

Name : Demitries Baskhara Rivaldo Tolla
NIM : 123180137
Class : B

EDA Assignment

Data Source :

Dataset didapatkan dari web <https://www.kaggle.com/> yang berisi data tentang film pada layanan streaming. Dataset ini bisa dibuat acuan untuk melakukan pencarian film pada layanan streaming tertentu. Dataset ini juga dapat dijadikan sebagai pencarian film berdasarkan rating pada situs IMDB , jenis , rating usia ataupun director yang memegang film tersebut. Dataset ini terdiri lebih dari 16 Ribu judul film yang berbeda.

Metadata :

- ID
- Title
- Year
- Age
- IMDb
- Rotten Tomatoes
- Netflix
- Hulu
- Prime Video
- Disney+
- Type
- Director
- Country
- Language
- Runtime

Basic Statistic

Summary

X	ID	Title	Year	
Min. : 0	Min. : 1	Length:16744	Min. :1902	
1st Qu.: 4186	1st Qu.: 4187	Class :character	1st Qu.:2000	
Median : 8372	Median : 8372	Mode :character	Median :2012	
Mean : 8372	Mean : 8372		Mean :2003	
3rd Qu.:12557	3rd Qu.:12558		3rd Qu.:2016	
Max. :16743	Max. :16744		Max. :2020	
Age	IMDb	Rotten.Tomatoes	Netflix	
Length:16744	Min. :0.000	Length:16744	Min. :0.0000	
Class :character	1st Qu.:5.100	Class :character	1st Qu.:0.0000	
Mode :character	Median :6.100	Mode :character	Median :0.0000	
	Mean :5.903		Mean :0.2126	
	3rd Qu.:6.900		3rd Qu.:0.0000	
	Max. :9.300		Max. :1.0000	
	NA's :571			
Hulu	Prime.Video	Disney.	Type	Directors
Min. :0.00000	Min. :0.0000	Min. :0.00000	Min. :0	Length:16744
1st Qu.:0.00000	1st Qu.:0.0000	1st Qu.:0.00000	1st Qu.:0	Class :character
Median :0.00000	Median :1.0000	Median :0.00000	Median :0	Mode :character
Mean :0.05393	Mean :0.7378	Mean :0.03368	Mean :0	
3rd Qu.:0.00000	3rd Qu.:1.0000	3rd Qu.:0.00000	3rd Qu.:0	
Max. :1.00000	Max. :1.0000	Max. :1.00000	Max. :0	
Genres	Country	Language	Runtime	
Length:16744	Length:16744	Length:16744	Min. : 1.00	
Class :character	Class :character	Class :character	1st Qu.: 82.00	
Mode :character	Mode :character	Mode :character	Median : 92.00	
			Mean : 93.41	
			3rd Qu.: 104.00	
			Max. :1256.00	
			NA's :592	

