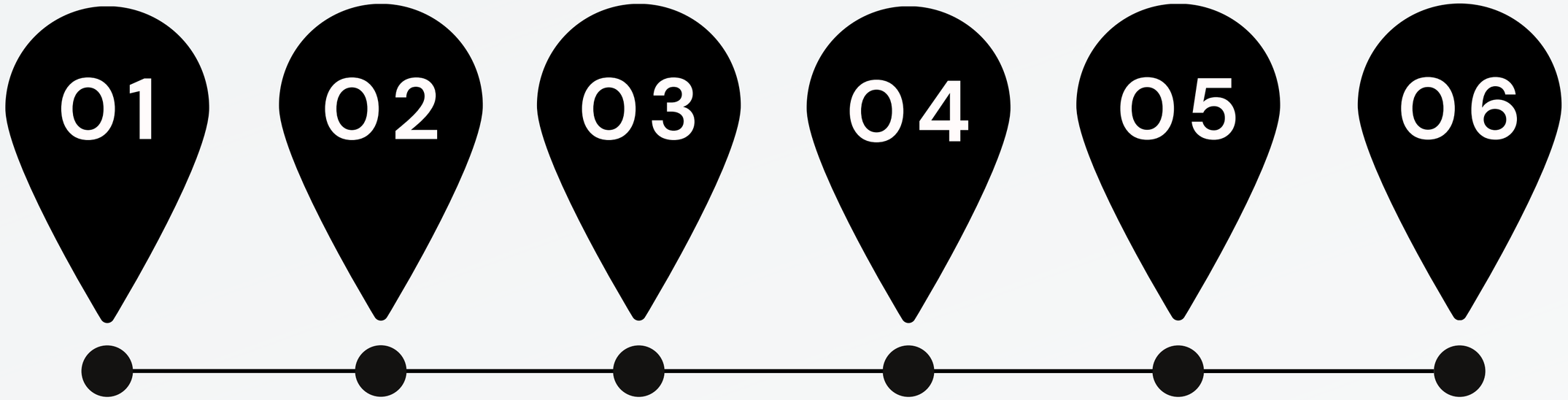
STREAMING SERVICE ANALYSIS - AZURE SQL

ABHISHEK TIKAM RAMCHANDANI – 002743745
JURREYAH FIRDAWS MOHAMMED – 002747514
SAMHITHA MEREDDY – 002796140
ATHARVA UPLENCHWAR – 002990536

CONTENT

Objectives



The project plans to focus on creating a relational database for storing and retrieving user, show, episode, subscription, and view information for streaming services.

Analyze user behavior to discover trends, viewing preferences, and watching patterns.

Determine the level of popularity for various forms of media, such as films, TV series, genres, and actors.

Offer streaming service providers business insights for selecting content licenses, devising ad campaigns, and enhancing user engagement.

In order to generate useful insights, the project will concentrate on analyzing data from streaming providers such as Netflix, Hulu, Disney+ etc.

Reports and data visualizations will be produced to highlight key findings for stakeholders.

