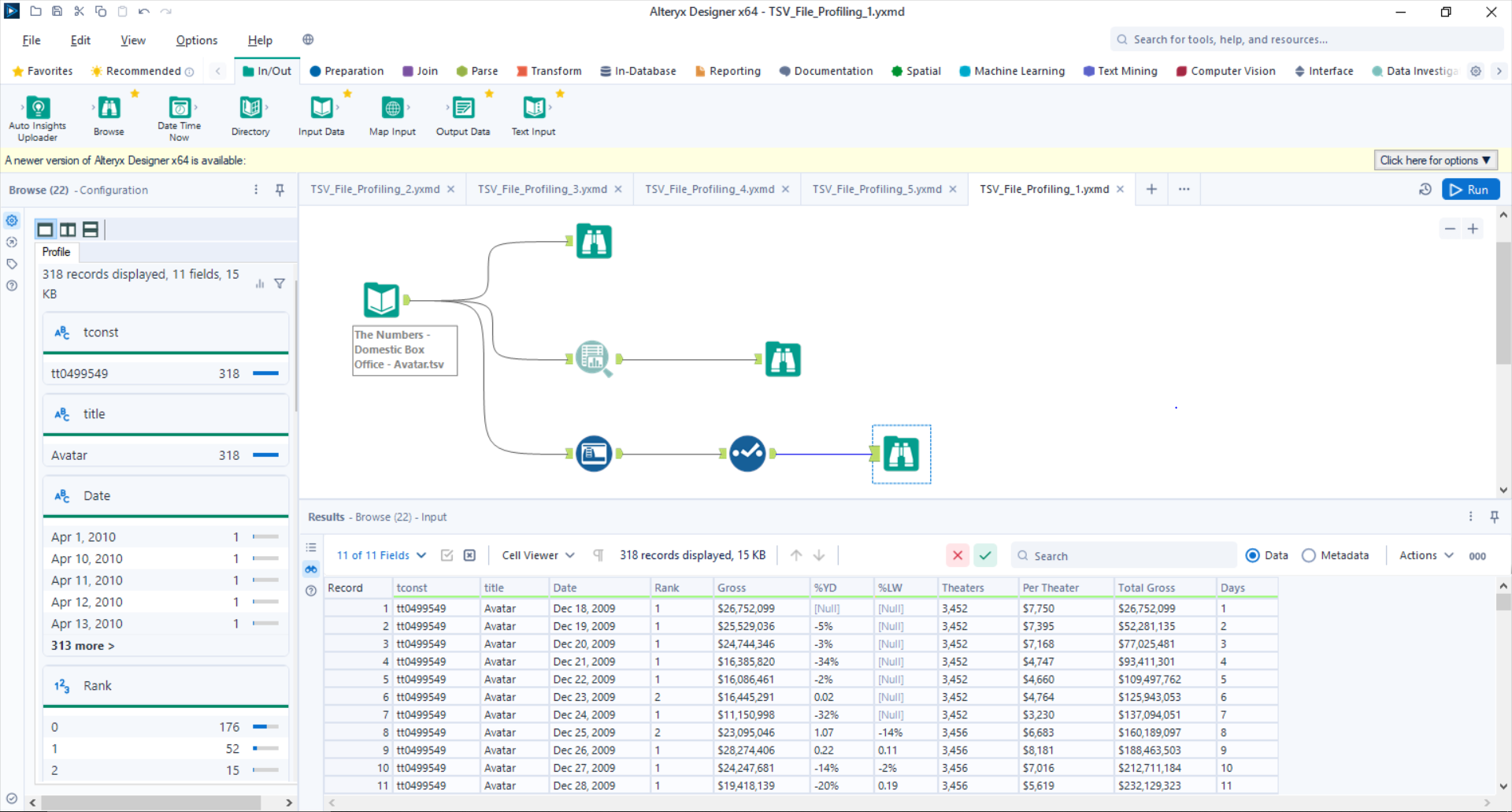
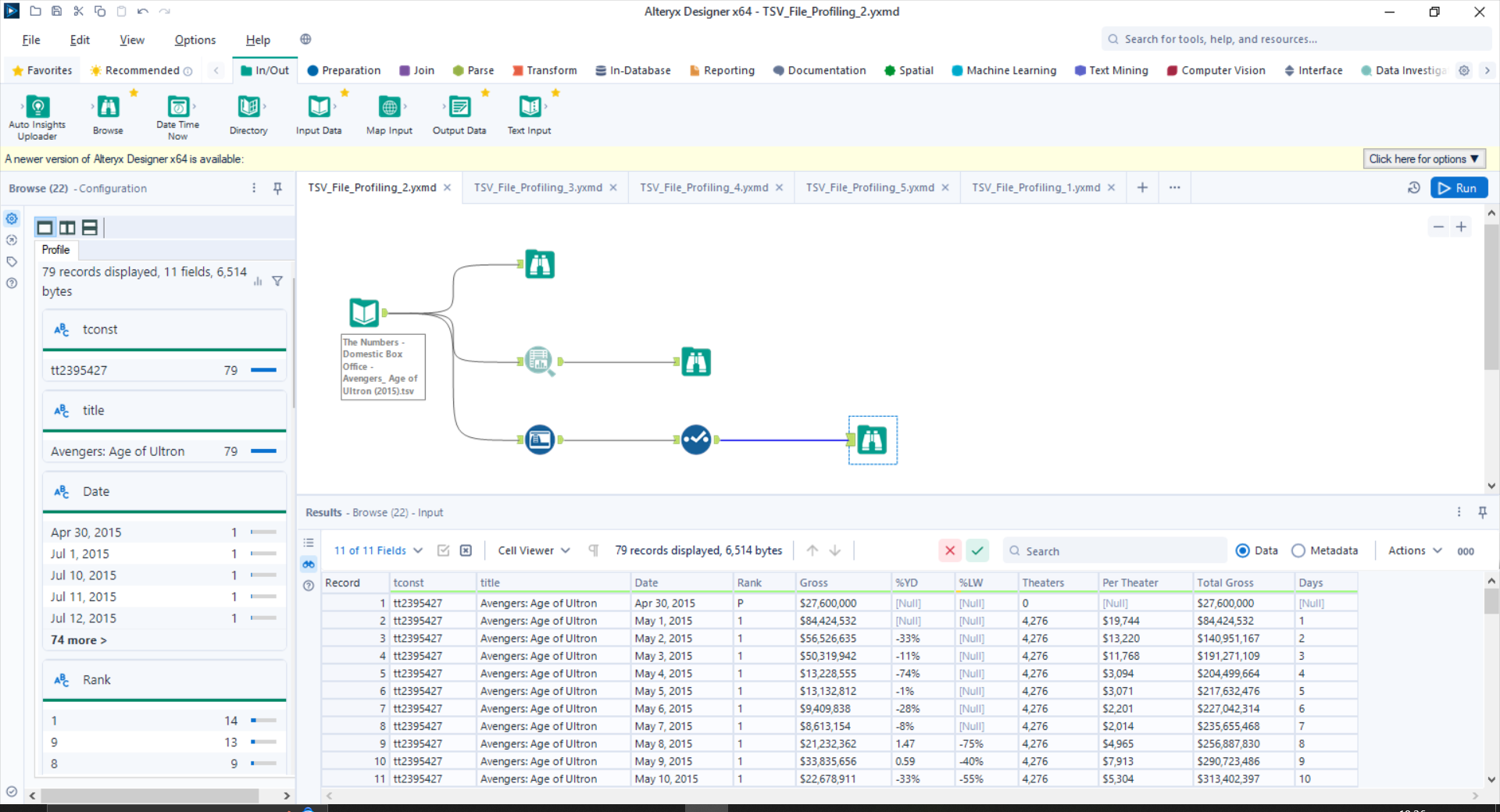
**TSV ALTREYX WORKFLOW**