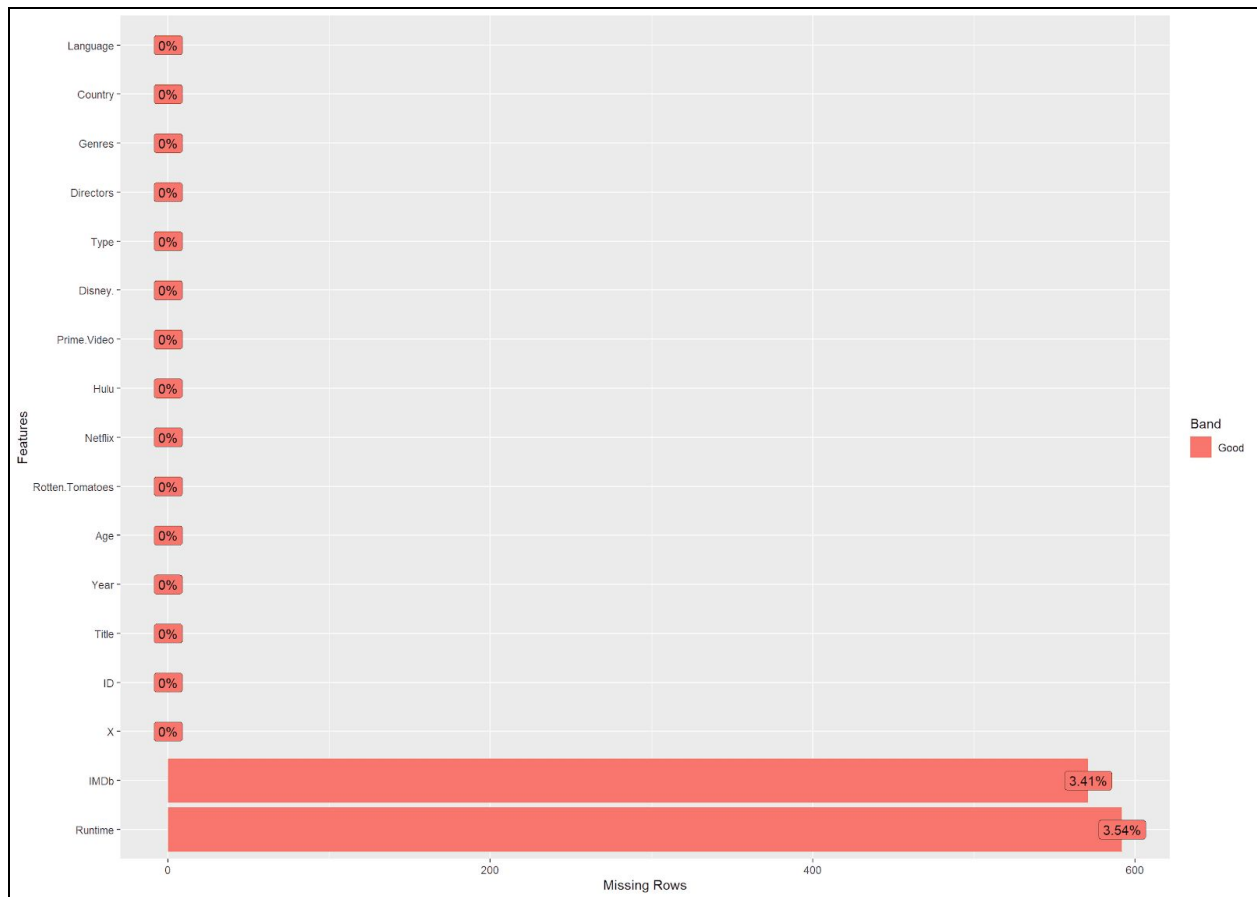
Dataset terdiri dari 16743 baris.

Raw Counts

Name	Value
Rows	16,744
Columns	17
Discrete columns	7
Continuous columns	10
All missing columns	0
Missing observations	1,163
Complete Rows	15,819
Total observations	284,648
Memory allocation	3.9 Mb

Dataset terdiri dari 17 kolom.

