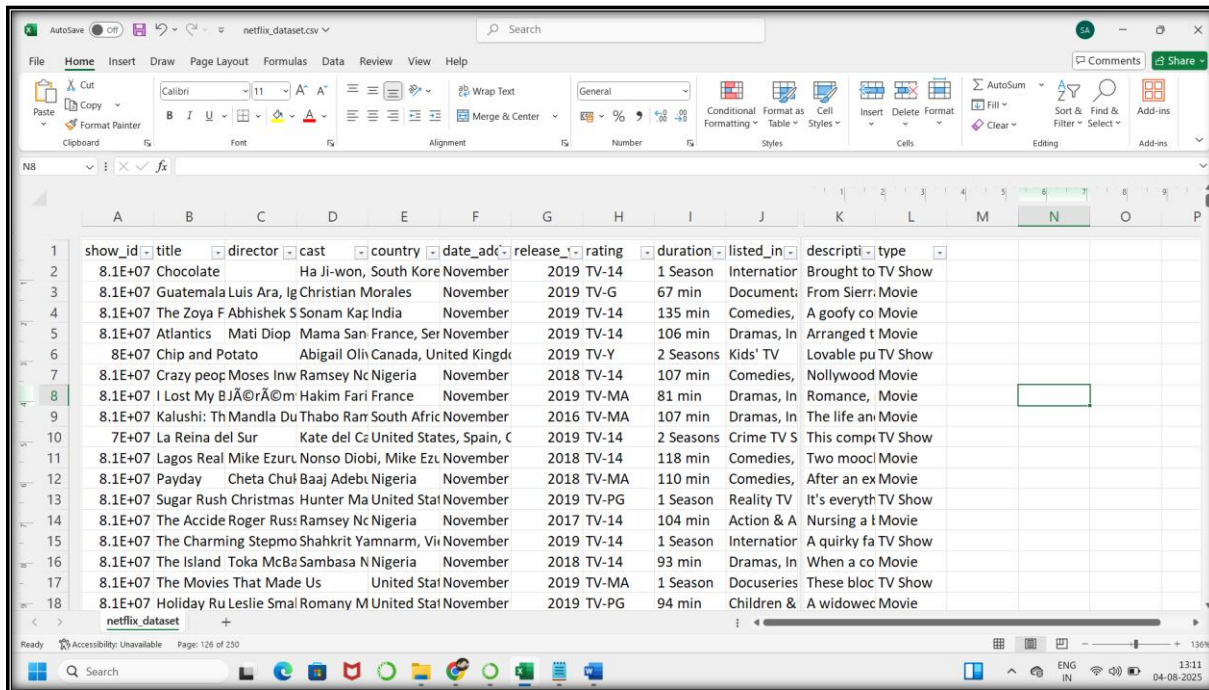


Netflix Dataset Cleaning – Excel

This document outlines the steps followed to clean the Netflix Movies and TV Shows dataset using Microsoft Excel. It includes method descriptions and placeholders for relevant screenshots.