ARCHITECTURE DIAGRAM

DATA SOURCE

alteryx

DATA CLEANING

Microsoft Azure
Blob Storage

STORING DATA

Azure
Data Factory

ETL

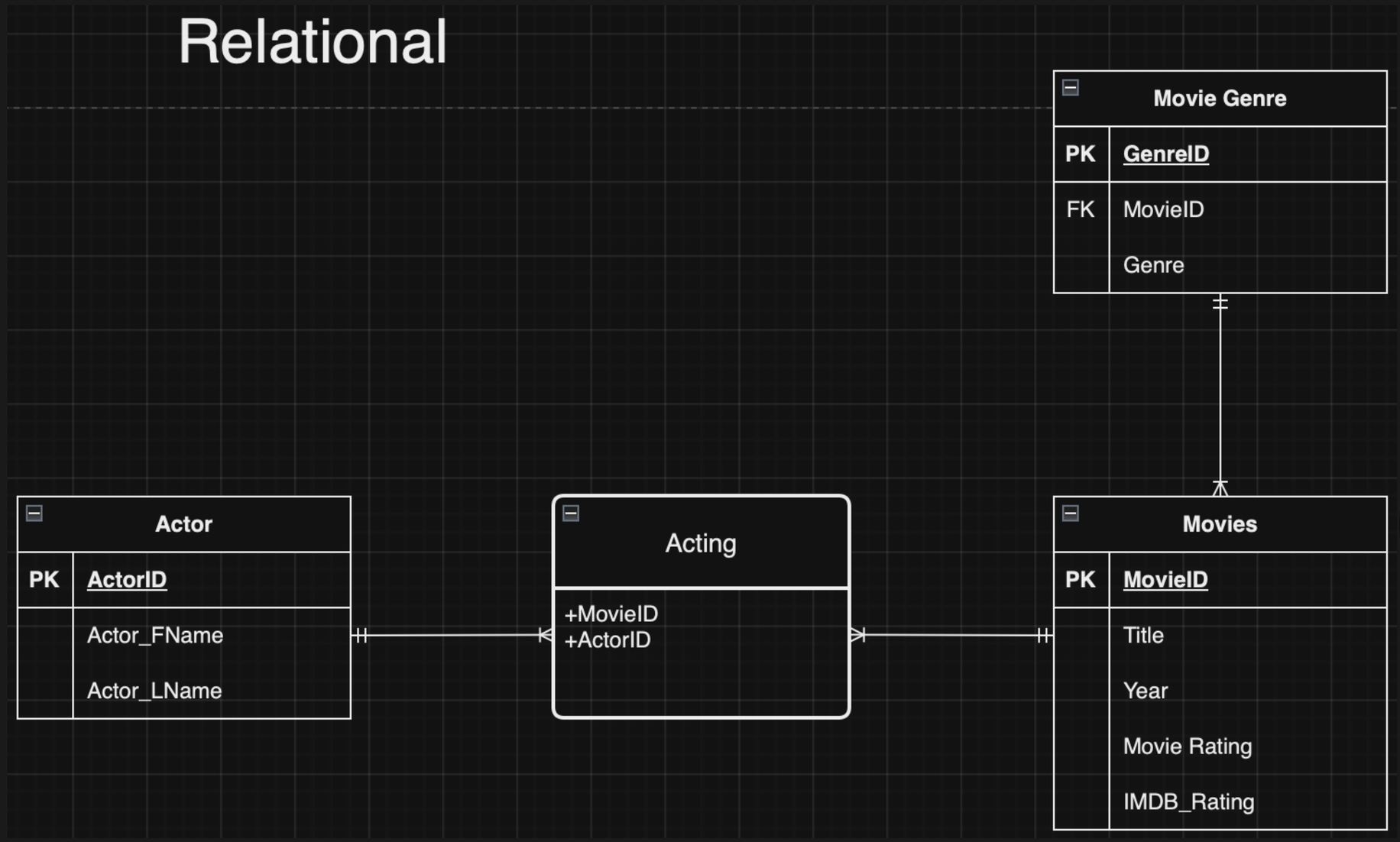
DATA REFRESH
& QUERYING

Azure SQL

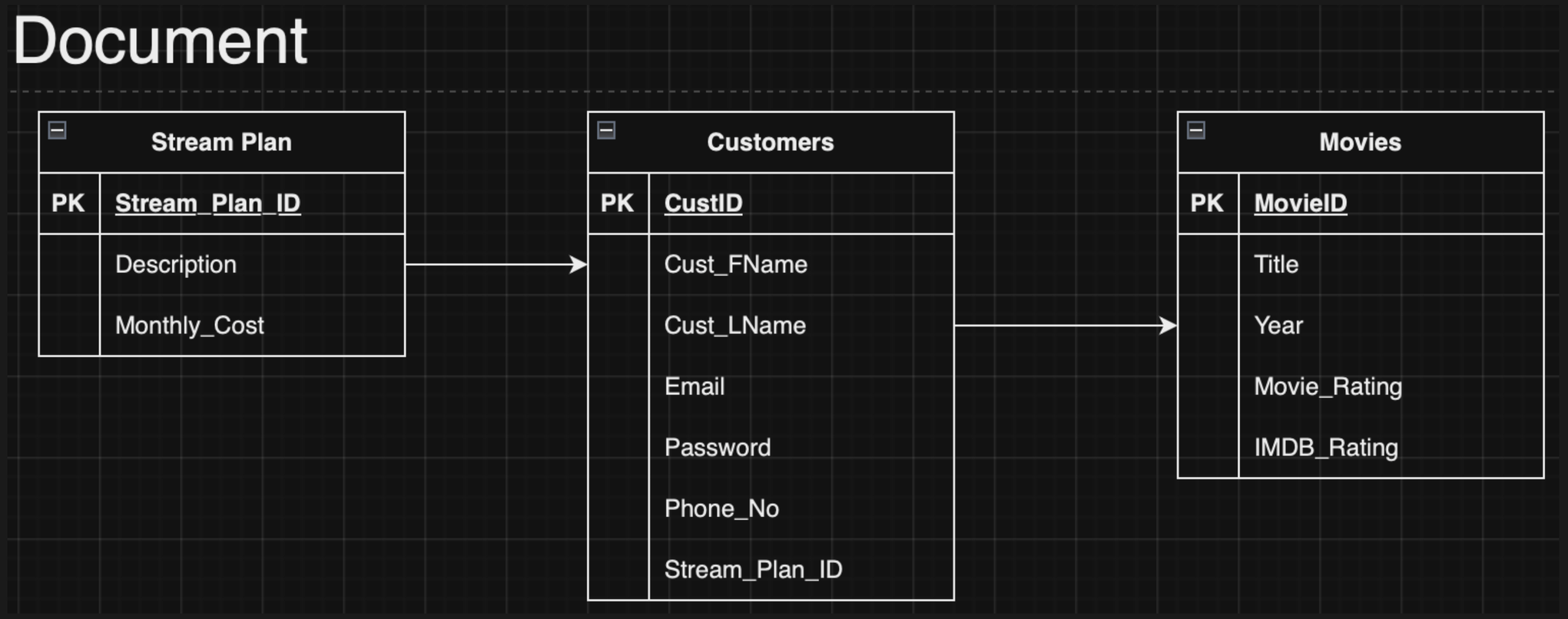
Power BI

VISUALIZATIONS

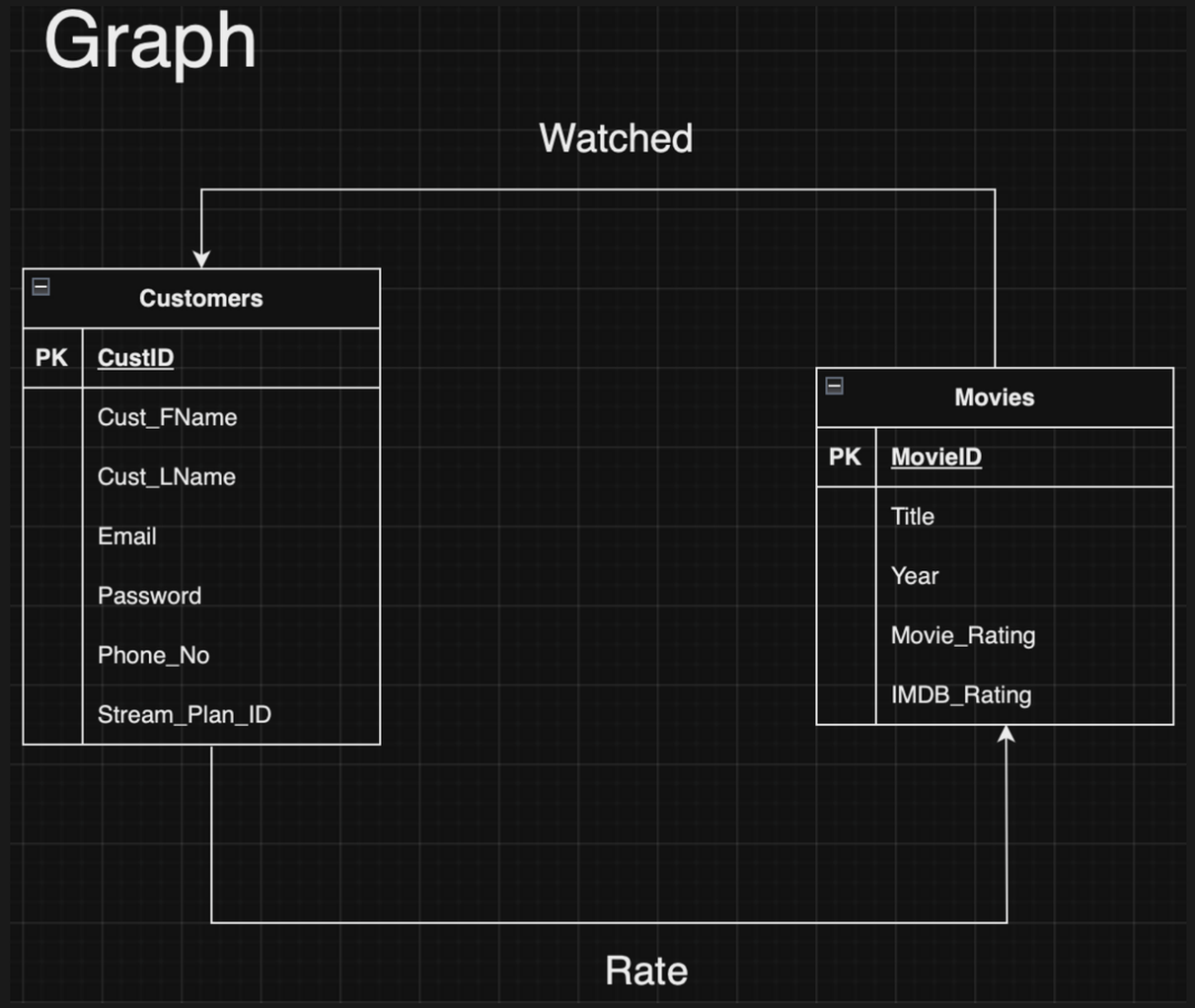
ERD-RELATIONAL DATABASE



ERD-DOCUMENT DATABASE



ERD-GRAPH DATABSE



RAW-DATASET

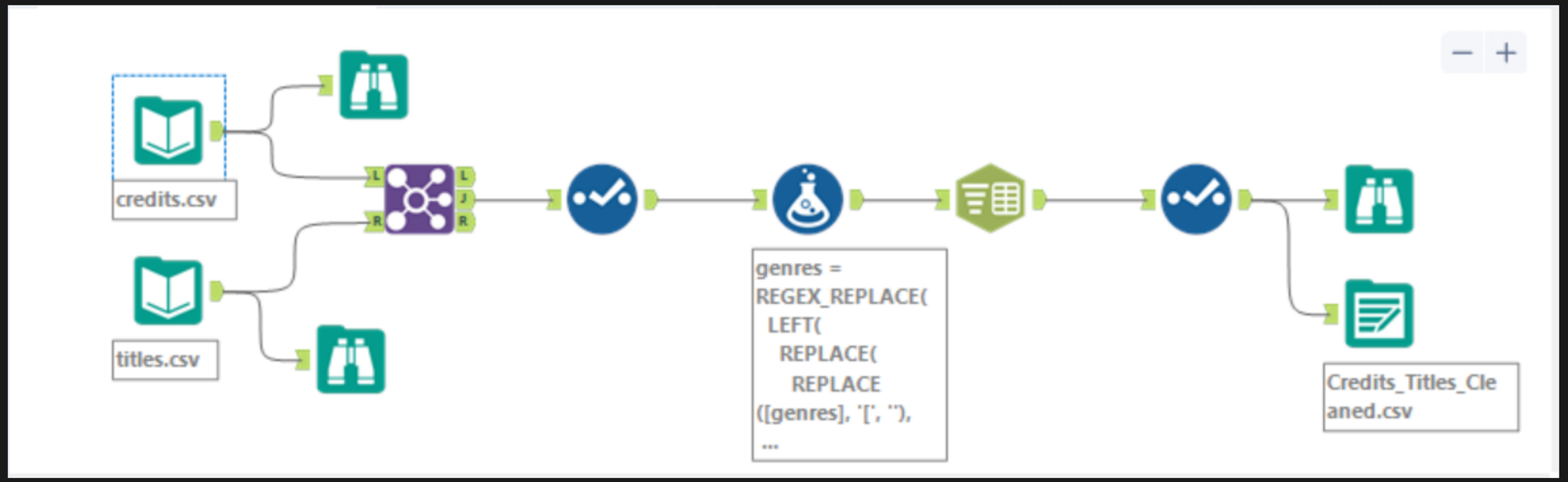
id	title	type	description	release_ye	age_certif	runtime	genres	productio	seasons	imdb_id	imdb_sco	imdb_vote	tmdb_pop	tmdb_sco
ts300399	Five Came Back: The Referer	SHOW	This collection i	1945	TV-MA	51	['documer	['US']	1				0.6	
tm84618	Taxi Driver	MOVIE	A mentally unst	1976	R	114	['drama',	['US']		tt0075314	8.2	808582	40.965	8.179
tm154986	Deliverance	MOVIE	Intent on seein	1972	R	109	['drama',	['US']		tt0068473	7.7	107673	10.01	7.3
tm127384	Monty Python and the Holy	MOVIE	King Arthur, acc	1975	PG	91	['fantasy',	['GB']		tt0071853	8.2	534486	15.461	7.811
tm120801	The Dirty Dozen	MOVIE	12 American mi	1967		150	['war', 'ac	['GB', 'US']		tt0061578	7.7	72662	20.398	7.6
ts22164	Monty Python's Flying Circus	SHOW	A British sketch	1969	TV-14	30	['comedy'	['GB']	4	tt0063929	8.8	73424	17.617	8.306
tm70993	Life of Brian	MOVIE	Brian Cohen is a	1979	R	94	['comedy'	['GB']		tt0079470	8	395024	17.77	7.8
tm14873	Dirty Harry	MOVIE	When a madma	1971	R	102	['thriller',	['US']		tt0066999	7.7	155051	12.817	7.5
tm119281	Bonnie and Clyde	MOVIE	In the 1930s, bc	1967	R	110	['crime', 'c	['US']		tt0061418	7.7	112048	15.687	7.5
tm98978	The Blue Lagoon	MOVIE	Two small child	1980	R	104	['romance	['US']		tt0080453	5.8	69844	50.324	6.156
tm44204	The Guns of Navarone	MOVIE	A team of allied	1961		158	['action', '	['GB', 'US']		tt0054953	7.5	50748	13.844	7.3
tm67378	The Professionals	MOVIE	An arrogant Tex	1966	PG-13	117	['western'	['US']		tt0060862	7.3	16446	13.123	7.1
tm69997	Richard Pryor: Live in Concer	MOVIE	Richard Pryor d	1979	R	78	['comedy'	['US']		tt0079807	8.1	5141	4.718	7.5
tm16479	White Christmas	MOVIE	Two talented sc	1954		115	['romance	['US']		tt0047673	7.5	42488	8.915	7.2
tm135083	Cairo Station	MOVIE	Qinawi, a physi	1958		77	['drama',	['EG']		tt0051390	7.5	4471	5.546	7.3
tm89386	Hitler: A Career	MOVIE	A keen chronicl	1977	PG	150	['history',	['DE']		tt0191182	7.5	2460	4.305	7.3
tm156453	FTA	MOVIE	A documentary	1972	R	97	['war', 'do	['US']		tt0068562	6.2	418	1.268	6.1

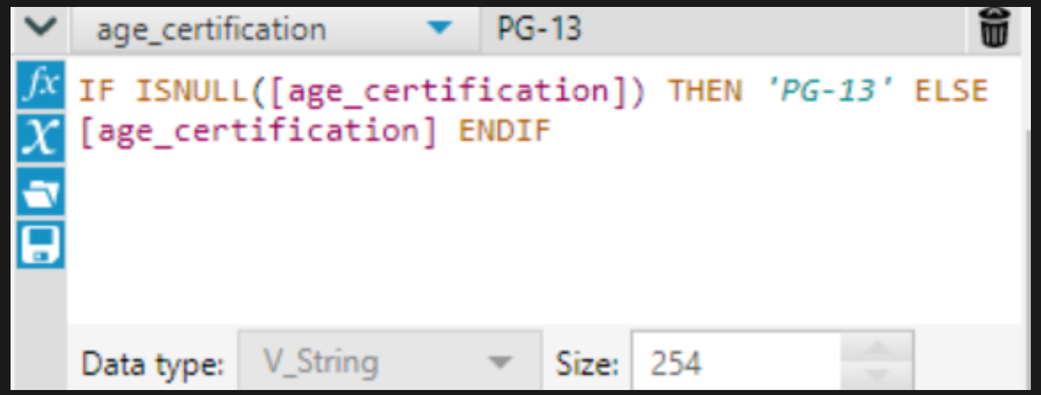
Titles Table

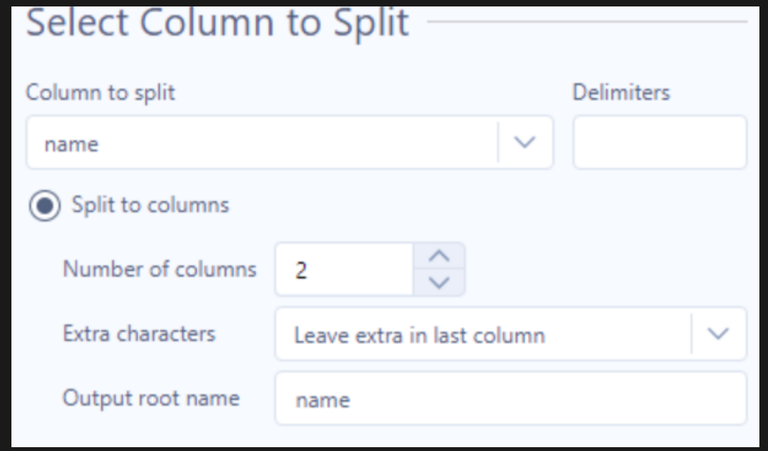
person_id	id	name	character	role
3748	tm84618	Robert De Niro	Travis Bickle	ACTOR
14658	tm84618	Jodie Foster	Iris Steensma	ACTOR
7064	tm84618	Albert Brooks	Tom	ACTOR
3739	tm84618	Harvey Keitel	Matthew 'Sport' Higgins	ACTOR
48933	tm84618	Cybill Shepherd	Betsy	ACTOR
32267	tm84618	Peter Boyle	Wizard	ACTOR
519612	tm84618	Leonard Harris	Senator Charles Palantine	ACTOR
29068	tm84618	Diahnne Abbott	Concession Girl	ACTOR
519613	tm84618	Gino Ardito	Policeman at Rally	ACTOR
3308	tm84618	Martin Scorsese	Passenger Watching Silhoue	ACTOR
43791	tm84618	Murray Moston	Iris' Time Keeper	ACTOR
519614	tm84618	Richard Higgs	Secret Service Agent	ACTOR
519615	tm84618	Bill Minkin	Tom's Assistant (uncreditec	ACTOR