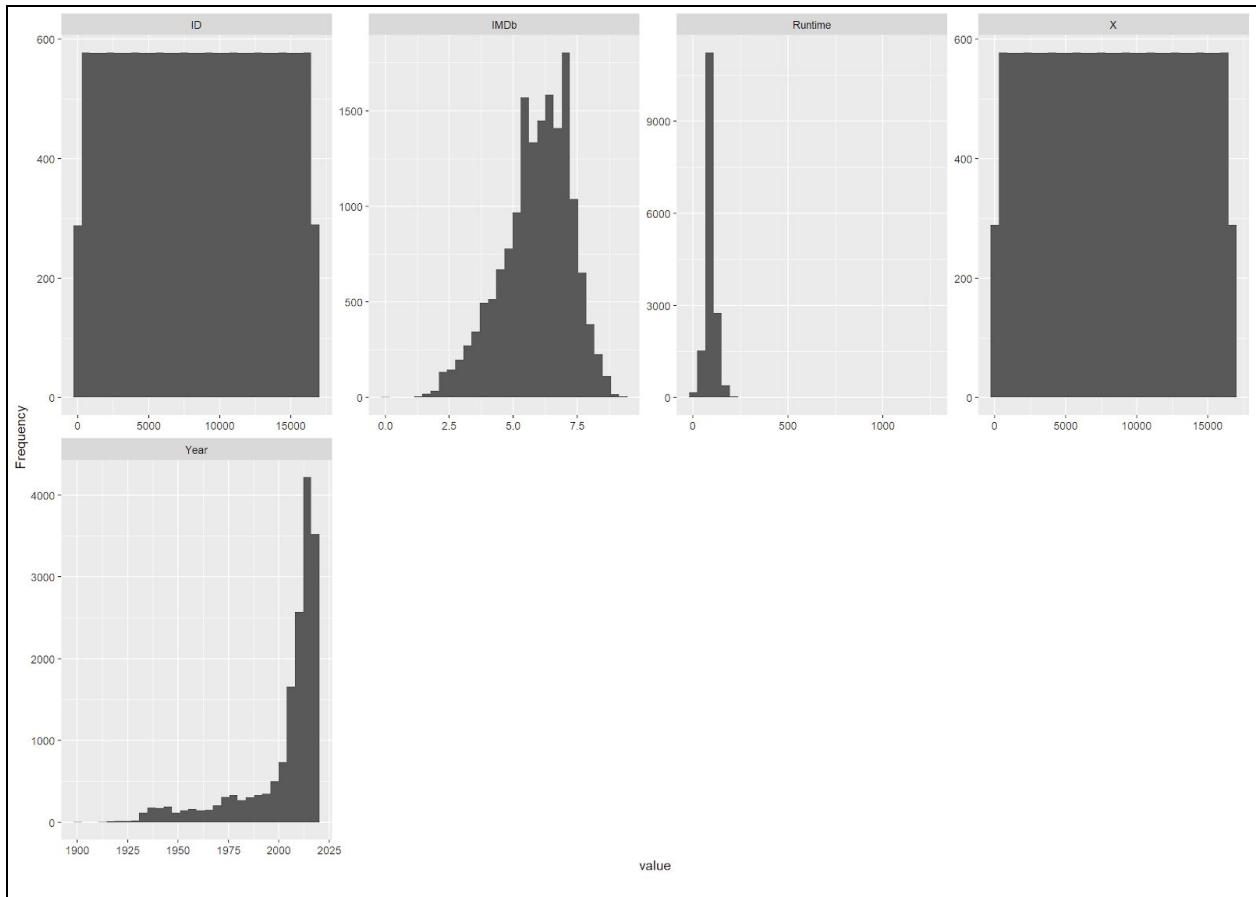
Data Structure



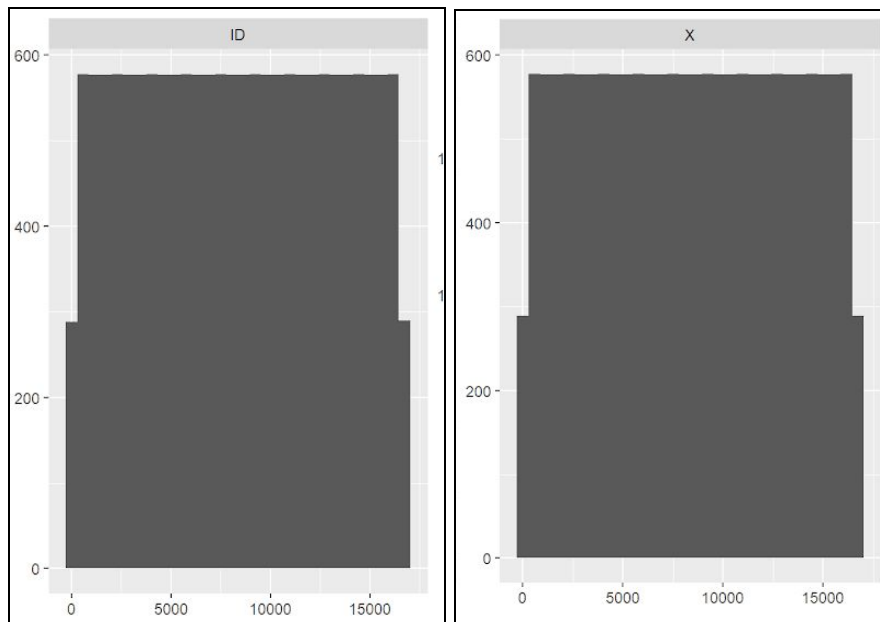
Ada beberapa data pada kolom IMDb dan runtime yang hilang.

