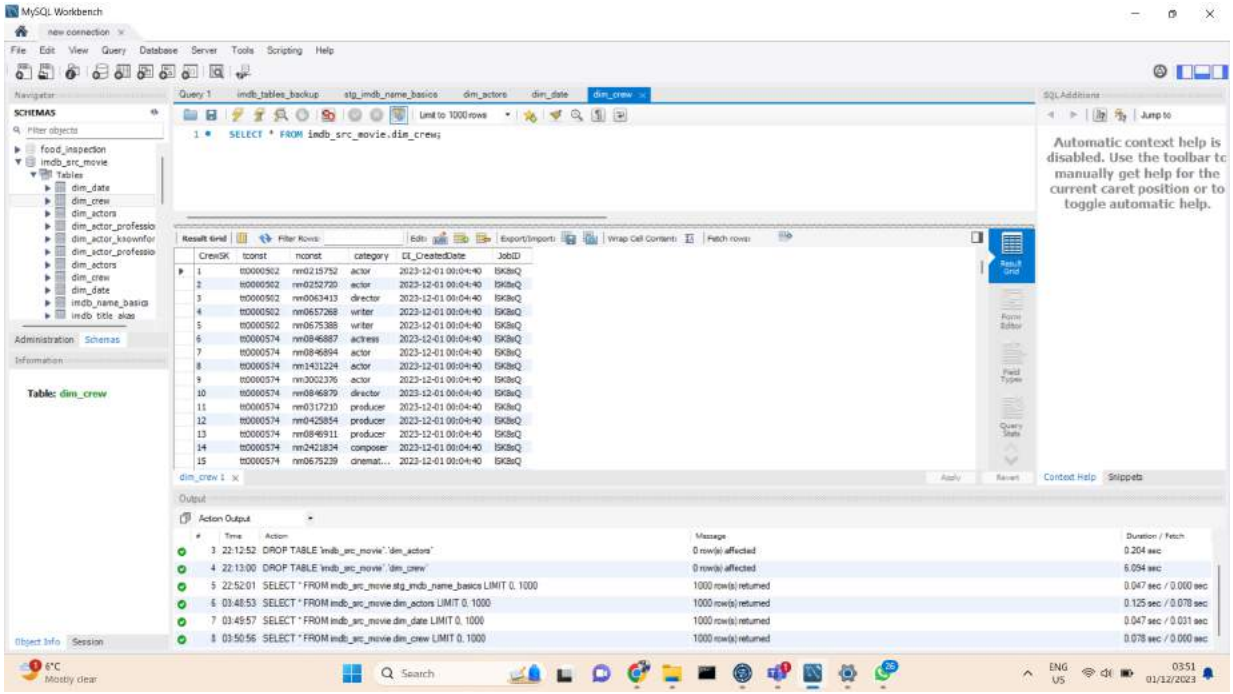


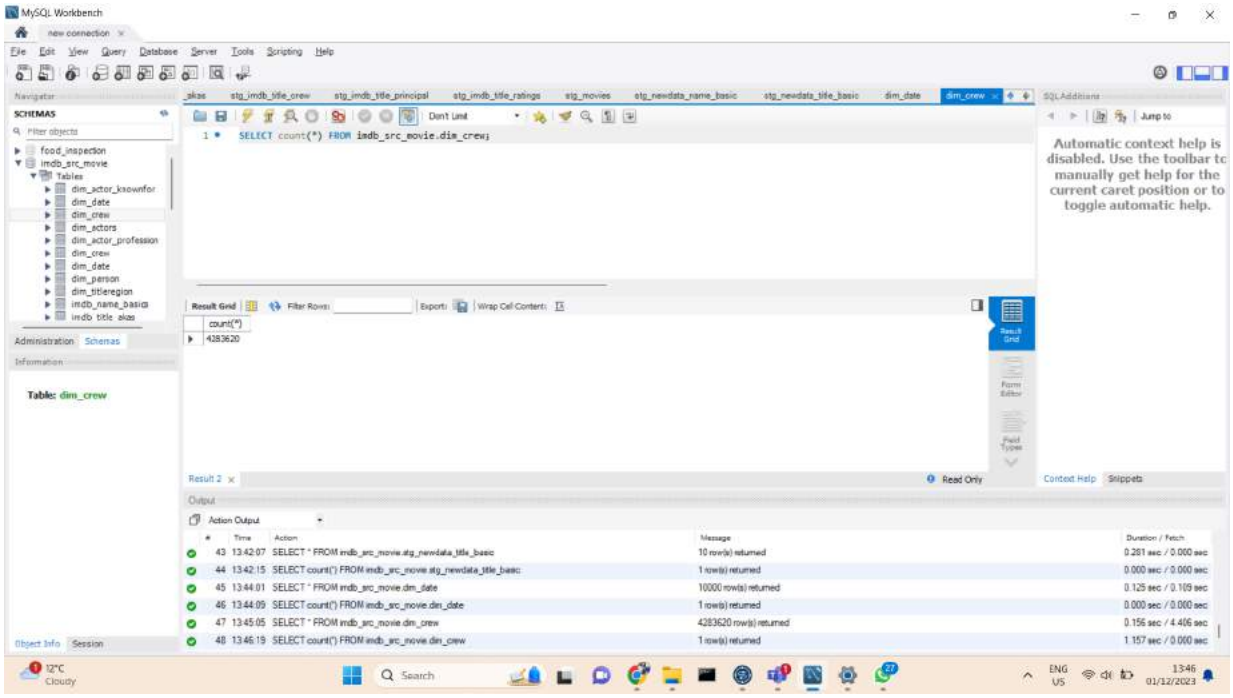



MySQL Workbench
SELECT count(*) FROM imdb_src_movie.dim_date;
Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.
Table: dim_date
Action Output

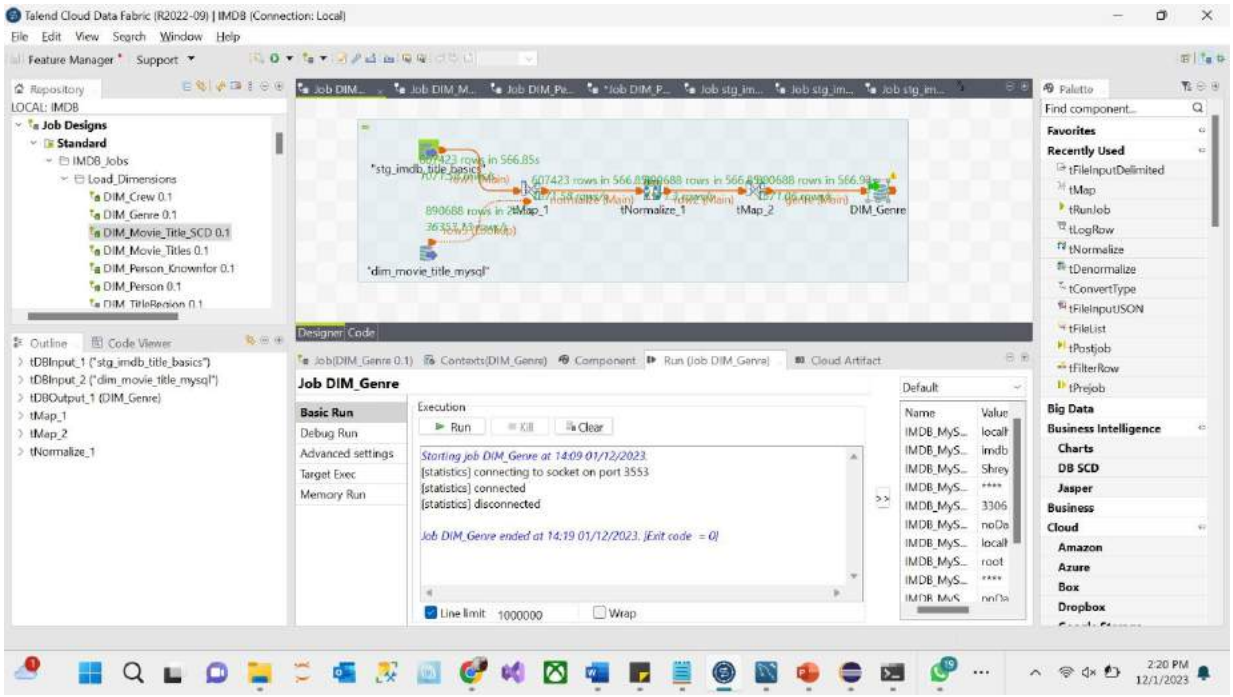
DIM Crew:

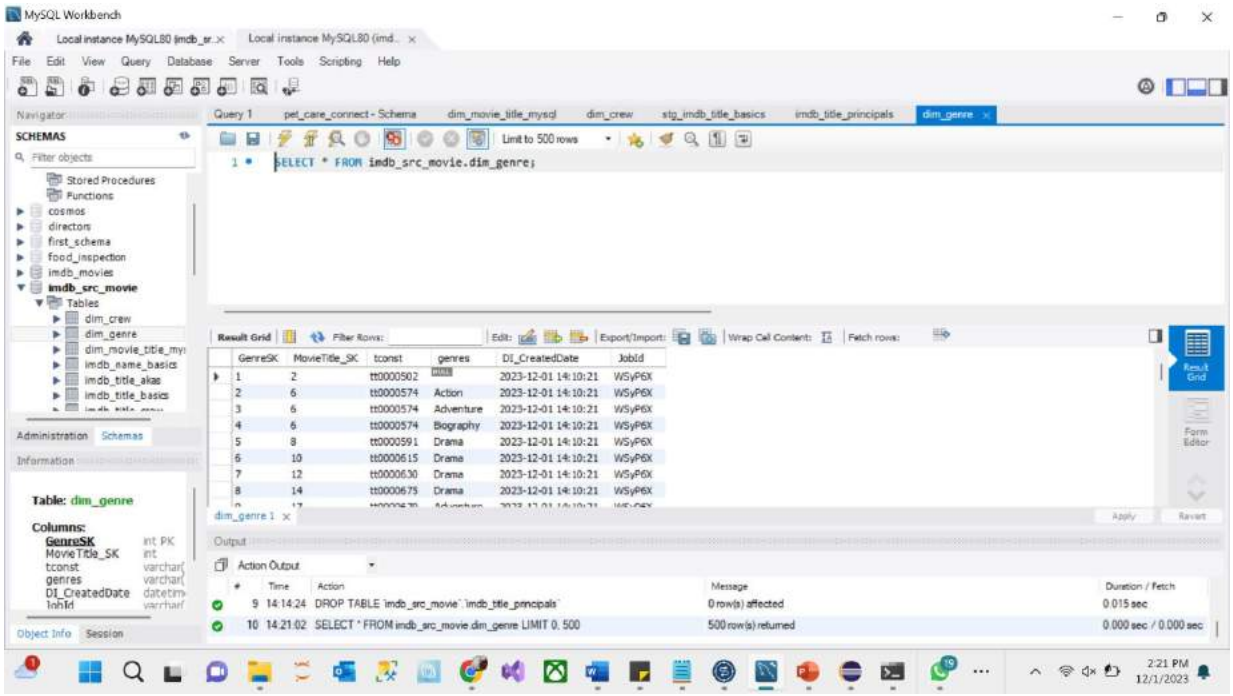

Talend Cloud Data Fabric (R2023-04) | IMDB (Connection: Rhea)
Job DIM_Crew
Basic Run
Debug Run
Advanced settings
Target Exec
Memory Run
Starting job DIM_Crew at 00:04 01/12/2023.
[statistics] connecting to socket on port 3875
[statistics] connected
[statistics] disconnected
Job DIM_Crew ended at 00:25 01/12/2023. [Exit code = 0]

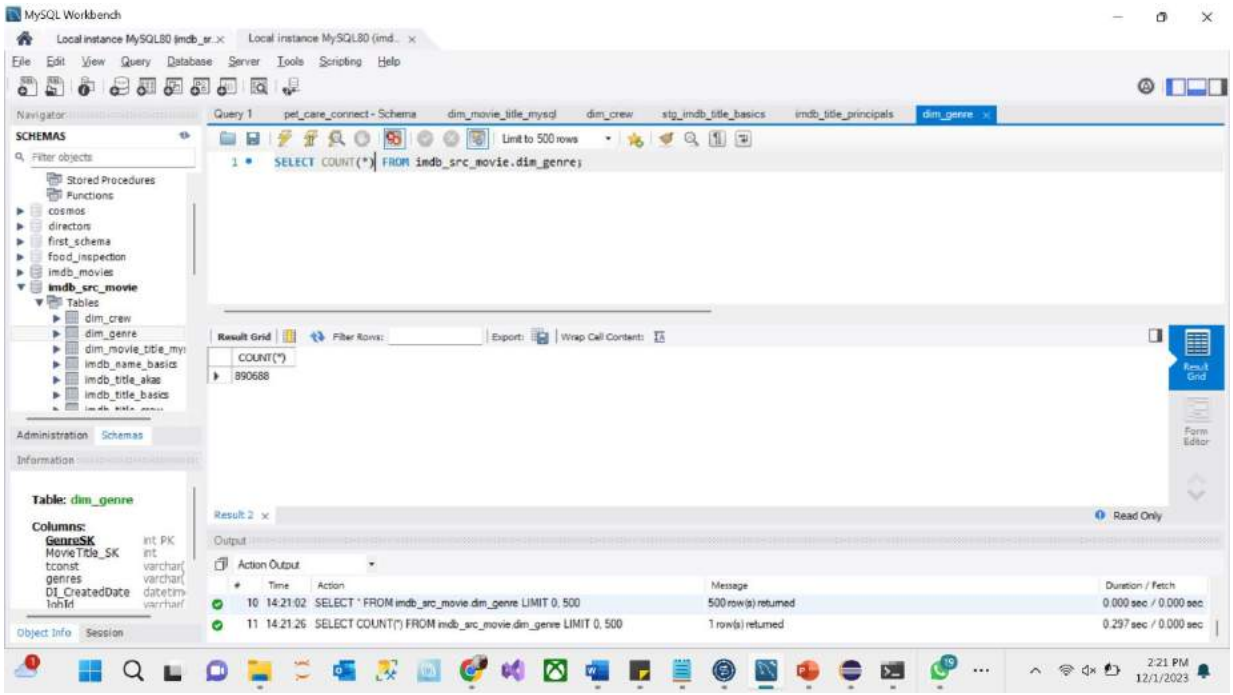

MySQL Workbench
SELECT * FROM imdb_src_movie.dim_crew;
Table: dim_crew


MySQL Workbench
SELECT count(*) FROM imdb_src_movie.dim_crew;
Table: dim_crew

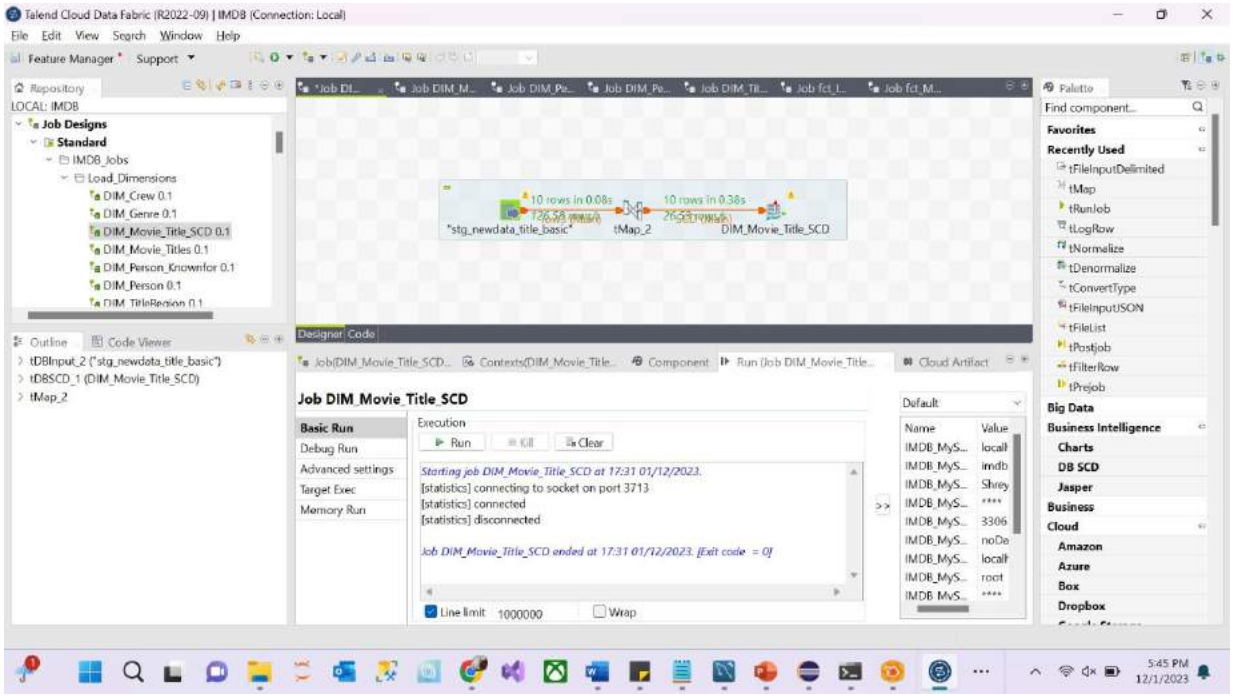
DIM Genre:

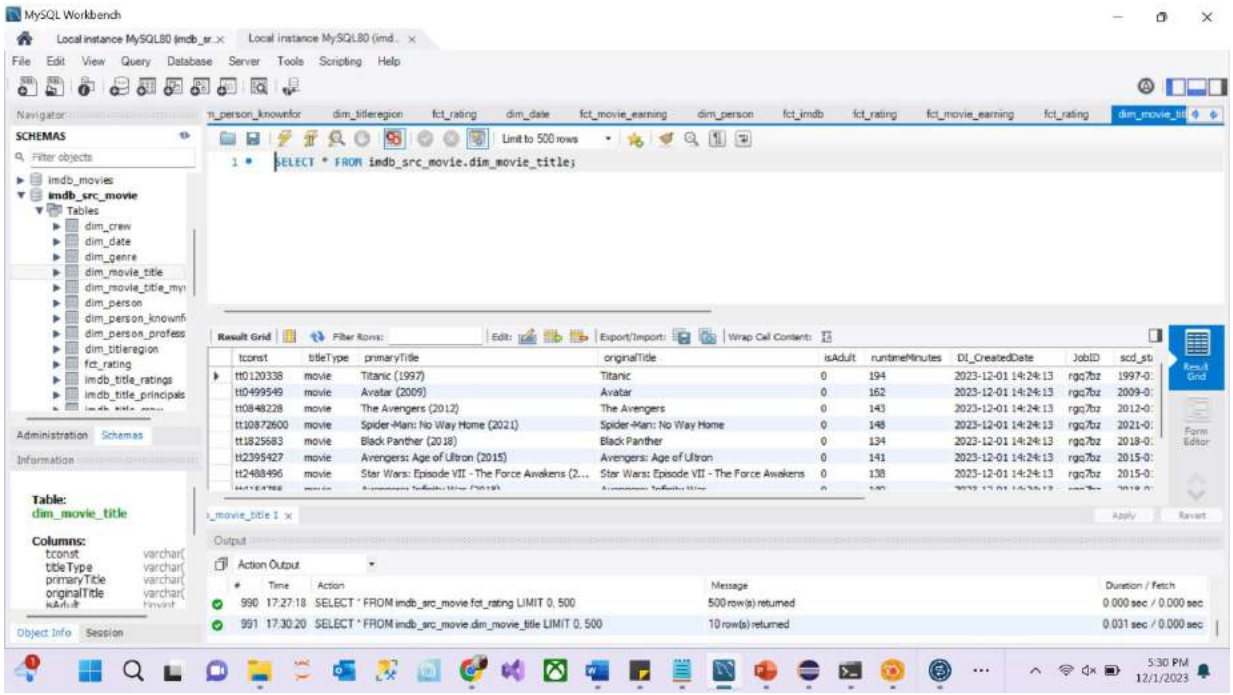

Talend Cloud Data Fabric (R2022-09) | IMDB (Connection: Local)
Job DIM_Genre
Basic Run
Debug Run
Advanced settings
Target Exec
Memory Run
Run
Kill
Clear
Starting job DIM_Genre at 14:09 01/12/2023.
[statistics] connecting to socket on port 3553
[statistics] connected
[statistics] disconnected
Job DIM_Genre ended at 14:19 01/12/2023. [Exit code = 0]
Line limit 1000000
Wrap
tDBInput_1 ("stg_imdb_title_basics")
tDBInput_2 ("dim_movie_title_mysql")
tDBOutput_1 (DIM_Genre)
tMap_1
tMap_2
tNormalize_1
DIM_Genre
"stg_imdb_title_basics"
"dim_movie_title_mysql"

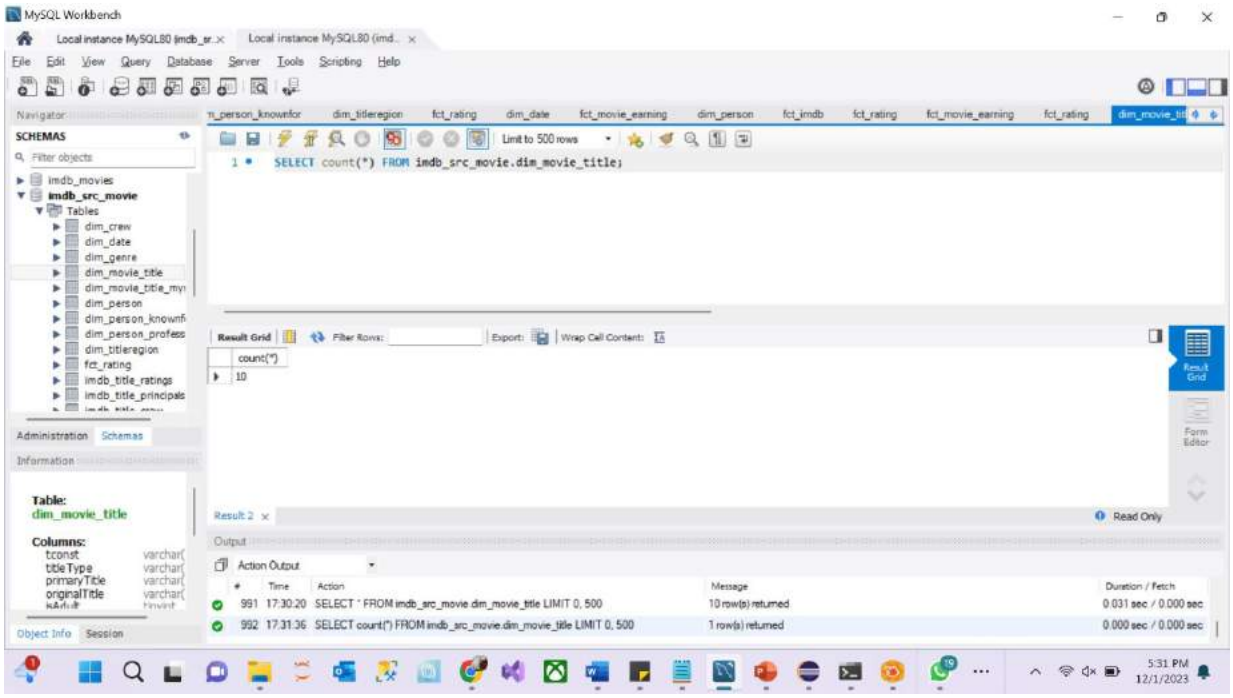

MySQL Workbench
SELECT * FROM imdb_src_movie.dim_genre;
Table: dim_genre
Columns:
GenreSK int PK
MovieTitle_SK int
tconst varchar(
genres varchar(
DI_CreatedDate datetim
JobId varchar(
GenreSK MovieTitle_SK tconst genres DI_CreatedDate JobId
1 2 tt0000502 2023-12-01 14:10:21 WSyP6X
2 6 tt0000574 Action 2023-12-01 14:10:21 WSyP6X
3 6 tt0000574 Adventure 2023-12-01 14:10:21 WSyP6X
4 6 tt0000574 Biography 2023-12-01 14:10:21 WSyP6X
5 8 tt0000591 Drama 2023-12-01 14:10:21 WSyP6X
6 10 tt0000615 Drama 2023-12-01 14:10:21 WSyP6X
7 12 tt0000630 Drama 2023-12-01 14:10:21 WSyP6X
8 14 tt0000675 Drama 2023-12-01 14:10:21 WSyP6X
9 14:14:24 DROP TABLE `imdb_src_movie`.`imdb_title_principals` 0 row(s) affected 0.015 sec
10 14:21:02 SELECT * FROM imdb_src_movie.dim_genre LIMIT 0, 500 500 row(s) returned 0.000 sec / 0.000 sec