***SCREENSHOTS AND ANALYSIS***

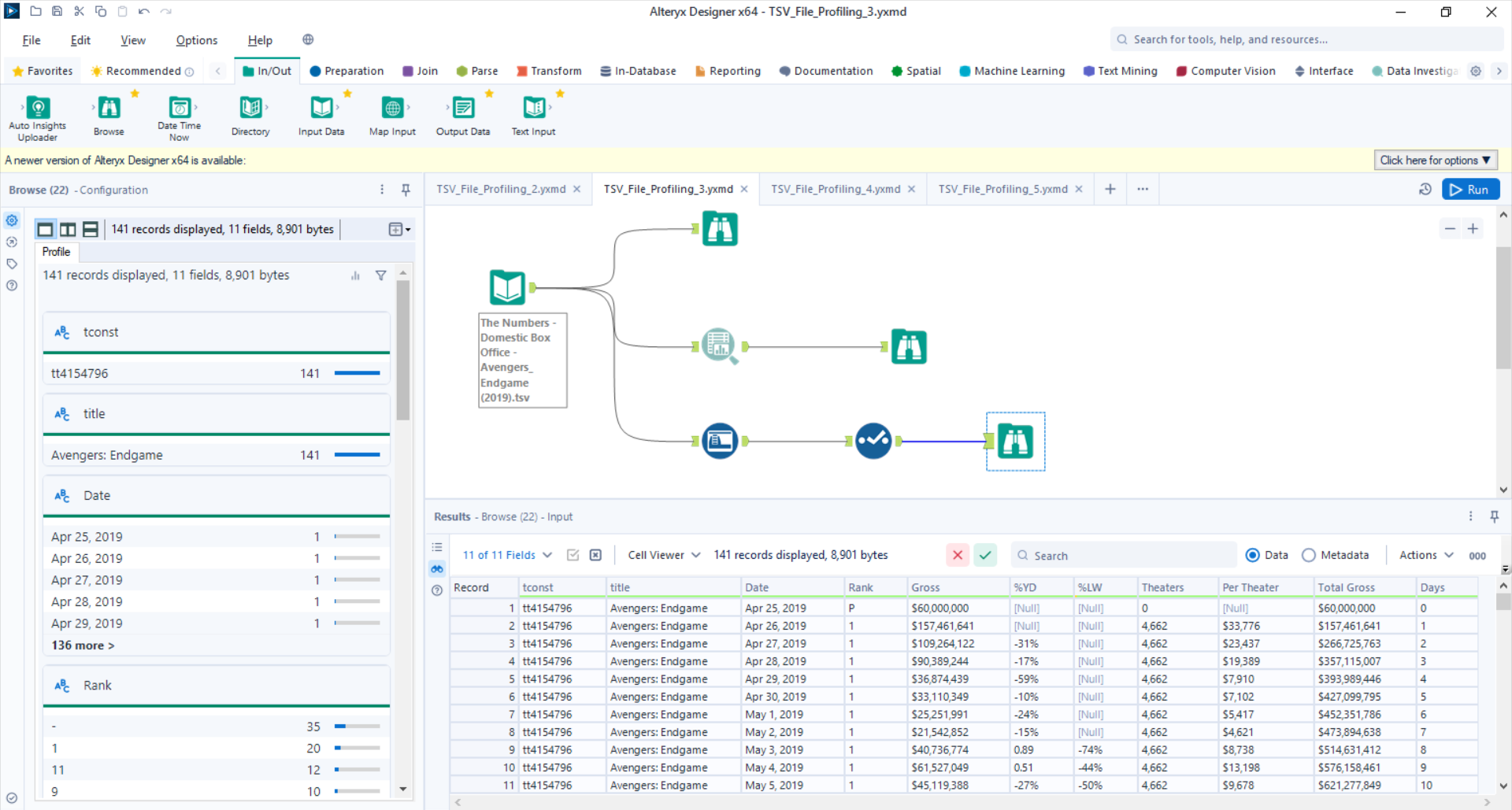
1. Avatar



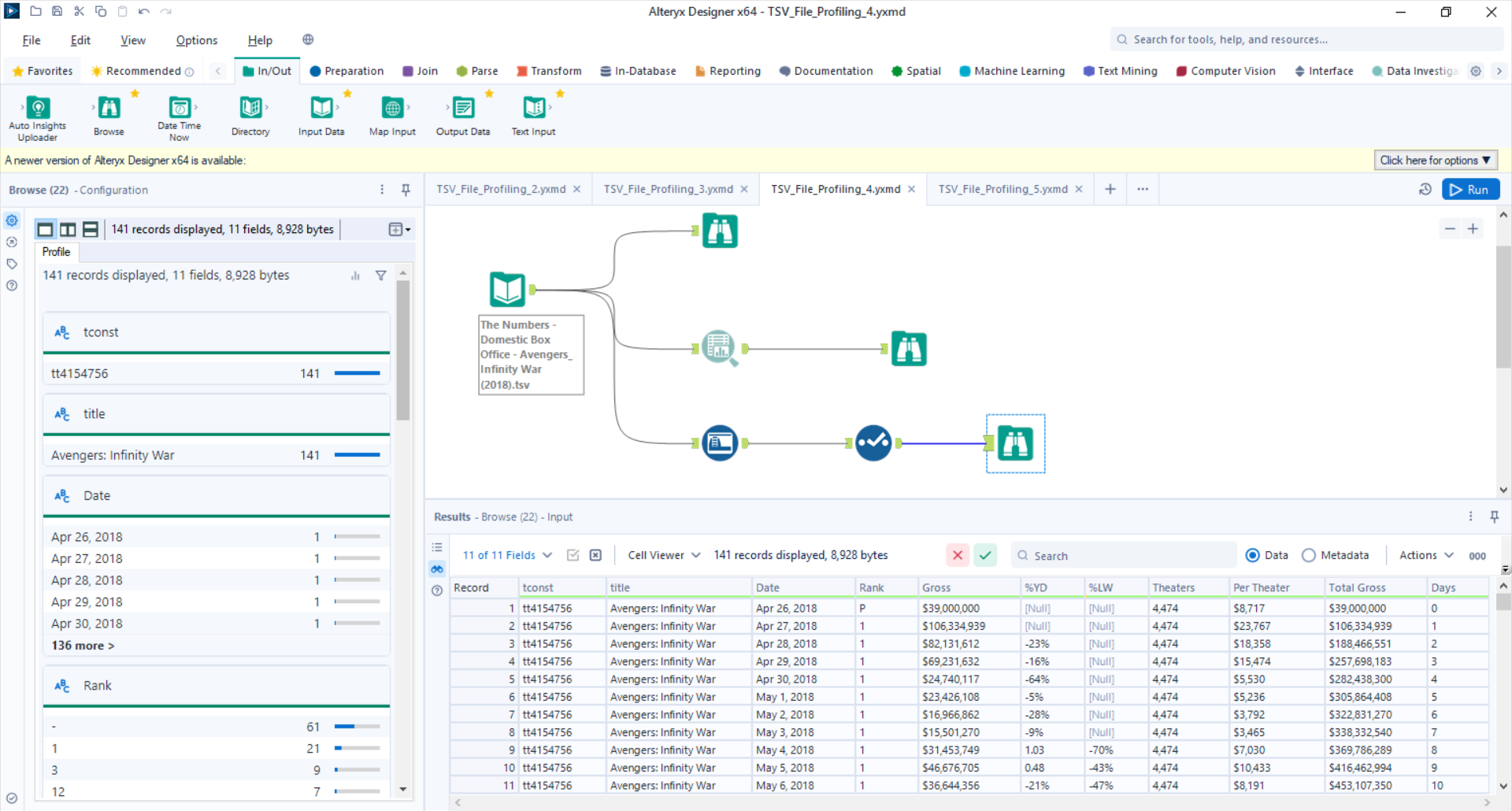
1. Avengers\_ Age of Ultron (2015)