Univariate Distribution

Histogram

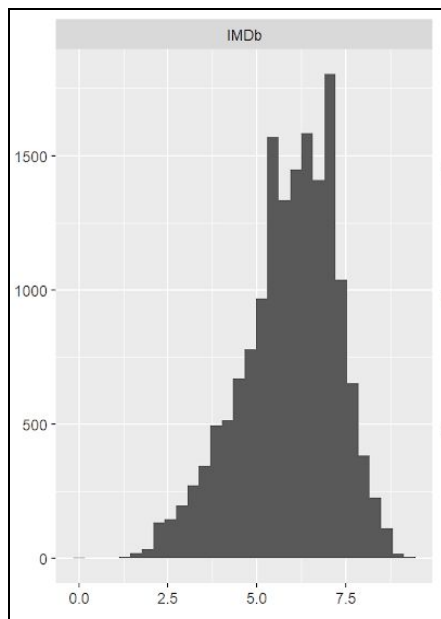


Histogram ID dan X



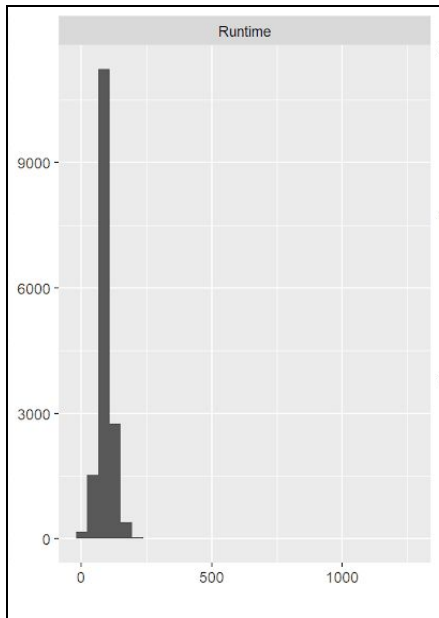
Data pada kolom ID dan X tidak mengandung informasi yang penting untuk dataset. Data pada kolom tersebut hanya berisi urutan angka yang berurutan sehingga membuat histogram berbentuk simetris.

Histogram IMDb



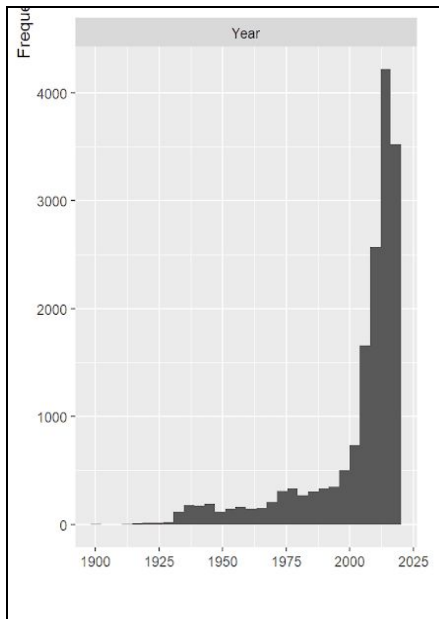
Data pada kolom IMDb berisi informasi mengenai rating film pada situs IMDb. Pada histogram dapat terlihat sebaran nilai rating. Nilai antara 5.0 dan 7.5 adalah nilai yang paling banyak. Nilai rata rata pada histogram IMDb adalah 5,903.