Credits Table

ALTERYX-CLEANING







Field	Type	Size	Rename
id	V_String	254	
name	V_String	254	
title	V_String	254	
release_year	V_String	254	
age_certification	V_String	254	
genres	V_String	254	
imdb_score	Float	4	
name1	V_String	254	first_name
name2	V_String	254	last_name
*Unknown	Unknown	0	

DATASET AFTER CLEANING

Titles Uncleaned

id	title	type	description	release_ye	age_certif	runtime	genres	productio	seasons	imdb_id	imdb_sco	imdb_vote	tmdb_pop	tmdb_score
ts300399	Five Came Back: The Referer	SHOW	This collection i	1945	TV-MA	51	['documer	['US']	1				0.6	
tm84618	Taxi Driver	MOVIE	A mentally unst	1976	R	114	['drama', '	['US']		tt0075314	8.2	808582	40.965	8.179
tm154986	Deliverance	MOVIE	Intent on seein	1972	R	109	['drama', '	['US']		tt0068473	7.7	107673	10.01	7.3
tm127384	Monty Python and the Holy	MOVIE	King Arthur, acc	1975	PG	91	['fantasy',	['GB']		tt0071853	8.2	534486	15.461	7.811
tm120801	The Dirty Dozen	MOVIE	12 American mi	1967		150	['war', 'ac	['GB', 'US']		tt0061578	7.7	72662	20.398	7.6
ts22164	Monty Python's Flying Circus	SHOW	A British sketch	1969	TV-14	30	['comedy'	['GB']	4	tt0063929	8.8	73424	17.617	8.306
tm70993	Life of Brian	MOVIE	Brian Cohen is a	1979	R	94	['comedy'	['GB']		tt0079470	8	395024	17.77	7.8
tm14873	Dirty Harry	MOVIE	When a madma	1971	R	102	['thriller',	['US']		tt0066999	7.7	155051	12.817	7.5
tm119281	Bonnie and Clyde	MOVIE	In the 1930s, bc	1967	R	110	['crime', 'c	['US']		tt0061418	7.7	112048	15.687	7.5
tm98978	The Blue Lagoon	MOVIE	Two small child	1980	R	104	['romance	['US']		tt0080453	5.8	69844	50.324	6.156

Titles Cleaned

id	Title	Year	Movie_Rating	Genre	Ratings
ts300399	NocturnalAnimals	2016	R	thriller	7.5
tm84618	Dharmakshetra	2014	TV-14	drama	8.3
tm154986	BoBurnham:Inside	2021	R	comedy	8.7
tm127384	RiseofEmpires:Ottoman	2020	PG-13	drama	7.9
tm120801	IO	2019	PG-13	scifi	4.7
ts22164	AndBreatheNormally	2018	PG-13	drama	6.9
tm70993	PutYourHeadonMyShoulder	2019	TV-Y7	drama	8
tm14873	RealRob	2015	TV-14	comedy	6.3
tm119281	TheDUFF	2015	PG-13	comedy	6.4
tm98978	TheLincolnLawyer	2022	TV-MA	crime	7.7
tm44204	YooHoototheRescue	2019	TV-Y	animation	6.8
tm67378	DI4RI	2022	PG-13	family	7.3
tm69997	AjabPremKiGhazabKahani	2009	PG	comedy	6.3
tm16479	RaginiMMS	2011	PG-13	drama	5.1
tm135083	Kaaval	2021	PG-13	thriller	5.1
tm89386	WelcometoWeddingHell	2022	TV-14	drama	6.9
tm156453	IntotheInferno	2016	PG-13	documentation	7.2

DATASET AFTER CLEANING

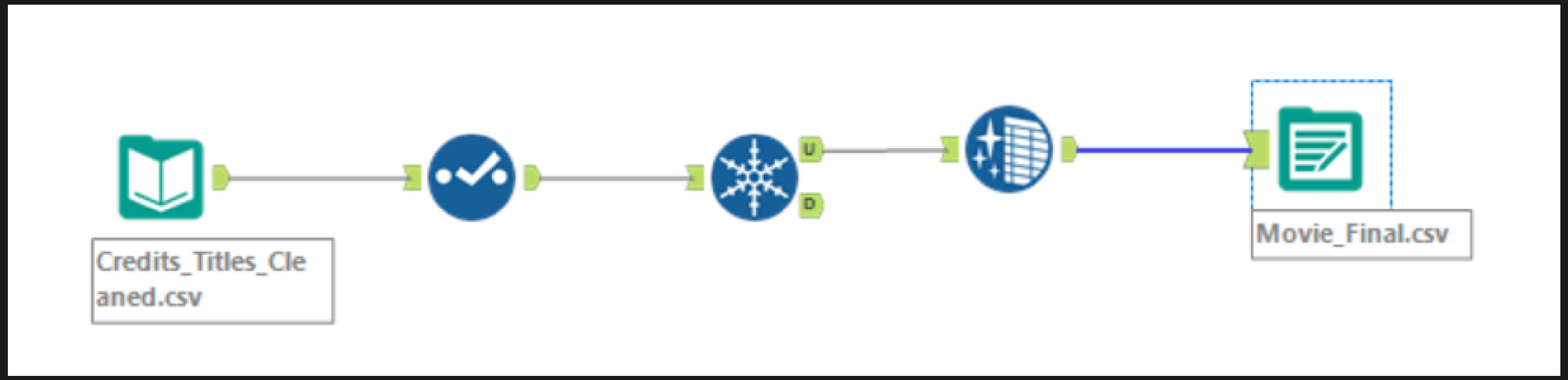
Credits Uncleaned

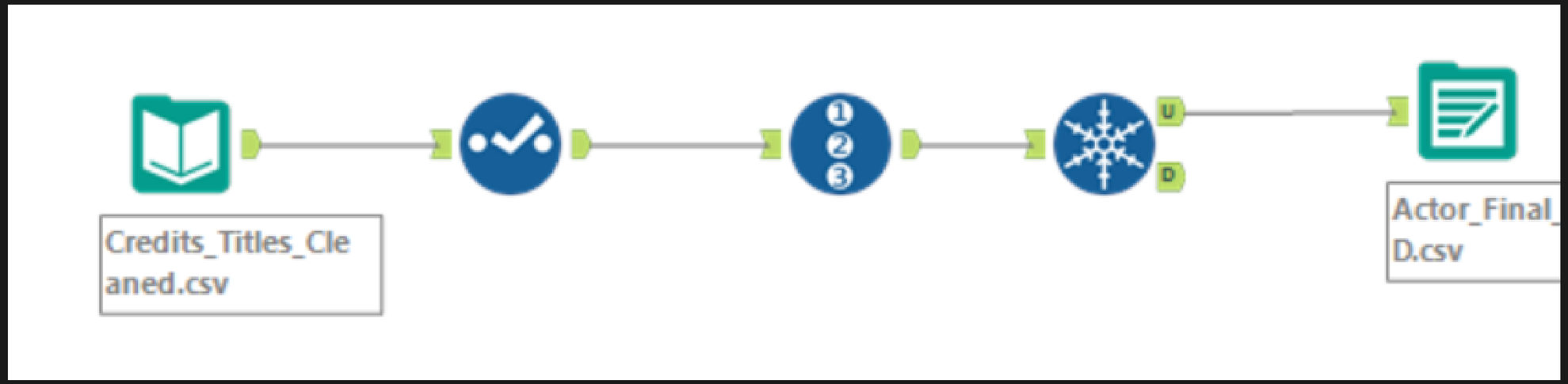
person_id	id	name	character	role
3748	tm84618	Robert De Niro	Travis Bickle	ACTOR
14658	tm84618	Jodie Foster	Iris Steensma	ACTOR
7064	tm84618	Albert Brooks	Tom	ACTOR
3739	tm84618	Harvey Keitel	Matthew 'Sport' Higgins	ACTOR
48933	tm84618	Cybill Shepherd	Betsy	ACTOR
32267	tm84618	Peter Boyle	Wizard	ACTOR
519612	tm84618	Leonard Harris	Senator Charles Palantine	ACTOR
29068	tm84618	Diahnne Abbott	Concession Girl	ACTOR
519613	tm84618	Gino Ardito	Policeman at Rally	ACTOR
3308	tm84618	Martin Scorsese	Passenger Watching Silhoue	ACTOR
43791	tm84618	Murray Moston	Iris' Time Keeper	ACTOR
519614	tm84618	Richard Higgs	Secret Service Agent	ACTOR
519615	tm84618	Bill Minkin	Tom's Assistant (uncreditec	ACTOR
82426	tm84618	Bob Maroff	Mafioso (uncredited)	ACTOR
20935	tm84618	Victor Argo	Melio, Delicatessen Owner	ACTOR
7753	tm84618	Joe Spinell	Personell Officer	ACTOR
43279	tm84618	Robinson Frank Adu	Angry Black Man (uncredite	ACTOR

Credits Cleaned

id	Actor_Fname	Actor_LName
tm84618	Robert	De Niro
tm84618	Jodie	Foster
tm84618	Albert	Brooks
tm84618	Harvey	Keitel
tm84618	Cybill	Shepherd
tm84618	Peter	Boyle
tm84618	Leonard	Harris
tm84618	Diahnne	Abbott
tm84618	Gino	Ardito
tm84618	Martin	Scorsese
tm84618	Murray	Moston
tm84618	Richard	Higgs
tm84618	Bill	Minkin
tm84618	Bob	Maroff
tm84618	Victor	Argo
tm84618	Joe	Spinell
tm84618	Robinson	Frank Adu