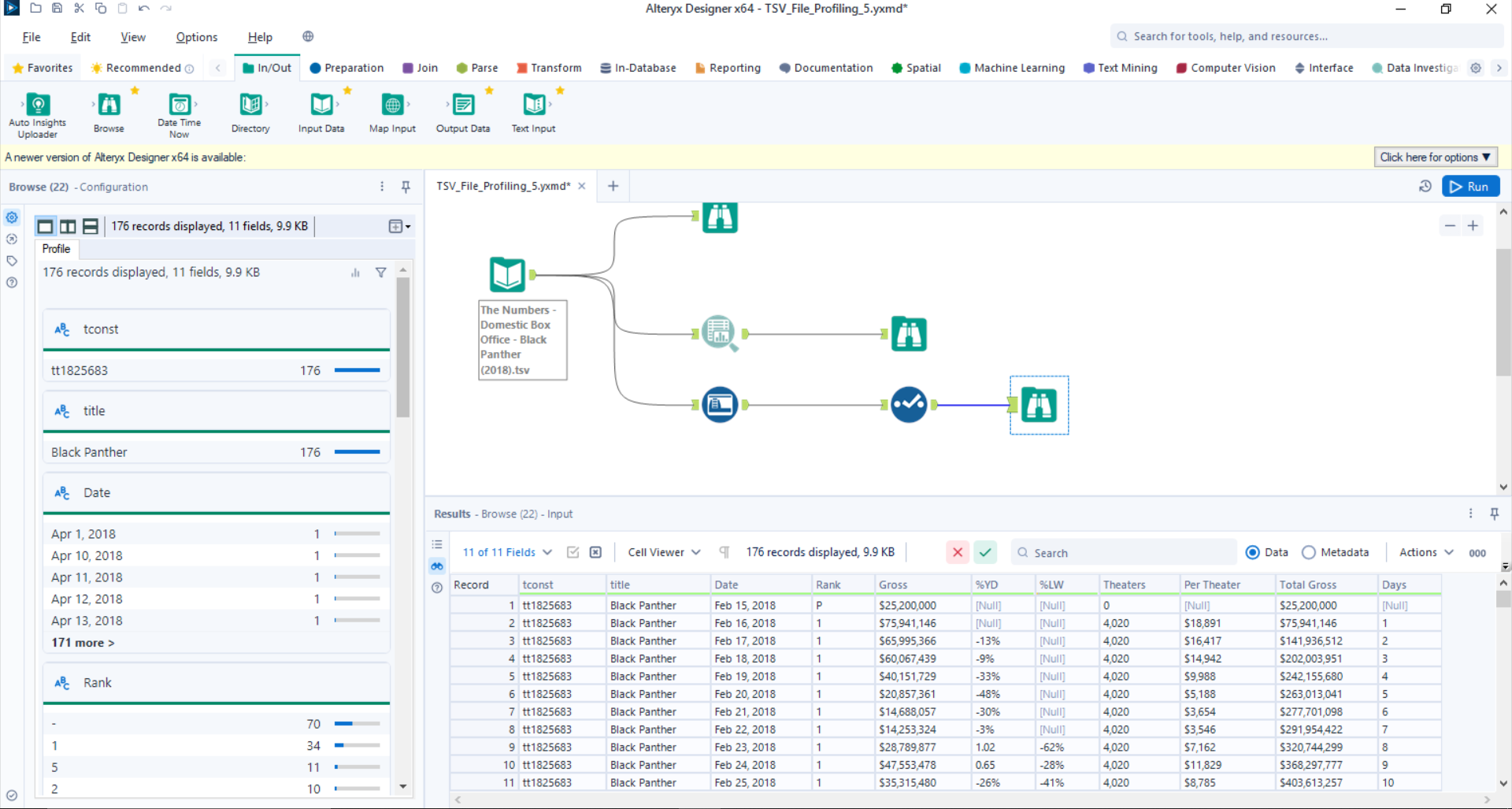
1. Avengers\_ Endgame (2019)



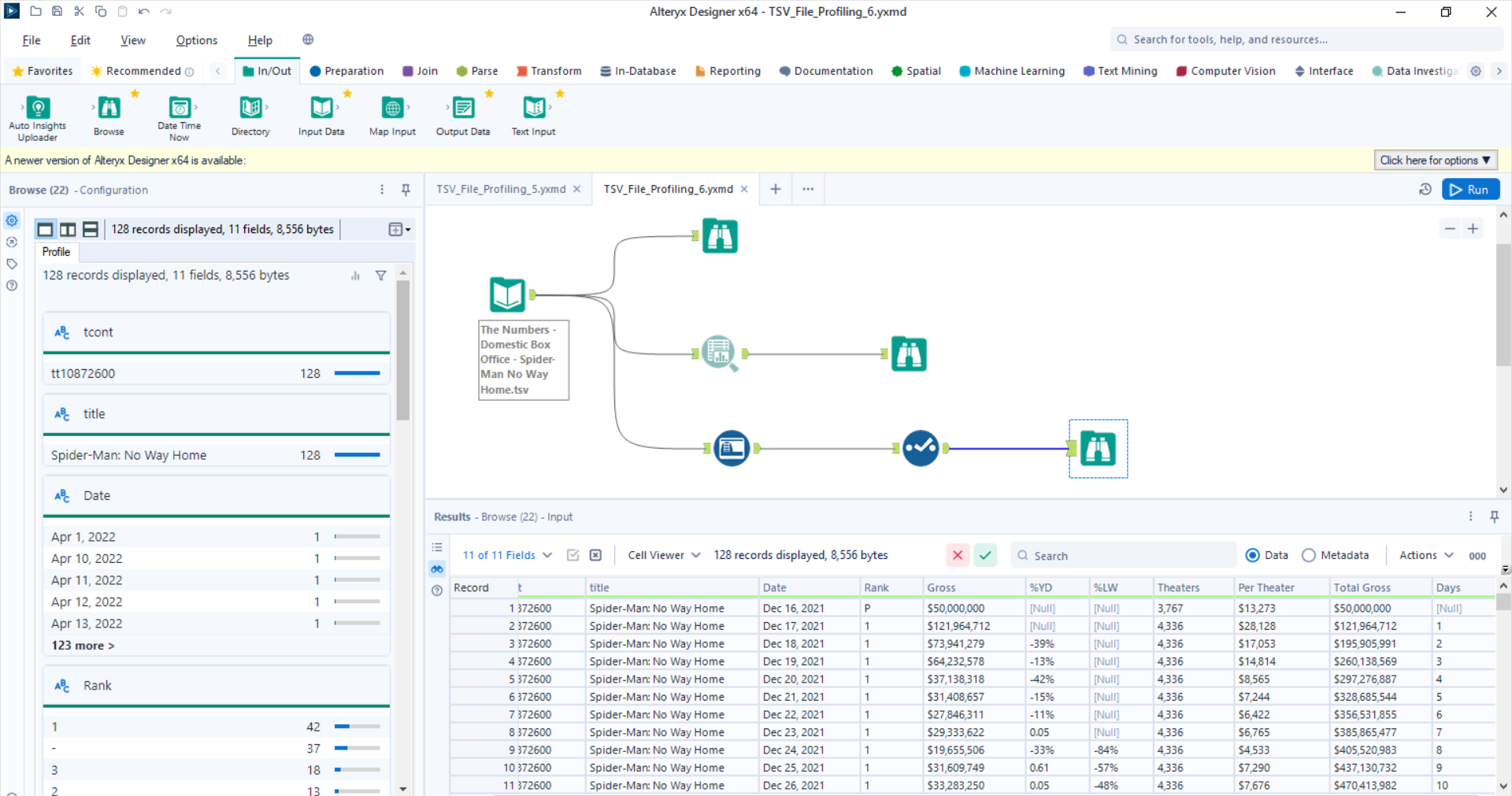
1. Avengers\_ Infinity War (2018)