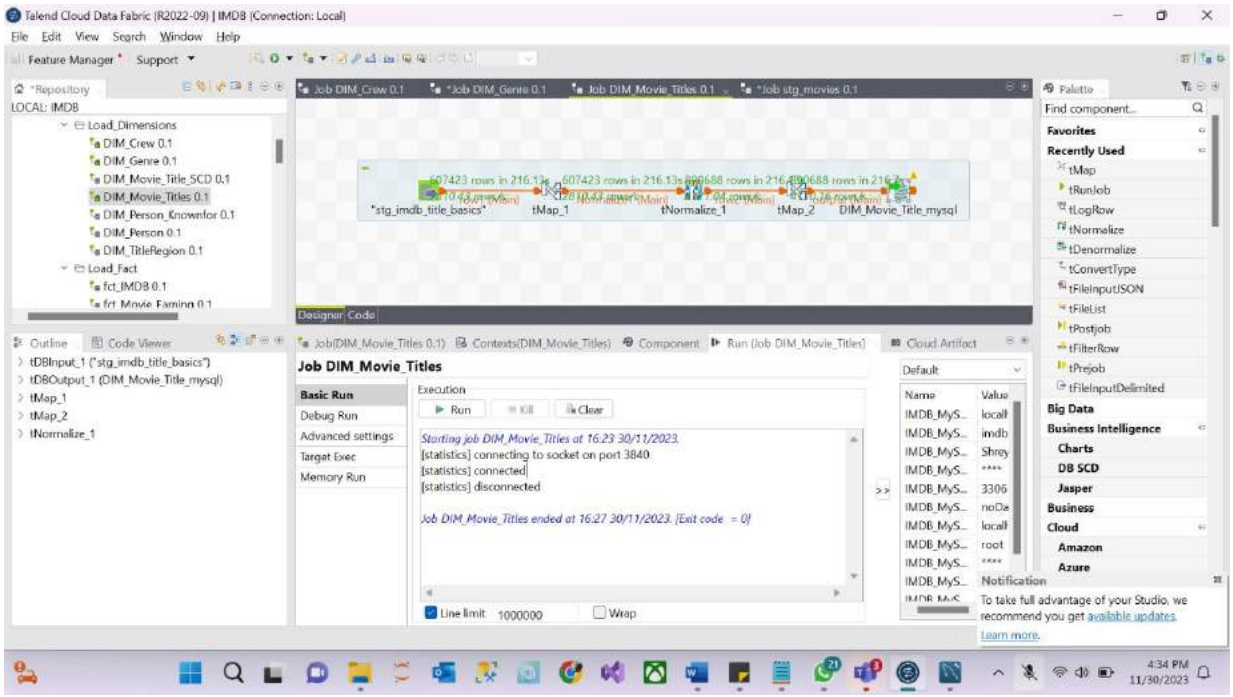
DIM Movie Title SCD:

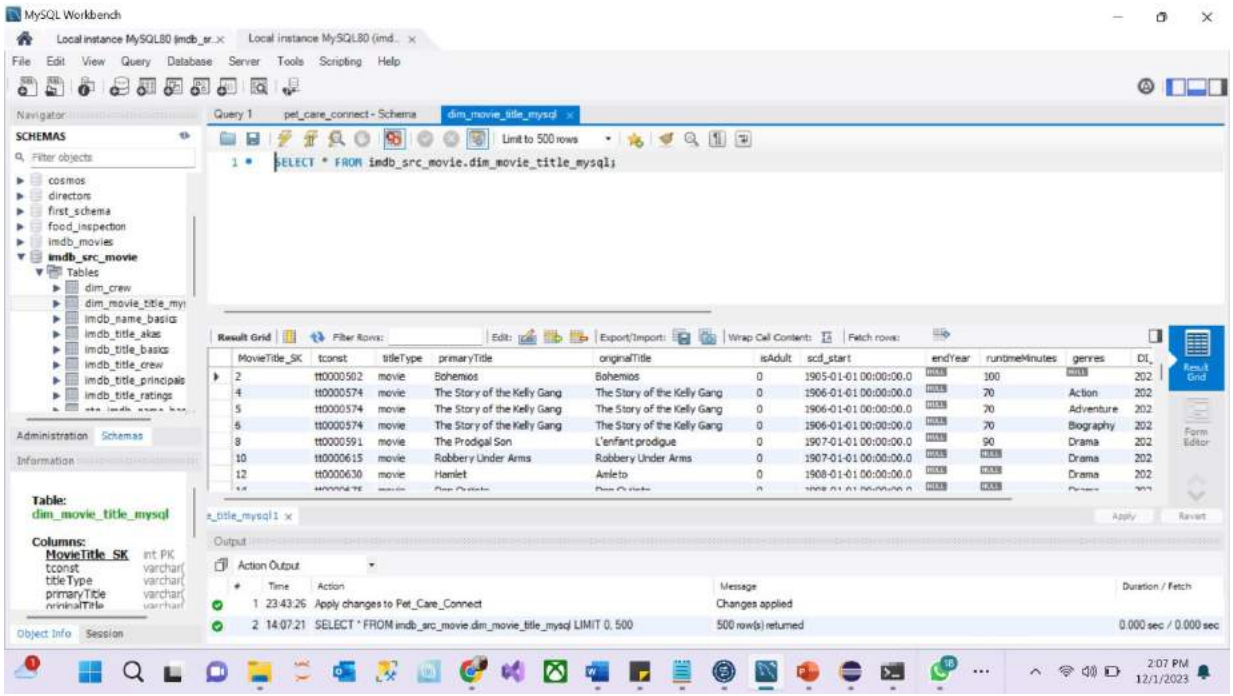



MySQL Workbench
Local instance MySQL80 (imdb_sr...
Local instance MySQL80 (imd...
File Edit View Query Database Server Tools Scripting Help
SCHEMAS
imdb_movies
imdb_src_movie
Tables
dim_crew
dim_date
dim_genre
dim_movie_title
dim_movie_title_myi
dim_person
dim_person_knownfo
dim_person_profess
dim_titleregion
fct_rating
imdb_title_ratings
imdb_title_principals
Administration Schemas
Information
Table:
dim_movie_title
Columns:
tconst varchar(
titleType varchar(
primaryTitle varchar(
originalTitle varchar(
Object Info Session
Limit to 500 rows
1 • SELECT * FROM imdb_src_movie.dim_movie_title;
Result Grid Filter Rows: Edit: Export/Import: Wrap Cell Content:
tconst titleType primaryTitle originalTitle isAdult runtimeMinutes DI_CreatedDate JobID scd_sti
tt0120338 movie Titanic (1997) Titanic 0 194 2023-12-01 14:24:13 rgq7bz 1997-0:
tt0499549 movie Avatar (2009) Avatar 0 162 2023-12-01 14:24:13 rgq7bz 2009-0:
tt0848228 movie The Avengers (2012) The Avengers 0 143 2023-12-01 14:24:13 rgq7bz 2012-0:
tt10872600 movie Spider-Man: No Way Home (2021) Spider-Man: No Way Home 0 148 2023-12-01 14:24:13 rgq7bz 2021-0:
tt1825683 movie Black Panther (2018) Black Panther 0 134 2023-12-01 14:24:13 rgq7bz 2018-0:
tt2395427 movie Avengers: Age of Ultron (2015) Avengers: Age of Ultron 0 141 2023-12-01 14:24:13 rgq7bz 2015-0:
tt2488496 movie Star Wars: Episode VII - The Force Awakens (2... Star Wars: Episode VII - The Force Awakens 0 138 2023-12-01 14:24:13 rgq7bz 2015-0:
Result Grid Form Editor
Apply Revert
Output
Action Output
Time Action Message Duration / Fetch
990 17:27:18 SELECT * FROM imdb_src_movie.fct_rating LIMIT 0, 500 500 row(s) returned 0.000 sec / 0.000 sec
991 17:30:20 SELECT * FROM imdb_src_movie.dim_movie_title LIMIT 0, 500 10 row(s) returned 0.031 sec / 0.000 sec
5:30 PM
12/1/2023

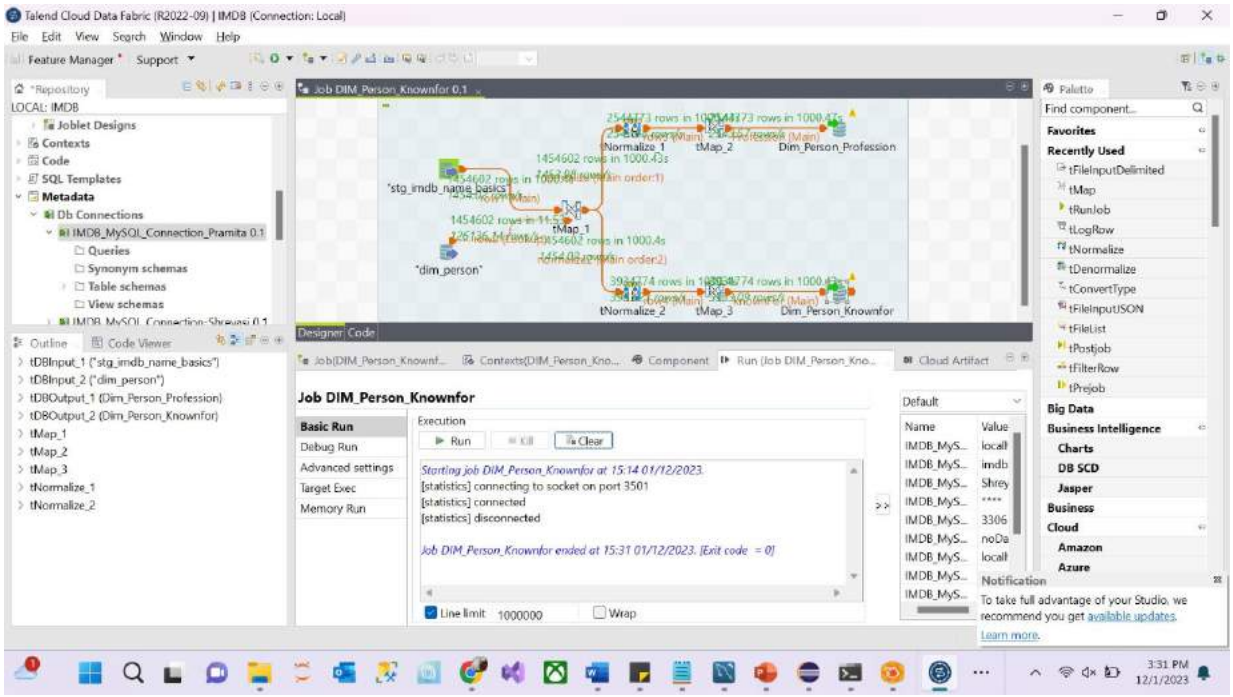

MySQL Workbench
Local instance MySQL80 (imdb_sr...
Local instance MySQL80 (imd...
File Edit View Query Database Server Tools Scripting Help
SCHEMAS
imdb_movies
imdb_src_movie
Tables
dim_crew
dim_date
dim_genre
dim_movie_title
dim_movie_title_myi
dim_person
dim_person_knownfo
dim_person_profess
dim_titleregion
fct_rating
imdb_title_ratings
imdb_title_principals
Administration Schemas
Information
Table:
dim_movie_title
Columns:
tconst varchar(
titleType varchar(
primaryTitle varchar(
originalTitle varchar(
Object Info Session
Limit to 500 rows
1 • SELECT count(*) FROM imdb_src_movie.dim_movie_title;
Result Grid Filter Rows: Export: Wrap Cell Content:
count(*)
10
Result Grid Form Editor
Result 2
Read Only
Output
Action Output
Time Action Message Duration / Fetch
991 17:30:20 SELECT * FROM imdb_src_movie.dim_movie_title LIMIT 0, 500 10 row(s) returned 0.031 sec / 0.000 sec
992 17:31:36 SELECT count(*) FROM imdb_src_movie.dim_movie_title LIMIT 0, 500 1 row(s) returned 0.000 sec / 0.000 sec
5:31 PM
12/1/2023

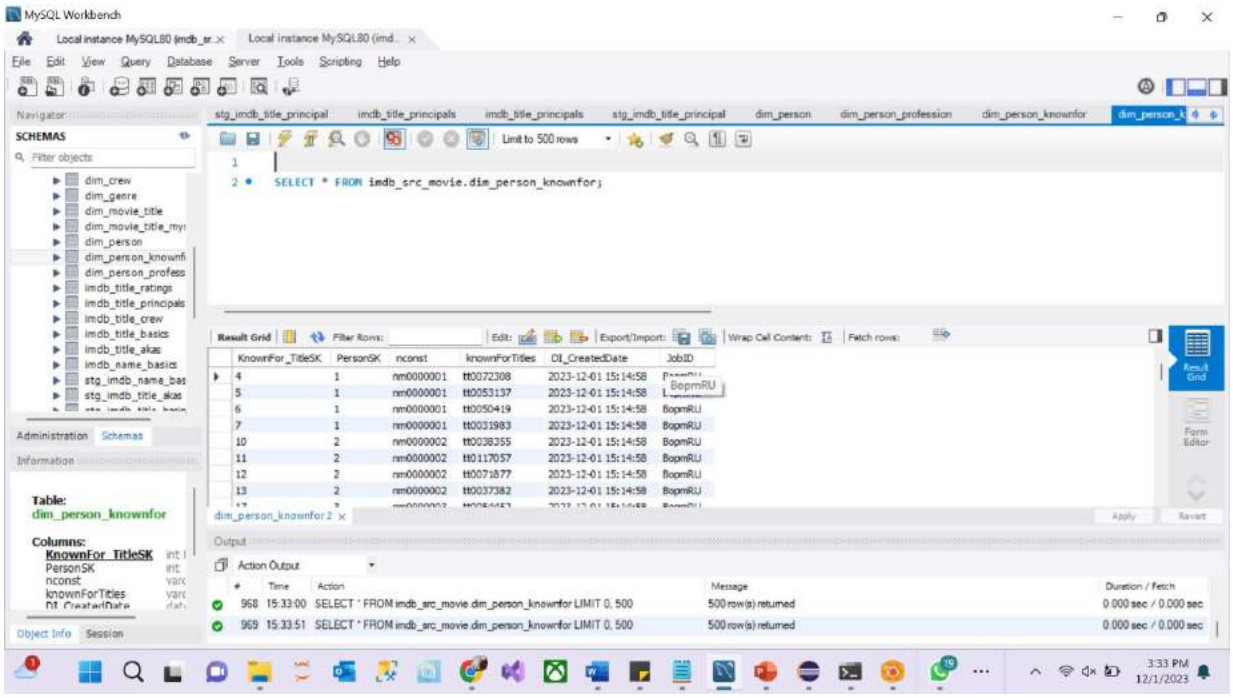
DIM Movie Titles:

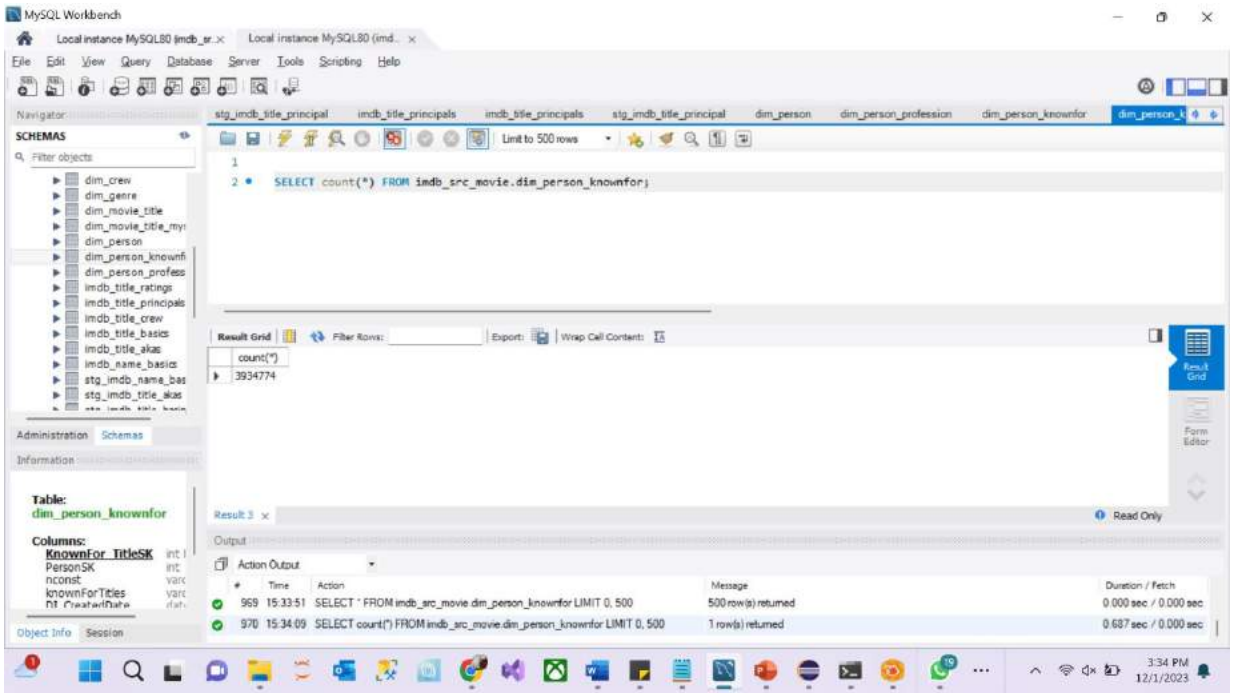




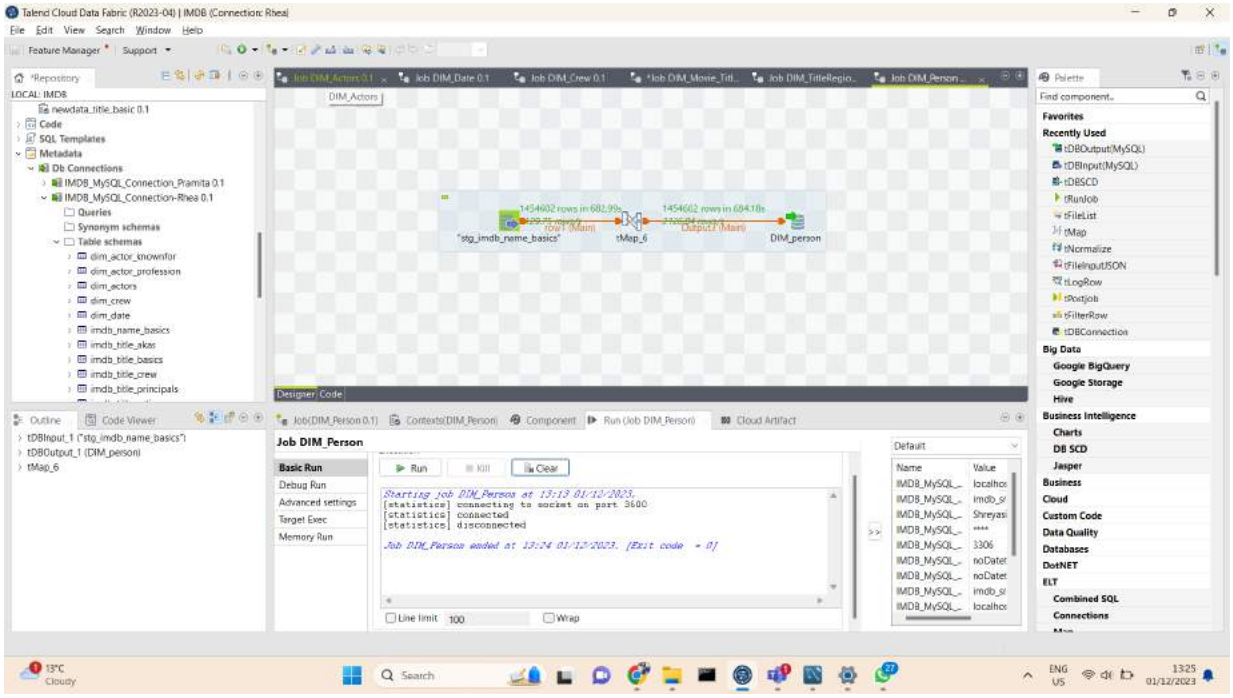
DIM Person Known for:

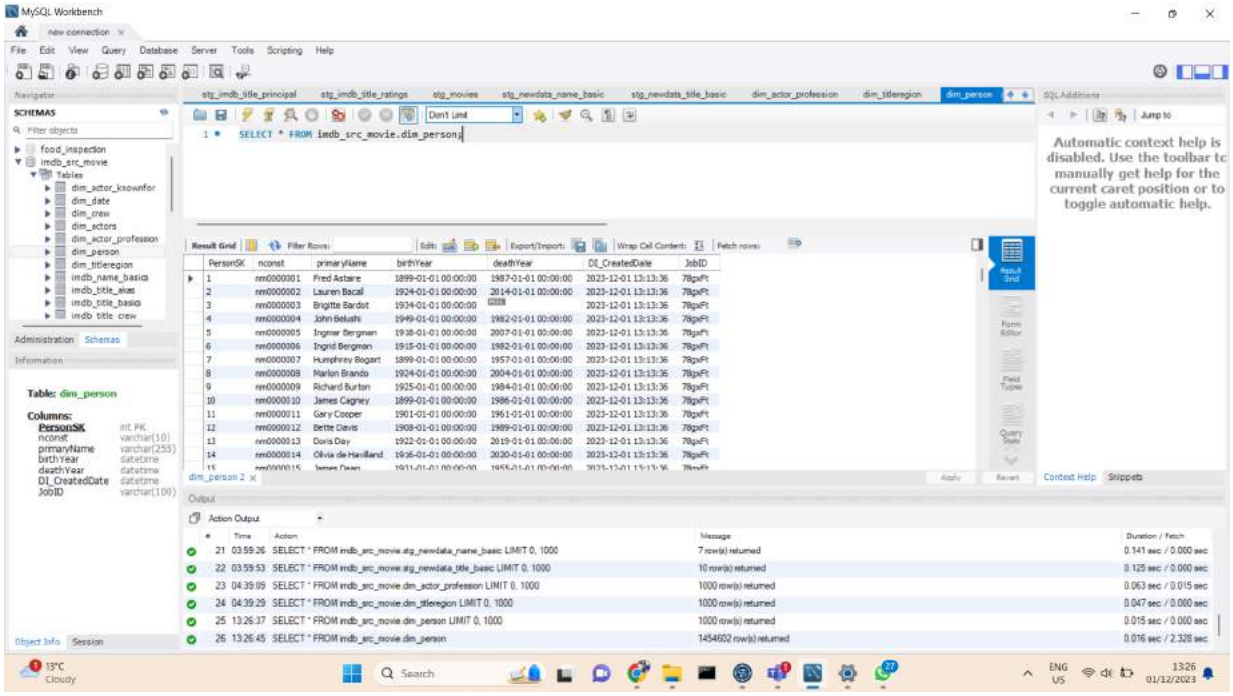


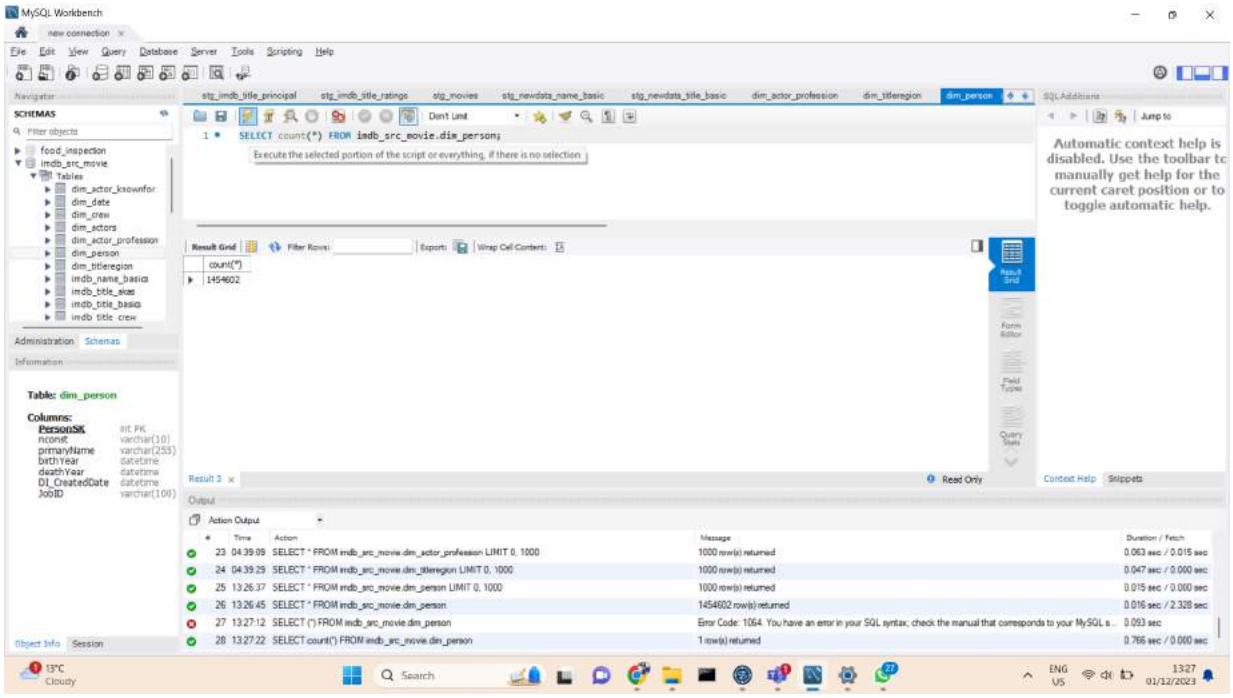



MySQL Workbench
SELECT count(*) FROM imdb_src_movie.dim_person_knownfor;
count(*)
3934774
Table:
dim_person_knownfor
Columns:
KnownFor_TitleSK
PersonSK
nconst
knownForTitles
Action Output
969 15:33:51 SELECT * FROM imdb_src_movie.dim_person_knownfor LIMIT 0, 500 500 row(s) returned 0.000 sec / 0.000 sec
970 15:34:09 SELECT count(*) FROM imdb_src_movie.dim_person_knownfor LIMIT 0, 500 1 row(s) returned 0.687 sec / 0.000 sec