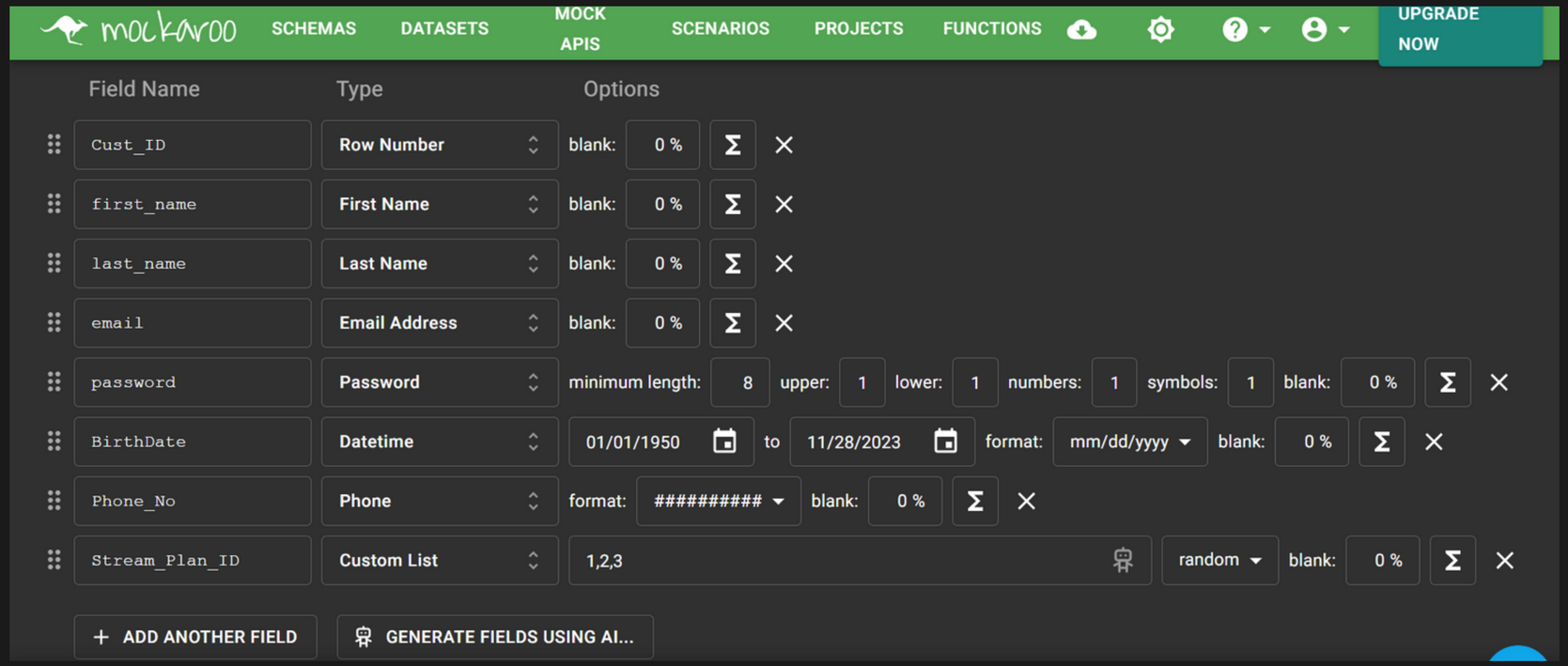
ALTERYX DATA FLOW





MOCK DATASET

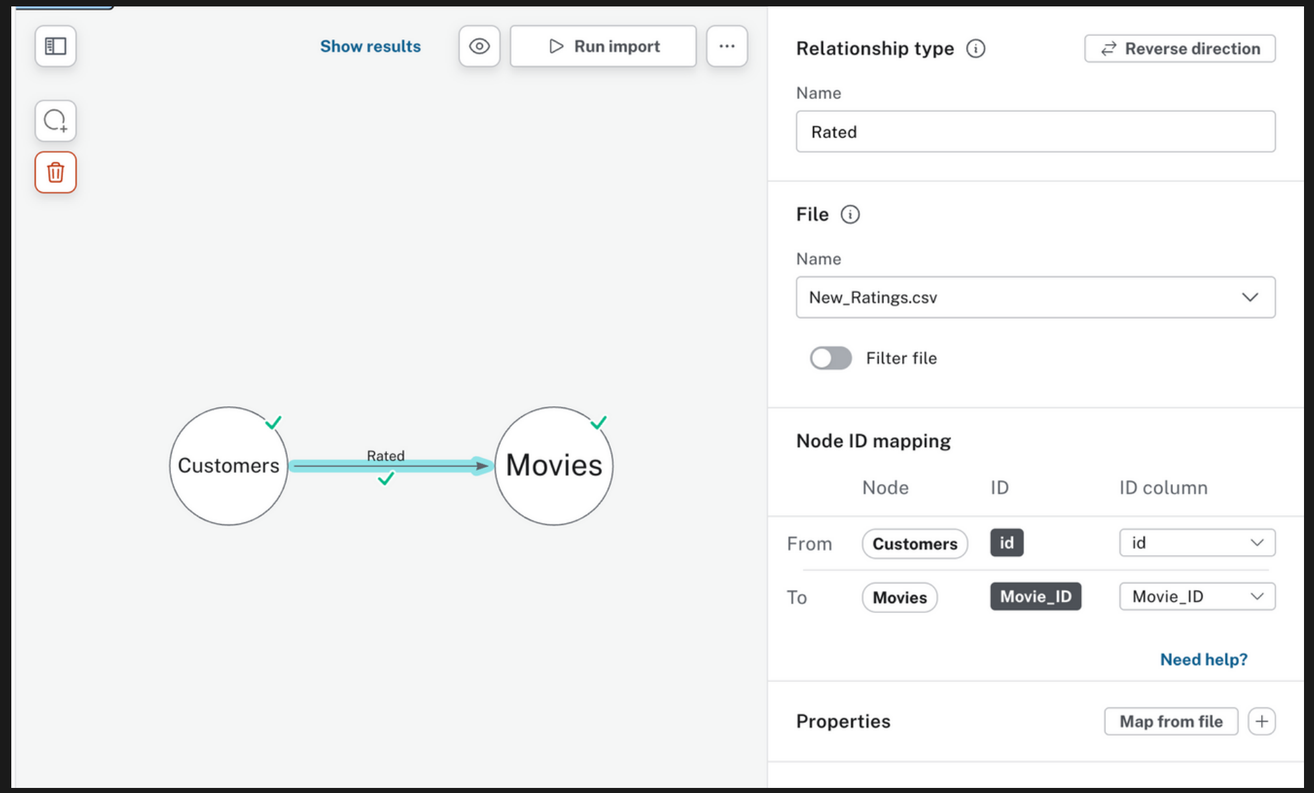
Customers Table

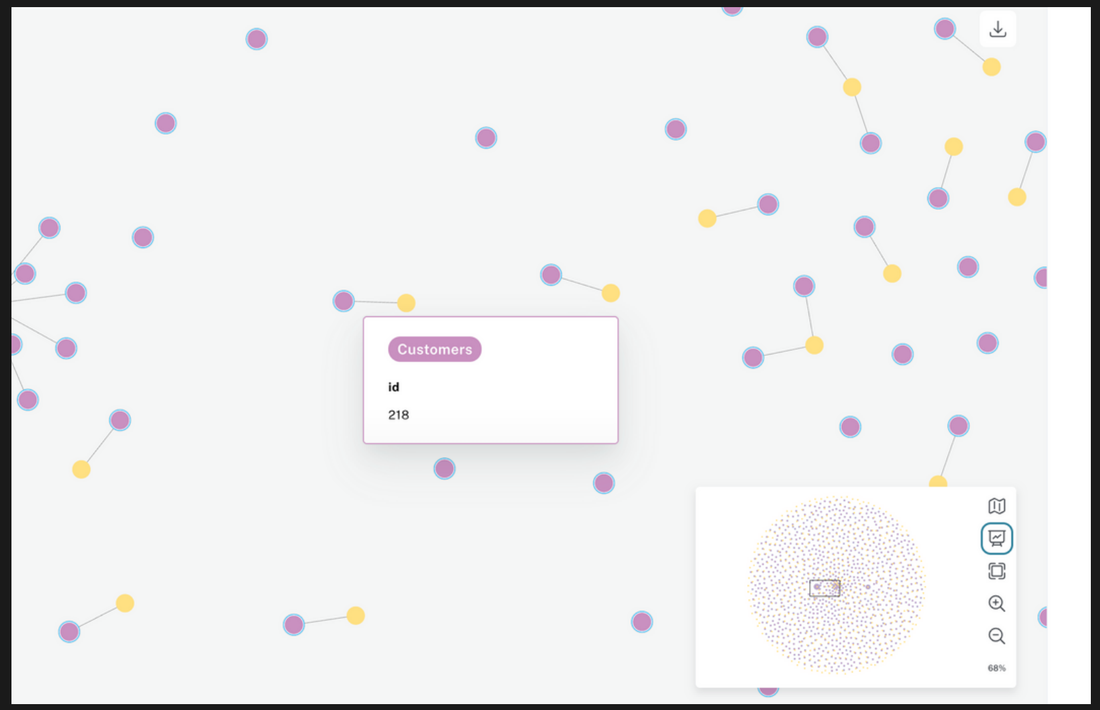


CustID	Cust_Fnar	Cust_Lnan	Email	Birthdate	Phone_No	Stream_Plan_ID
1	Goddard	Glassup	gglassup0@u	5/31/2016	9285113000	2
2	Vito	Newlands	vnewlands1@	5/31/2019	5669242007	3
3	Bealle	Dobbinsor	bdobbinson2	8/13/1953	9021631876	1
4	Pat	Morton	pmorton3@c	9/27/1959	1832061480	3
5	Lucinda	Kenelevsic	lkenelevside4	9/3/2018	9058120542	3