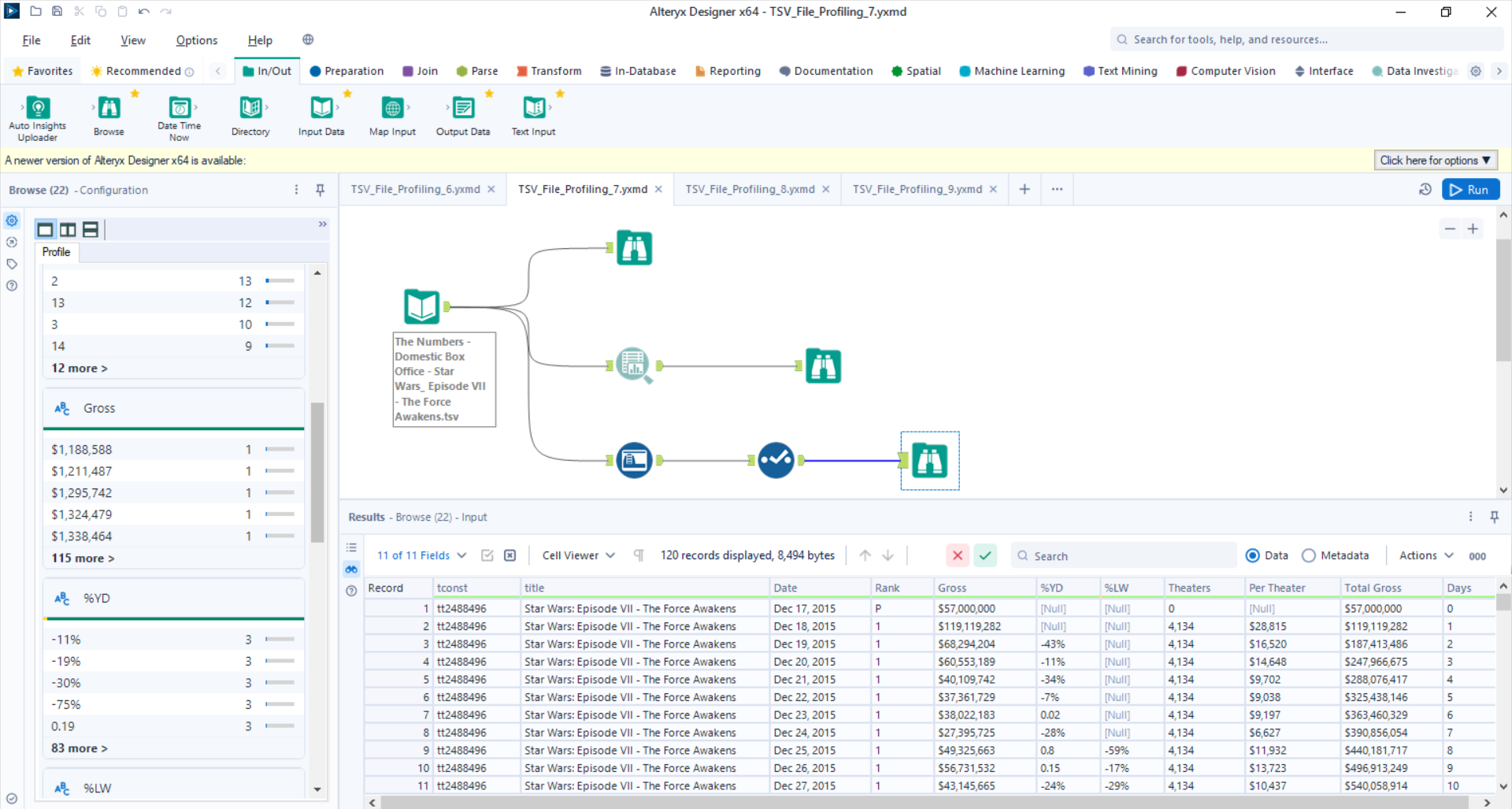
1. Black Panther (2018)



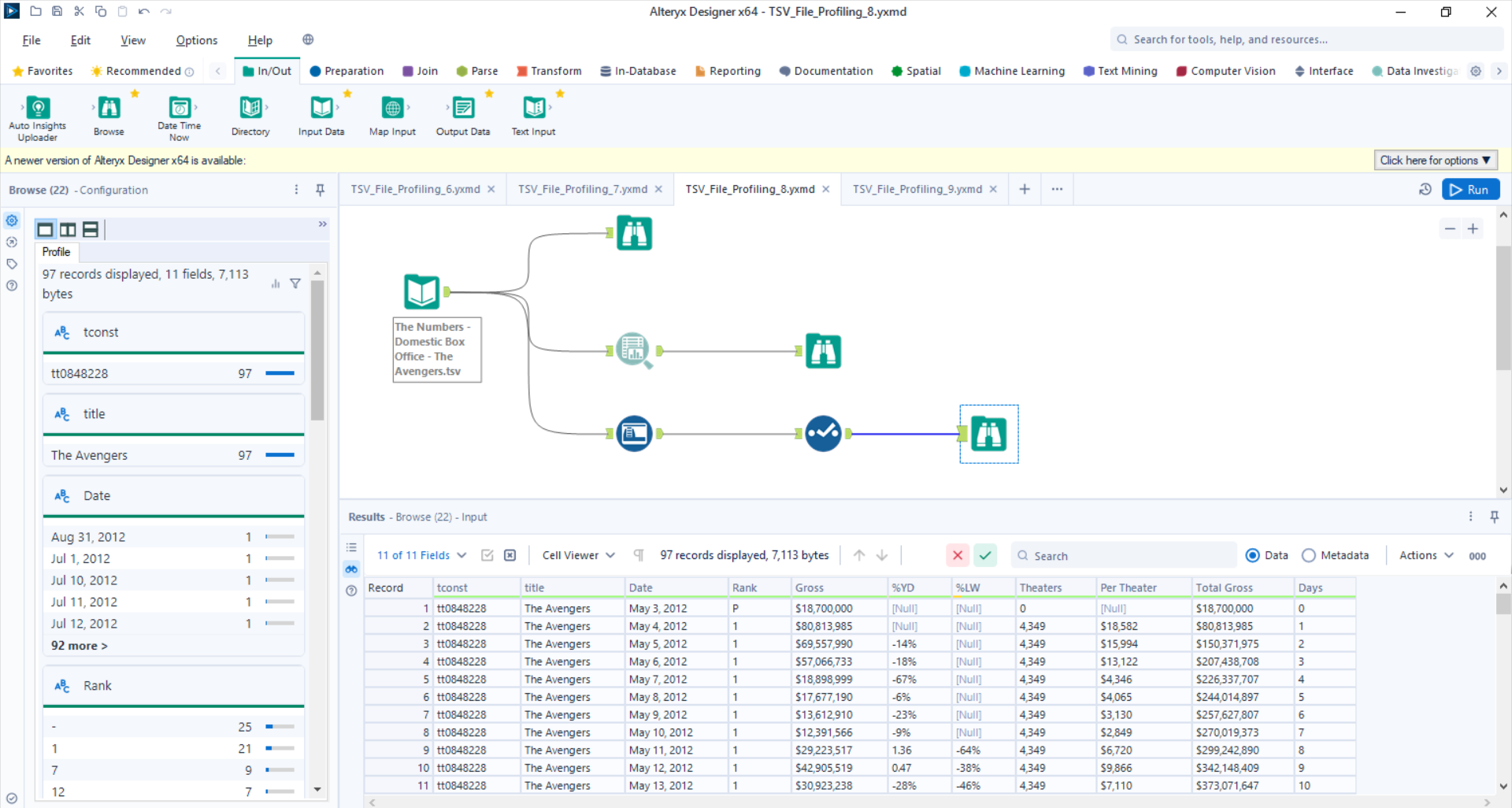
1. Spider-Man No Way Home



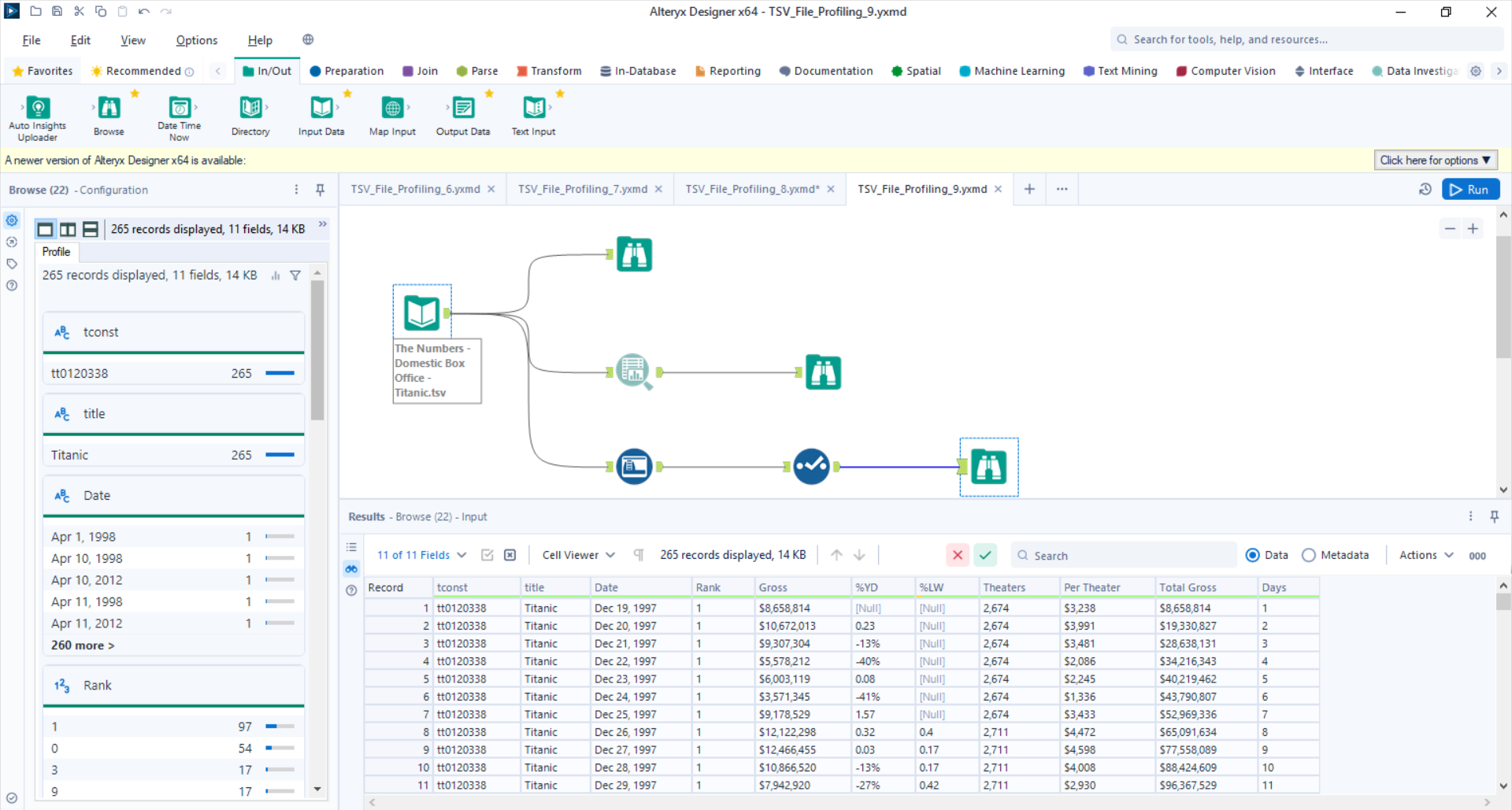
1. Star Wars\_ Episode VII - The Force Awakens