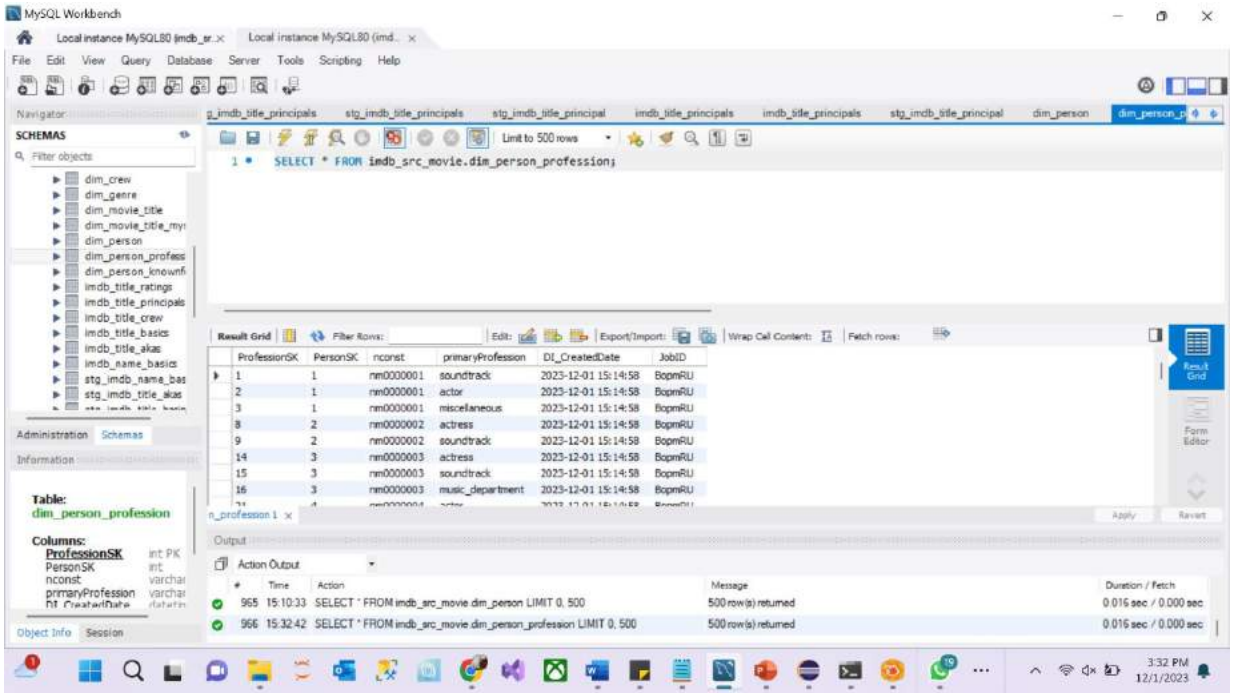
DIM Person:

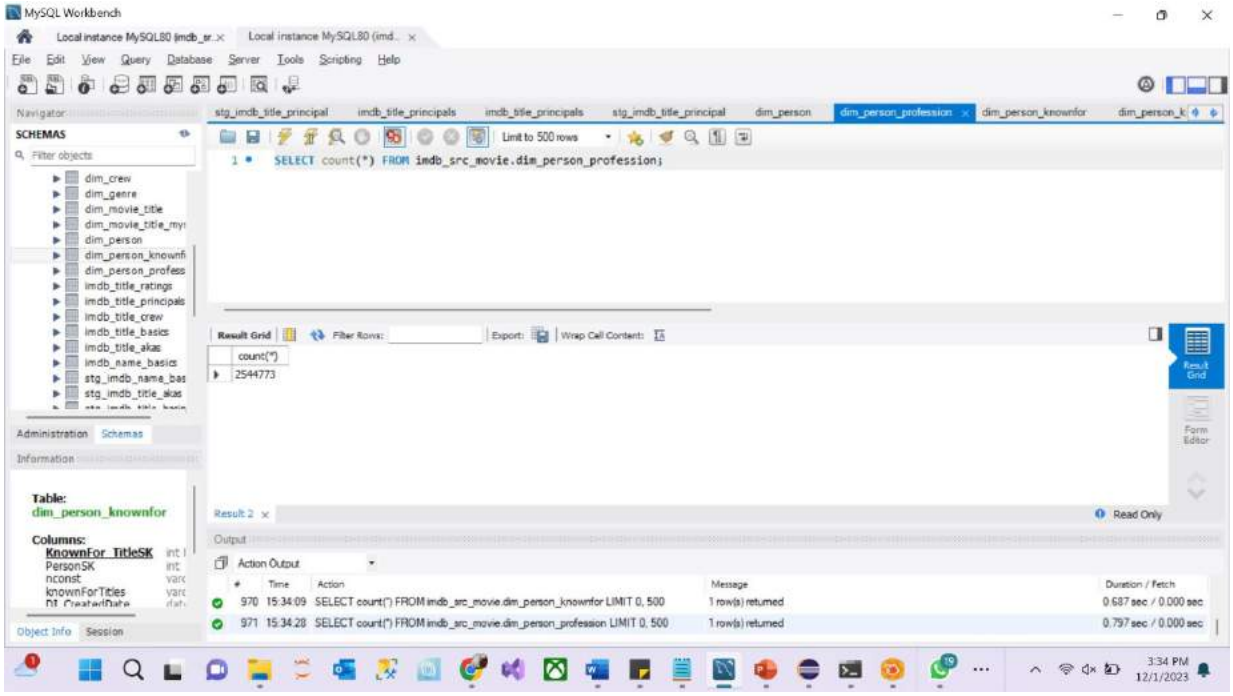

Talend Cloud Data Fabric (R2023-04) | IMDB (Connection: Rhea)
Job DIM_Person
Basic Run
Debug Run
Advanced settings
Target Exec
Memory Run
Starting job DIM_Person at 13:13 01/12/2023.
[statistics] connecting to socket on port 3600
[statistics] connected
[statistics] disconnected
Job DIM_Person ended at 13:24 01/12/2023. [Exit code = 0]
"stg_imdb_name_basics"
tMap_6
DIM_person


MySQL Workbench
SELECT * FROM imdb_src_movie.dim_person;
Table: dim_person
Columns:
PersonSK int PK
nconst varchar(10)
primaryName varchar(255)
birthYear datetime
deathYear datetime
DI_CreatedDate datetime
JobID varchar(100)

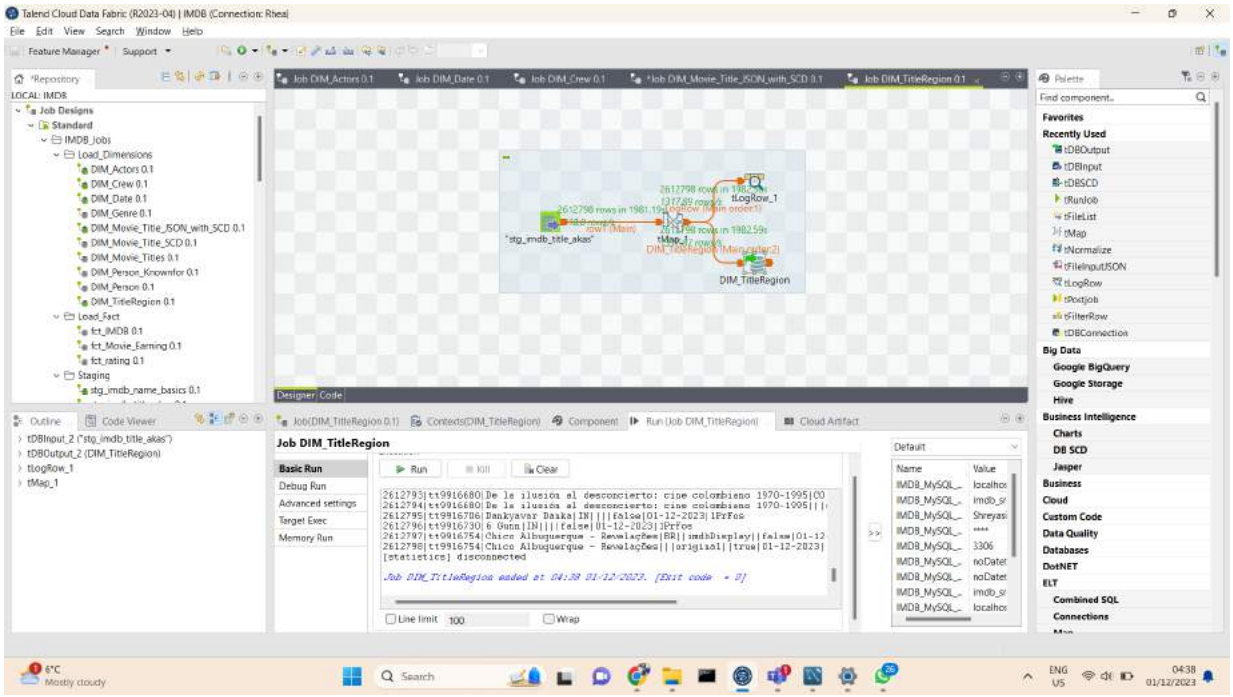

MySQL Workbench
SELECT count(*) FROM imdb_src_movie.dim_person;
Execute the selected portion of the script or everything, if there is no selection
Table: dim_person
Columns:
PersonSK int PK
nconst varchar(10)
primaryName varchar(255)
birthYear datetime
deathYear datetime
DI_CreatedDate datetime
JobID varchar(100)

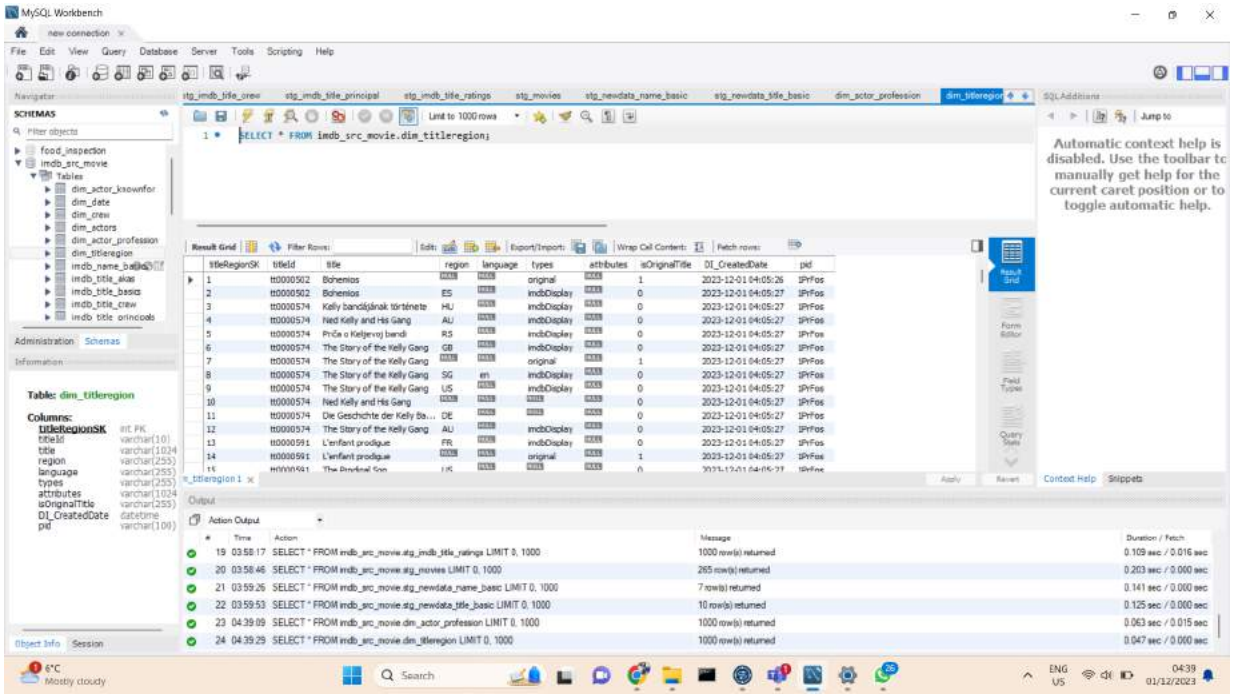
DIM Person Profession:

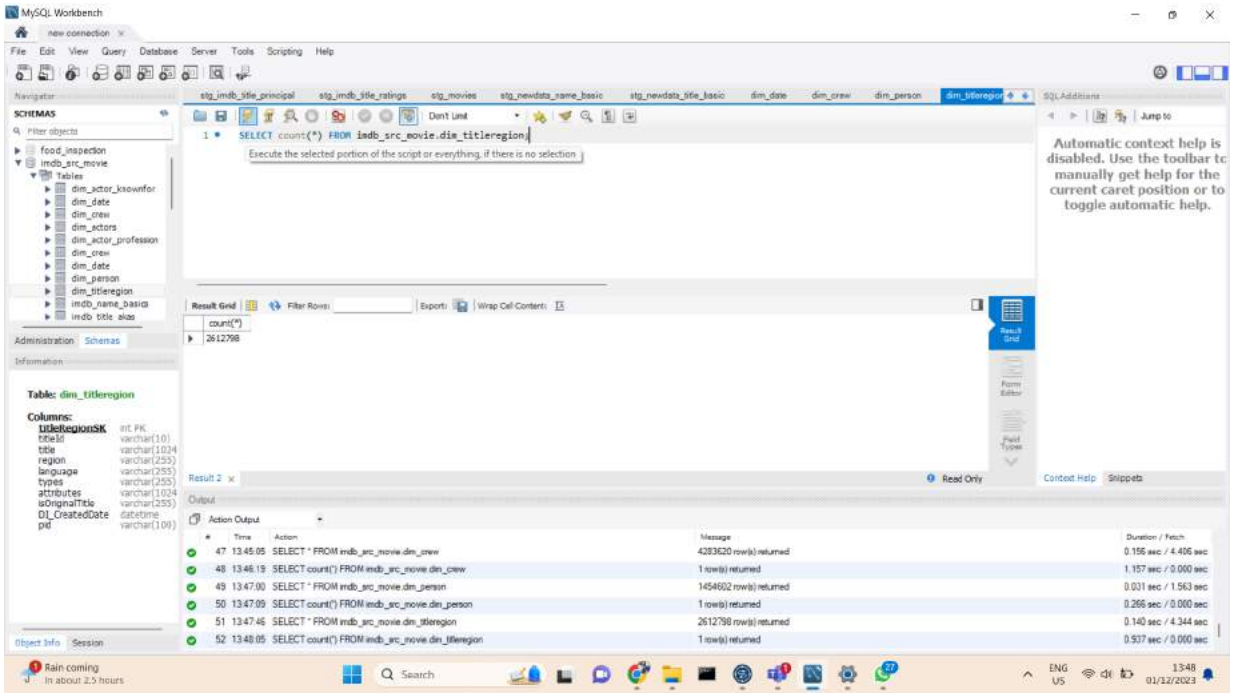




DIM Title Region:

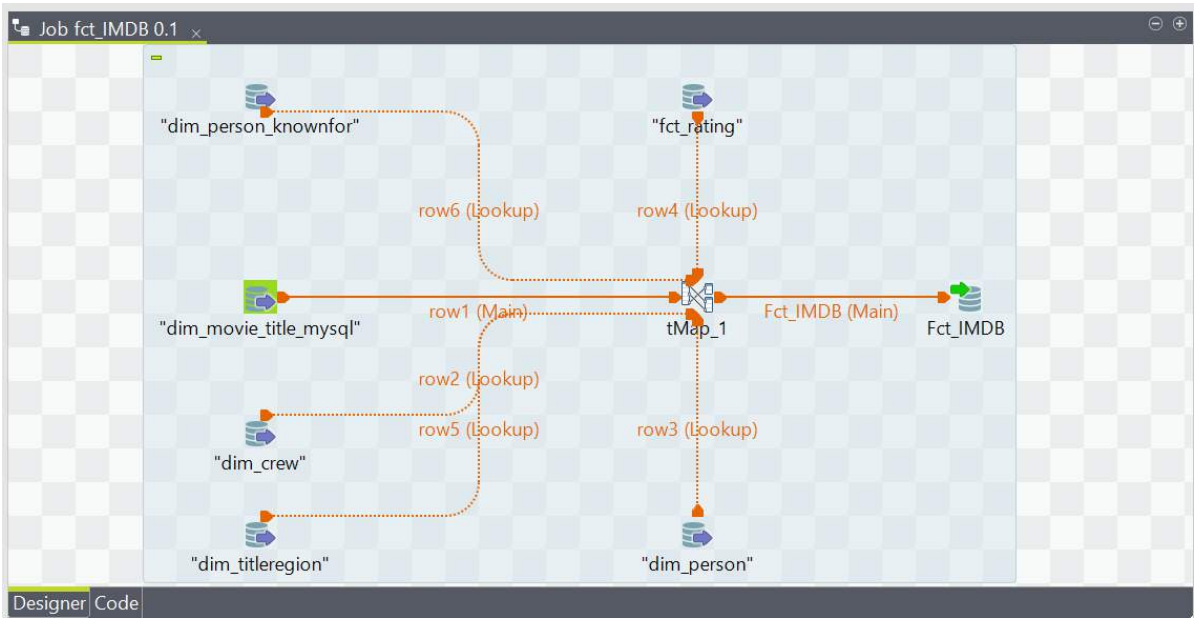






Fact Tables:

Fact IMDB:



```
1 •  select * from fct_imdb;
```

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content: | Fetch rov

IMDB_SK	CrewSK	PersonSK	titleRegionSK	MovieTitle_SK	RatingSK	KnownFor_TitleSK	DI_CreatedDate	JobID
1	1	78737	2	1	1	273677	2023-12-02 23:45:08	932oOL
2	2	90925	2	1	1	315725	2023-12-02 23:45:08	932oOL
3	3	27130	2	1	1	0	2023-12-02 23:45:08	932oOL
4	4	220366	2	1	1	765605	2023-12-02 23:45:08	932oOL
5	5	225962	2	1	1	785128	2023-12-02 23:45:08	932oOL
6	6	278664	12	2	2	968730	2023-12-02 23:45:08	932oOL
7	7	278665	12	2	2	968731	2023-12-02 23:45:08	932oOL
8	8	616427	12	2	2	1713276	2023-12-02 23:45:08	932oOL
9	9	836037	12	2	2	2394060	2023-12-02 23:45:08	932oOL
10	10	278660	12	2	2	968717	2023-12-02 23:45:08	932oOL
11	11	111817	12	2	2	388148	2023-12-02 23:45:08	932oOL

fct_imdb 1 ×

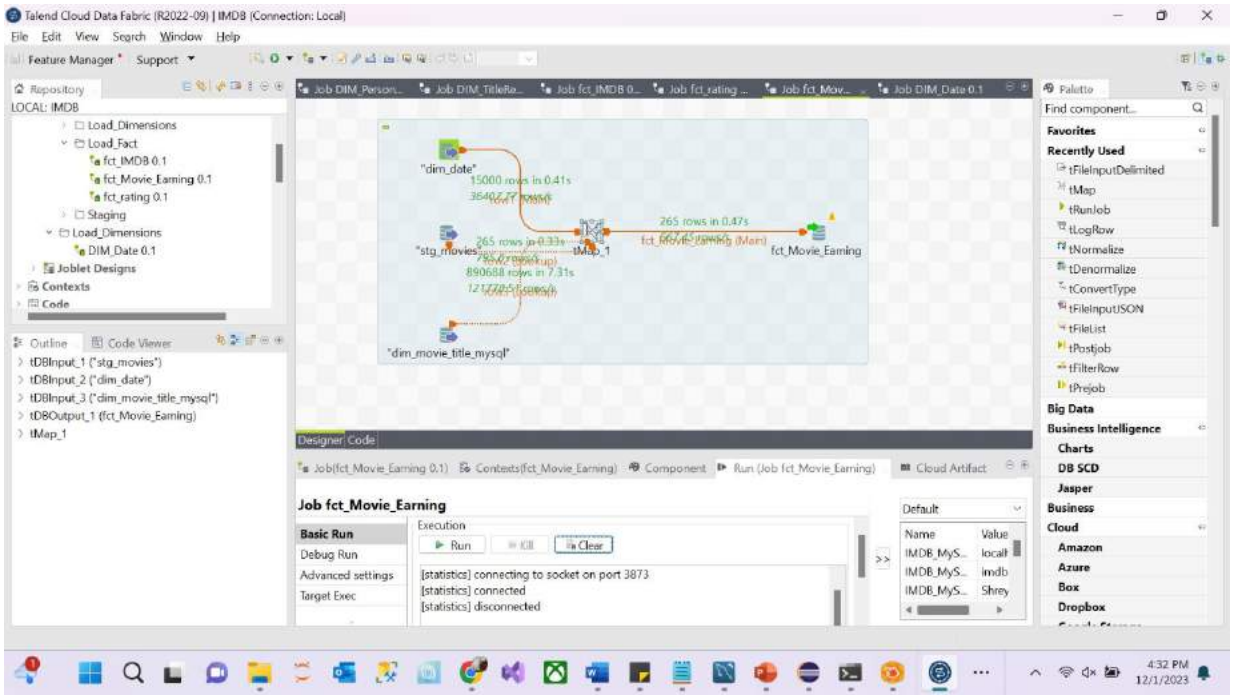
Apply

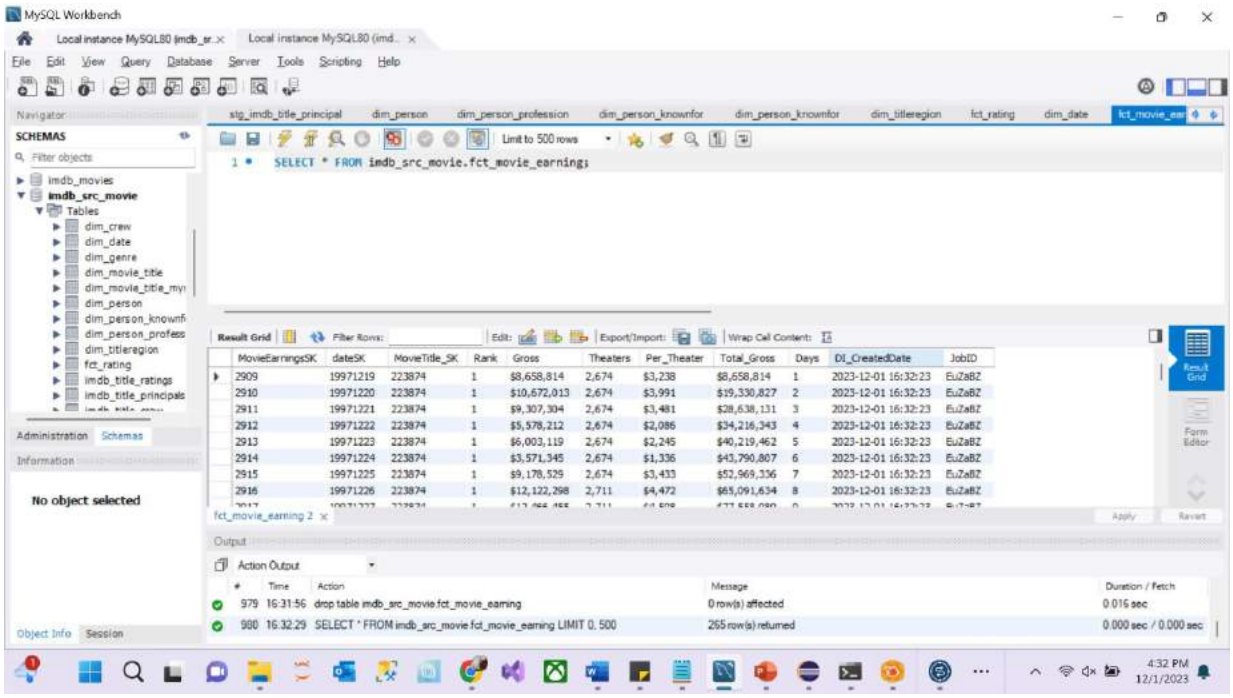
```
1 •  select COUNT(*) from fct_imdb;
```