Histogram Runtime



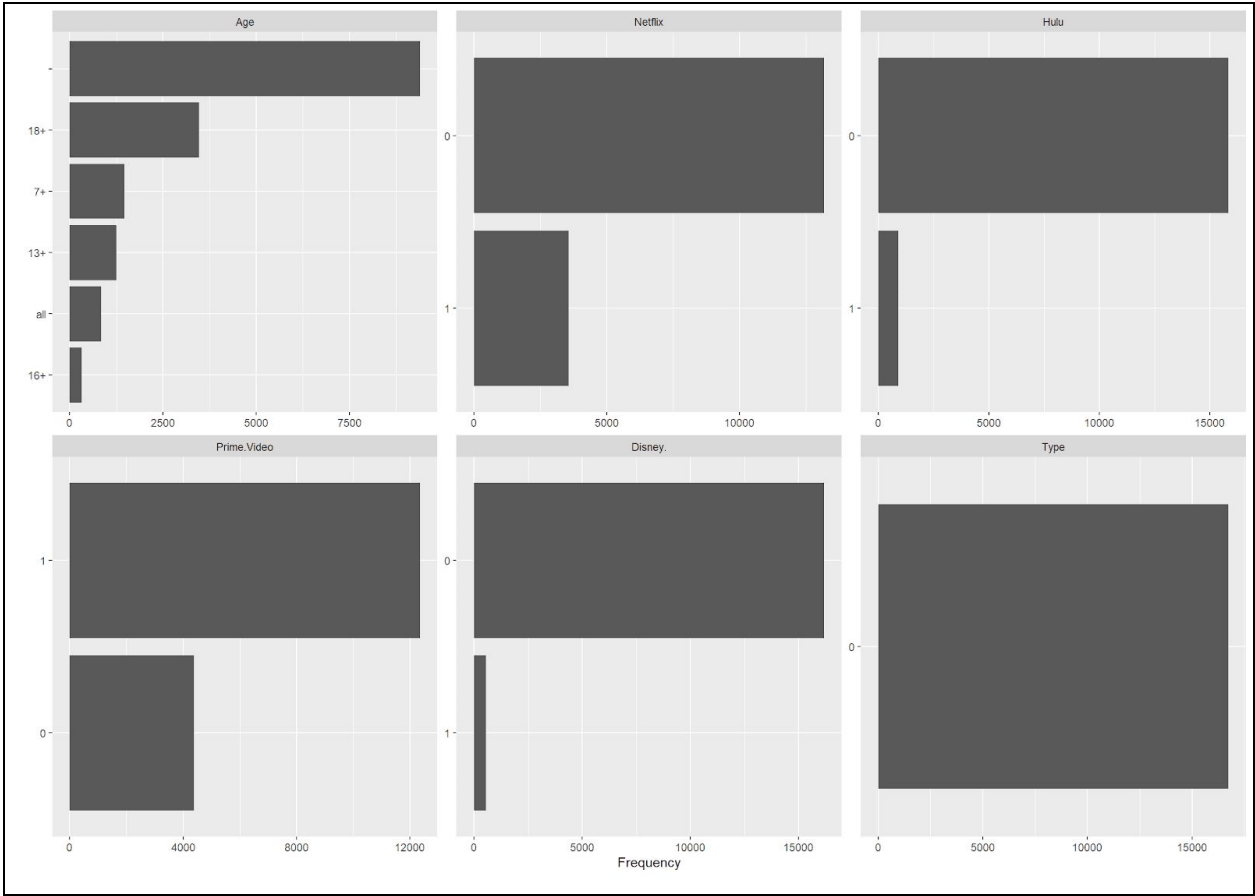
Data pada kolom runtime berisi informasi durasi film. Pada histogram tidak terlalu terlihat secara detail dikarenakan terdapat beberapa data dengan runtime tinggi tapi berjumlah sedikit. Rata rata nilai runtime adalah 93,41 , dengan nilai terendah yaitu 1 dan nilai tertinggi yaitu 1256.

Histogram Year

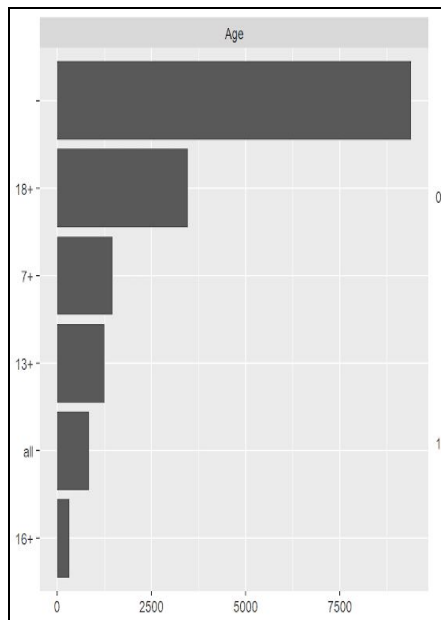


Data pada kolom year berisi informasi tentang tahun dimana film dibuat. Pada histogram mulai tahun 2000 terjadi peningkatan produksi film secara signifikan jika dibandingkan dengan tahun sebelumnya.