1. Domestic Box Office - The Avengers



1. Domestic Box Office – Titanic



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

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

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

**Gross-** Remove $ and comma

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

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

**Theaters-** Remove comma

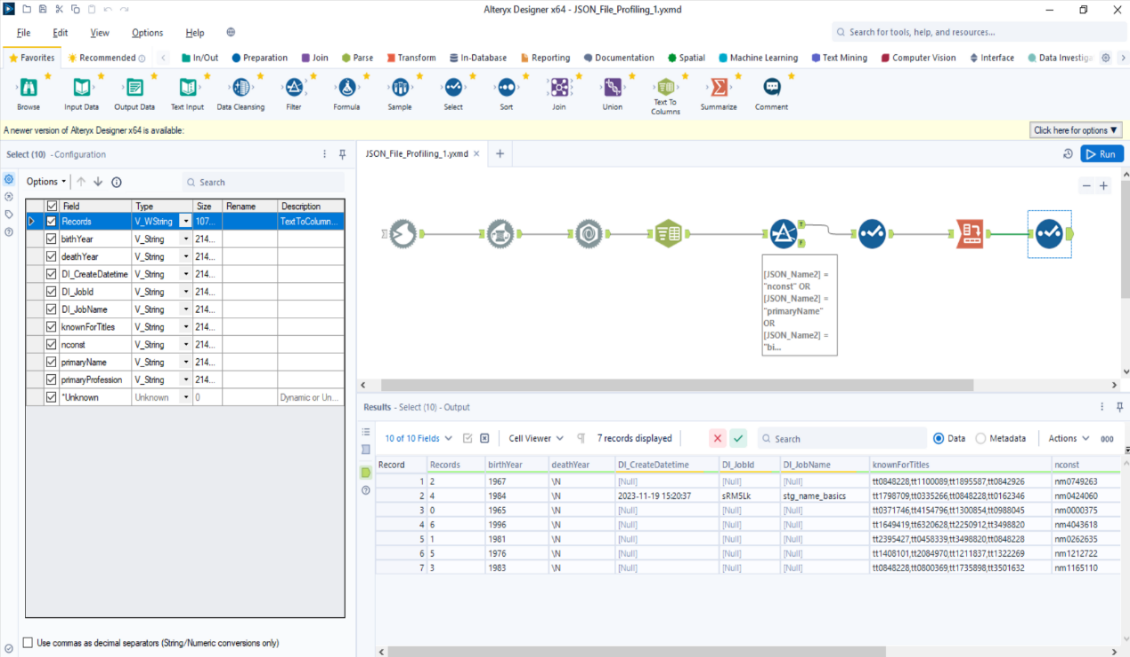
**Per Theater-** Remove $ and comma

**Total Gross-** Remove $ and comma for calculation

**JSON ALTREYX WORKFLOW**

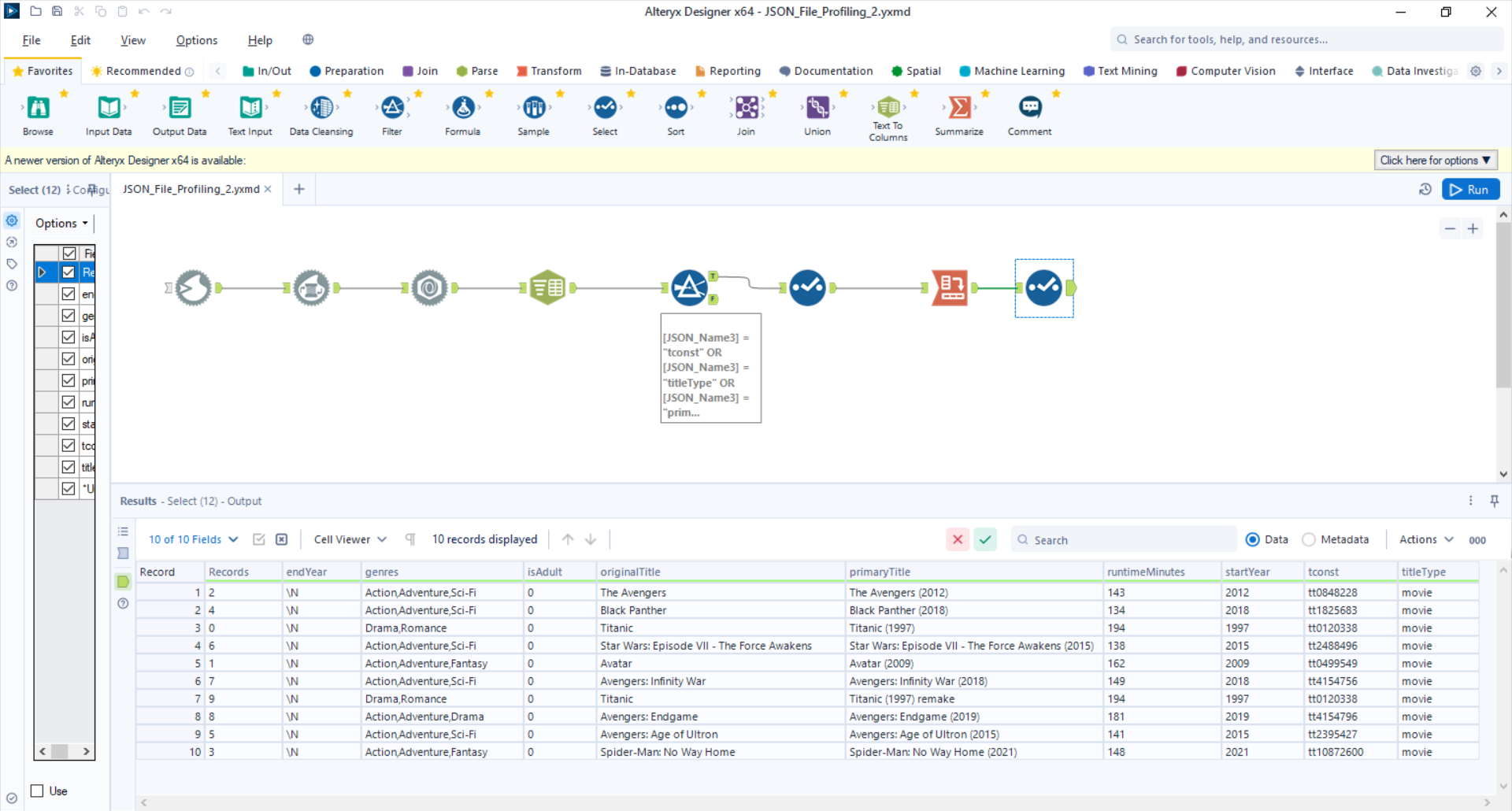
***SCREENSHOTS AND ANALYSIS***

1. **Used Blob Input**
2. **Convert to text using Blob Converter**
3. **Parse out the JSON**
4. **Transfer individual column we get into separate columns/split using Text to Column**
5. **Filter data by taking required field and using Select**
6. **Using CrossTab for transpose**
7. newdata\_name\_basics