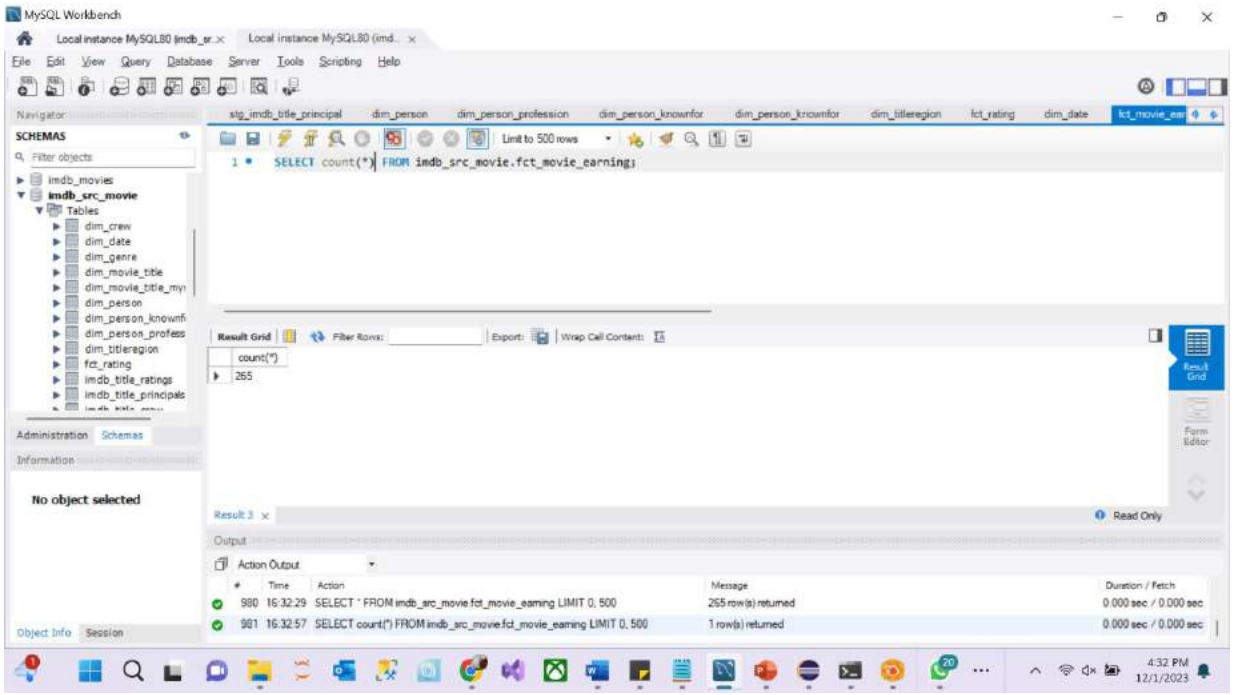
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

COUNT(*)
4299743

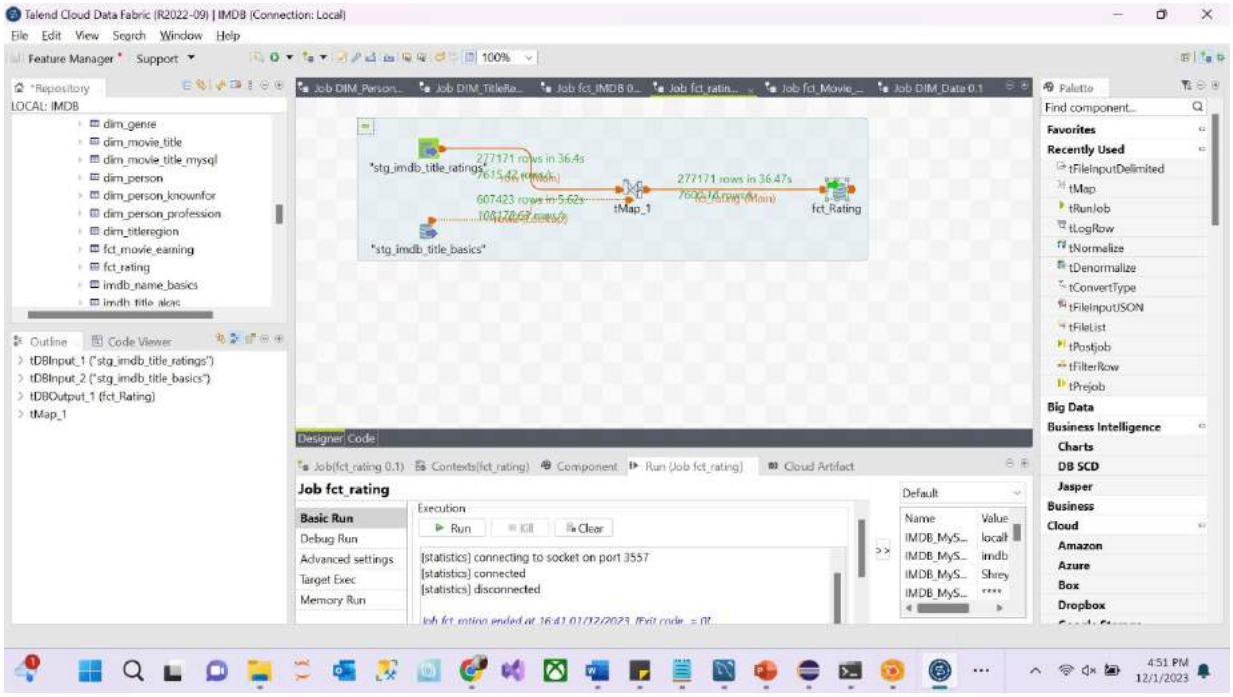
Fact Movie Earnings:

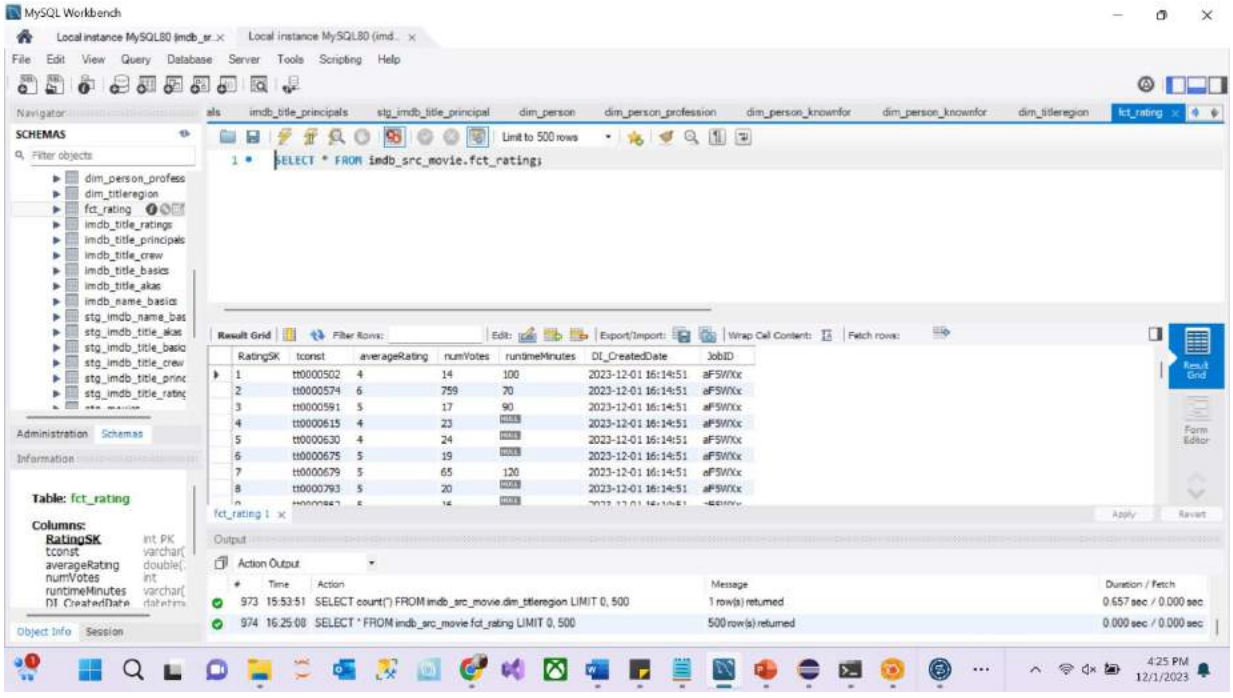


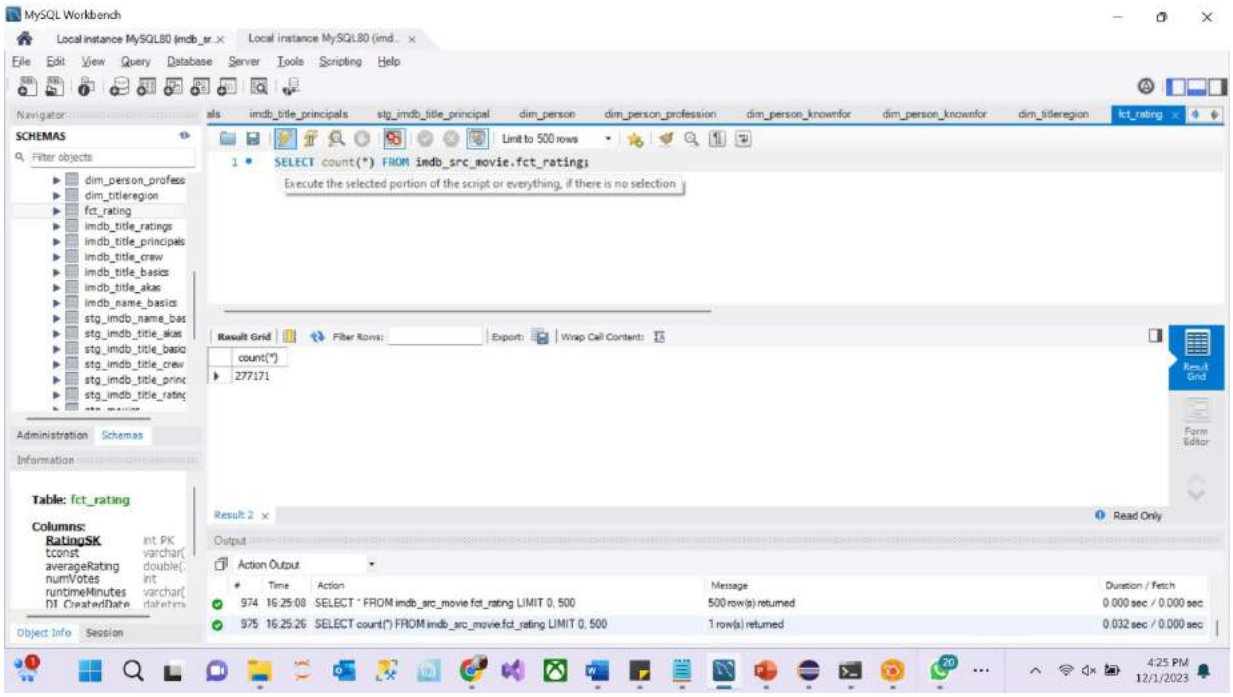



MySQL Workbench
File Edit View Query Database Server Tools Scripting Help
SCHEMAS
imdb_movies
imdb_src_movie
Tables
dim_crew
dim_date
dim_genre
dim_movie_title
dim_person
fct_rating
imdb_title_ratings
imdb_title_principals
SELECT count(*) FROM imdb_src_movie.fct_movie_earning;
count(*)
265
No object selected
Action Output
980 16:32:29 SELECT * FROM imdb_src_movie.fct_movie_earning LIMIT 0, 500 265 row(s) returned 0.000 sec / 0.000 sec
981 16:32:57 SELECT count(*) FROM imdb_src_movie.fct_movie_earning LIMIT 0, 500 1 row(s) returned 0.000 sec / 0.000 sec
4:32 PM
12/1/2023

Fact Rating:


Talend Cloud Data Fabric (R2022-09) | IMDB (Connection: Local)
File Edit View Search Window Help
Feature Manager Support
Repository
LOCAL: IMDB
dim_genre
dim_movie_title
dim_movie_title_mysql
dim_person
dim_person_knownfor
dim_person_profession
dim_titleregion
fct_movie_earning
fct_rating
imdb_name_basics
imdb_title_akas
Outline
Code Viewer
tDBInput_1 ("stg_imdb_title_ratings")
tDBInput_2 ("stg_imdb_title_basics")
tDBOutput_1 (fct_Rating)
tMap_1
"stg_imdb_title_ratings"
277171 rows in 36.4s
607423 rows in 5.62s
"stg_imdb_title_basics"
tMap_1
277171 rows in 36.47s
fct_Rating
Designer Code
Job(fct_rating 0.1)
Contexts(fct_rating)
Component
Run (Job fct_rating)
Cloud Artifact
Job fct_rating
Basic Run
Debug Run
Advanced settings
Target Exec
Memory Run
Execution
Run
Kill
Clear
[statistics] connecting to socket on port 3557
[statistics] connected
[statistics] disconnected
Default
Name
Value
Palette
Find component...
Favorites
Recently Used
tFileInputDelimited
tMap
tRunJob
tLogRow
tNormalize
tDenormalize
tConvertType
tFileInputJSON
tFileList
tPostjob
tFilterRow
tPrejob
Big Data
Business Intelligence
Charts
DB SCD
Jasper
Business
Cloud
Amazon
Azure
Box
Dropbox
4:51 PM
12/1/2023


MySQL Workbench
Local instance MySQL80 (imdb_sr...
Local instance MySQL80 (imd...
File Edit View Query Database Server Tools Scripting Help
Navigator
SCHEMAS
Filter objects
dim_person_profess
dim_titleregion
fct_rating
imdb_title_ratings
imdb_title_principals
imdb_title_crew
imdb_title_basics
imdb_title_akas
imdb_name_basics
stg_imdb_name_bas
stg_imdb_title_akas
stg_imdb_title_basic
stg_imdb_title_crew
stg_imdb_title_princ
stg_imdb_title_rating
Administration Schemas
Information
Table: fct_rating
Columns:
RatingSK int PK
tconst varchar(
averageRating double(
numVotes int
runtimeMinutes varchar(
DI_CreatedDate datetime
Object Info Session
imdb_title_principals
stg_imdb_title_principal
dim_person
dim_person_profession
dim_person_knownfor
dim_person_knownfor
dim_titleregion
fct_rating
Limit to 500 rows
1 SELECT * FROM imdb_src_movie.fct_rating;
Result Grid
Filter Rows:
Edit:
Export/Import:
Wrap Cell Content:
Fetch rows:
RatingSK tconst averageRating numVotes runtimeMinutes DI_CreatedDate JobID
1 tt0000502 4 14 100 2023-12-01 16:14:51 aFSWXx
2 tt0000574 6 759 70 2023-12-01 16:14:51 aFSWXx
3 tt0000591 5 17 90 2023-12-01 16:14:51 aFSWXx
4 tt0000615 4 23 2023-12-01 16:14:51 aFSWXx
5 tt0000630 4 24 2023-12-01 16:14:51 aFSWXx
6 tt0000675 5 19 2023-12-01 16:14:51 aFSWXx
7 tt0000679 5 65 120 2023-12-01 16:14:51 aFSWXx
8 tt0000793 5 20 2023-12-01 16:14:51 aFSWXx
fct_rating 1
Apply
Revert
Output
Action Output
Time Action Message Duration / Fetch
973 15:53:51 SELECT count(*) FROM imdb_src_movie.dim_titleregion LIMIT 0, 500 1 row(s) returned 0.657 sec / 0.000 sec
974 16:25:08 SELECT * FROM imdb_src_movie.fct_rating LIMIT 0, 500 500 row(s) returned 0.000 sec / 0.000 sec
4:25 PM
12/1/2023



SQL Queries -

Trend Analysis:

1. Average movie runtime change over the years:

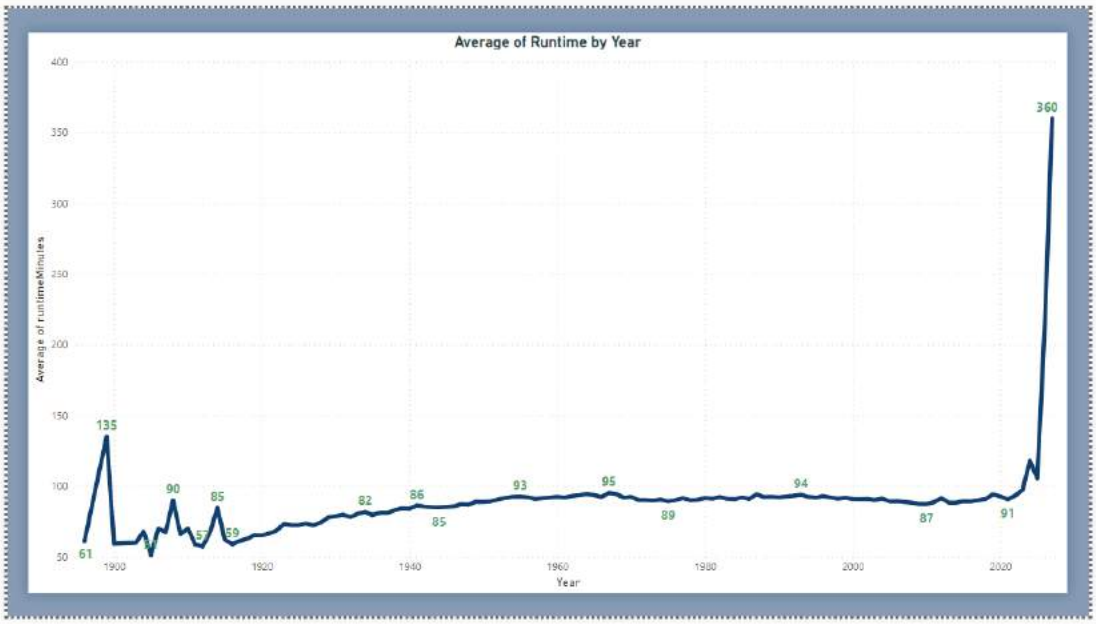
a. For every year what is the average run-time and plot it on the graph.

select startYear, avg(runtimeMinutes) as AvgRuntime from dim_movie_title_mysql
group by startYear
order by startYear;

```
2 •  select startYear, avg(runtimeMinutes) as AvgRuntime from dim_movie_title_mysql
3    group by startYear
4    order by startYear;
5
```

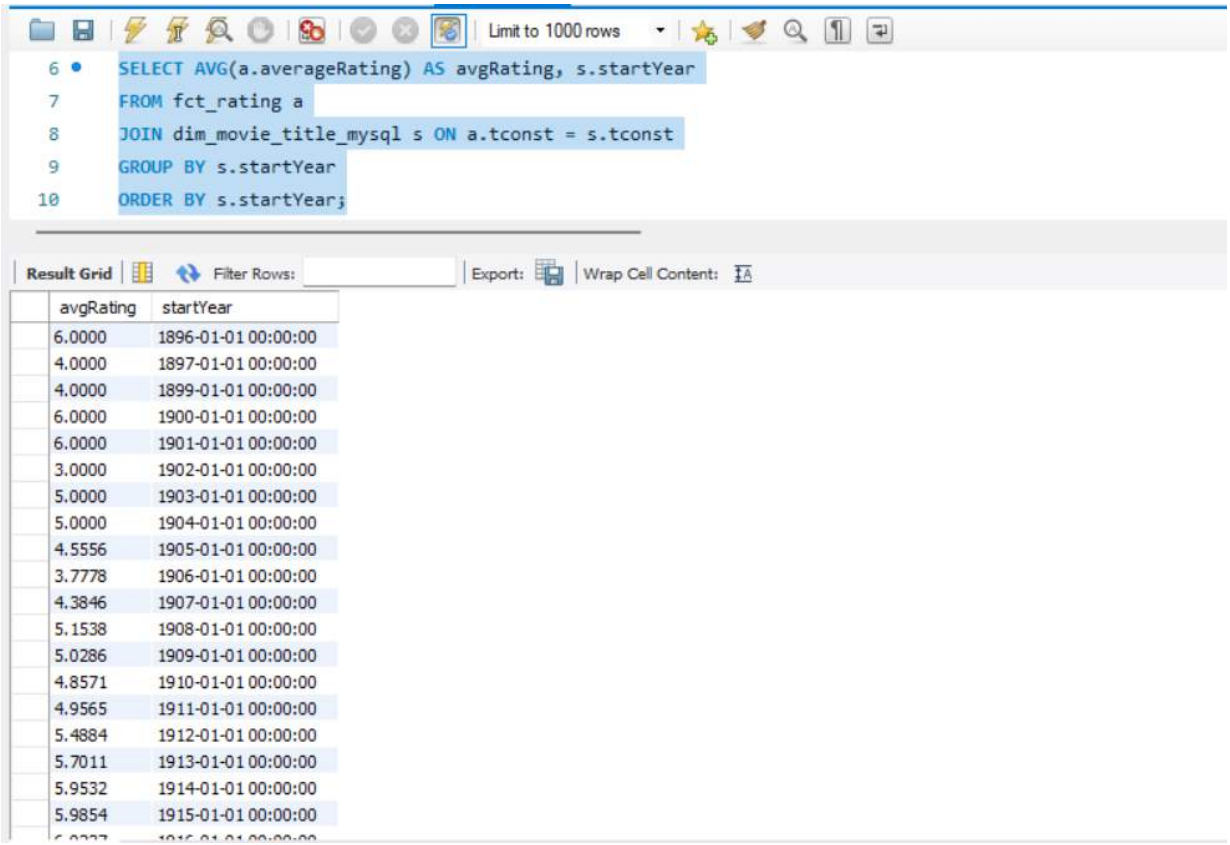
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

startYear	AvgRuntime
1896-01-01 00:00:00	61
1897-01-01 00:00:00	NULL
1898-01-01 00:00:00	NULL
1899-01-01 00:00:00	135
1900-01-01 00:00:00	59.5
1901-01-01 00:00:00	NULL
1902-01-01 00:00:00	NULL
1903-01-01 00:00:00	60
1904-01-01 00:00:00	68
1905-01-01 00:00:00	51
1906-01-01 00:00:00	70
1907-01-01 00:00:00	67.5
1908-01-01 00:00:00	90
1909-01-01 00:00:00	66.25
1910-01-01 00:00:00	69.93333333333334
1911-01-01 00:00:00	58.91304347826087
1912-01-01 00:00:00	57.285714285714285
1913-01-01 00:00:00	67.33333333333333
1914-01-01 00:00:00	84.84065934065934

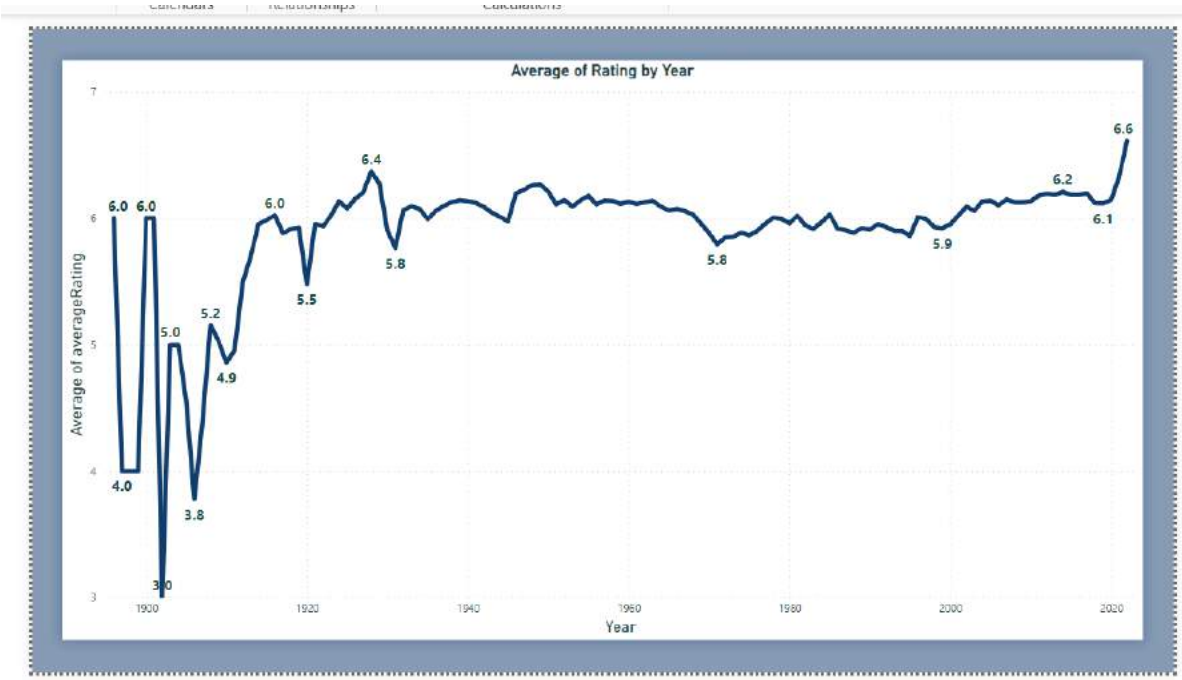


2. Is there a correlation between the average movie rating and the year of release?
a. The goal is to identify in which year we had good movies.

```
SELECT AVG(a.averageRating) AS avgRating, s.startYear
FROM fct_rating a
JOIN dim_movie_title_mysql s ON a.tconst = s.tconst
GROUP BY s.startYear
ORDER BY s.startYear;
```



avgRating	startYear
6.0000	1896-01-01 00:00:00
4.0000	1897-01-01 00:00:00
4.0000	1899-01-01 00:00:00
6.0000	1900-01-01 00:00:00
6.0000	1901-01-01 00:00:00
3.0000	1902-01-01 00:00:00
5.0000	1903-01-01 00:00:00
5.0000	1904-01-01 00:00:00
4.5556	1905-01-01 00:00:00
3.7778	1906-01-01 00:00:00
4.3846	1907-01-01 00:00:00
5.1538	1908-01-01 00:00:00
5.0286	1909-01-01 00:00:00
4.8571	1910-01-01 00:00:00
4.9565	1911-01-01 00:00:00
5.4884	1912-01-01 00:00:00
5.7011	1913-01-01 00:00:00
5.9532	1914-01-01 00:00:00
5.9854	1915-01-01 00:00:00