Step 1: Load the Dataset

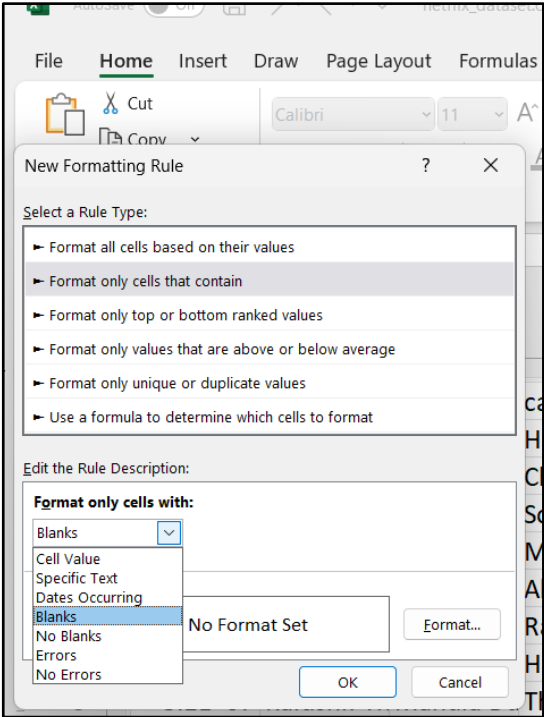
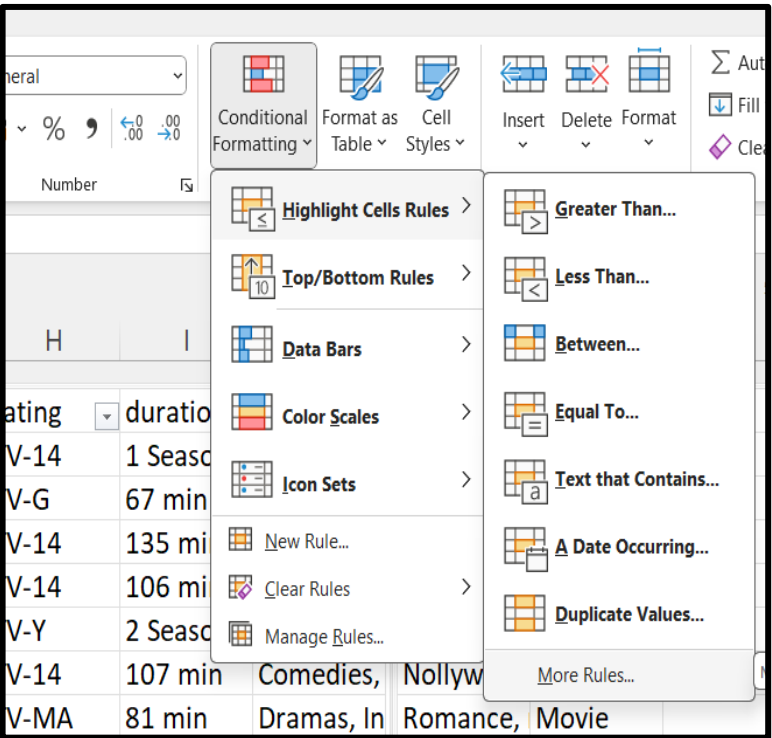
Opened the raw Netflix dataset in Microsoft Excel.