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

1. newdata\_title\_basics

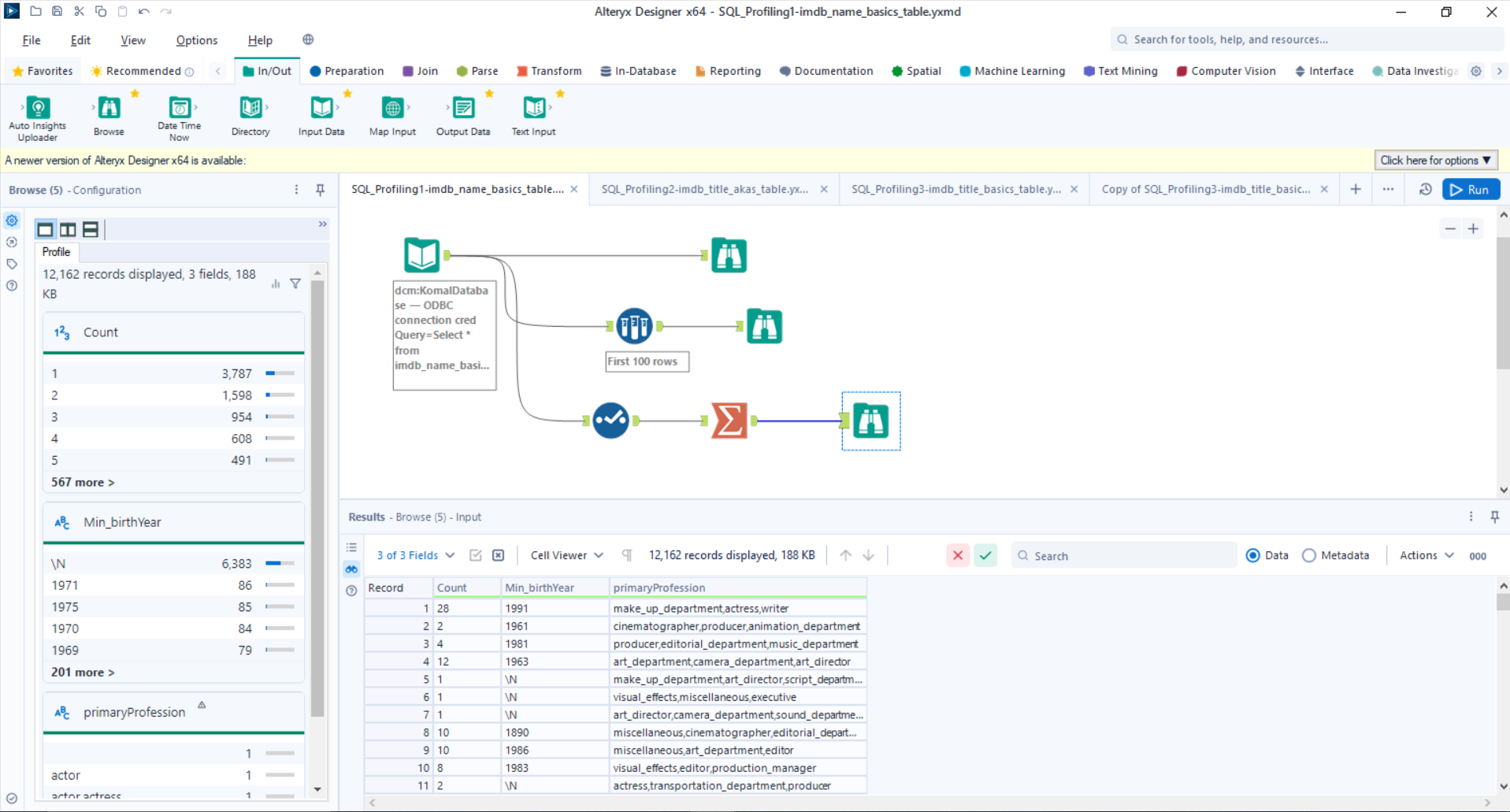


* The endYear column has NULL (\N) that is not adding value and is also not necessary for any visualization, so we can ignore the column or replace with 0