3. Visualization of the distribution of movie releases per year

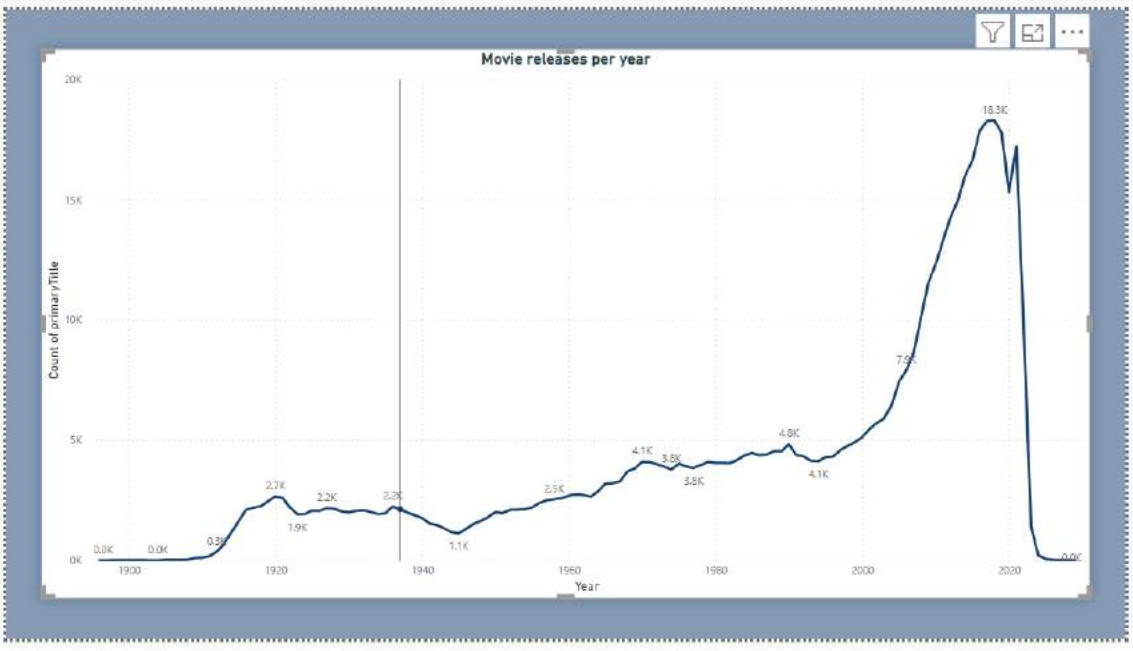
a. No. of movies released every year

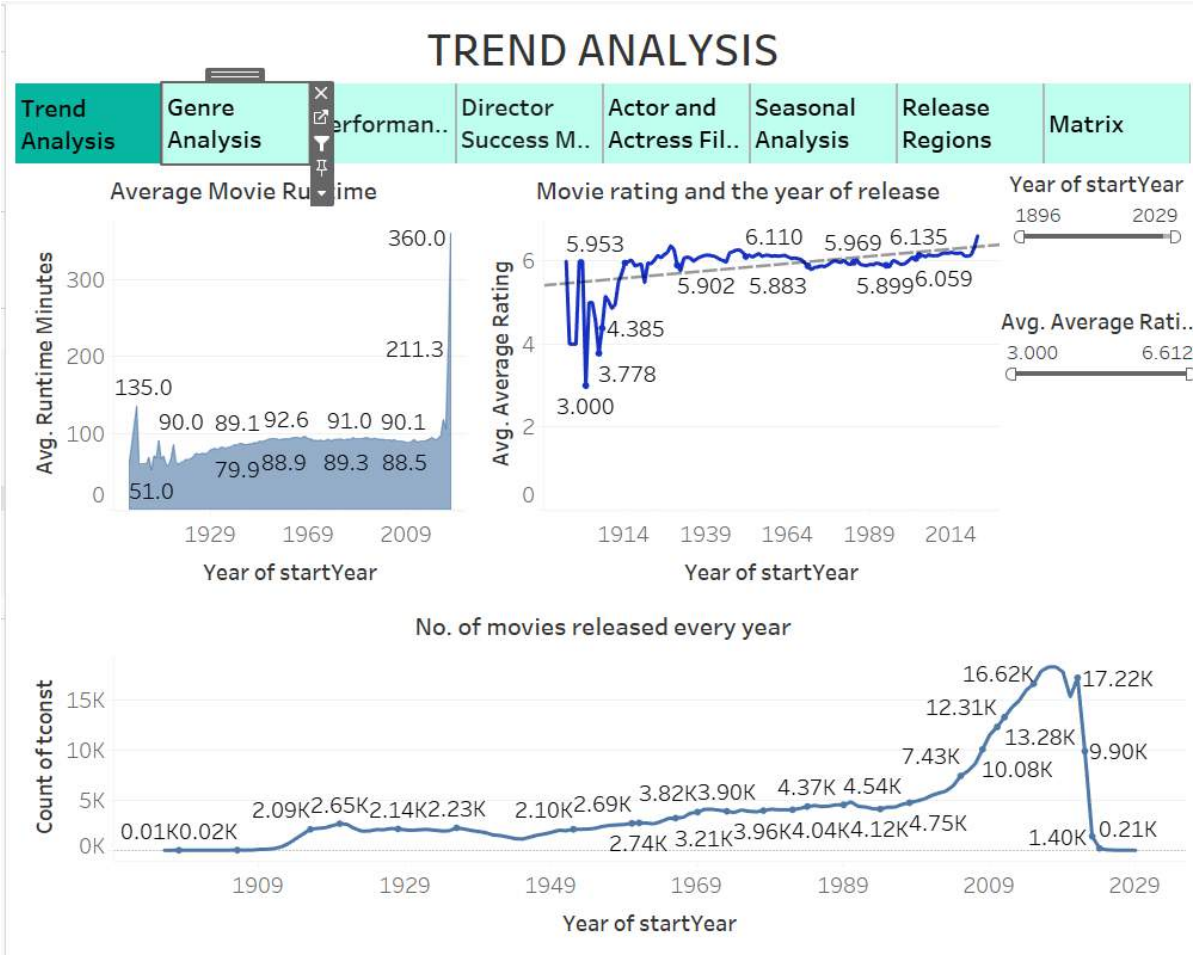
```
SELECT COUNT(tconst) AS movieCount, startYear
FROM dim_movie_title_mysql
GROUP BY startYear;
```

```
11
12 •  SELECT COUNT(tconst) AS movieCount, startYear
13     FROM dim_movie_title_mysql
14     GROUP BY startYear;
15
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

movieCount	startYear
18	1905-01-01 00:00:00
23	1906-01-01 00:00:00
15	1907-01-01 00:00:00
21	1908-01-01 00:00:00
92	1909-01-01 00:00:00
105	1910-01-01 00:00:00
347	1912-01-01 00:00:00
165	1911-01-01 00:00:00
673	1913-01-01 00:00:00
1633	1915-01-01 00:00:00
1156	1914-01-01 00:00:00
2456	1919-01-01 00:00:00
2093	1916-01-01 00:00:00
2188	1917-01-01 00:00:00
2231	1936-01-01 00:00:00
2064	1925-01-01 00:00:00
2244	1918-01-01 00:00:00
2652	1920-01-01 00:00:00
2192	1922-01-01 00:00:00





Genre Analysis:

1. Which genres have seen the most significant ratings in popularity over the past decade?
a. The goal is to identify the popularity of the genres.

```
SELECT
  g.genres,avg(r.averageRating) as avg_rating
FROM
  dim_genre g
JOIN
  dim_movie_title_mysql t ON g.tconst = t.tconst
JOIN
  fct_imdb i ON t.MovieTitle_SK = i.MovieTitle_SK
JOIN
fct_rating r ON r.RatingSK = i.RatingSK
WHERE
  t.startYear BETWEEN '2014-01-01' AND '2023-12-31'

GROUP BY g.genres
ORDER BY
avg_rating DESC;
```

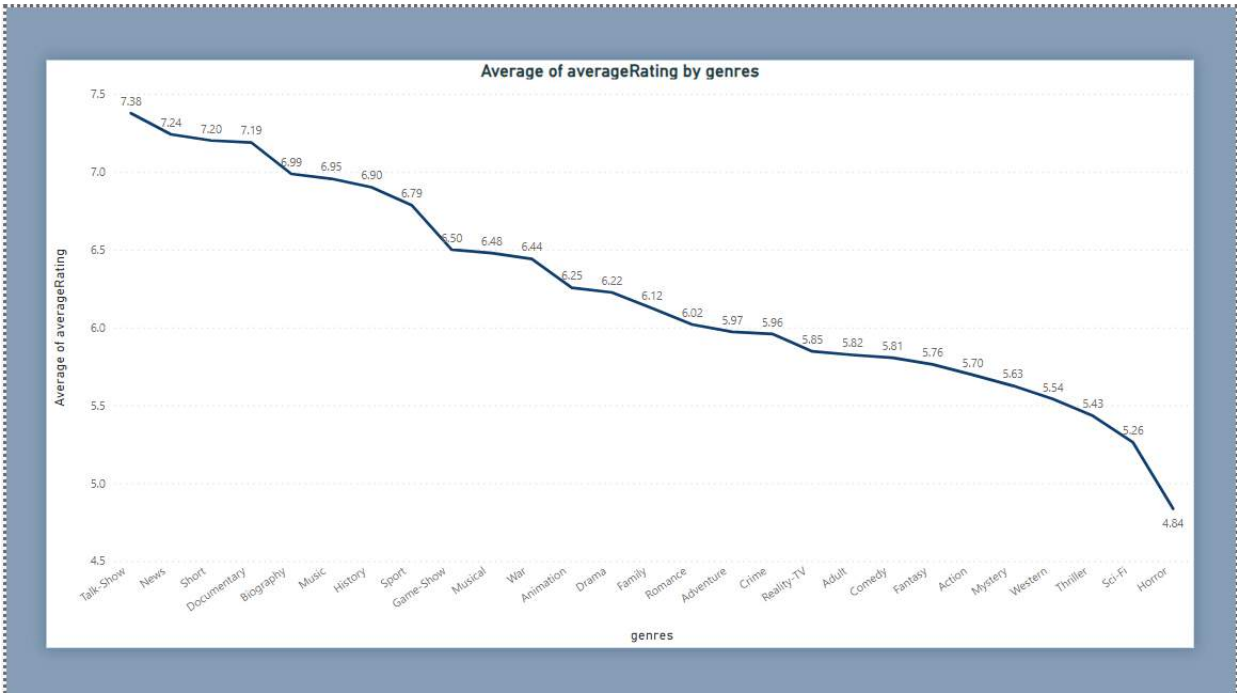
```
56
57 •  SELECT
58      g.genres,avg(r.averageRating) as avg_rating
59    FROM
60      dim_genre g
61    JOIN
62      dim_movie_title_mysql t ON g.tconst = t.tconst
63    JOIN
64      fct_imdb i ON t.MovieTitle_SK = i.MovieTitle_SK
65    JOIN
66    fct_rating r ON r.RatingSK = i.RatingSK
67    WHERE
68      t.startYear BETWEEN '2014-01-01' AND '2023-12-31'
69
70    GROUP BY g.genres
71    ORDER BY
72    avg_rating DESC;
73
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

genres	avg_rating
Talk-Show	7.5714
News	7.2368
Short	7.2000
Documentary	7.1882
Biography	6.9849
Music	6.9549
History	6.8990
Sport	6.7898
Game-Show	6.5000
Musical	6.4798

Result 24

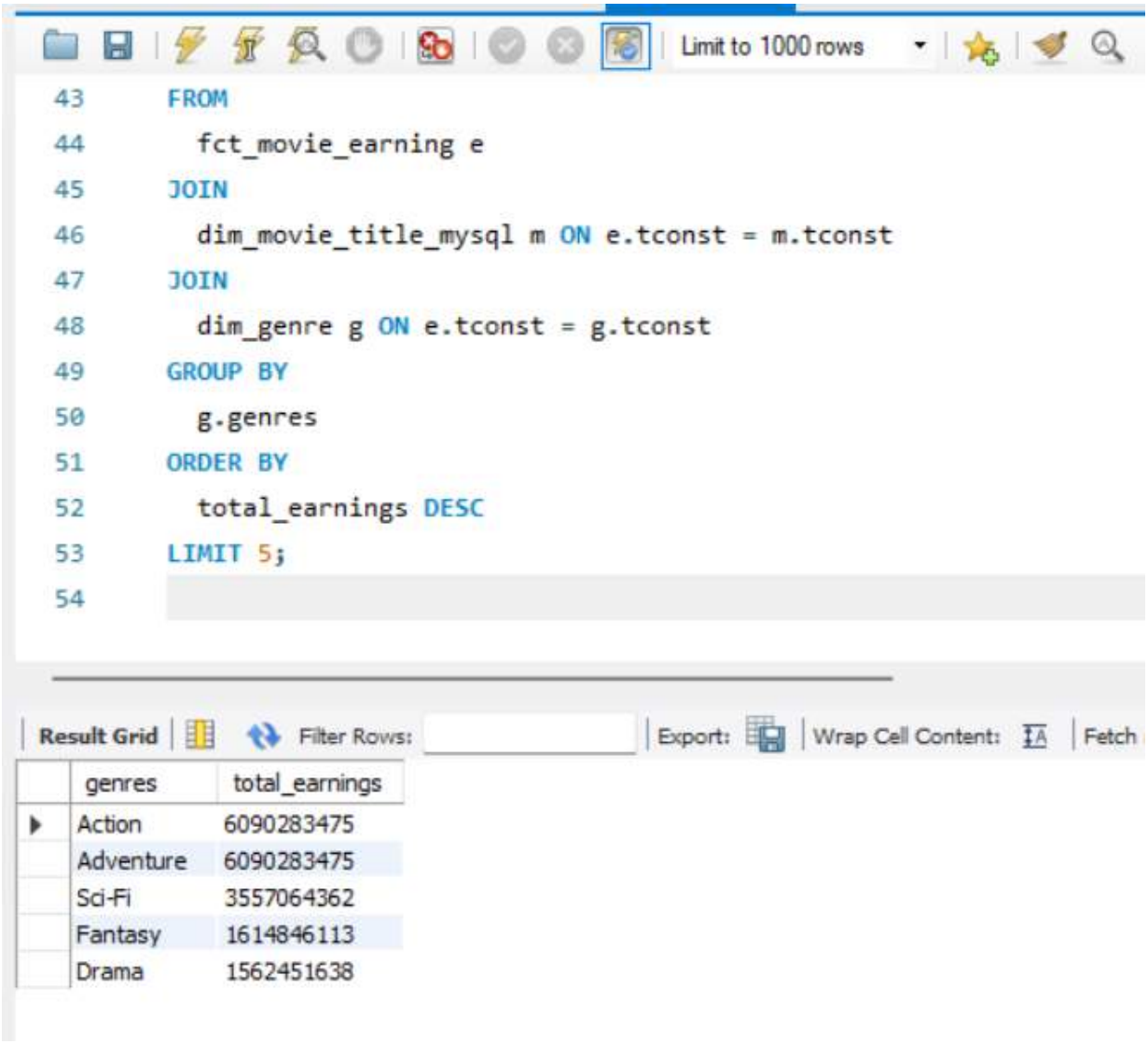
Output



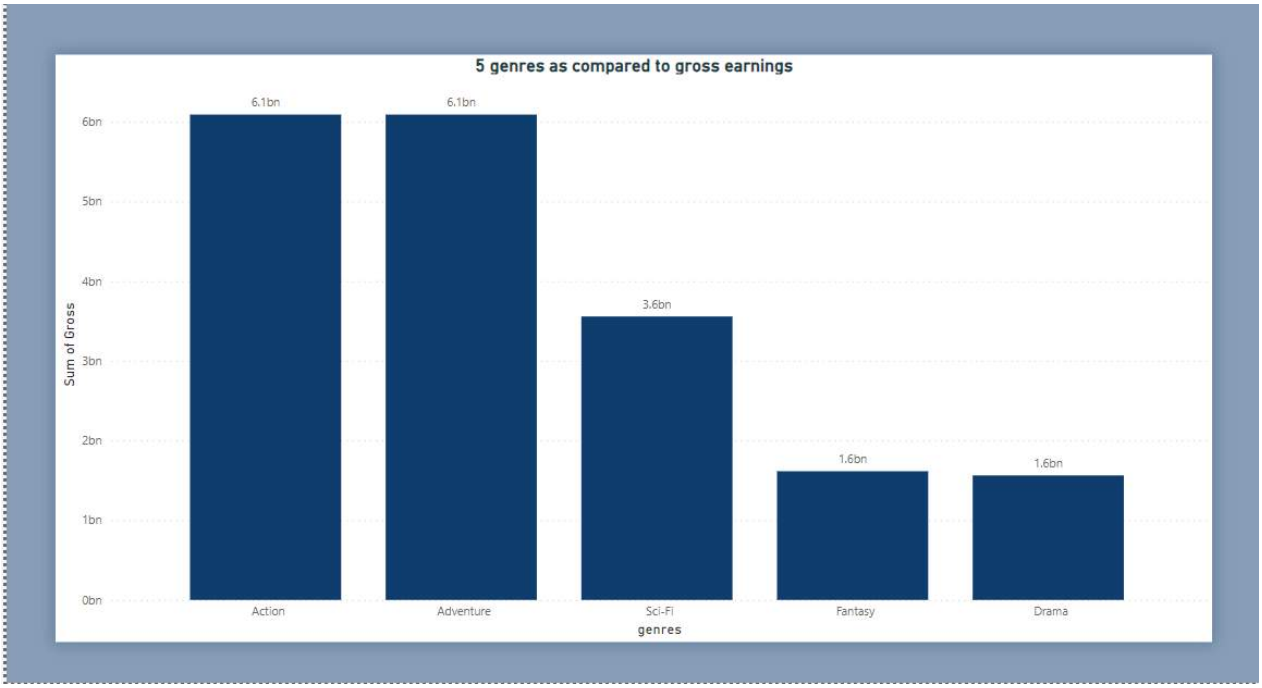
2. Show the top 5 genres as compared to gross earnings for the 9 box office movies.

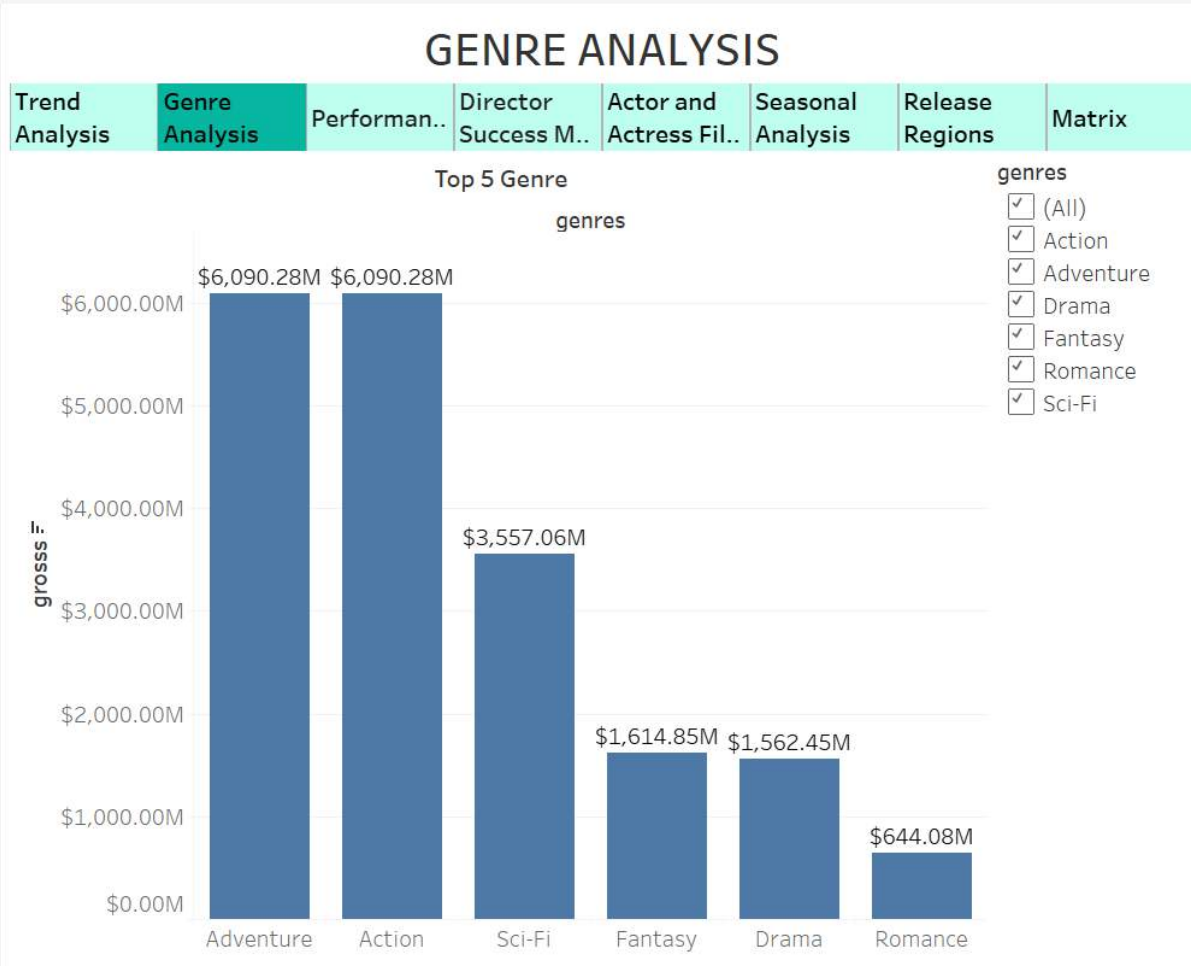
```
SELECT
  g.genres,
  SUM(e.Gross) AS total_earnings
FROM
```

```
  fct_movie_earning e
JOIN
  dim_movie_title_mysql m ON e.tconst = m.tconst
JOIN
  dim_genre g ON e.tconst = g.tconst
GROUP BY
  g.genres
ORDER BY
  total_earnings DESC
LIMIT 5;
```



genres	total_earnings
Action	6090283475
Adventure	6090283475
Sci-Fi	3557064362
Fantasy	1614846113
Drama	1562451638





Performance Metrics:

1. Correlation between the movie's runtime and its average rating

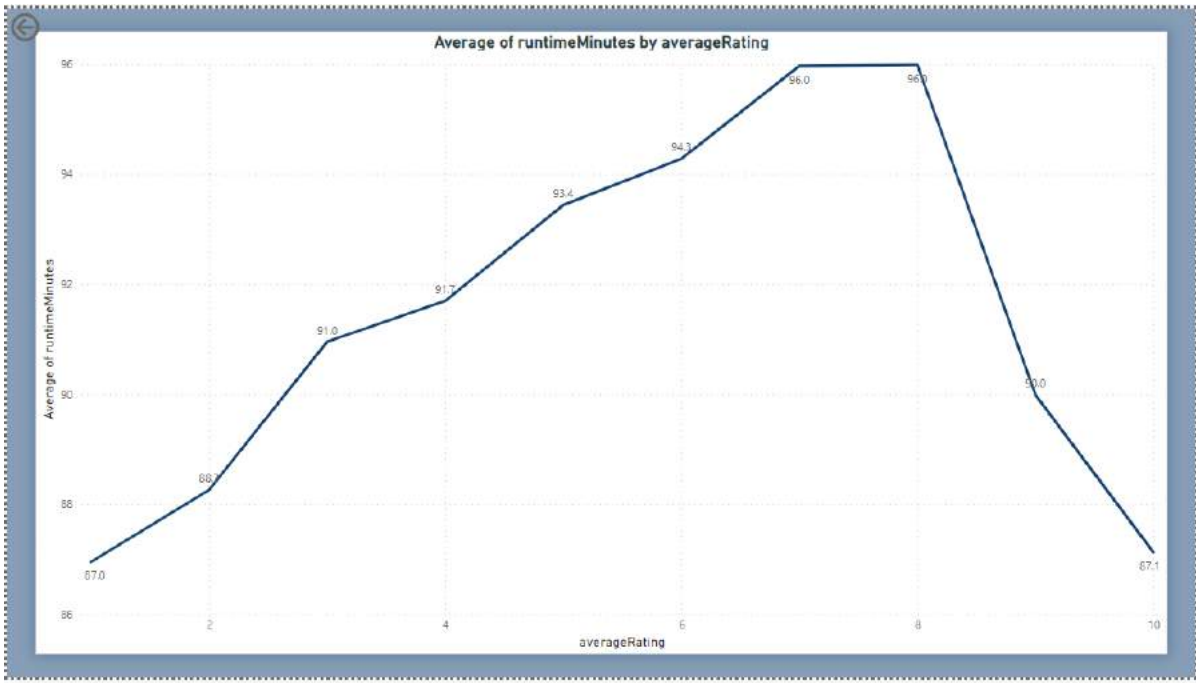
a. The goal is to understand the impact on the ratings due to the movie's runtime.

```
select averageRating, avg(t.runtimeMinutes)
from fct_rating r
join fct_imdb i on i.RatingSK = r.RatingSK
join dim_movie_title_mysql t on t.MovieTitle_SK = i.MovieTitle_SK
group by averageRating;
```

```
52
53 •  select averageRating, avg(t.runtimeMinutes)
54    from fct_rating r
55    join fct_imdb i on i.RatingSK = r.RatingSK
56    join dim_movie_title_mysql t on t.MovieTitle_SK = i.MovieTitle_SK
57    group by averageRating;
58
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

averageRating	avg(t.runtimeMinutes)
4	91.72483520840561
6	94.28492464309858
5	93.4542158379714
7	95.9889024994896
8	95.99926023983804
3	90.97819989801121
9	90.00511789435204
2	88.28345250255363
1	86.85284280936455
10	87.1255230125523



2. Correlation between the movie's runtime and its average gross
a. The goal is to understand the impact on the gross earnings due to the movie's runtime.

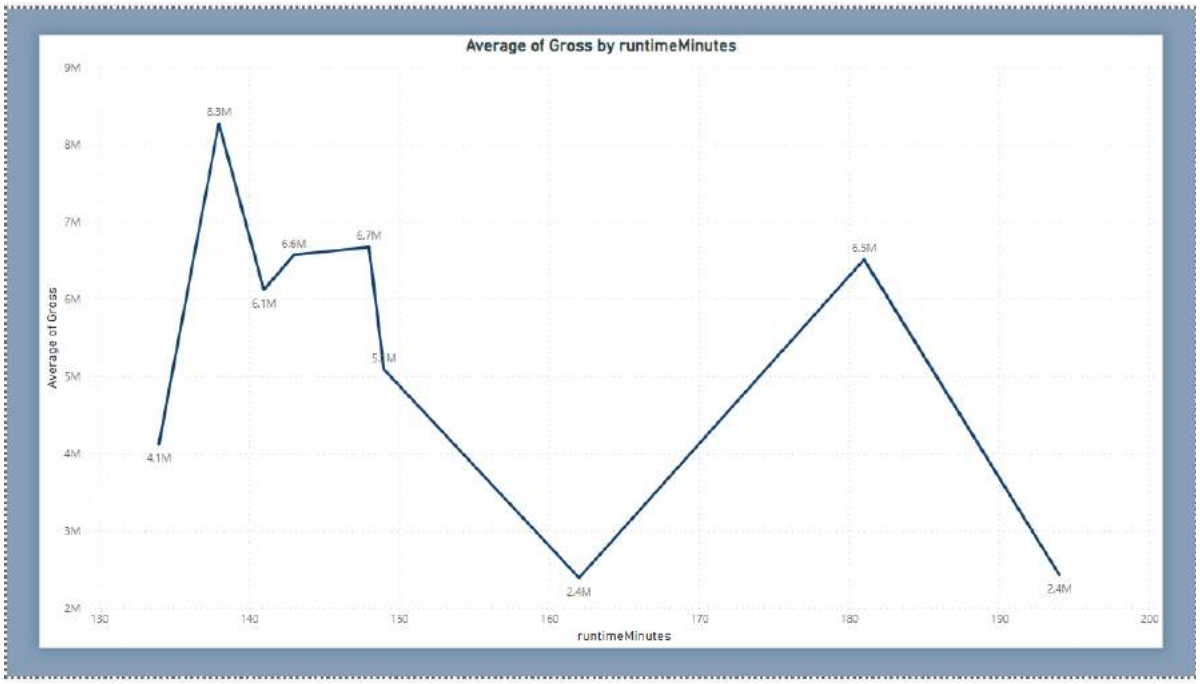
```
select primaryTitle, sum(Gross),runTimeMinutes
from dim_movie_title_mysql t
```

join fct_movie_earning e on t.MovieTitle_SK = e.MovieTitle_SK
group by primaryTitle,runTimeMinutes;

```
52
53 •  select primaryTitle, sum(Gross),runTimeMinutes
54    from dim_movie_title_mysql t
55    join fct_movie_earning e on t.MovieTitle_SK = e.MovieTitle_SK
56    group by primaryTitle,runTimeMinutes;
57
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

primaryTitle	sum(Gross)	runTimeMinutes
Avatar	760508625	162
Avengers: Age of Ultron	483598418	141
Avengers: Endgame	918373000	181
Avengers: Infinity War	717815482	149
Black Panther	725259566	134
Spider-Man: No Way Home	854337488	148
Star Wars: Episode VII - The Force Awakens	992642689	138
The Avengers	637748207	143
Titanic	644078638	194

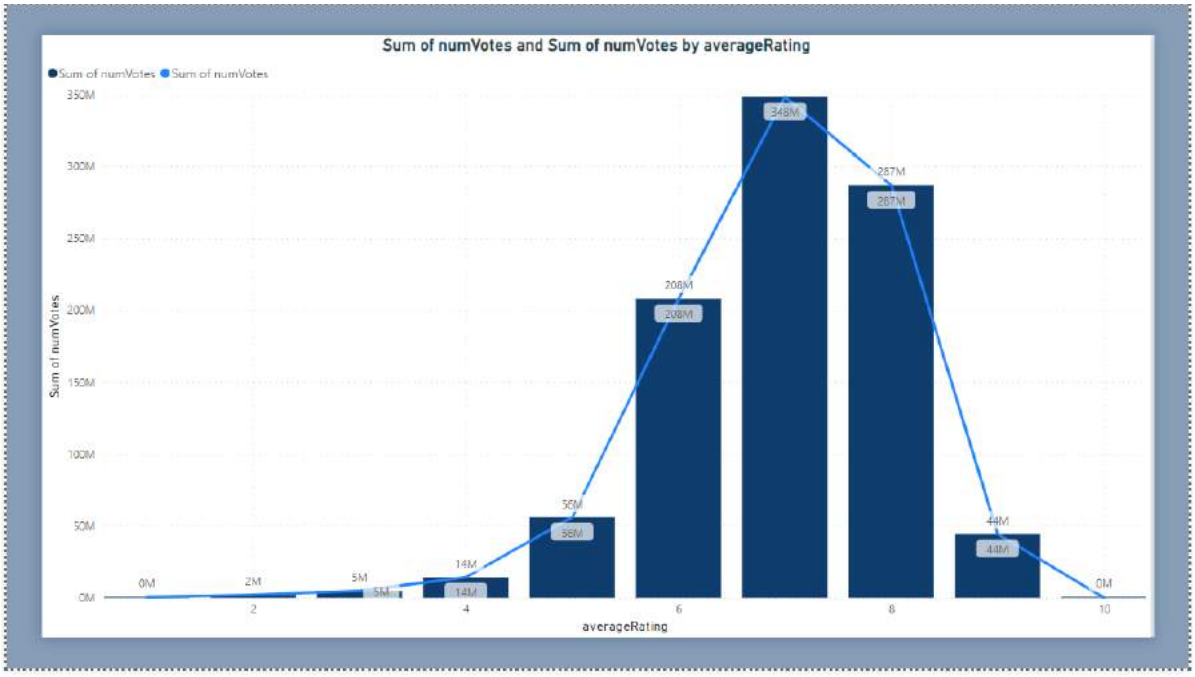


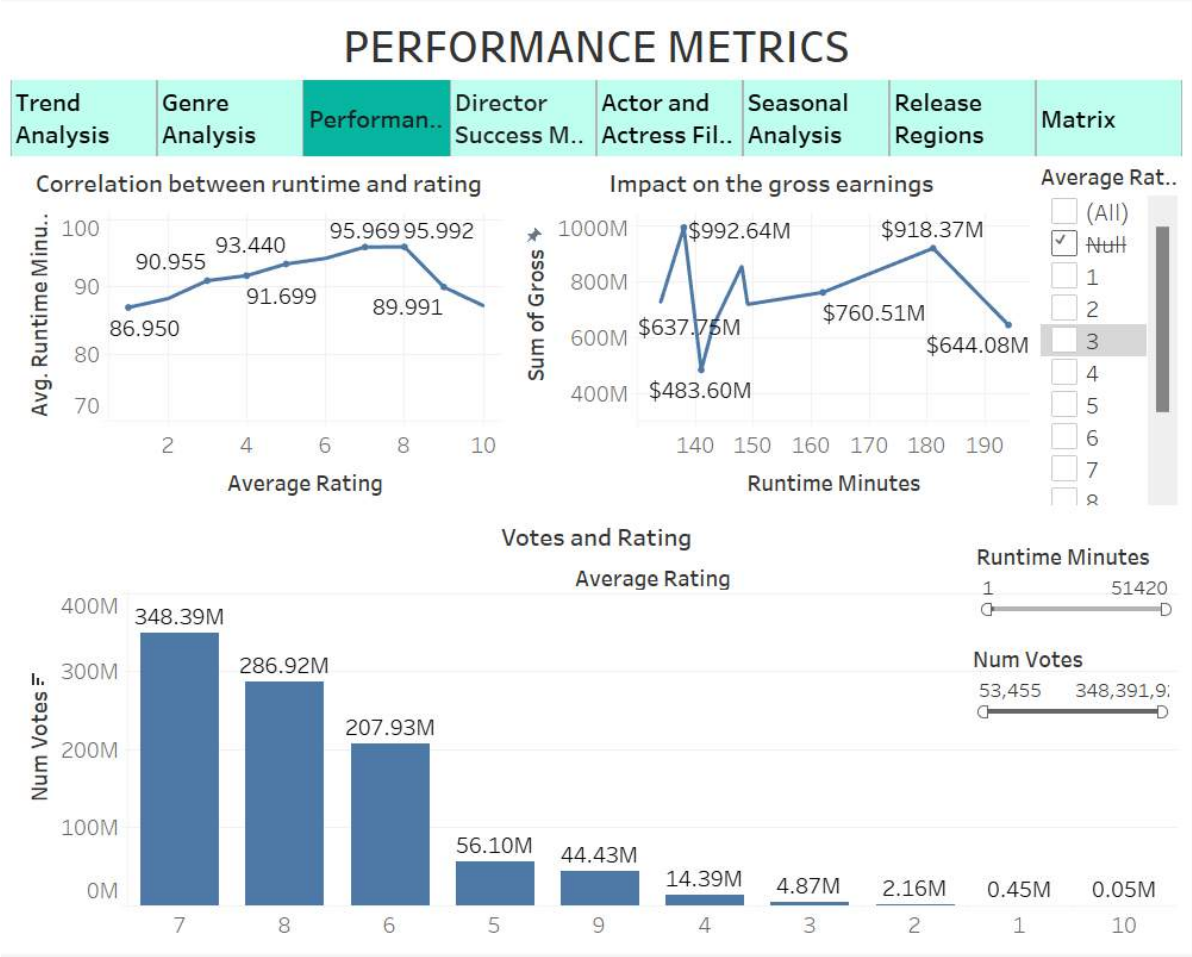
3. Relationship between the number of votes and the average rating of movies.
a. The goal is to identify the correlation between votes and rating
select averageRating,sum(numVotes) from fct_rating
group by averageRating;

```
58 • select averageRating,sum(numVotes) from fct_rating
59   group by averageRating;
60
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

averageRating	sum(numVotes)
4	14389566
6	207928332
5	56097207
7	348392088
8	286926542
3	4872904
9	44427452
2	2159728
1	451294
10	53463





Director Success Metrics:

1. Identify directors with the most films rated above 7. Sort them in descending order

a. The goal is to identify which directors have made films that were received well by the audience

```
select p.PrimaryName, count(c.tconst) from dim_crew c
join fct_imdb i on i.CrewSK = c.CrewSK
join fct_rating r on r.RatingSK = i.RatingSK
join dim_person p on p.PersonSK = i.PersonSK
where c.category = 'director' and r.averageRating > 7
group by p.PrimaryName ;
```

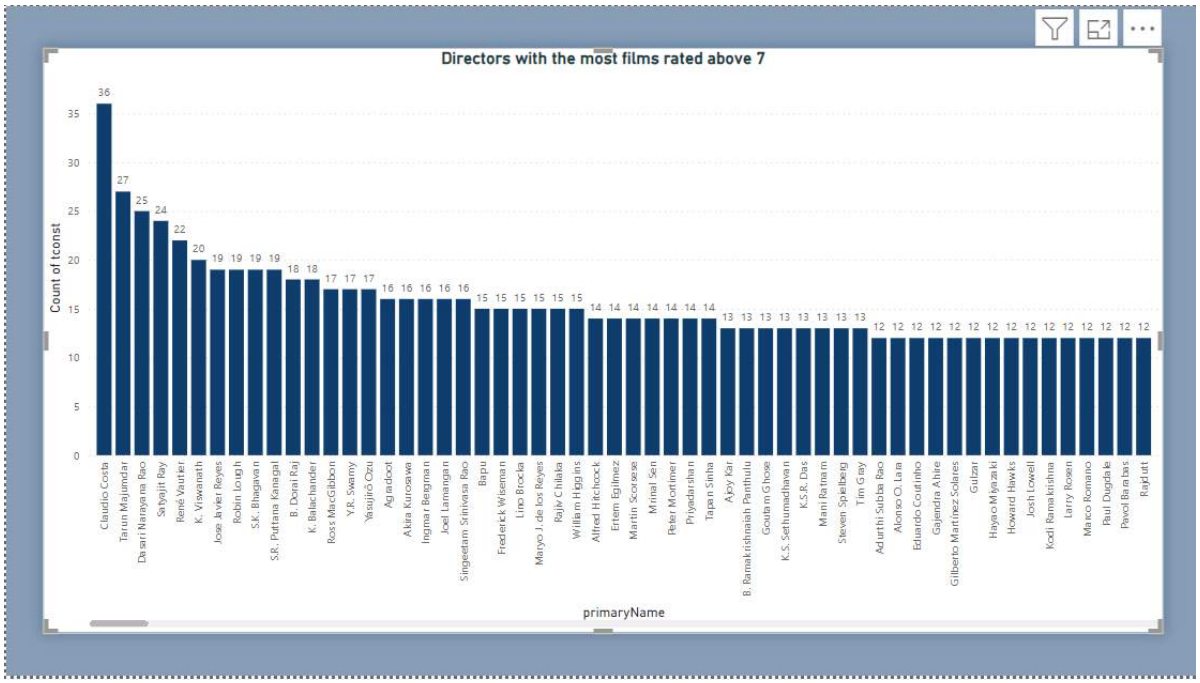
SQL File 8* SQL File 4* dim_person_profession SQL File 6*

Don't Limit

```
6  select p.PrimaryName, count(c.tconst) from dim_crew c
7  join fct_imdb i on i.CrewSK = c.CrewSK
8  join fct_rating r on r.RatingSK = i.RatingSK
9  join dim_person p on p.PersonSK = i.PersonSK
10 where c.category = 'director' and r.averageRating > 7
11 group by p.PrimaryName ;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

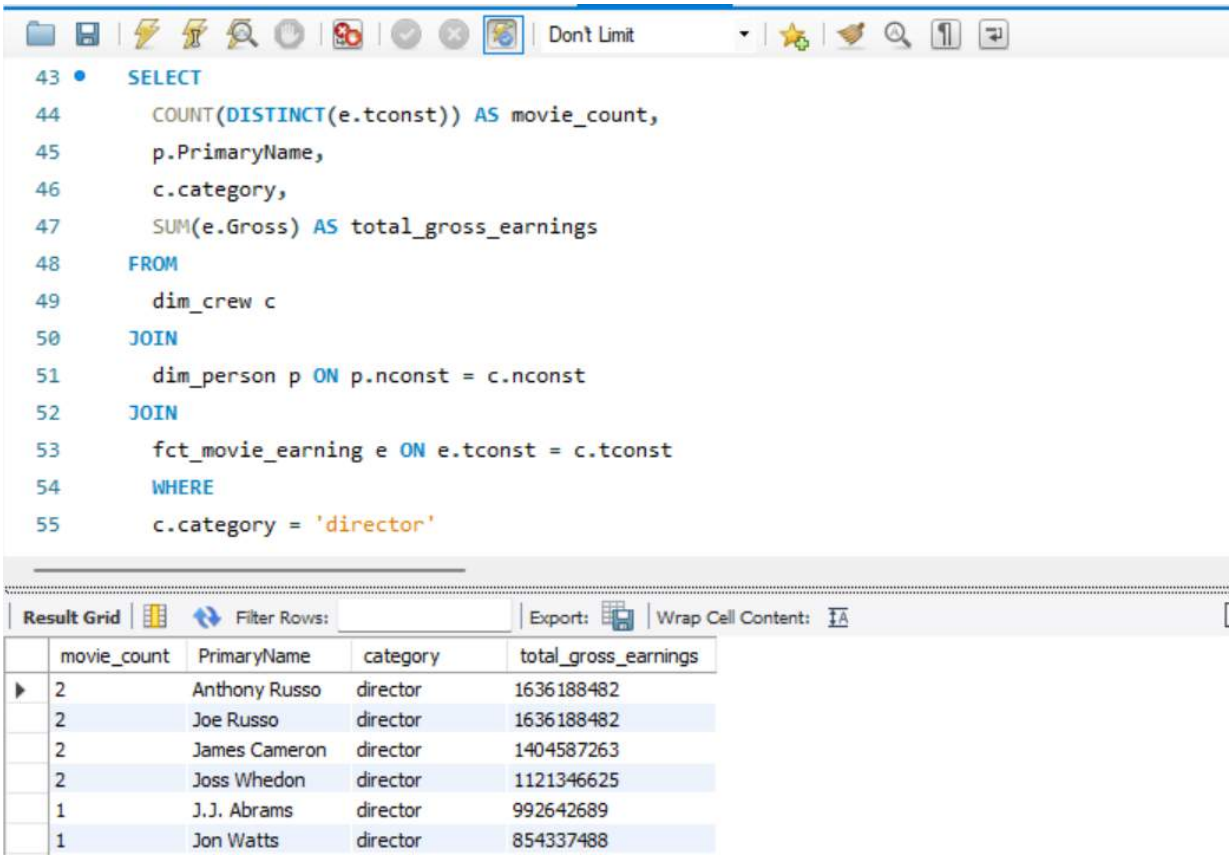
PrimaryName	count(c.tconst)
Marco Romano	12
René Vautier	11
Dasari Narayana Rao	9
Goutam Ghose	6
Louise Lemoine	6
Neal Hutcheson	5
Larry Rosen	5
Jamie Benning	5
Anand Patwardhan	4
Greg Stump	4



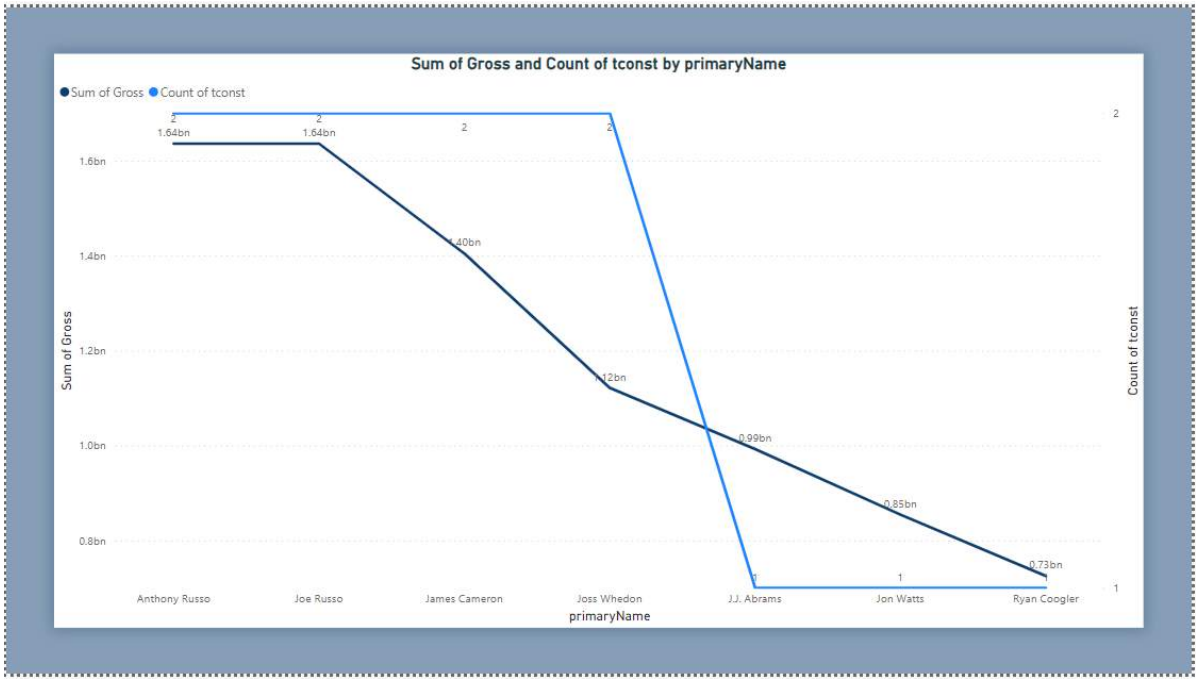
2. Determine the directors who have directed the highest number of films overall and their respective gross earnings trends.
a. The goal is to identify if there are any notable directors who consistently contribute to high-grossing movies

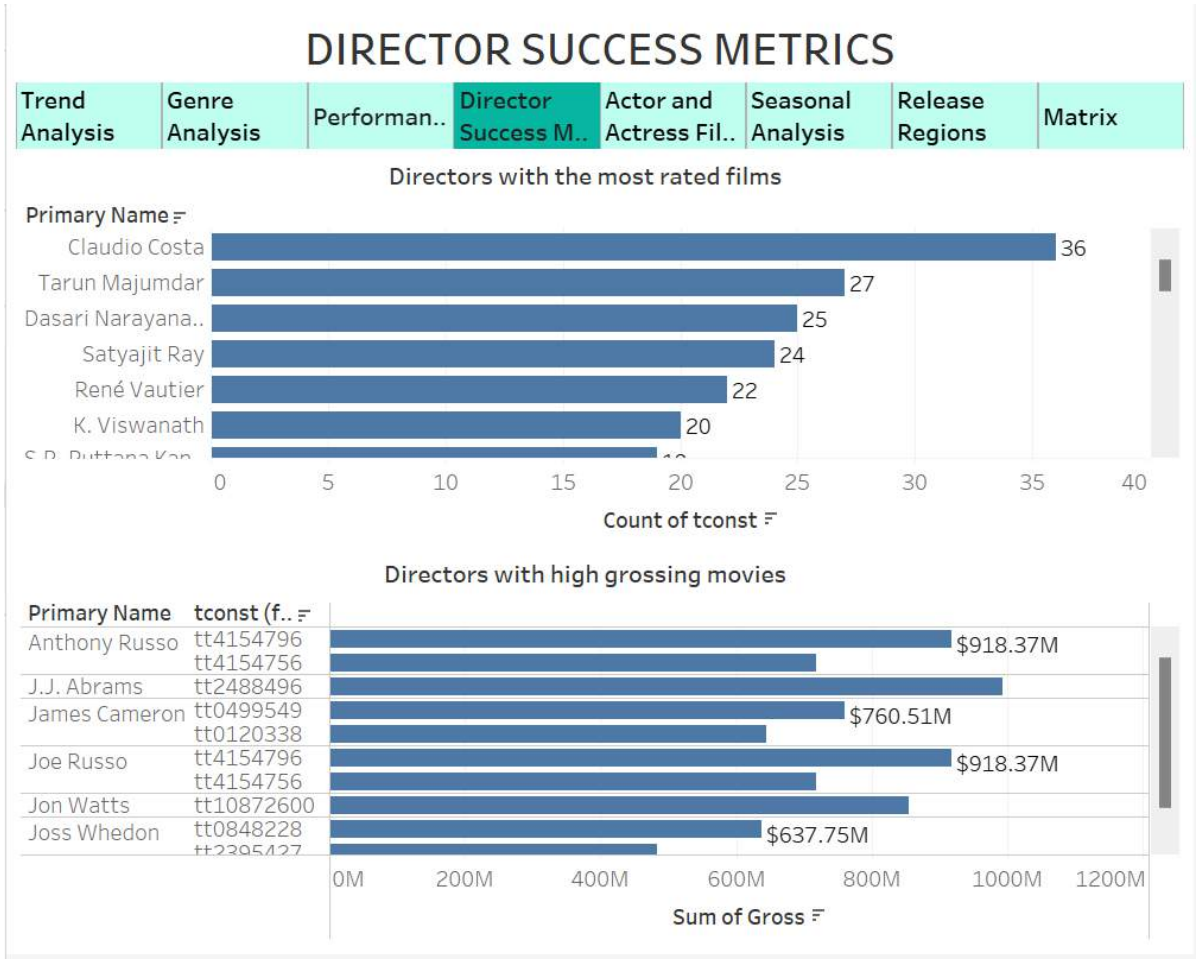
```
SELECT
  COUNT(DISTINCT(e.tconst)) AS movie_count,
  p.PrimaryName,
  c.category,
  SUM(e.Gross) AS total_gross_earnings
FROM
  dim_crew c
```

```
JOIN
  dim_person p ON p.nconst = c.nconst
JOIN
  fct_movie_earning e ON e.tconst = c.tconst
  WHERE
  c.category = 'director'
GROUP BY
  p.PrimaryName, c.category
ORDER BY
  total_gross_earnings DESC;
```



movie_count	PrimaryName	category	total_gross_earnings
2	Anthony Russo	director	1636188482
2	Joe Russo	director	1636188482
2	James Cameron	director	1404587263
2	Joss Whedon	director	1121346625
1	J.J. Abrams	director	992642689
1	Jon Watts	director	854337488

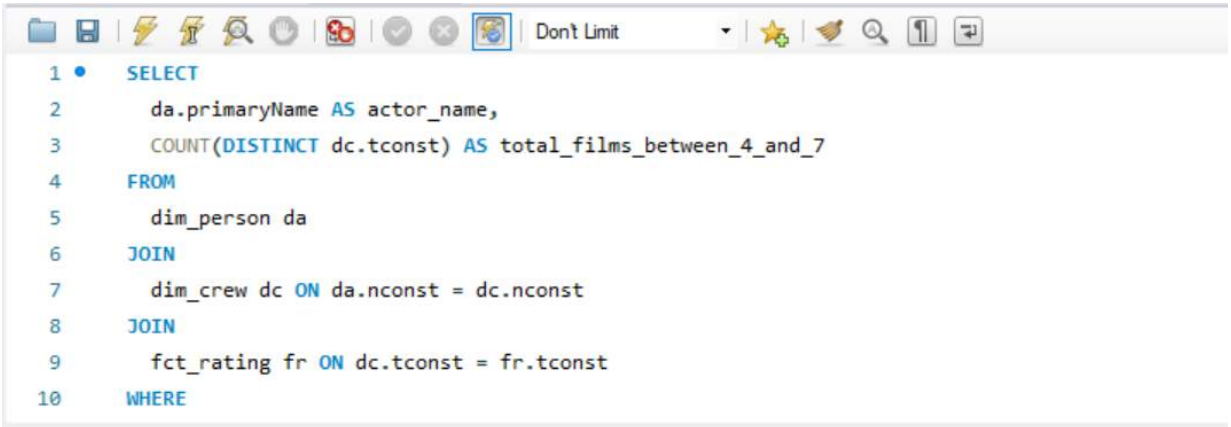




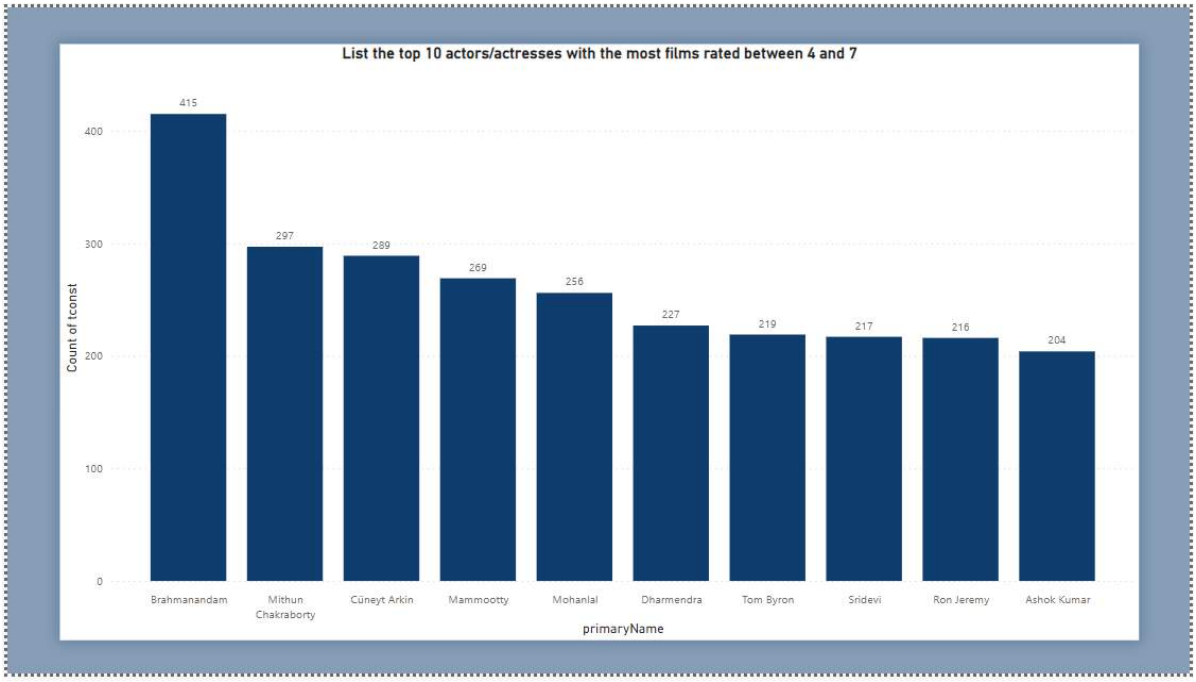
Actor and Actress Film Records:

1. List the top 10 actors/actresses with the most films rated between 4 and 7.
a. The goal is to identify the popular actors/actresses

```
SELECT
  da.primaryName AS actor_name,
  COUNT(DISTINCT dc.tconst) AS total_films_between_4_and_7
FROM
  dim_person da
JOIN
  dim_crew dc ON da.nconst = dc.nconst
JOIN
  fct_rating fr ON dc.tconst = fr.tconst
WHERE
  fr.averageRating BETWEEN 4 AND 7 and dc.category ="actor" or dc.category = "actress"
GROUP BY
  da.primaryName
ORDER BY
  total_films_between_4_and_7 DESC
  Limit 10;
```



actor_name	total_films_between_4_and_7
Brahmanandam	415
Mithun Chakraborty	297
Cüneyt Arkin	289
Mammootty	269
Mohanlal	256
Sridevi	232
Dharmendra	227
Tom Byron	219
Ron Jeremy	216
Aachi Manorama	213



2. Compare the top 5 actors and actresses based on movie ratings.
a. The goal is to compare the top 5 actor and actresses

```
select primaryName,avg(r.averageRating) from dim_person_profession pp
join dim_person p on p.nconst = pp.nconst
join fct_imdb i on i.PersonSK = p.PersonSK
join dim_crew c on c.CrewSK = i.CrewSK
join fct_rating r on r.RatingSK = i.RatingSK
where c.category IN ('actor','actress')
group by primaryName;
```

```sql
-- Compare the top 5 actors and actresses based on movie ratings
 select primaryName,avg(r.averageRating) from dim_person_profession pp
join dim_person p on p.nconst = pp.nconst
join fct_imdb i on i.PersonSK = p.PersonSK
join dim_crew c on c.CrewSK = i.CrewSK
join fct_rating r on r.RatingSK = i.RatingSK
where c.category IN ('actor','actress')
group by primaryName;

SELECT
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

primaryName	avg(r.averageRating)
Srdjan Jablanovic	10.0000
Pierre Mirel	10.0000
Paul Roraback	10.0000
Donovan Bailey	10.0000
Sophia Laia	10.0000
Jugoslav Josipovic	10.0000
Davi Freitas	10.0000
Orin Friesen	10.0000
Sam DeRose	10.0000
Brad	10.0000
Margareta Spânu-Cemârtan	10.0000
Hercules Nikolopoulos	10.0000
Raj Tirandasu	10.0000
Ana Blake	10.0000
Taylor Van Doorne	10.0000
Martin Nadin	10.0000
Deeksha Welekar	10.0000
Pulakesh Bhattacharya	10.0000

Result 3

Output

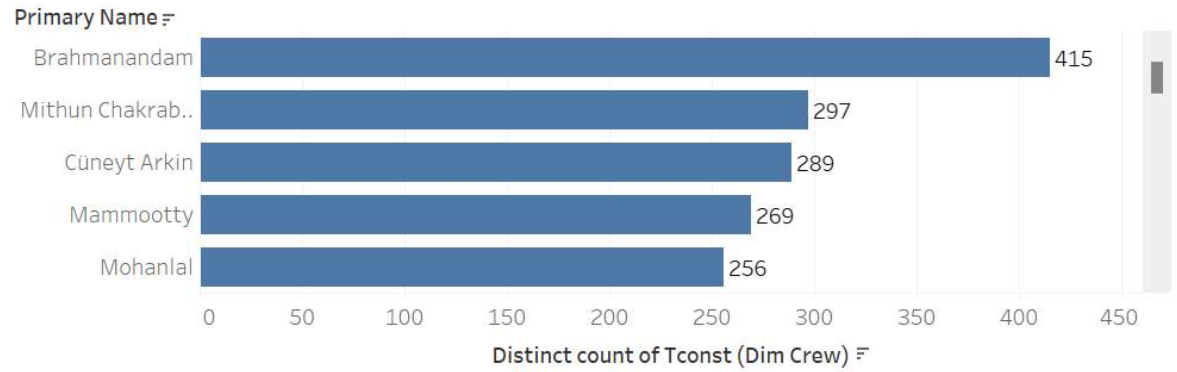
Action Output

#	Time	Action	Message
13	02:54:56	select primaryName,avg(r.averageRating) from dim_person_profession pp join dim_person p on p.nconst = pp.nconst join fct_imdb i on i.Pe...	Error Code: 2013. Lost connection to MySQL server during query
14	03:06:07	select primaryName,avg(r.averageRating) from dim_person_profession pp join dim_person p on p.nconst = pp.nconst join fct_imdb i on i.Pe...	379252 row(s) returned

ACTOR AND ACTRESS FILM RECORDS

Trend Analysis | Genre Analysis | Performan.. | Director Success M.. | Actor and Actress Fil.. | Seasonal Analysis | Release Regions | Matrix

Top 10 Actors

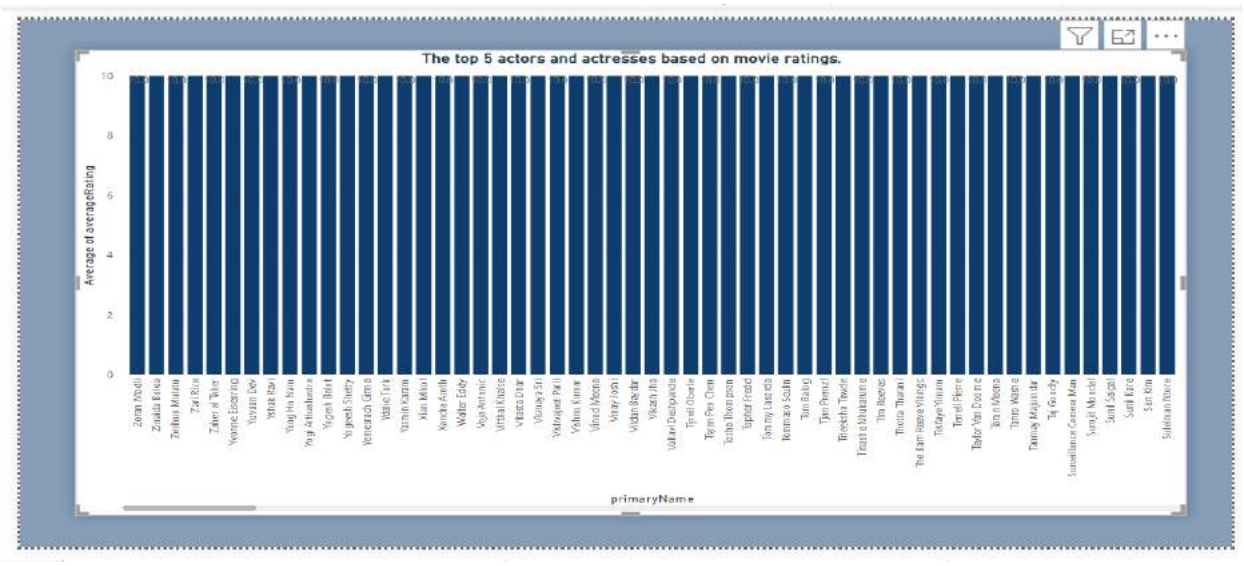


Top 5 by rating

Primary Name	Avg. ..	
Zoran Modli	10	Abc
Zinaida Bolea	10	Abc
Zerihun Mulatu	10	Abc
Zari Rice	10	Abc
Zaher al Taher	10	Abc
Yvonne Eisenring	10	Abc
Yuvaan Dev	10	Abc
Yshak Ravi	10	Abc
Yong Ho Nam	10	Abc
Yogi Arthashastra	10	Abc
Yogesh Balot	10	Abc

Category

- [ ] (All)
- [ ] Null
- [x] actor
- [x] actress
- [ ] archive_fo...
- [ ] archive_so...
- [ ] cinematogr...
- [ ] composer
- [ ] director
- [ ] editor



Seasonal Analysis:

1. How does movie performance vary across different seasons of the year? (based on gross earnings)

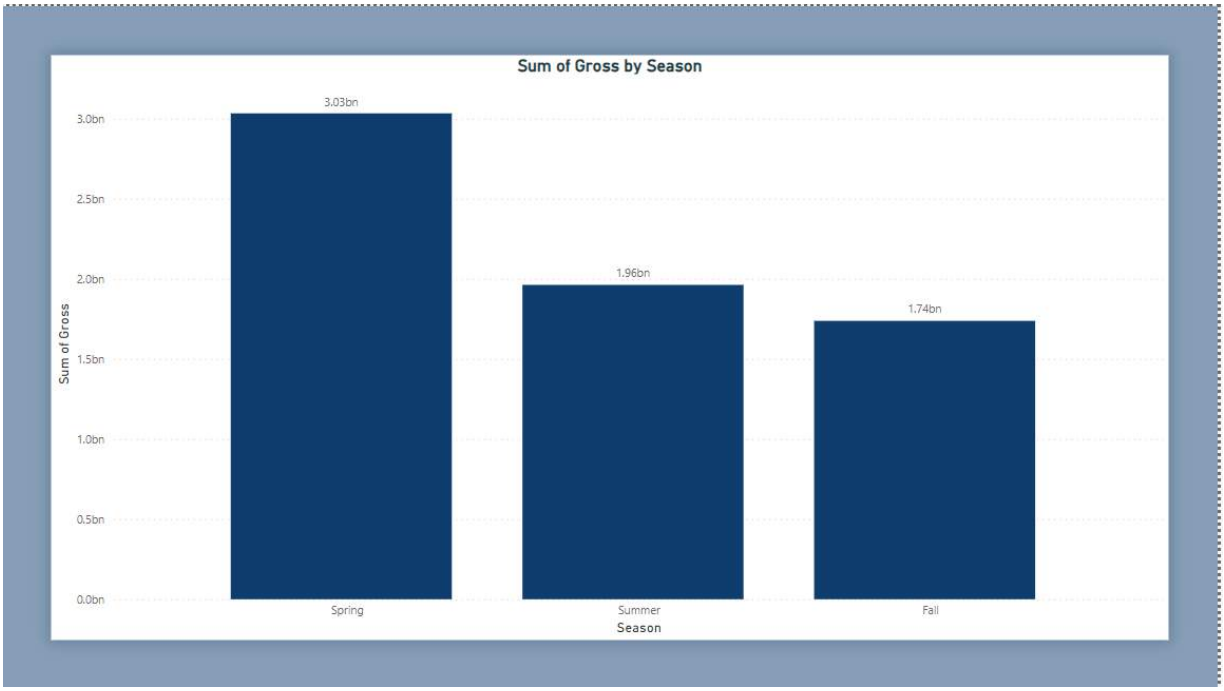
a. The goal is to identify which season has an impact on the revenue.

```
SELECT
  CASE
    WHEN d.Month_Num IN (0, 1, 2, 3) THEN 'Spring'
    WHEN d.Month_Num IN (4, 5, 6, 7) THEN 'Summer'
    WHEN d.Month_Num IN (8, 9, 10, 11) THEN 'Fall'
  END AS season,
  SUM(e.Gross) AS total_gross_earnings
FROM
  fct_movie_earning e
JOIN
  dim_date d ON e.dateSK = d.dateSK
GROUP BY
  season
ORDER BY
  MIN(d.Month_Num);
```

```sql
SELECT
  CASE
    WHEN d.Month_Num IN (0, 1, 2, 3) THEN 'Spring'
    WHEN d.Month_Num IN (4, 5, 6, 7) THEN 'Summer'
    WHEN d.Month_Num IN (8, 9, 10, 11) THEN 'Fall'
  END AS season,
  SUM(e.Gross) AS total_gross_earnings
FROM
  fct_movie_earning e
JOIN
  dim_date d ON e.dateSK = d.dateSK
GROUP BY
  season
ORDER BY
  MIN(d.Month Num);
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

season	total_gross_earnings
Spring	3033650624
Summer	1962604141
Fall	1738107348



2. Top 3 movies based on the season (spring, summer, fall)
a. Spring - Jan to April
b. Summer - May to August
c. Fall - September to December

```
SELECT
  e.primaryTitle,
  e.originalTitle,
  e.startYear,
  e.runtimeMinutes,
  e.Gross,
  e.season
FROM (
  SELECT
    t.primaryTitle,
    t.originalTitle,
    t.startYear,
    t.runtimeMinutes,
    e.Gross,
    CASE
      WHEN d.Month_Num BETWEEN 1 AND 4 THEN 'Spring'
      WHEN d.Month_Num BETWEEN 5 AND 8 THEN 'Summer'
      WHEN d.Month_Num BETWEEN 9 AND 12 THEN 'Fall'
      ELSE 'Unknown'
    END AS season,
    ROW_NUMBER() OVER (PARTITION BY
                  CASE
                    WHEN d.Month_Num BETWEEN 1 AND 4 THEN 'Spring'
                    WHEN d.Month_Num BETWEEN 5 AND 8 THEN 'Summer'
                    WHEN d.Month_Num BETWEEN 9 AND 12 THEN 'Fall'
                    ELSE 'Unknown'
                  END
                  ORDER BY e.Gross DESC) AS season_rank
  FROM
    fct_movie_earning e
  JOIN
    dim_date d ON e.dateSK = d.dateSK
  JOIN
    dim_movie_title_mysql t ON e.tconst = t.tconst
  WHERE
    d.Month_Num BETWEEN 1 AND 12
) e
WHERE
  e.season_rank <= 3
ORDER BY
  e.season, e.Gross DESC;
```

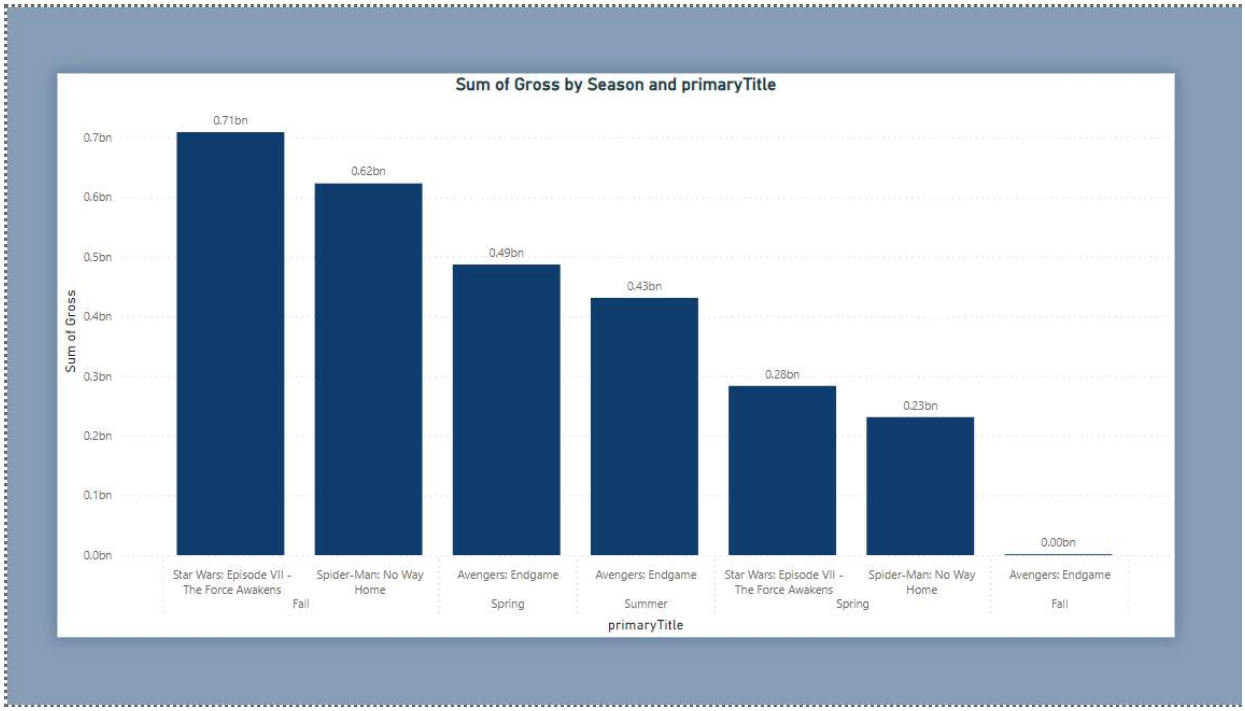
Query 1 | Administration - Data Import/Res... | SQL File 1*