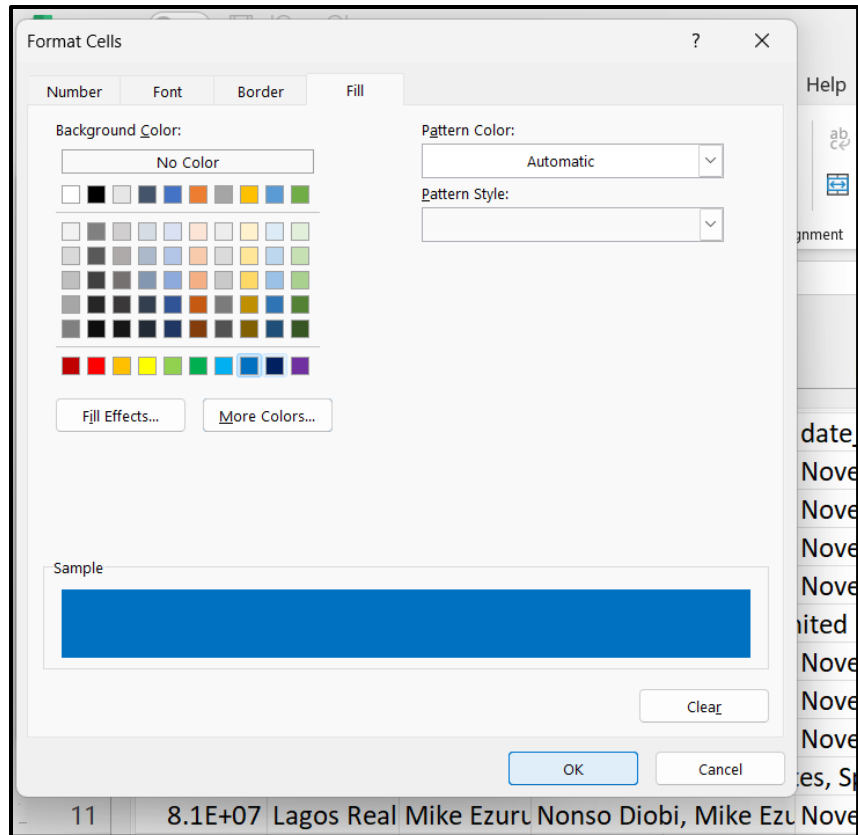


show_id	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description	type
8.1E+07	Chocolate	Ha Ji-won,	South Korea	November	2019	TV-14	1 Season	International	Brought to TV Show		
8.1E+07	Guatemala	Luis Ara, Ig	Christian Morales	November	2019	TV-G	67 min	Document	From Sierr	Movie	
8.1E+07	The Zoya F	Abhishek S	Sonam Kap	India	November	2019	TV-14	135 min	Comedies,	A goofy co	Movie
8.1E+07	Atlantics	Mati Diop	Mama San	France, Ser	November	2019	TV-14	106 min	Dramas, In	Arranged t	Movie
8E+07	Chip and Potato	Abigail Oli	Canada, United	Kingdo	2019	TV-Y	2 Seasons	Kids' TV	Lovable pu	TV Show	
8.1E+07	Crazy peop	Moses Inw	Ramsey Nc	Nigeria	November	2018	TV-14	107 min	Comedies,	Nollywood	Movie
8.1E+07	I Lost My B	JÃ©rÃ©m	Hakim Fari	France	November	2019	TV-MA	81 min	Dramas, In	Romance,	Movie
8.1E+07	Kalushi: Th	Mandla Du	Thabo Ran	South Afric	November	2016	TV-MA	107 min	Dramas, In	The life an	Movie
7E+07	La Reina del Sur	Kate del C	United States, Spain, C	2019	TV-14	2 Seasons	Crime TV S	This compi	TV Show		
8.1E+07	Lagos Real	Mike Ezur	Nonso Diobi, Mike Ez	November	2018	TV-14	118 min	Comedies,	Two moocl	Movie	
8.1E+07	Payday	Cheta Chu	Baaj Adeb	Nigeria	November	2018	TV-MA	110 min	Comedies,	After an ex	Movie
8.1E+07	Sugar Rush	Christmas	Hunter Ma	United Stat	November	2019	TV-PG	1 Season	Reality TV	It's everyth	TV Show
8.1E+07	The Accide	Roger Russ	Ramsey Nc	Nigeria	November	2017	TV-14	104 min	Action & A	Nursing a l	Movie
8.1E+07	The Charming	Stepmo	Shahkrit Yamnarm, Vi	November	2019	TV-14	1 Season	Internation	A quirky fa	TV Show	
8.1E+07	The Island	Toka McBa	Sambasa N	Nigeria	November	2018	TV-14	93 min	Dramas, In	When a co	Movie
8.1E+07	The Movies That	Made Us	United Stat	November	2019	TV-MA	1 Season	Docuseries	These bloc	TV Show	
8.1E+07	Holiday Ru	Leslie Smal	Romany M	United Stat	November	2019	TV-PG	94 min	Children &	A widowec	Movie

Step 2: Identify and Highlight Missing Values

Used Conditional Formatting to highlight blank cells in columns like 'director', 'cast', and 'country'.