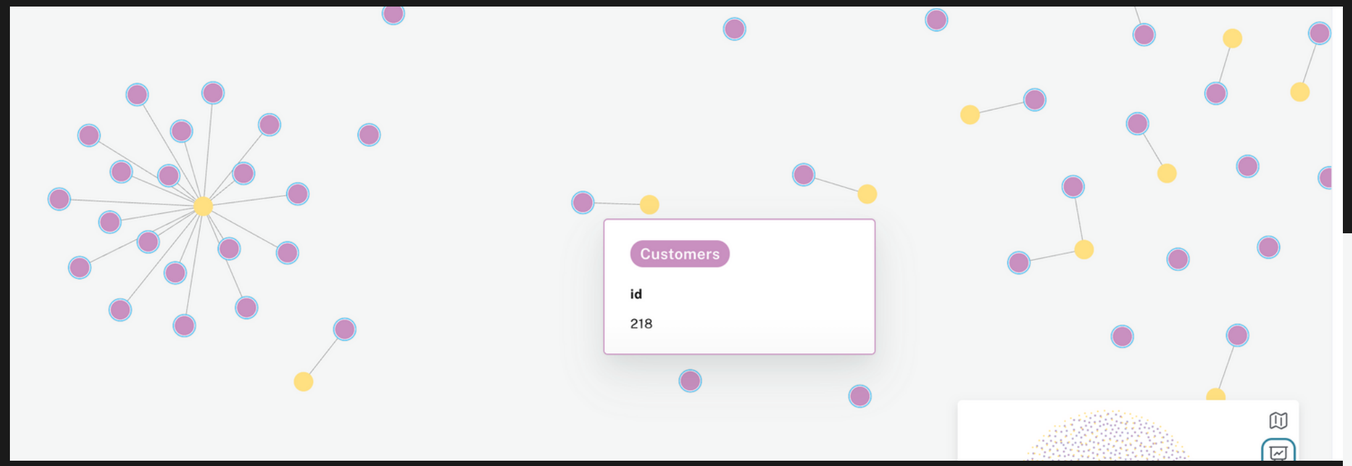
Graph Model Implementation

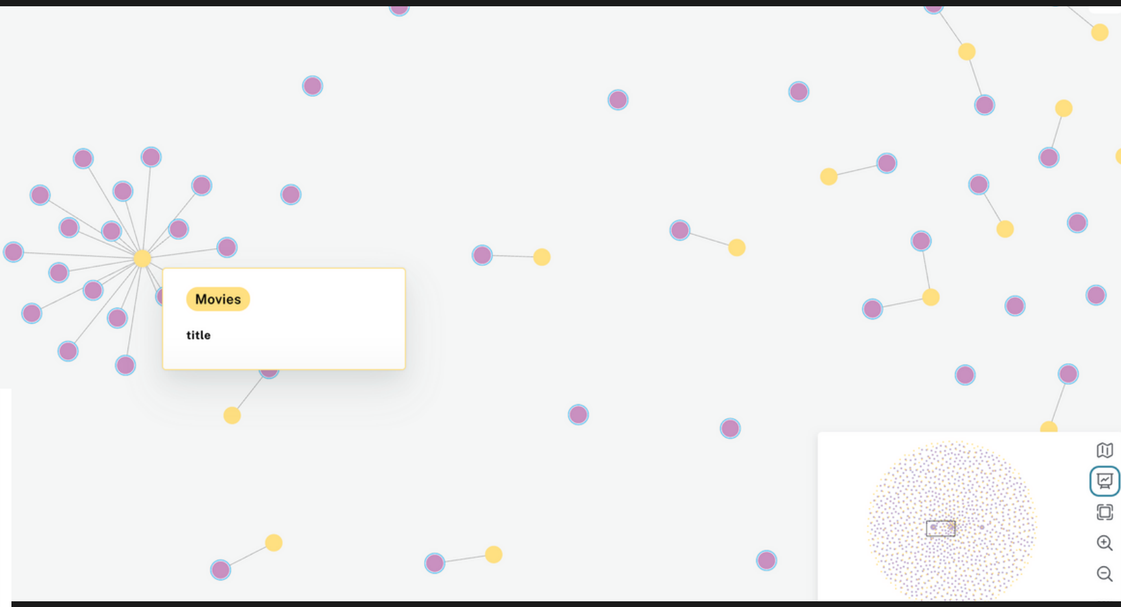
Using Neo4J

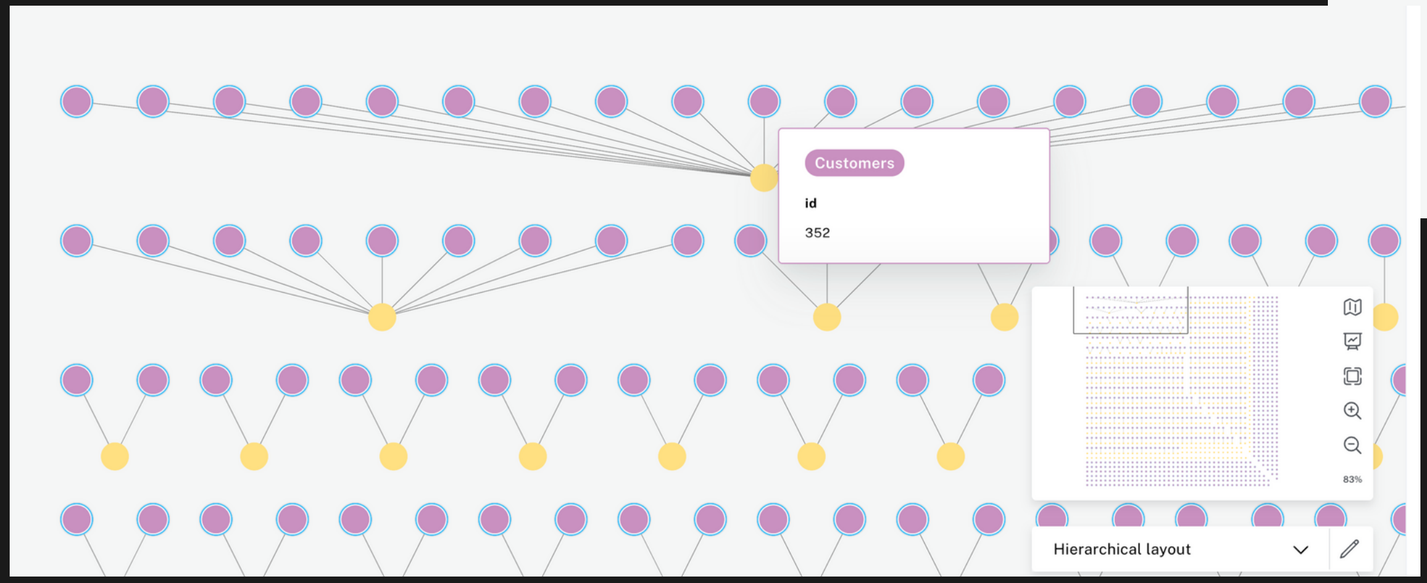




Graph Visualisations: Customers (Rated->) Movies

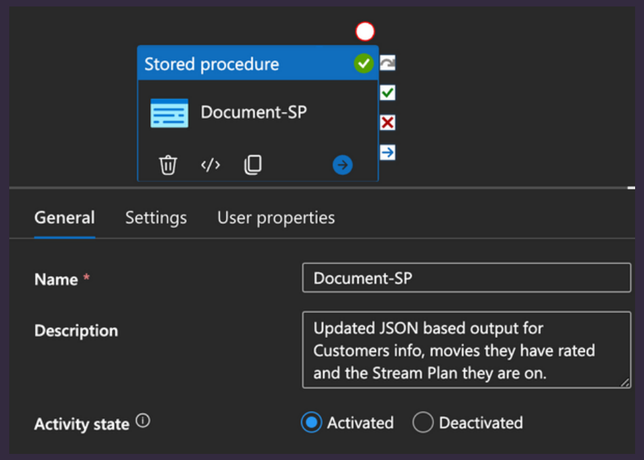






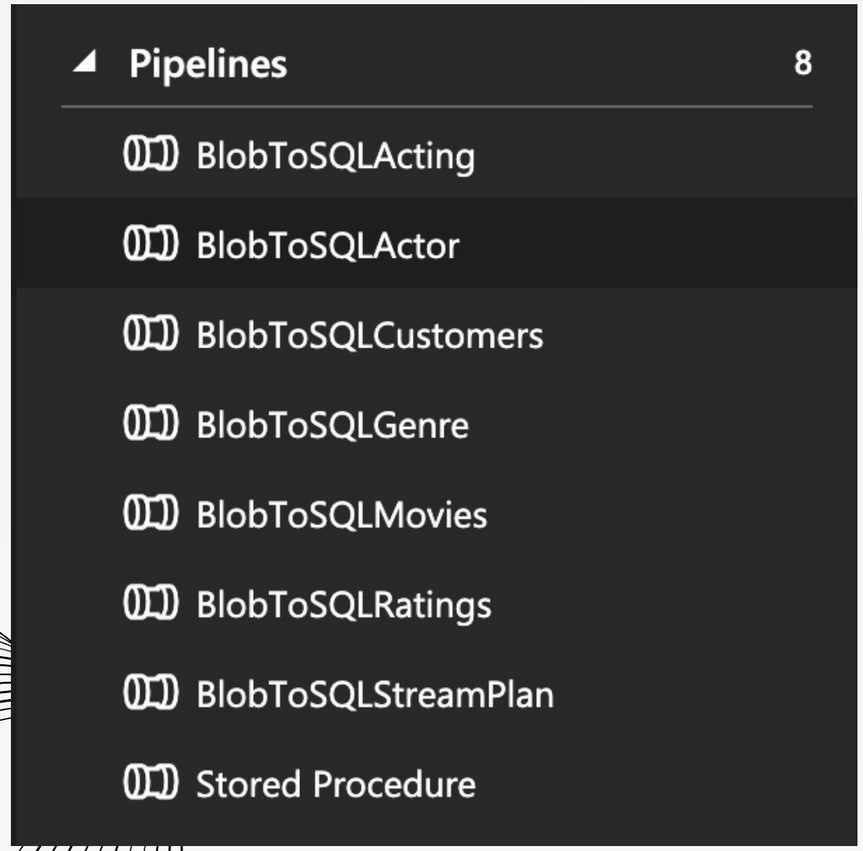
DOCUMENT MODEL

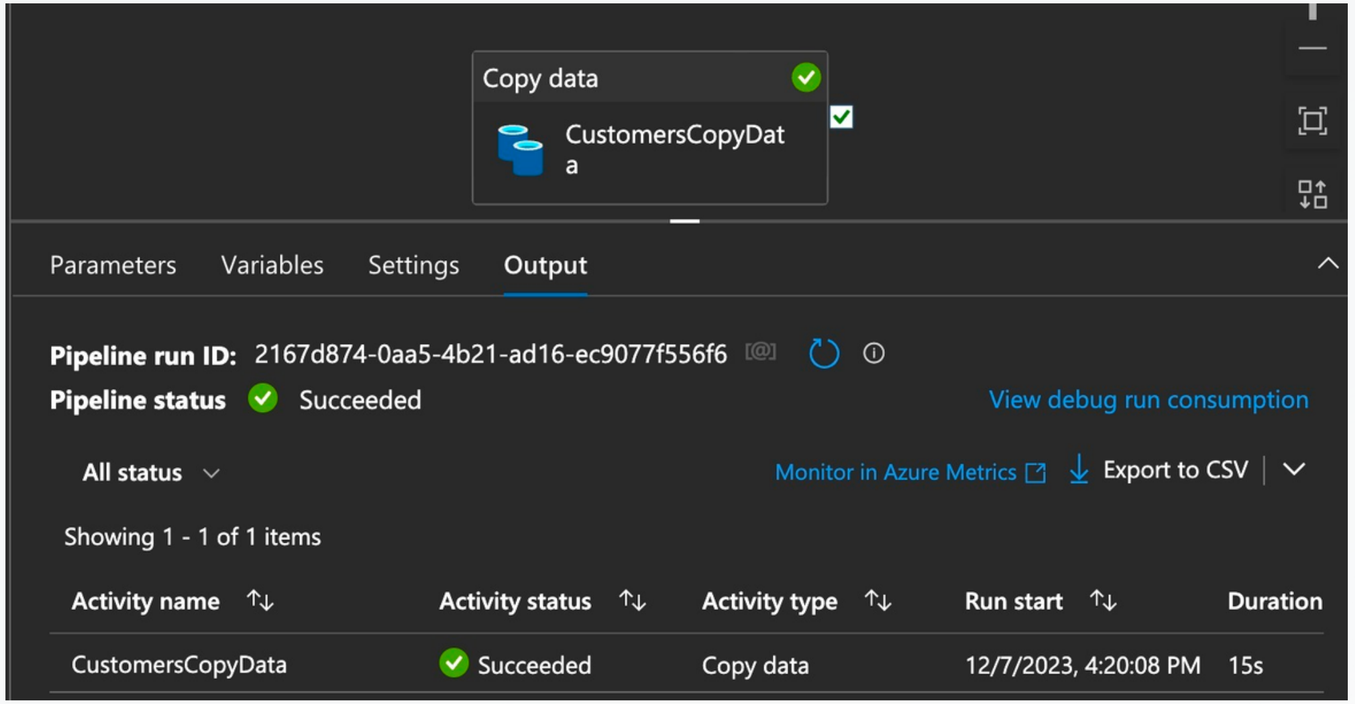
```sql
CREATE PROCEDURE GetCustomerDetails
AS
BEGIN
    SELECT
        c.CustID as id,
        c.Cust_FName as firstName,
        c.Cust_LName as lastName,
        c.Email as email,
        sp.Description as streamPlanDescription,
        sp.Monthly_Cost as streamPlanMonthlyCost,
        (SELECT
             r.MovieID as movieID,
             r.IMDB_Rating as rating
         FROM Ratings r
         WHERE r.CustID = c.CustID
         FOR JSON PATH) as ratedMovies
    FROM Customers c
    INNER JOIN Stream_Plan sp ON sp.Stream_Plan_ID = c.Stream_Plan_ID
    FOR JSON PATH;
END;
```