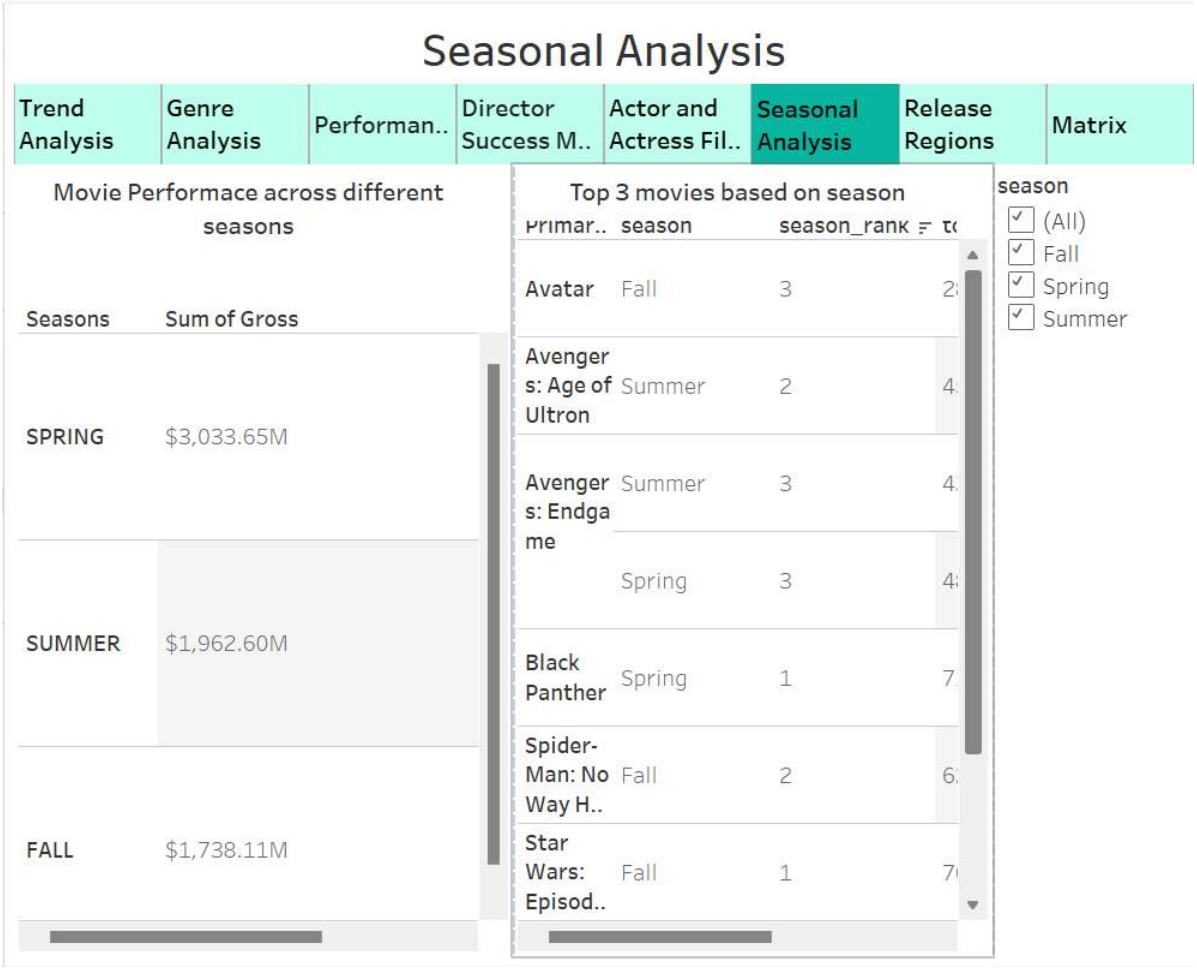
Limit to 1000 rows

```
307     WHERE
308       d.Month_Num BETWEEN 1 AND 12
309   ) e
310   WHERE
311     e.season_rank <= 3
312   ORDER BY
313     e.season, e.Gross DESC;
314
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

primaryTitle	originalTitle	startYear	runtimeMinutes	Gross	season
Spider-Man: No Way Home	Spider-Man: No Way Home	2021-01-01 00:00:00	148	121964712	Fall
Star Wars: Episode VII - The Force Awakens	Star Wars: Episode VII - The Force Awakens	2015-01-01 00:00:00	138	119119282	Fall
Spider-Man: No Way Home	Spider-Man: No Way Home	2021-01-01 00:00:00	148	73941279	Fall
Avengers: Endgame	Avengers: Endgame	2019-01-01 00:00:00	181	157461641	Spring
Avengers: Endgame	Avengers: Endgame	2019-01-01 00:00:00	181	109264122	Spring
Avengers: Infinity War	Avengers: Infinity War	2018-01-01 00:00:00	149	106334939	Spring
The Avengers	The Avengers	2012-01-01 00:00:00	143	9017913	Summer
The Avengers	The Avengers	2012-01-01 00:00:00	143	5826863	Summer
The Avengers	The Avengers	2012-01-01 00:00:00	143	5641642	Summer

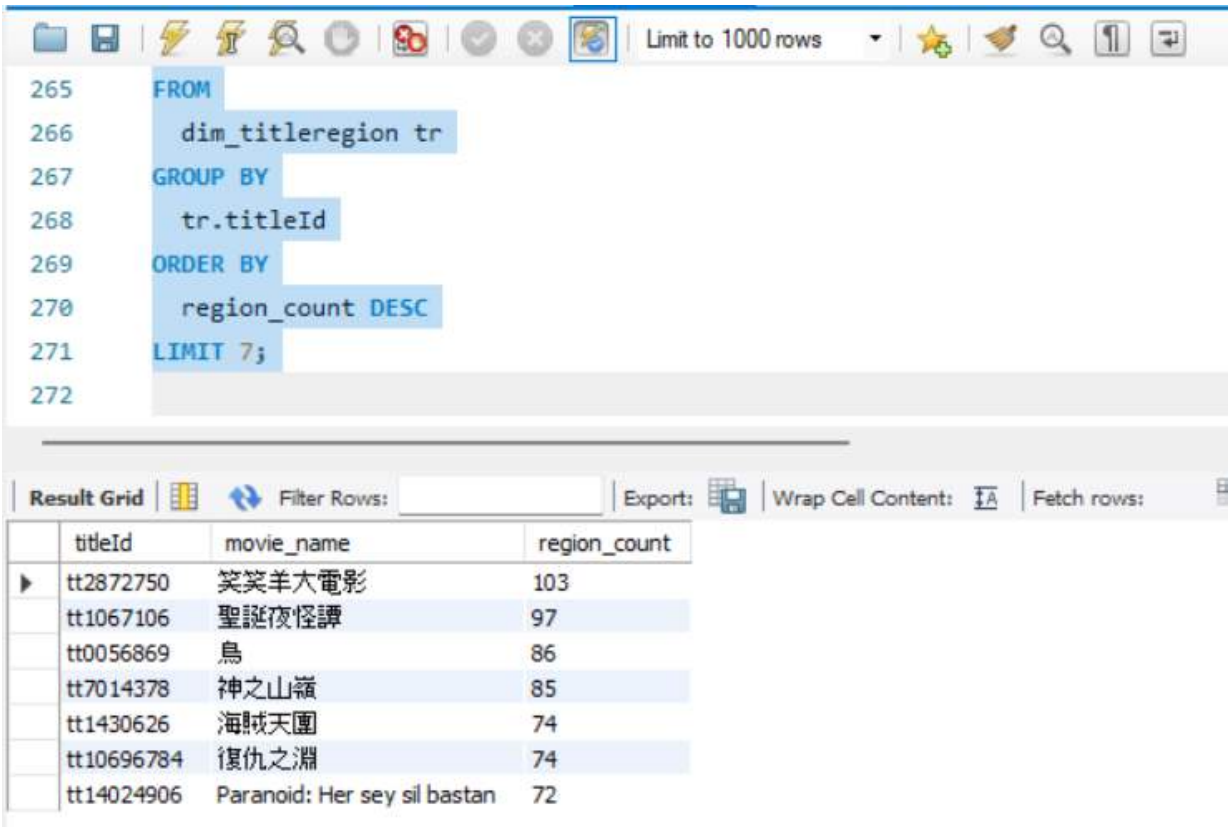




Release Regions:

1. Identify movies that have had the widest release across multiple regions.
a. The goal is to identify which movie was released in most regions.

```
SELECT
  tr.titleId,
  MAX(tr.title) AS movie_name,
  COUNT(DISTINCT tr.region) AS region_count
FROM
  dim_titleregion tr
GROUP BY
  tr.titleId
ORDER BY
  region_count DESC
LIMIT 7;
```



```
265     FROM
266       dim_titleregion tr
267     GROUP BY
268       tr.titleId
269     ORDER BY
270       region_count DESC
271     LIMIT 7;
272
```

Limit to 1000 rows

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

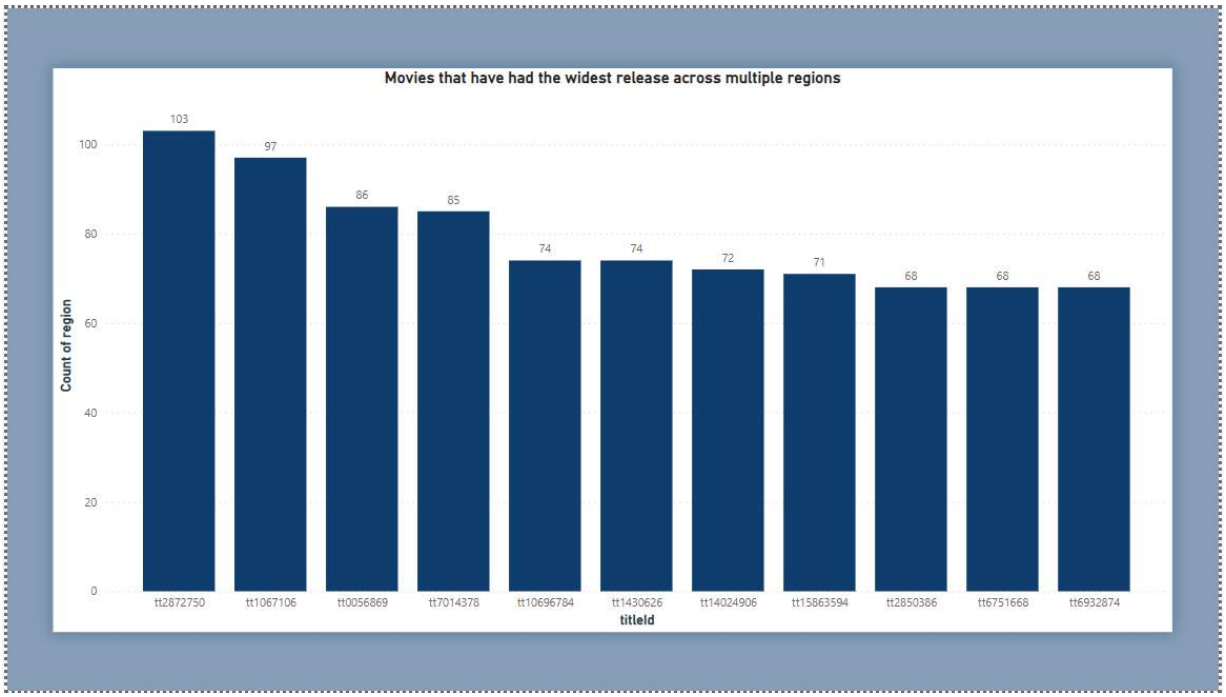
titleId	movie_name	region_count
tt2872750	笑笑羊大電影	103
tt1067106	聖誕夜怪譚	97
tt0056869	鳥	86
tt7014378	神之山嶺	85
tt1430626	海賊天團	74
tt10696784	復仇之淵	74
tt14024906	Paranoid: Her sey sil bastan	72

RELEASE REGIONS

Trend Analysis | Genre Analysis | Performan.. | Director Success M.. | Actor and Actress Fil.. | Seasonal Analysis | Release Regions | Matrix

Widest release across multiple regions

Title Id	Max. Title	Distinct count of Region	
tt7014378	Le Sommet des dieux	1	Abc
tt6932874	Босс-молокосос 2	1	Abc
tt6751668	Parazit	1	Abc
tt4302938	Kubo in dve struni	1	Abc
tt2872750	Xiao Yang Xiao En	1	Abc
tt2850386	The Croods 2: A New Age	1	Abc
tt2386490	How to Train Your Dragon: The Hid..	1	Abc
tt2262227	Книга жизни	1	Abc
tt1667889	Ice Age: Th4w	1	Abc
tt15863594	Just Short of Perfect	1	Abc
tt1430626	Pirates!	1	Abc
tt14024906	Paranoid: Everything Wipe from th..	1	Abc
tt10696784	Vực Sâu Thù Hận	1	Abc
tt1067106	Navidad divertida de Scrooge	1	



Matrix:

Create a dashboard for the 9 box office movies that shows movie-wise:

1. No. of directors

```
SELECT
  me.tconst,
  me.primaryTitle AS movie_title,
  GROUP_CONCAT(DISTINCT dp.primaryName) AS director_names,
  COUNT(DISTINCT c.nconst) AS num_directors
FROM
  dim_crew c
JOIN
  fct_movie_earning fe ON c.tconst = fe.tconst
JOIN
  dim_movie_title_mysql me ON fe.tconst = me.tconst
JOIN
  dim_person dp ON c.nconst = dp.nconst
  where
  c.category = "Director"
GROUP BY
  me.tconst, me.primaryTitle;
```

```
534     fct_movie_earning fe ON c.tconst = fe.tconst
535   JOIN
536     dim_movie_title_mysql me ON fe.tconst = me.tconst
537   JOIN
538     dim_person dp ON c.nconst = dp.nconst
539     where
540     c.category = "Director"
541   GROUP BY
542     me.tconst, me.primaryTitle;
543
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

tconst	movie_title	director_names	num_directors
tt0120338	Titanic	James Cameron	1
tt0499549	Avatar	James Cameron	1
tt0848228	The Avengers	Joss Whedon	1
tt10872600	Spider-Man: No Way Home	Jon Watts	1
tt1825683	Black Panther	Ryan Coogler	1
tt2395427	Avengers: Age of Ultron	Joss Whedon	1
tt2488496	Star Wars: Episode VII - The Force Awakens	J.J. Abrams	1
tt4154756	Avengers: Infinity War	Anthony Russo,Joe Russo	2
tt4154796	Avengers: Endgame	Anthony Russo,Joe Russo	2

2. No. of actors/actresses

```
  SELECT
  me.tconst,
  me.primaryTitle AS movie_title,
  GROUP_CONCAT(DISTINCT dp.primaryName) AS actor_names,
  COUNT(DISTINCT c.nconst) AS num_actors
FROM
  dim_crew c
JOIN
  fct_movie_earning fe ON c.tconst = fe.tconst
JOIN
  dim_movie_title_mysql me ON fe.tconst = me.tconst
JOIN
  dim_person dp ON c.nconst = dp.nconst
  where
  c.category = "actor" or c.category = "actress"
GROUP BY
  me.tconst, me.primaryTitle;
```

```
526 •    SELECT
527      me.tconst,
528      me.primaryTitle AS movie_title,
529      GROUP_CONCAT(DISTINCT dp.primaryName) AS actors_names,
530      COUNT(DISTINCT c.nconst) AS num_actors
531    FROM
532      dim_crew c
533    JOIN
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

tconst	movie_title	actors_names	num_actors
tt0120338	Titanic	Billy Zane,Kate Winslet,Kathy Bates,Leonardo DiCaprio	4
tt0499549	Avatar	Michelle Rodriguez,Sam Worthington,Sigourney Weaver,Zoe Saldana	4
tt0848228	The Avengers	Chris Evans,Jeremy Renner,Robert Downey Jr.,Scarlett Johansson	4
tt10872600	Spider-Man: No Way Home	Benedict Cumberbatch,Jacob Batalon,Tom Holland,Zendaya	4
tt1825683	Black Panther	Chadwick Boseman,Danai Gurira,Lupita Nyong'o,Michael B. Jordan	4
tt2395427	Avengers: Age of Ultron	Chris Evans,Chris Hemsworth,Mark Ruffalo,Robert Downey Jr.	4
tt2488496	Star Wars: Episode VII - The Force Awakens	Daisy Ridley,Domhnall Gleeson,John Boyega,Oscar Isaac	4
tt4154756	Avengers: Infinity War	Chris Evans,Chris Hemsworth,Mark Ruffalo,Robert Downey Jr.	4
tt4154796	Avengers: Endgame	Chris Evans,Chris Hemsworth,Mark Ruffalo,Robert Downey Jr.	4

3. No. of writers

```
  SELECT
  me.tconst,
  me.primaryTitle AS movie_title,
  GROUP_CONCAT(DISTINCT dp.primaryName) AS writer_names,
  COUNT(DISTINCT c.nconst) AS num_writers
FROM
  dim_crew c
JOIN
  fct_movie_earning fe ON c.tconst = fe.tconst
JOIN
  dim_movie_title_mysql me ON fe.tconst = me.tconst
JOIN
  dim_person dp ON c.nconst = dp.nconst
  where
  c.category = "writer"
GROUP BY
  me.tconst, me.primaryTitle;
```

```
553     JOIN
554       fct_movie_earning fe ON c.tconst = fe.tconst
555     JOIN
556       dim_movie_title_mysql me ON fe.tconst = me.tconst
557     JOIN
558       dim_person dp ON c.nconst = dp.nconst
559       where
560       c.category = "writer"
561     GROUP BY
562       me.tconst, me.primaryTitle;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

tconst	movie_title	writer_names	num_writers
tt0848228	The Avengers	Zak Penn	1
tt10872600	Spider-Man: No Way Home	Chris McKenna,Erik Sommers,Stan Lee,Steve Ditko	4
tt1825683	Black Panther	Jack Kirby,Joe Robert Cole,Stan Lee	3
tt2395427	Avengers: Age of Ultron	Jack Kirby,Jim Starlin,Joe Simon,Stan Lee	4
tt2488496	Star Wars: Episode VII - The Force Awakens	George Lucas,Lawrence Kasdan,Michael Arndt	3
tt4154756	Avengers: Infinity War	Christopher Markus,Jack Kirby,Stan Lee,Stephe...	4
tt4154796	Avengers: Endgame	Christopher Markus,Jack Kirby,Stan Lee,Stephe...	4

4. Total worldwide gross earnings

```
SELECT
  COUNT(DISTINCT fr.tconst) AS movie_count,
  GROUP_CONCAT(DISTINCT me.primaryTitle ORDER BY me.primaryTitle ASC) AS
      movie_names,
  SUM(fr.Gross) AS total_gross_earnings
FROM
  dim_crew c
JOIN
  fct_movie_earning fr ON c.tconst = fr.tconst
JOIN
  dim_movie_title_mysql me ON fr.tconst = me.tconst;
```

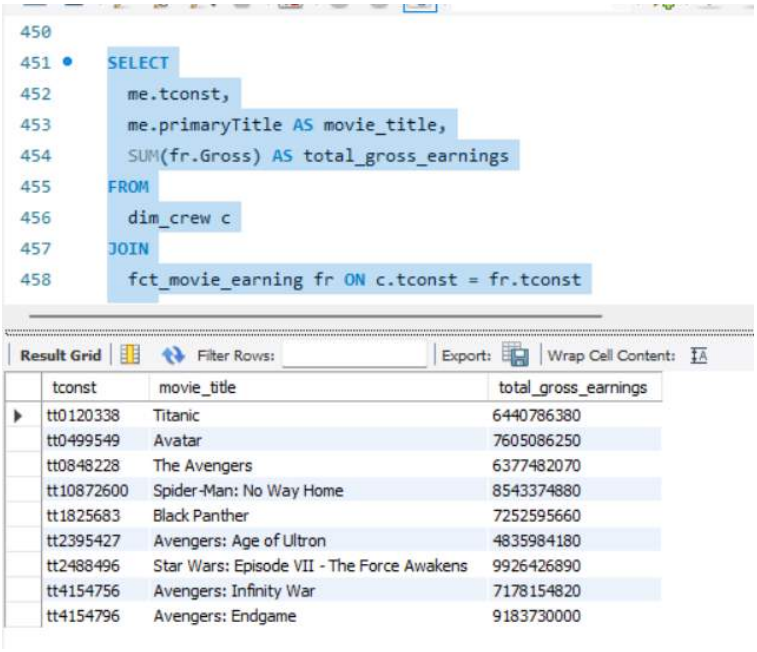
```
373 •   SELECT
374       COUNT(DISTINCT fr.tconst) AS movie_count,
375       GROUP_CONCAT(DISTINCT me.primaryTitle ORDER BY me.primaryTitle ASC) AS movie_names,
376       SUM(fr.Gross) AS total_gross_earnings
377     FROM
378       dim_crew c
379     JOIN
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

movie_count	movie_names	total_gross_earnings
9	Avatar,Avengers: Age of Ultron,Avengers: Endgame,Avengers: Infinity War,Black Panther,Spider-Man: No Way Home,S...	67343621130

```
SELECT
  me.tconst,
  me.primaryTitle AS movie_title,
  SUM(fr.Gross) AS total_gross_earnings
FROM
  dim_crew c
JOIN
  fct_movie_earning fr ON c.tconst = fr.tconst
JOIN
```

```
  dim_movie_title_mysql me ON fr.tconst = me.tconst
GROUP BY
  me.tconst, me.primaryTitle;
```



```
450
451 •  SELECT
452      me.tconst,
453      me.primaryTitle AS movie_title,
454      SUM(fr.Gross) AS total_gross_earnings
455    FROM
456      dim_crew c
457    JOIN
458      fct_movie_earning fr ON c.tconst = fr.tconst
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

tconst	movie_title	total_gross_earnings
tt0120338	Titanic	6440786380
tt0499549	Avatar	7605086250
tt0848228	The Avengers	6377482070
tt10872600	Spider-Man: No Way Home	8543374880
tt1825683	Black Panther	7252595660
tt2395427	Avengers: Age of Ultron	4835984180
tt2488496	Star Wars: Episode VII - The Force Awakens	9926426890
tt4154756	Avengers: Infinity War	7178154820
tt4154796	Avengers: Endgame	9183730000

5. Average ratings

```
SELECT
  me.tconst,
  me.primaryTitle AS movie_title,
  AVG(rt.averageRating) AS average_rating
FROM
  dim_crew c
JOIN
  fct_movie_earning fe ON c.tconst = fe.tconst
JOIN
  dim_movie_title_mysql me ON fe.tconst = me.tconst
JOIN
  fct_rating rt ON fe.tconst = rt.tconst
GROUP BY
  me.tconst, me.primaryTitle;
```

```
466     me.tconst,
467     me.primaryTitle AS movie_title,
468     AVG(rt.averageRating) AS average_rating
469   FROM
470     dim_crew c
471   JOIN
472     fct_movie_earning fe ON c.tconst = fe.tconst
473   JOIN
474     dim_movie_title_mysql me ON fe.tconst = me.tconst
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

tconst	movie_title	average_rating
tt0120338	Titanic	8.0000
tt0499549	Avatar	8.0000
tt0848228	The Avengers	8.0000
tt10872600	Spider-Man: No Way Home	8.0000
tt1825683	Black Panther	7.0000
tt2395427	Avengers: Age of Ultron	7.0000
tt2488496	Star Wars: Episode VII - The Force Awakens	8.0000
tt4154756	Avengers: Infinity War	8.0000
tt4154796	Avengers: Endgame	8.0000

6. No. of regions, it was released

```
SELECT
  me.tconst,
  me.primaryTitle AS movie_title,
  COUNT(DISTINCT tr.region) AS num_regions_released
FROM
  dim_crew c
JOIN
  fct_movie_earning fe ON c.tconst = fe.tconst
JOIN
  dim_movie_title_mysql me ON fe.tconst = me.tconst
JOIN
  dim_titleregion tr ON me.tconst = tr.titleId
GROUP BY
  me.tconst, me.primaryTitle;
```

```
479
480 •  SELECT
481      me.tconst,
482      me.primaryTitle AS movie_title,
483      COUNT(DISTINCT tr.region) AS num_regions_released
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

tconst	movie_title	num_regions_released
tt0120338	Titanic	48
tt0499549	Avatar	44
tt0848228	The Avengers	50
tt10872600	Spider-Man: No Way Home	53
tt1825683	Black Panther	42
tt2395427	Avengers: Age of Ultron	50
tt2488496	Star Wars: Episode VII - The Force Awakens	51
tt4154756	Avengers: Infinity War	46
tt4154796	Avengers: Endgame	49

7. No. of genres

```
SELECT
  me.tconst,
  me.primaryTitle AS movie_title,
  COUNT(DISTINCT dg.genres) AS num_genres
FROM
  dim_crew c
JOIN
  fct_movie_earning fe ON c.tconst = fe.tconst
JOIN
  dim_movie_title_mysql me ON fe.tconst = me.tconst
JOIN
  dim_genre dg ON me.tconst = dg.tconst
GROUP BY
  me.tconst, me.primaryTitle;
```

```
500     dim_crew c
501   JOIN
502     fct_movie_earning fe ON c.tconst = fe.tconst
503   JOIN
504     dim_movie_title_mysql me ON fe.tconst = me.tconst
505   JOIN
506     dim_genre dg ON me.tconst = dg.tconst
507   GROUP BY
508     me.tconst, me.primaryTitle;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

tconst	movie_title	num_genres
tt0120338	Titanic	2
tt0499549	Avatar	3
tt0848228	The Avengers	3
tt10872600	Spider-Man: No Way Home	3
tt1825683	Black Panther	3
tt2395427	Avengers: Age of Ultron	3
tt2488496	Star Wars: Episode VII - The Force Awakens	3
tt4154756	Avengers: Infinity War	3
tt4154796	Avengers: Endgame	3

Dashboard for the 9 box office movies

primaryTitle	Count of genres	Average of averageRating	Sum of Gross	Count of region	Actor/Actress	Director	Writer	Count of region	Count of titleId
Titanic	2	8.00	644078638	1	4	1		1	1
The Avengers	3	8.00	637748207	1	4	1	1	1	1
Star Wars: Episode VII - The Force Awakens	3	8.00	992642689	1	4	1	3	1	1
Spider-Man: No Way Home	3	8.00	854337488	1	4	1	4	1	1
Black Panther	3	7.00	725259566	1	4	1	3	1	1
Avengers: Infinity War	3	8.00	717815482	1	4	2	4	1	1
Avengers: Endgame	3	8.00	918373000	1	4	2	4	1	1
Avengers: Age of Ultron	3	7.00	483598418	1	4	1	4	1	1
Avatar	3	8.00	760508625	1	4	1		1	1

MATRIX

Trend Analysis	Genre Analysis	Performan..	Director Success M..	Actor and Actress Fil..	Seasonal Analysis	Release Regions	Matrix

Movie-Wise Director, Actor, Actress, Writer, Gross Earnings, Average Rating and Region

tconst (fct_movie..	Primary Title	Primary Name	Sum of Gross	Average
tt2488496	Star Wars: Episode VII - The Force Awakens	Daisy Ridley	$992.64M	8
		Domhnall Gleeson	$992.64M	8
		John Boyega	$992.64M	8
		Oscar Isaac	$992.64M	8
tt4154796	Avengers: Endgame	Chris Evans	$918.37M	8
		Chris Hemsworth	$918.37M	8
		Mark Ruffalo	$918.37M	8
		Robert Downey Jr.	$918.37M	8
tt10872600	Spider-Man: No Way Home	Benedict Cumberbatch	$854.34M	8
		Jacob Batalon	$854.34M	8
		Tom Holland	$854.34M	8
		Zendaya	$854.34M	8
tt0499549	Avatar	Michelle Rodriguez	$760.51M	8
		Sam Worthington	$760.51M	8
		Sigourney Weaver	$760.51M	8
		Zoe Saldana	$760.51M	8
tt1825683	Black Panther	Chadwick Boseman	$725.26M	7
		Danai Gurira	$725.26M	7
		Lupita Nyong'o	$725.26M	7
		Michael B. Jordan	$725.26M	7
tt4154756	Avengers: Infinity War	Chris Evans	$717.82M	8
		Chris Hemsworth	$717.82M	8
		Mark Ruffalo	$717.82M	8

Category

- [ ] (All)
- [x] actor
- [x] actress
- [ ] archive_foo...
- [ ] archive_sou...
- [ ] cinematogr...
- [ ] composer
- [ ] director
- [ ] editor
- [ ] producer
- [ ] production...
- [ ] self
- [ ] writer