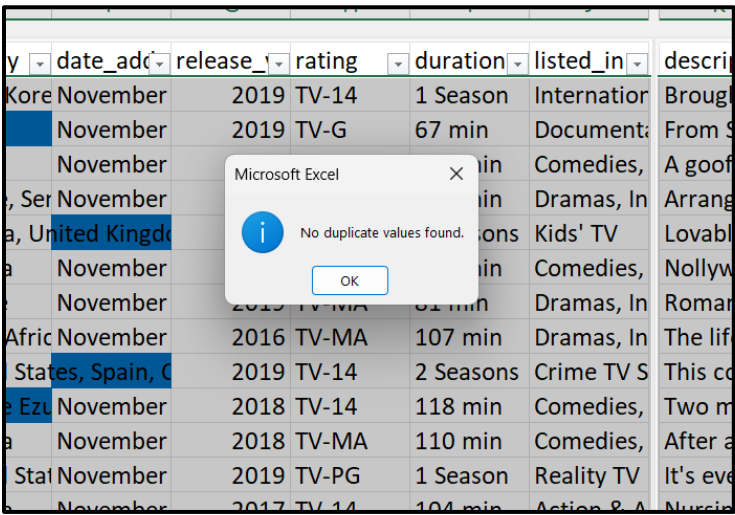
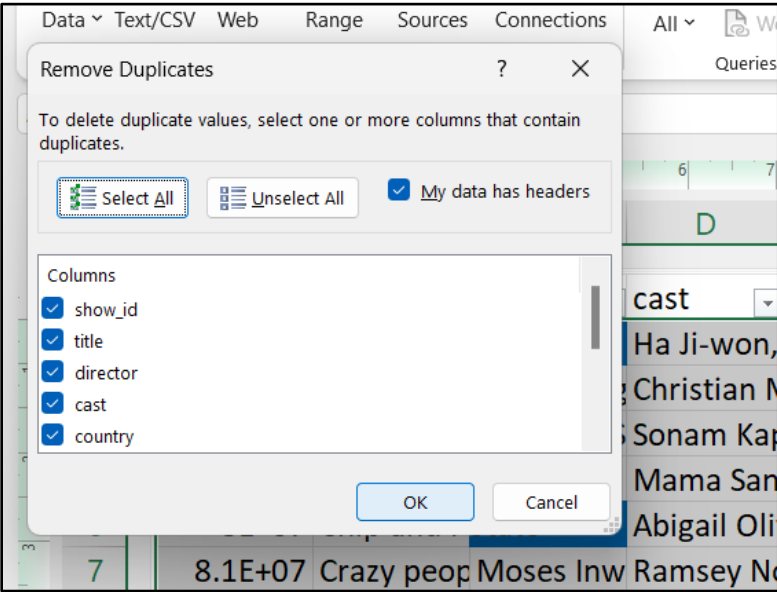
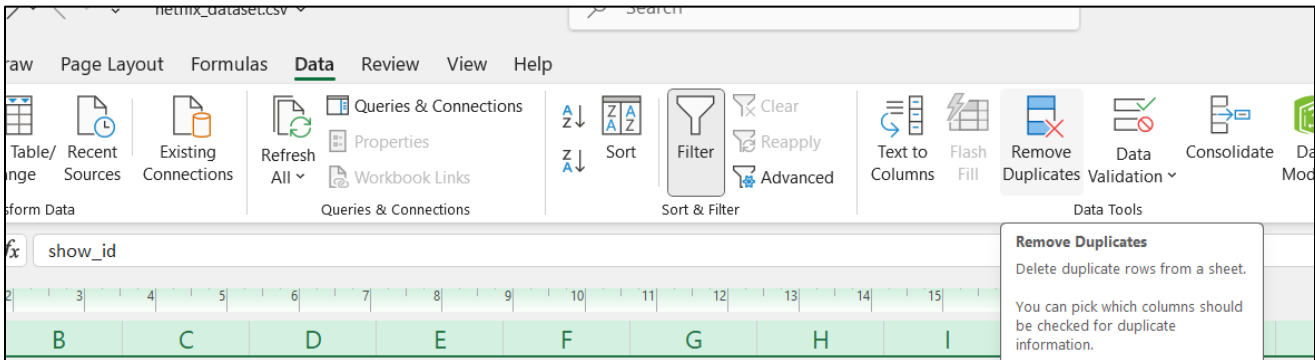




	A	B	C	D	E	F	G
1	show_id	title	director	cast	country	date_added	release_year
2	8.1E+07	Chocolate		Ha Ji-won,	South Korea	November	2019
3	8.1E+07	Guatemala	Luis Ara, Ig	Christian Morales		November	2019
4	8.1E+07	The Zoya F	Abhishek S	Sonam Kap	India	November	2019
5	8.1E+07	Atlantics	Mati Diop	Mama San	France, Ser	November	2019
6	8E+07	Chip and Potato		Abigail Oli	Canada, United Kingdo		2019
7	8.1E+07	Crazy peop	Moses Inw	Ramsey Nc	Nigeria	November	2018
8	8.1E+07	I Lost My B	JÃ©rÃ©m	Hakim Fari	France	November	2019
9	8.1E+07	Kalushi: Th	Mandla Du	Thabo Rarr	South Afric	November	2016
10	7E+07	La Reina del Sur		Kate del Ca	United States, Spain, C		2019
11	8.1E+07	Lagos Real	Mike Ezuru	Nonso Diobi,	Mike Ezu	November	2018
12	8.1E+07	Payday	Cheta Chuk	Baaj Adeb	Nigeria	November	2018
13	8.1E+07	Sugar Rush	Christmas	Hunter Ma	United Stat	November	2019
14	8.1E+07	The Accide	Roger Russ	Ramsey Nc	Nigeria	November	2017
15	8.1E+07	The Charming	Stepmo	Shahkrit Yam	narm, Vi	November	2019
16	8.1E+07	The Island	Toka McBa	Sambasa N	Nigeria	November	2018
17	8.1E+07	The Movies	That Made Us		United Stat	November	2019
18	8.1E+07	Holiday Ru	Leslie Smal	Romany M	United Stat	November	2019

