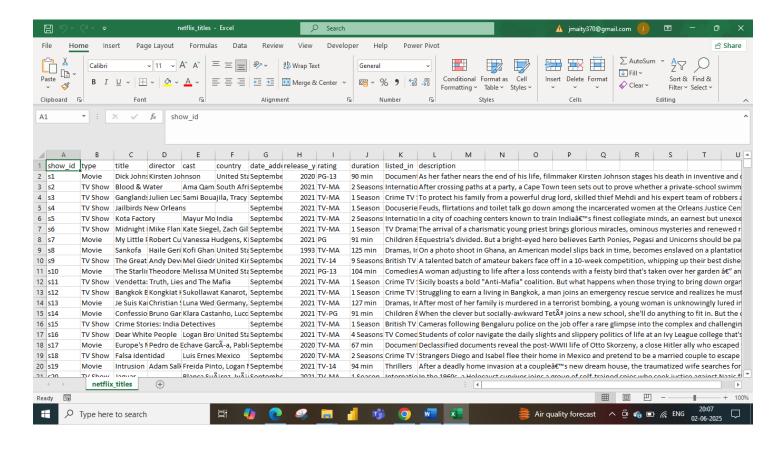
DATA ANALYST INTERNSHIP

NETFLIX MOVIES AND TV SHOWS

Jit Maity | jmaity370@gmail.com

Raw Data:



Variables:

- 1. show_id: Identify the id name of particular movie/tv show
- 2. type: Is it Movie or TV show
- 3. title: Title of movie/tv show
- 4. director: Name of Director
- 5. cast: who are part of acting
- 6. country: Release in which countries
- 7. date_added: date of streaming
- 8. release year: release year of that movie/tv show
- 9. rating: Likewise of that show
- 10. duration: duration of show in mins/seasons
- 11. listed in: category of the TV show/Movie
- 12. description: short summary

Data Cleaning:

Tool: Excel and Power Query

Brief Description:

Check Null Values using filters

Findings

director: 2,634

cast: 825

country: 831 → Filled with title-cased placeholder 'Nan' (as string)

date_added: 10

rating: 4 duration: 3

Replace by Director, Cast, Country by "Unknown"

And date_added by 01/01/2020 as per release year

Checking Duplicate Value

No Duplicate Values

- Convert Date-time in dd/mm/yyyy format using change type of power query
- Standardize the string element by proper type casting
- Uniform the Headers

Bold and standardize with first Letter Uppercase and rest of Lower case and space by "_"

Cleaned dataset View:

