



Netflix Dataset – Exploratory Data Analysis (Week 3)

Data Science Internship – FutureXcel

Student Name: Sumbul Abdul.Wahid

DATASET OVERVIEW

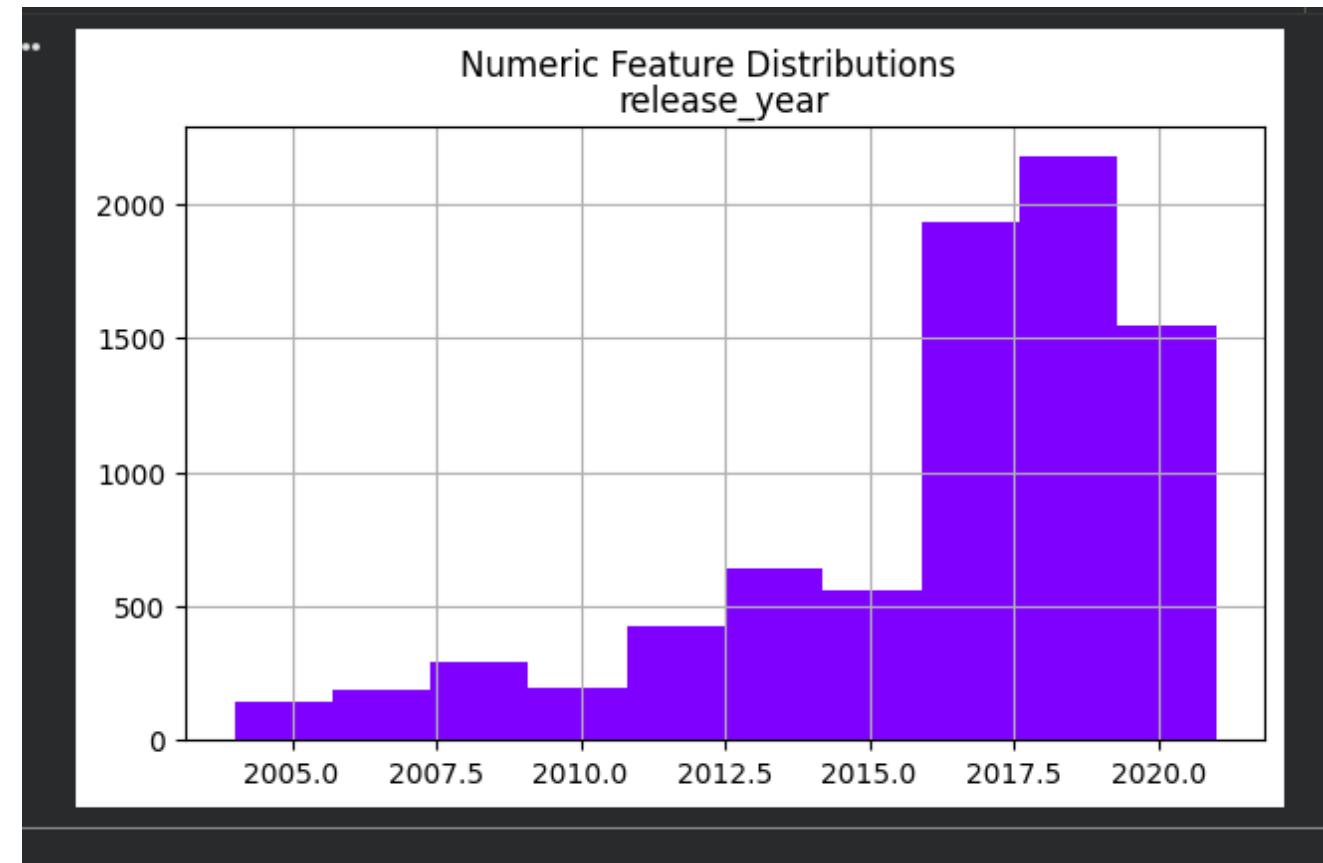
- Netflix movies and TV shows dataset
- Cleaned dataset from Week 2
- Contains content type, release year, rating, country, and duration

```
*** <class 'pandas.core.frame.DataFrame'>
RangeIndex: 8088 entries, 0 to 8087
Data columns (total 12 columns):
 #   Column      Non-Null Count Dtype  
--- 
 0   show_id     8088 non-null   object  
 1   type        8088 non-null   object  
 2   title       8088 non-null   object  
 3   director    8088 non-null   object  
 4   cast         8088 non-null   object  
 5   country     8088 non-null   object  
 6   date_added  8088 non-null   object  
 7   release_year 8088 non-null   float64 
 8   rating      8088 non-null   object  
 9   duration    8088 non-null   object  
 10  listed_in   8088 non-null   object  
 11  description 8088 non-null   object  
dtypes: float64(1), object(11)
memory usage: 758.4+ KB
```