**SQL ALTREYX WORKFLOW**

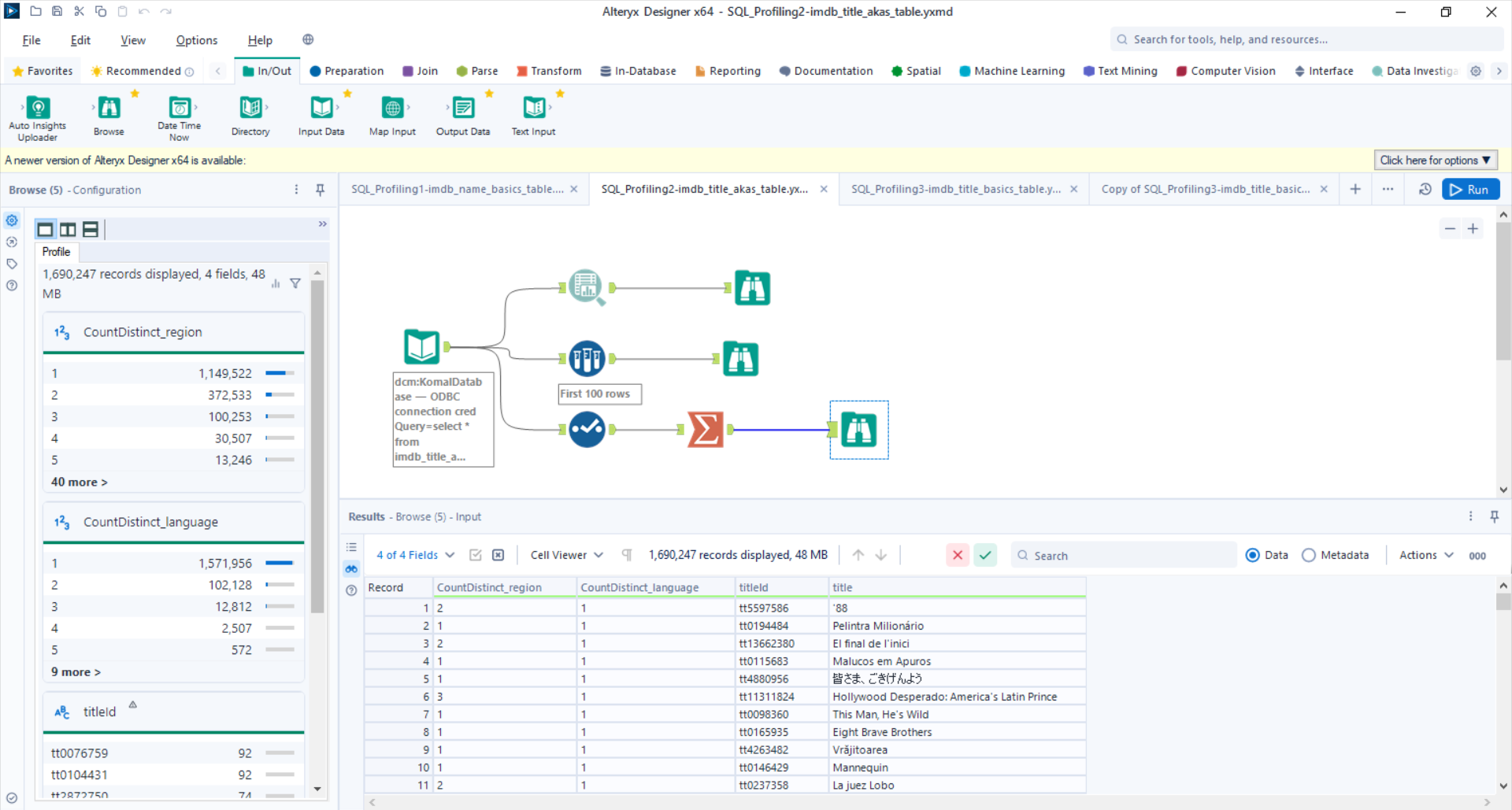
***SCREENSHOTS AND ANALYSIS***

1. imdb\_name\_basics



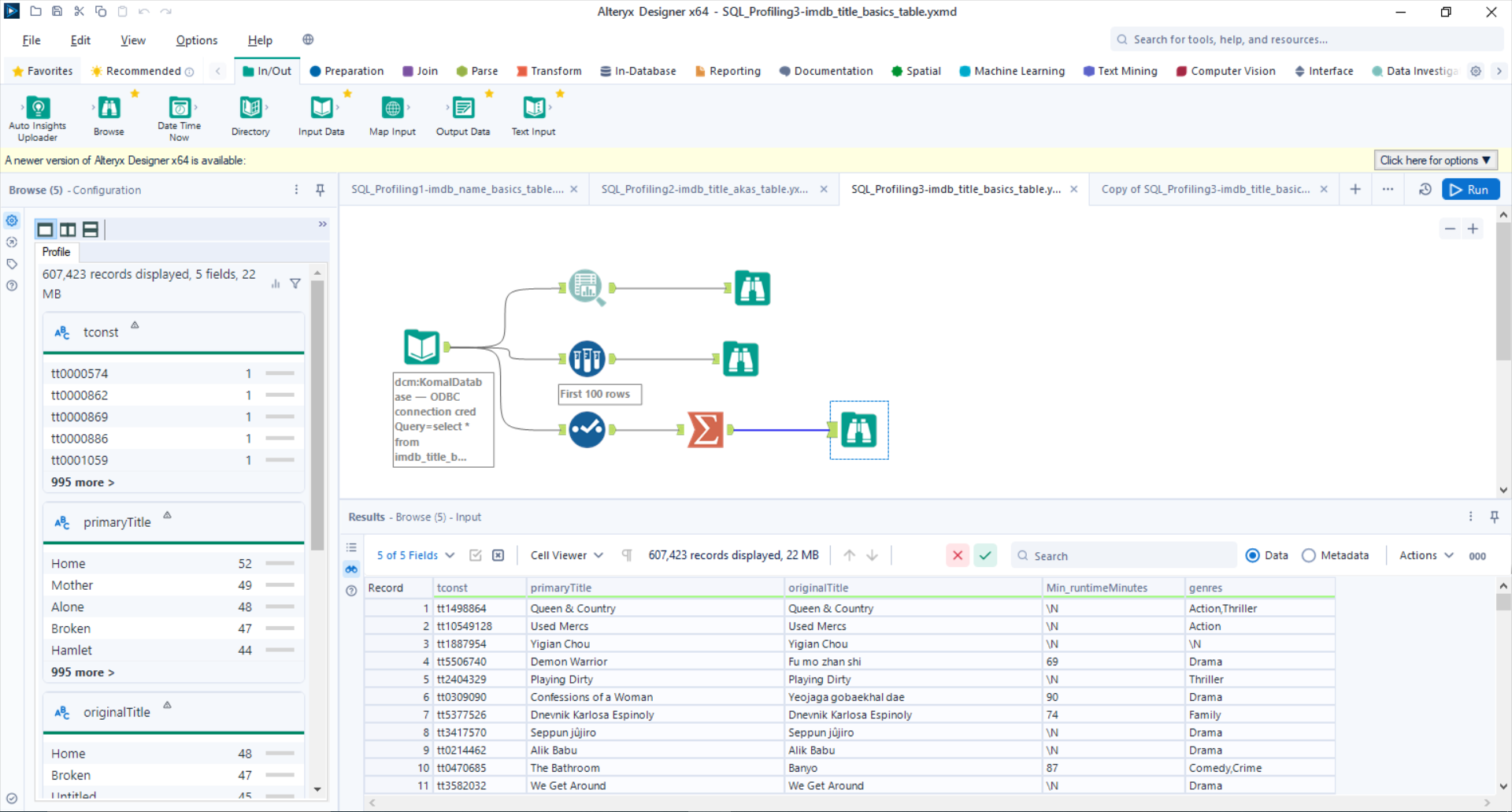
* **nconst, primaryName, birthYear, primaryProfession, knownForTitles, DI\_Create\_DT -** Everything is in proper format

1. imdb\_title\_akas



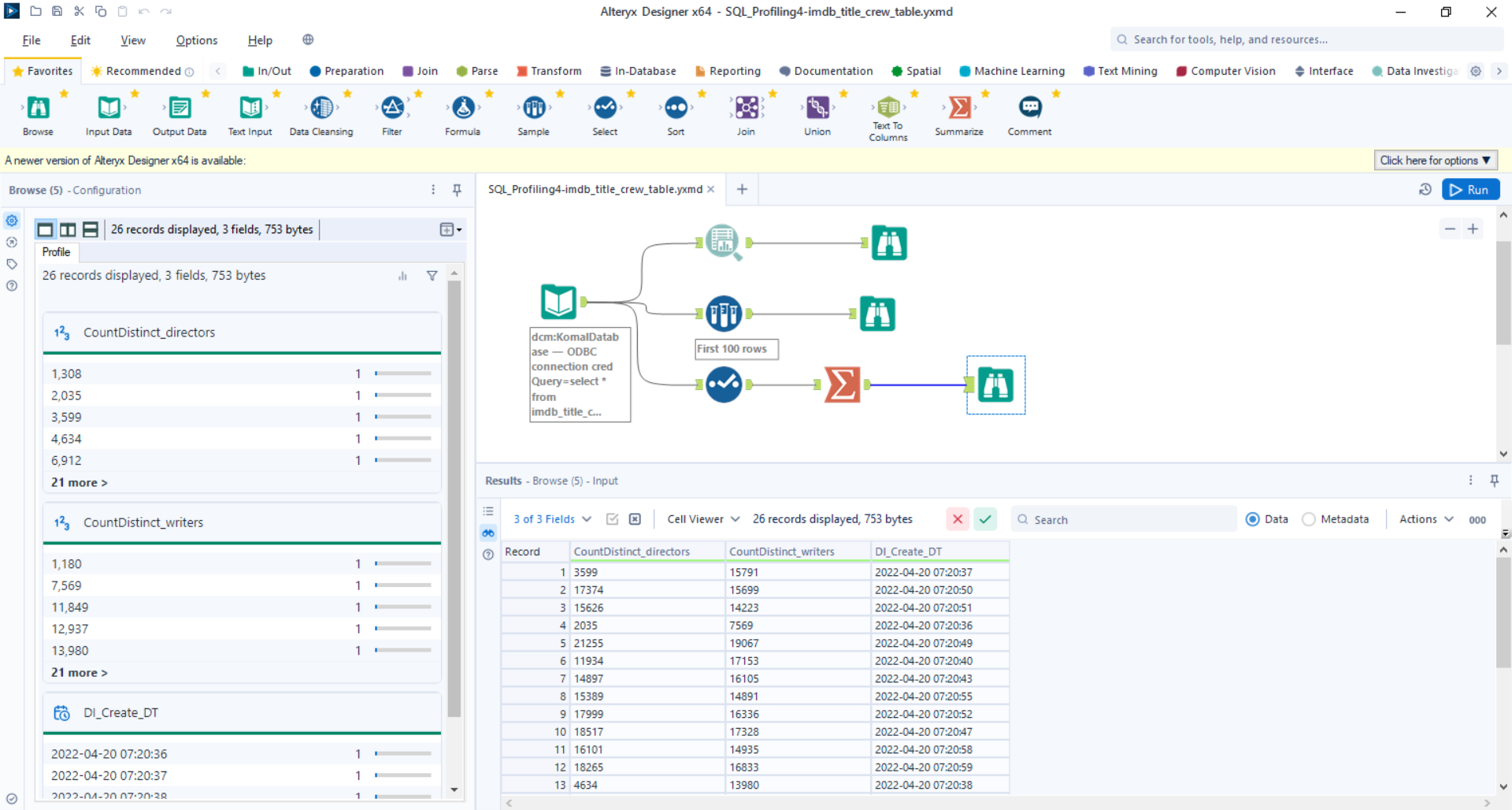
* **titleId, ordering, title, DI\_Create\_DT –** Everything is in proper format
* **region, language, types, attributes, isOriginalTitle –** Replace \N with 0/-1/99