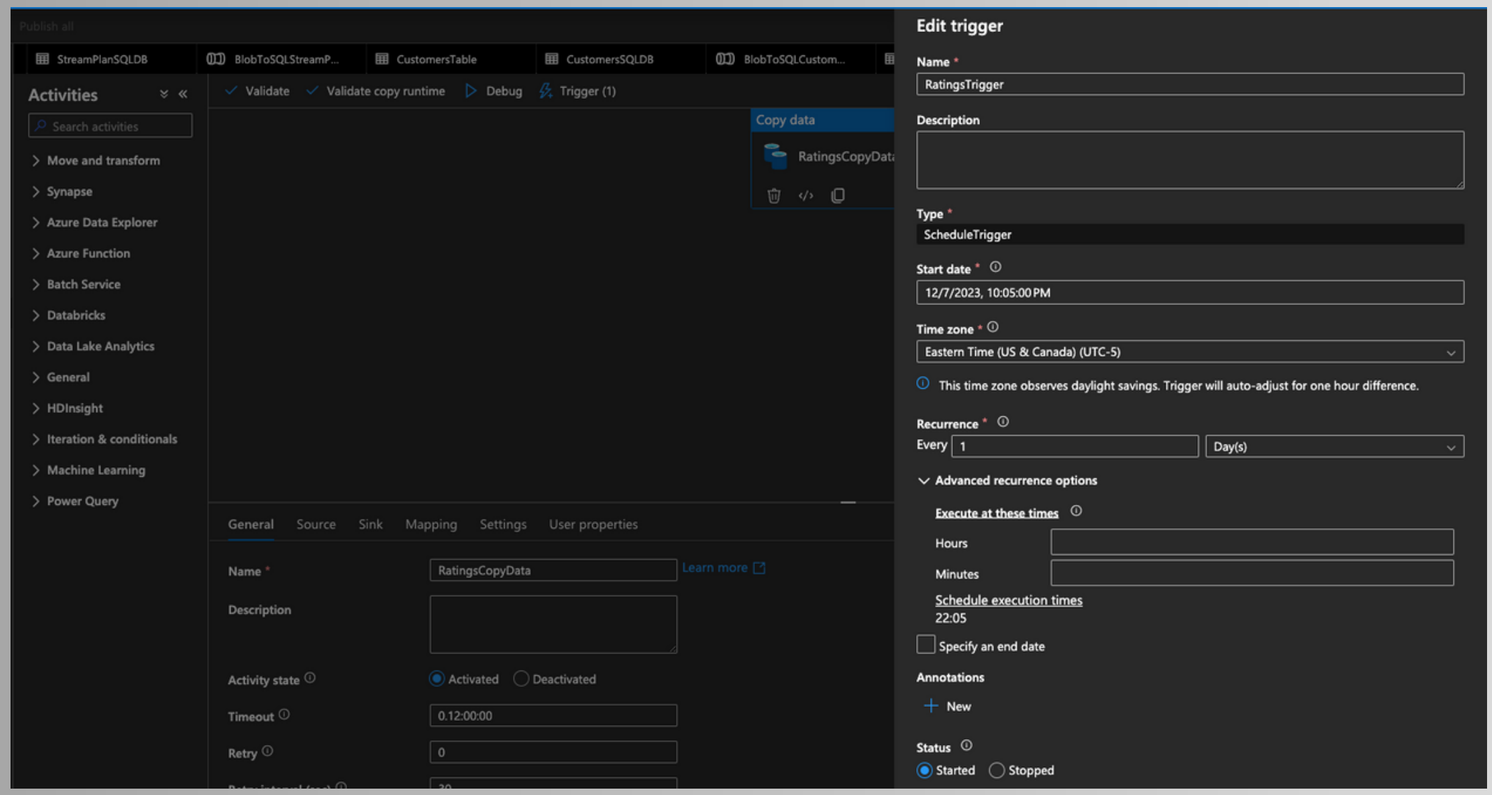

```
    "id": 1,
    "firstName": "Goddard",
    "lastName": "Glassup",
    "email": "gglassup0@un.org",
    "streamPlanDescription": "Standard: Unlimited ad-free movies and TV shows, watch on 2 devices at a time, Full HD, download
    "streamPlanMonthlyCost": 15.49,
    "ratedMovies": [
        {
            "movieID": "ts20358",
            "rating": 6.2
        },
        {
            "movieID": "ts273318",
            "rating": 5.9
        }
    ]
},
{
    "id": 2,
    "firstName": "Vito",
    "lastName": "Newlands",
    "email": "vnewlands1@e-recht24.de",
    "streamPlanDescription": "Premium: Unlimited ad-free movies and TV shows, watch on 4 devices at a time, Ultra HD, download
    "streamPlanMonthlyCost": 22.99
},
{
    "id": 3,
    "firstName": "Bealle",
    "lastName": "Dobbinson",
    "email": "bdobbinson2@slashdot.org",
    "streamPlanDescription": "Standard with ads: Ad-supported, watch on 2 devices at a time, Full HD, download on 2 devices",
    "streamPlanMonthlyCost": 6.99
},
```