Step 3: Remove Duplicates

Selected all columns and used Data → Remove Duplicates



Step 4: Standardize Text Fields

INSERT NEW COLUMN NEAR THE SELECTED COLUMN

ENTER THE FORMULA TO TRIM AND TO LOWERCASE THE SELECTED COLUMN

ONCE THE FORMULAS IS ENTERED IT REMOVE THE SPACES AND CONVERT IT INTO LOWERCASE

C	D	E	F	G	H	I	J
Director	cast		country	date_added	release_year	rating	duration
	Ha Ji-won, Yoon Kye-sang, Jang Seung-jo, Kang Bu-ja, Lee Jae-ryong, Min Jin-woong, Kim Won-hae, Yoo Teo		South Korea	November 2019	TV-14		1 Season
Ara, Ig	Christian Morales			November 2019	TV-G		67 min
	Sonam Kapoor, Dulquer Salmaan, Sanjay Kapoor, Sikander Kher, Angad Bedi, Koel Purie, Pooja Bhamrah, Manu						
shek S	Rishi Chadha		India	November 2019	TV-14		135 min
	Ibrahima Traore, Nicole Sougou, Amina Kane, Mariama Gassama, Coumba Dieng, Ibrahima Mbaye, Diankou						
Diop	Sembene		France, Senegal	November 2019	TV-14		106 min
	Abigail Oliver, Andrea Libman, Briana Buckmaster, Brian Dobson, Chance Hurstfield, Dominic Good, Emma Jayne Maas, Evan Byarushengo, Scotia Anderson, Alessandro Juliani		Canada, United Kingdom		2019	TV-Y	2 Seasons
	Ramsey Nouah, Chigul, Sola Sobowale, Ireti Doyle, Ben Tolutou, Francis Onwochei, Ememobong Nkana, Emem						
es Inwi	Inwang, Patrick Onyeke		Nigeria	November 2018	TV-14		107 min
dataset							

Director	cast	
	Ha Ji-won, Yoon Kye-sang, Jang Seung-jo, Kang Bu-ja, Lee Jae-ryong, Min Jin-woong, Kim Won-hae, Yoo Teo	=IF(D2="", "", LOWER(TRIM(D2)))
Ara, Ig	Christian Morales	LOWER(text)

ng, Jang		
ee Jae-		
Kim Won-	ha ji-won, yoon kye-sang, jang seung-jo, ka	
	christian morales	
edi, Koel		
Manu	sonam kapoor, dulquer salmaan, sanjay ka	
mbow,		
e Sougou,		
Gassama,		
a Mbaye,	mama sane, amadou mbow, ibrahima traor	
ibman,		
an		
eld.		

Step 5: Rename Column Header

CREATE A NEW ROW ABOVE THE HEADER IN THE A1 CELL USER THE FORMULA =UPPER(A2) TO CONVERT THE HEADER INTO UPPERCASE .

	A	B	C	D	E
1	SHOW_ID	TITLE	DIRECTOR CAST		
2	show_id	title	director	cast	
			Ha Ji-won, Yoon Kye-sang, Jang		

Step 9: Final Cleaned Dataset Overview

Reviewed the cleaned dataset for consistency and saved as 'Netflix_Cleaned.xlsx'.

	A	B	C	D	E	F	G	H	I	J	K	L
1	SHOW_ID	TITLE	DIRECTOR	CAST		COUNTRY	DATE_ADD	RELEASE_Y	RATING	DURATION	LISTED_IN	DESCRIPTION
2	show_id	title	director	cast		country	date_add	release_yr	rating	duration	listed_in	description ty
				Ha Ji-won, Yoon Kye-sang, Jang Seung-jo, Kang Bu-ja, Lee Jae-ryong, Min Jin-woong, Kim Won-hae, Yoo Teo	ha ji-won, yoon kye-sang, jang seung-jo,	South Korea	November	2019	TV-14	1 Season	International	Brought to TV
3	8.1E+07	Chocolate										
4	8.1E+07	Guatemala: Luis Ara, I	Christian Morales		christian morales		November	2019	TV-G	67 min	Document	From Sierra M
				Sonam Kapoor, Dulquer Salmaan, Sanjay Kapoor, Sikander Kher, Angad Bedi, Koel Purie, Pooja Bhamrah, Manu								
5	8.1E+07	The Zoya	Abhishek	Rishi Chadha	sonam kapoor, dulquer salmaan, sanjay	India	November	2019	TV-14	135 min	Comedies, A	goofy cc M
				Ibrahima Traore, Nicole Sougou, Amina Kane, Mariama Gassama, Coumba Dieng, Ibrahima Traore, Dieng								