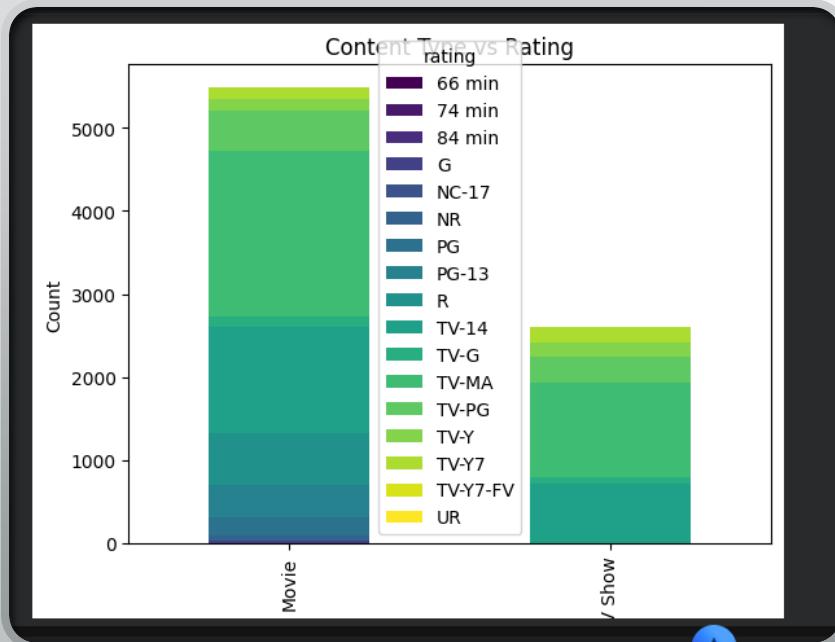
	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	Not Available	United States	2021-09-25	2020.0	PG-13	90 min	Documentaries	As her father nears the end of his life, film...
1	s2	TV Show	Blood & Water	Unknown	Ama Qamata, Khosi Ngema, Gall Mbokane, Thabani...	South Africa	2021-09-24	2021.0	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotos, Samuel Joey, Nabi...	Unknown	2021-09-24	2021.0	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...
3	s4	TV Show	Jailbirds	Unknown	Not Available	Unknown	2021-09-24	2021.0	TV-MA	1 Season	Documentaries, Reality TV	Feuds, flirtations and toilet talk go down amo...
4	s5	TV Show	New Orleans	Unknown	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	2021-09-24	2021.0	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train i...

UNIVARIATE ANALYSIS

- Analyzed individual variables
- Content type and ratings distribution
- Numeric feature distributions explored



BIVARIATE ANALYSIS



- Relationship between content type and rating
- Used tables and visual comparison

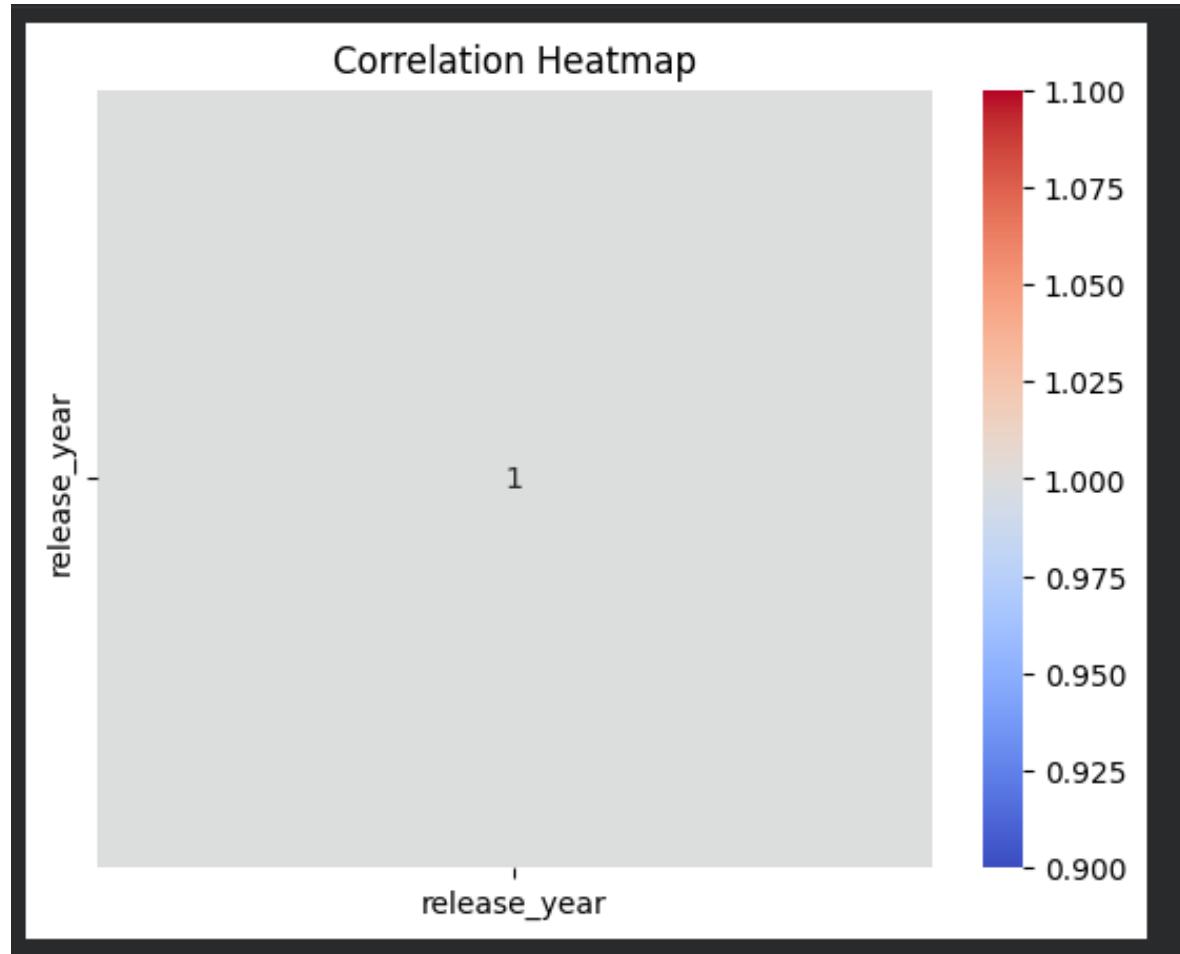
Trivariate Analysis

- Trivariate analysis was performed to understand the combined effect of multiple variables.
- The relationship between content type, country, and rating was analyzed.
- This analysis provides deeper insights compared to univariate and bivariate analysis.

type	country	rating	66 min	74 min	84 min	G	NC-17	NR	PG	PG-13	R	TV-14	TV-G	TV-MA	TV-PG	TV-Y	TV-Y7	TV-Y7-FV	UR
Movie	, France, Algeria	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	
	Argentina	0	0	0	0	0	1	0	0	1	0	5	2	24	2	0	0	0	
	Argentina, Brazil, France, Poland, Germany, Denmark	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	
	Argentina, Chile	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	
	Argentina, Chile, Peru	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
	
TV Show	United States, Sweden	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	
	United States, United Kingdom	0	0	0	0	0	0	0	0	0	0	1	0	4	3	1	0	0	
	United States, United Kingdom, Australia	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	
	Unknown	0	0	0	0	0	0	0	0	0	1	108	11	137	54	33	41	0	
	Uruguay, Germany	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	

CORRELATION

- Correlation analysis was performed on numerical features. Due to the presence of only one numerical variable in the dataset, limited correlation was observed.



KEY INSIGHTS

- Movies make up the majority of Netflix content.
- Content releases increased significantly after 2015.
- Ratings differ between movies and TV shows.
- A limited number of countries contribute most of the content.
- Numerical feature correlation is limited.

CONCLUSION

- This exploratory data analysis helped understand Netflix content trends and relationships between different variables. The insights gained from this analysis can support further data-driven decisions and advanced modeling.