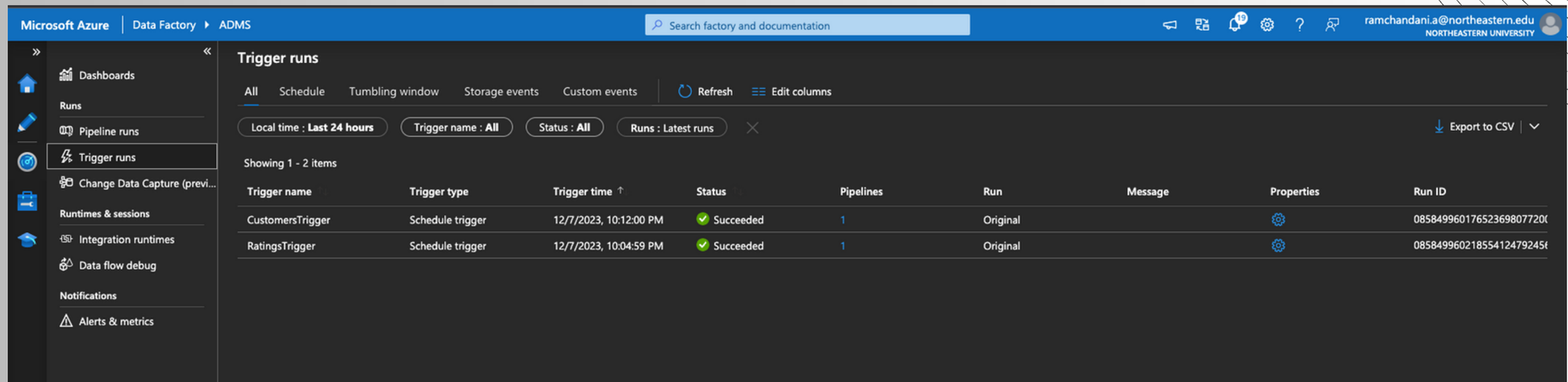
DATA PIPELINING



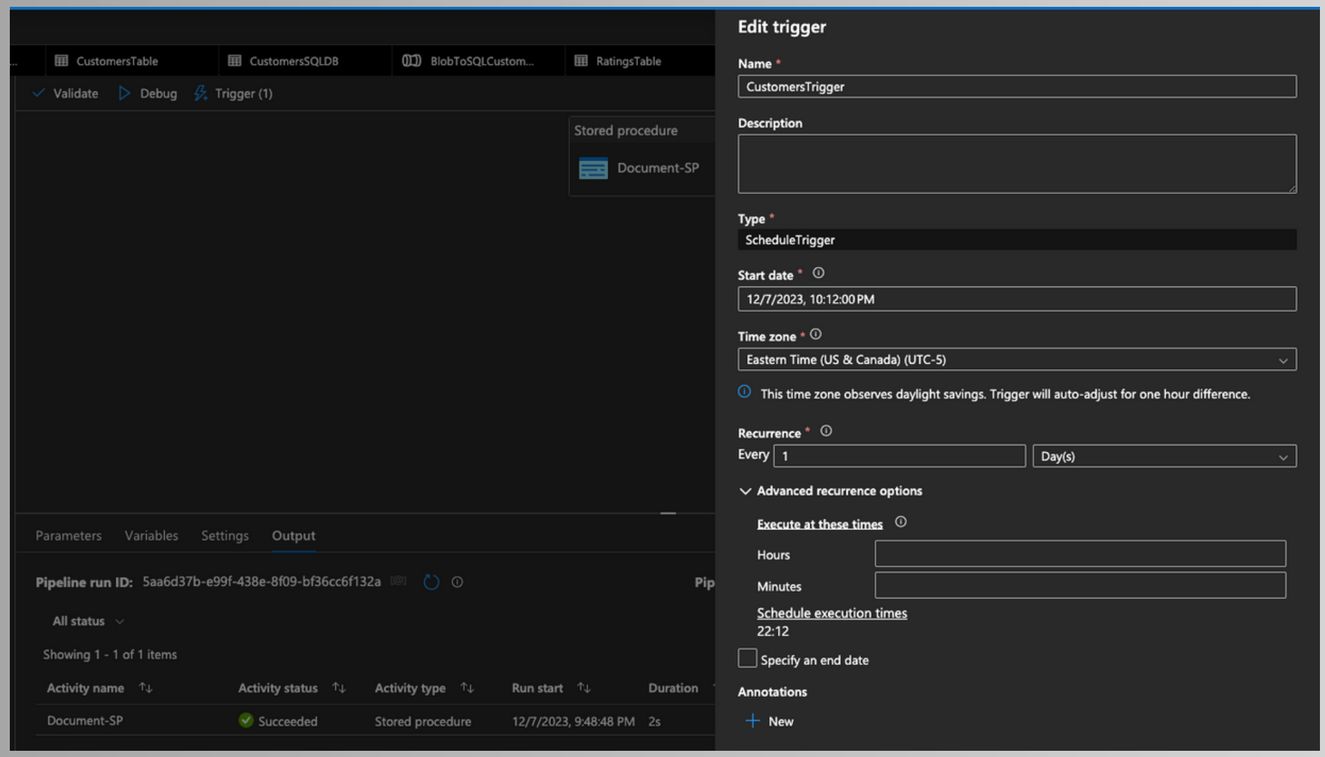


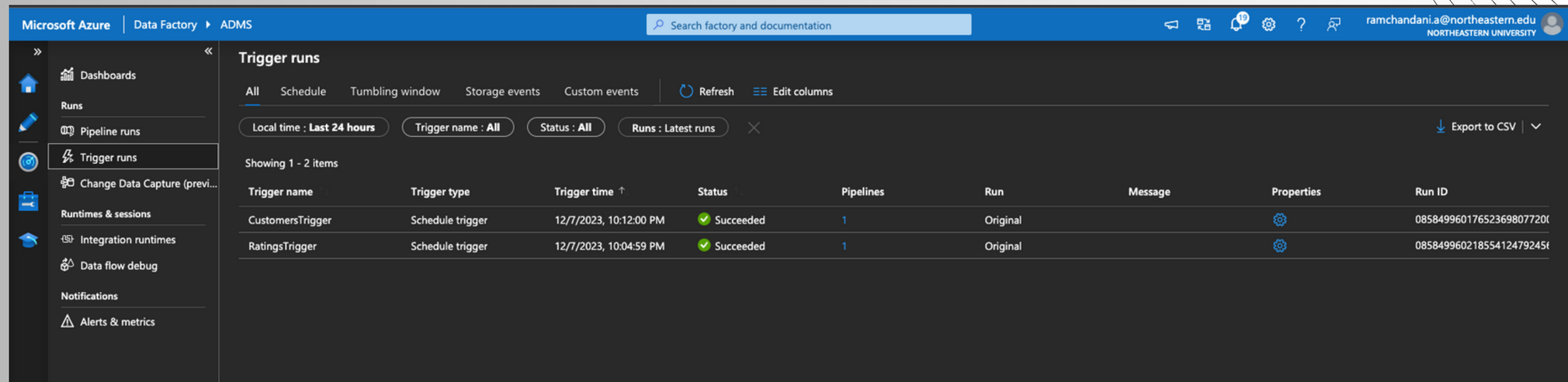
DATA REFRESH-RATINGS



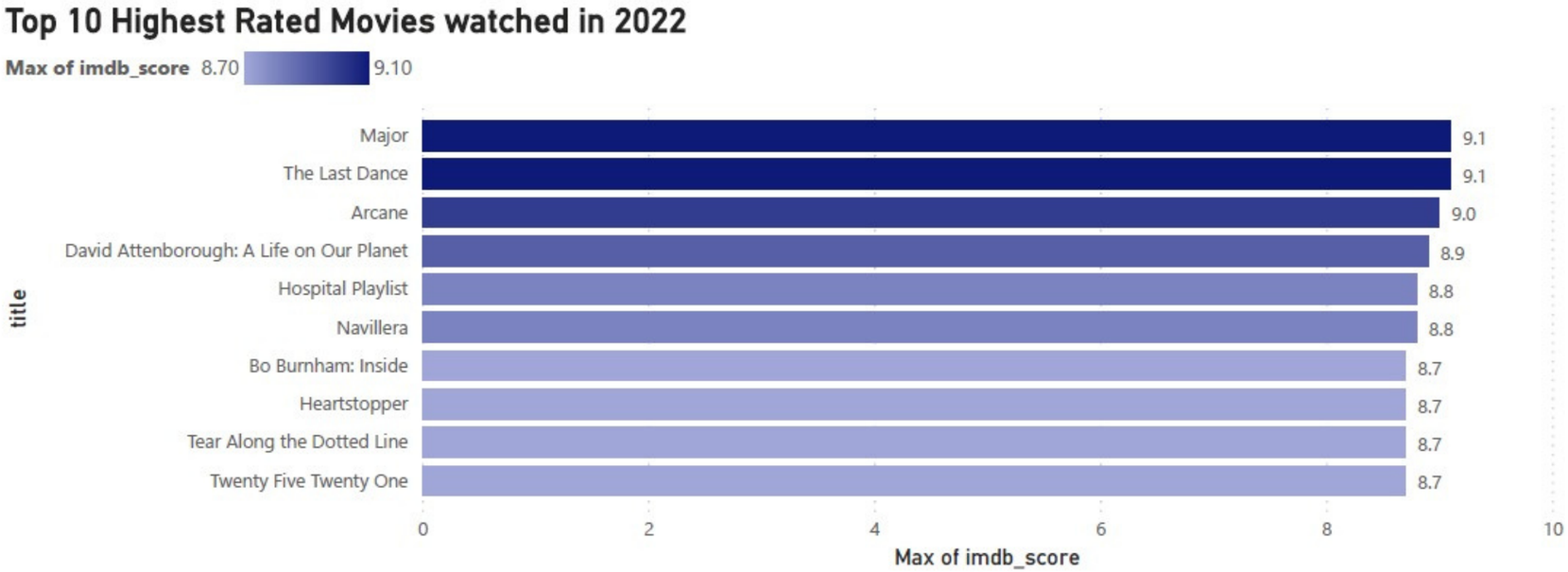


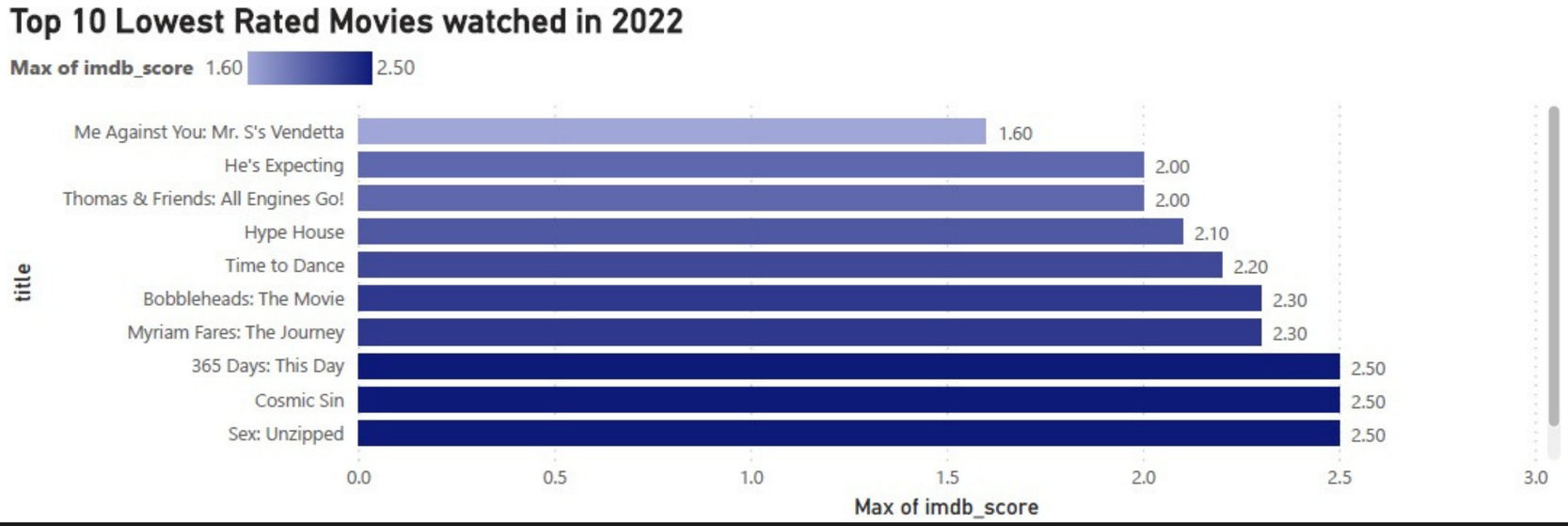
DATA REFRESH-CUSTOMERS



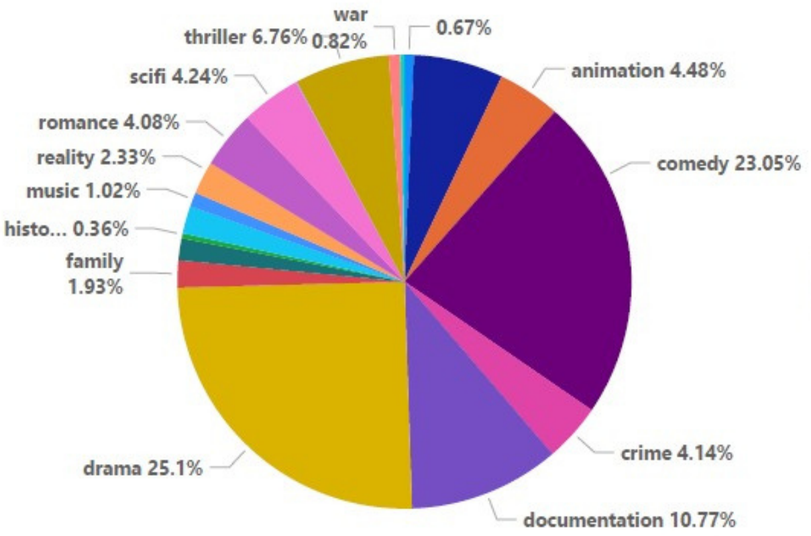


VISUALIZATIONS





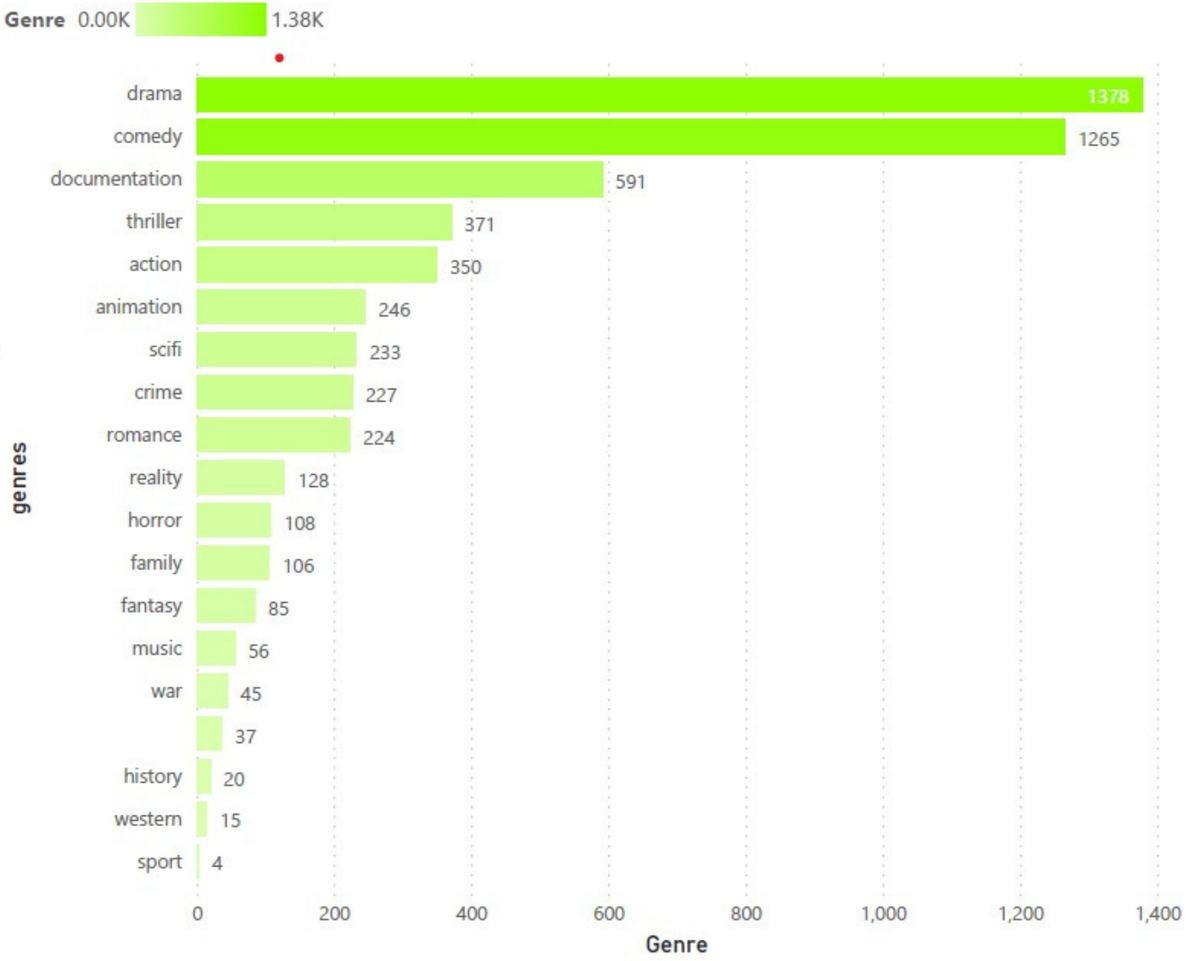
Top Rated Genre's



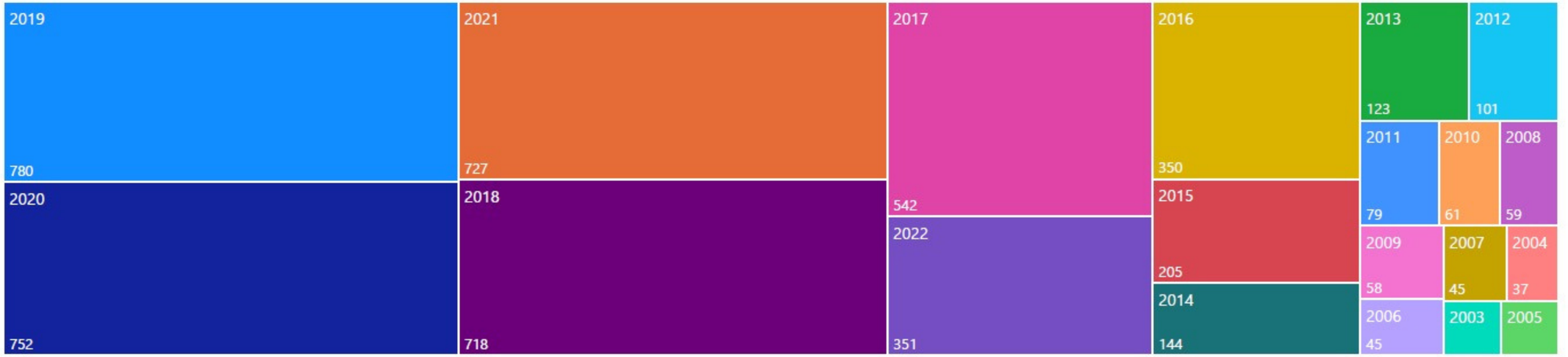
19

Count of genres

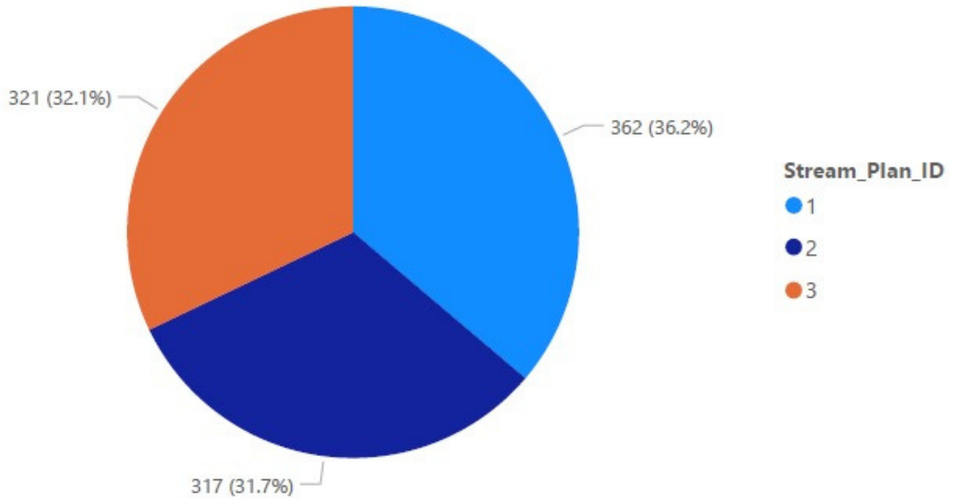
Number of Movies by Genre



Number of Movies Released Every Year



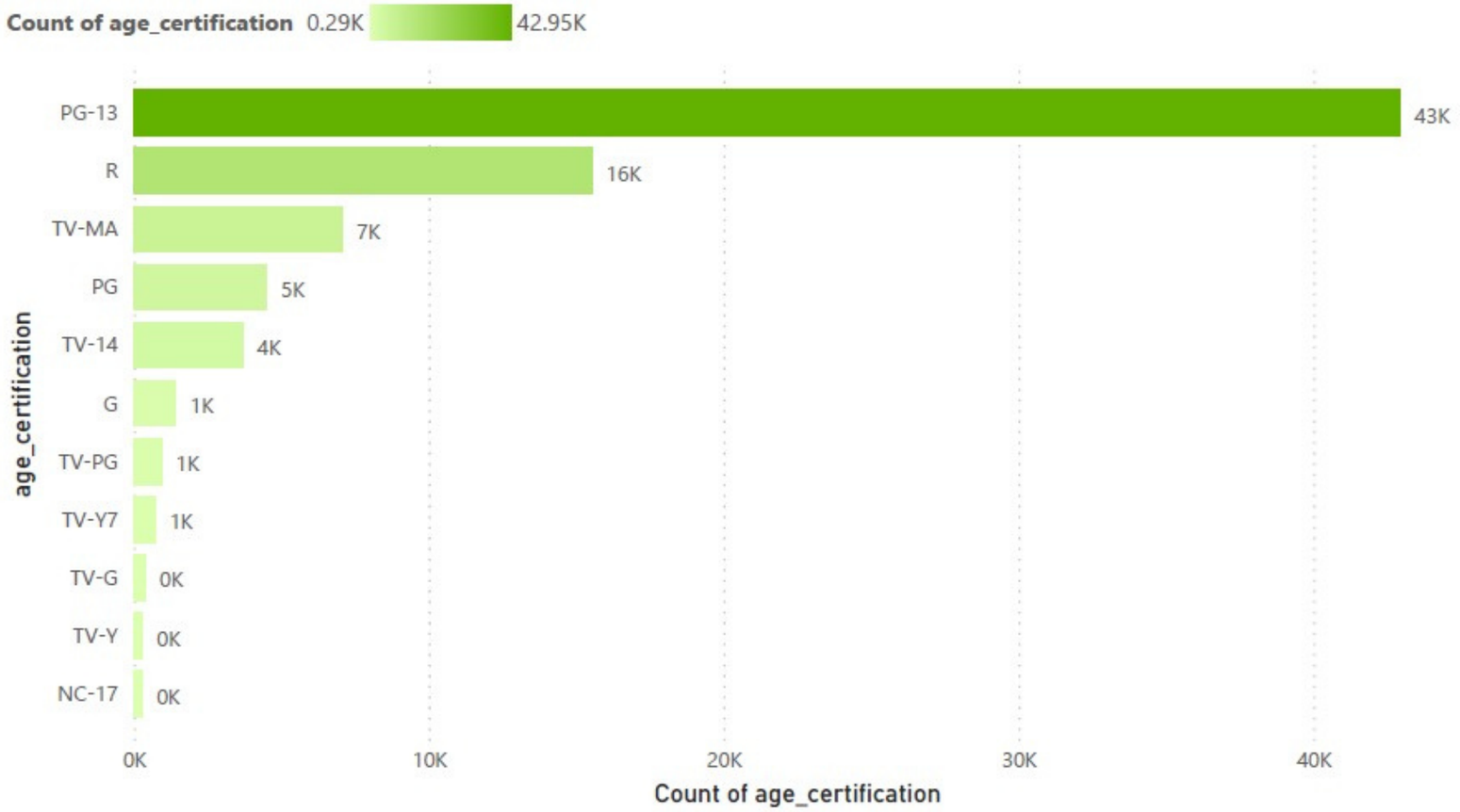
Total Percentage of Customers with Each Stream Plan



1000

Total Number of Customers

Number of Movies with Age Certification



Top 3 Actors with Highest Movie Count

Count of id	first_name	last_name
25	Boman	Irani
25	Kareena	Kapoor Khan
23	Shah	Rukh Khan

77.80K

Total Number of Actors

5489

Total Number of Movies

FUTURE PLANS

Implement more sophisticated data analytics methods to gain deeper insights from the streaming service data. This could involve using machine learning models to predict user behavior or preferences.

ADVANCED DATA ANALYTICS INTEGRATION

Develop more complex and interactive data visualizations to better understand user trends and patterns. This can help in making data-driven decisions more effectively..

ENHANCED DATA VISUALIZATION

Focus on scaling the database and data processing pipelines to handle an increased volume of data efficiently. This includes optimizing existing processes for better performance.

SCALABILITY AND PERFORMANCE OPTIMIZATION

THANK YOU!

QUESTIONS?