1. imdb\_title\_basics



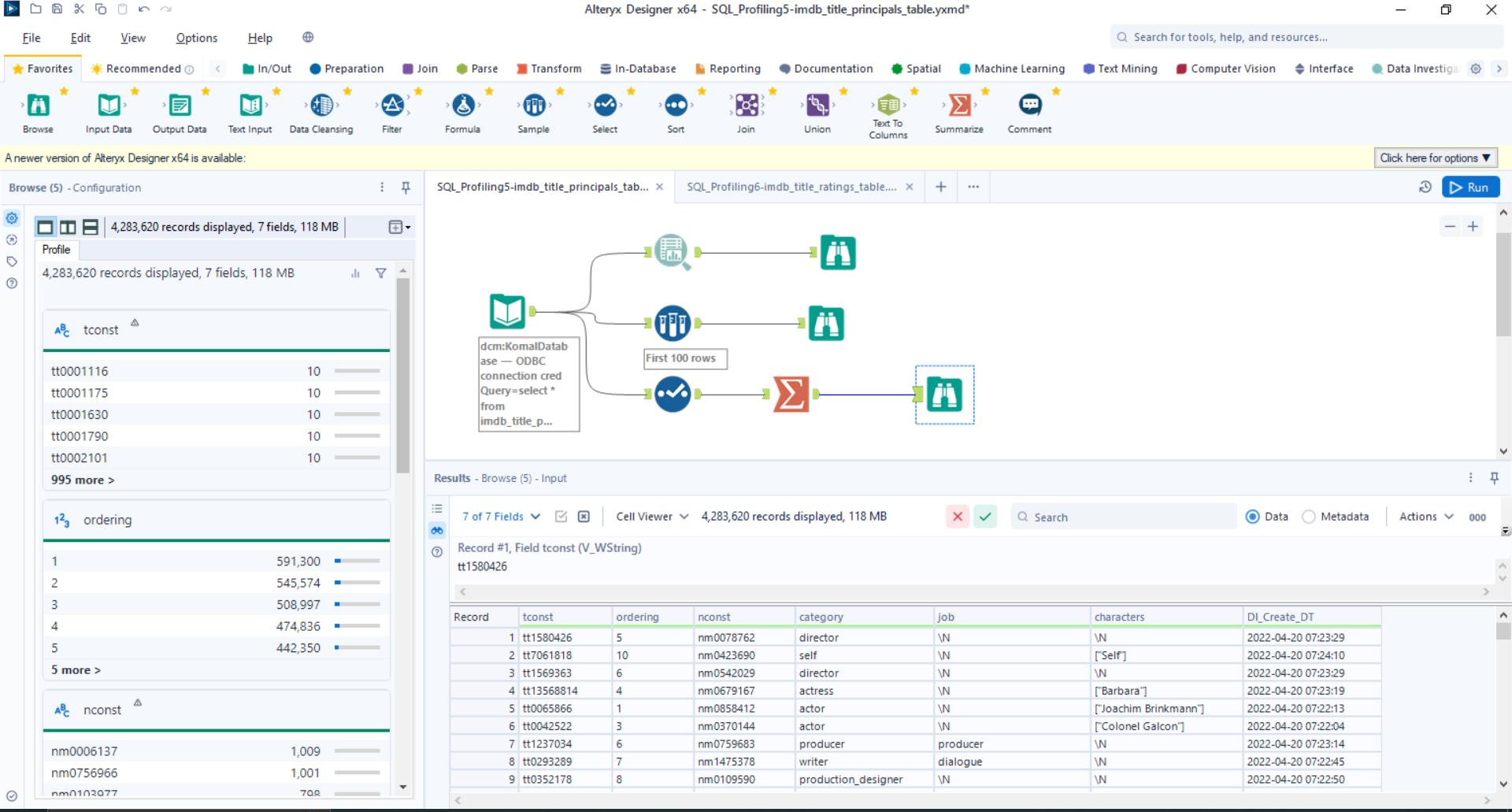
* **tconst, titleType, primaryTitle, originalTitle, isAdult, DI\_Create\_DT-** Everything is in proper format
* **startYear, endYear, runtimeMinutes, genres-** Replace \N values with 0 or -1

1. imdb\_title\_crew



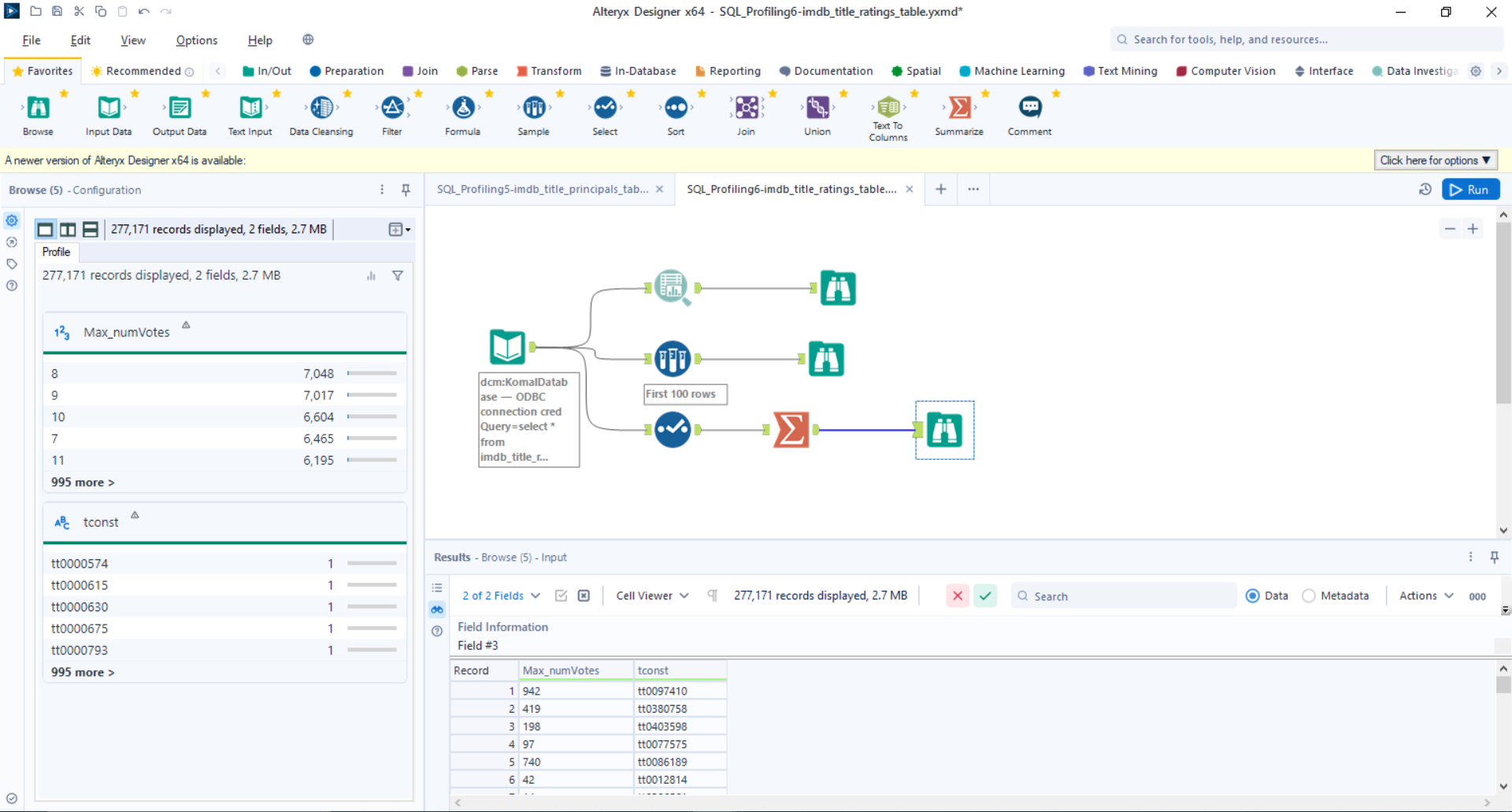
* **tconst ,DI\_Create\_DT –** Everything is in proper format
* **directors,writers -** Replace \N values with 0 or -1

1. imdb\_title\_principals



* **tconst, ordering, nconst, category, DI\_Create\_DT -** Everything is in proper format
* **job** -Replace \N with 0
* **characters –** Replace [“”] and \N with 0

1. imdb\_title\_ratings



* **tconst, averageRatings, DI\_Create\_DT -** Everything is in proper format
* **numVotes –** remove comma