Bar Chart

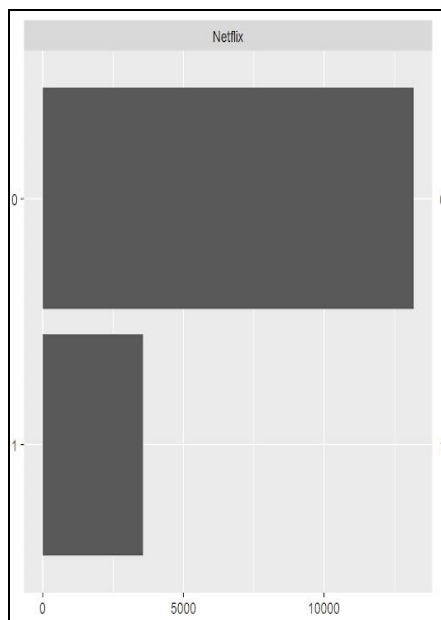


Bar Chart Age



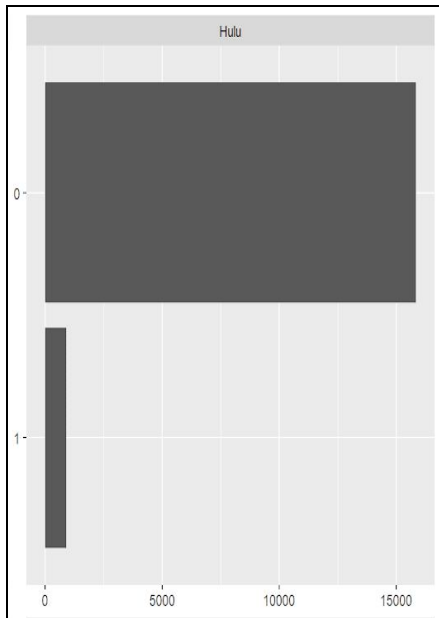
Data pada kolom age berisi informasi mengenai batasan umur untuk setiap film. Data terbagi menjadi 6 kategori batasan umur.

Bar Chart Netflix



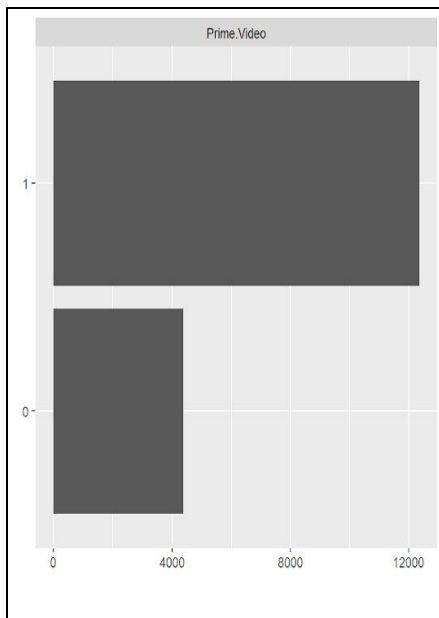
Data pada kolom Netflix berisi informasi mengenai data film yang tersedia di Netflix. Pada Bar Chart terlihat film yang tersedia pada Netflix berjumlah kurang dari 5000.

Bar Chart Hulu



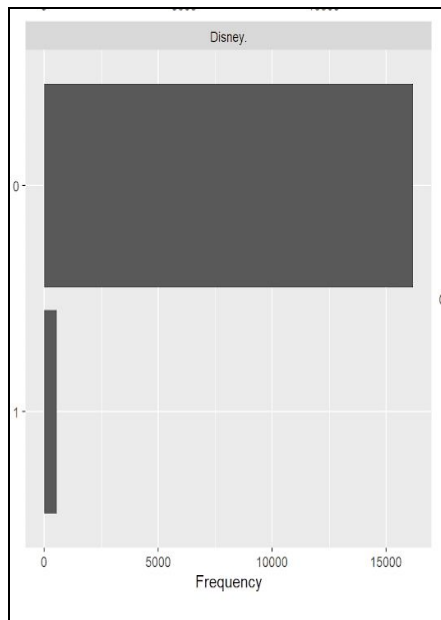
Data pada kolom Hulu berisi informasi mengenai data film yang tersedia di Hulu. Pada Bar Chart terlihat film yang tersedia pada Hulu hanya berjumlah beberapa ratus film saja.

Bar Chart Prime Video



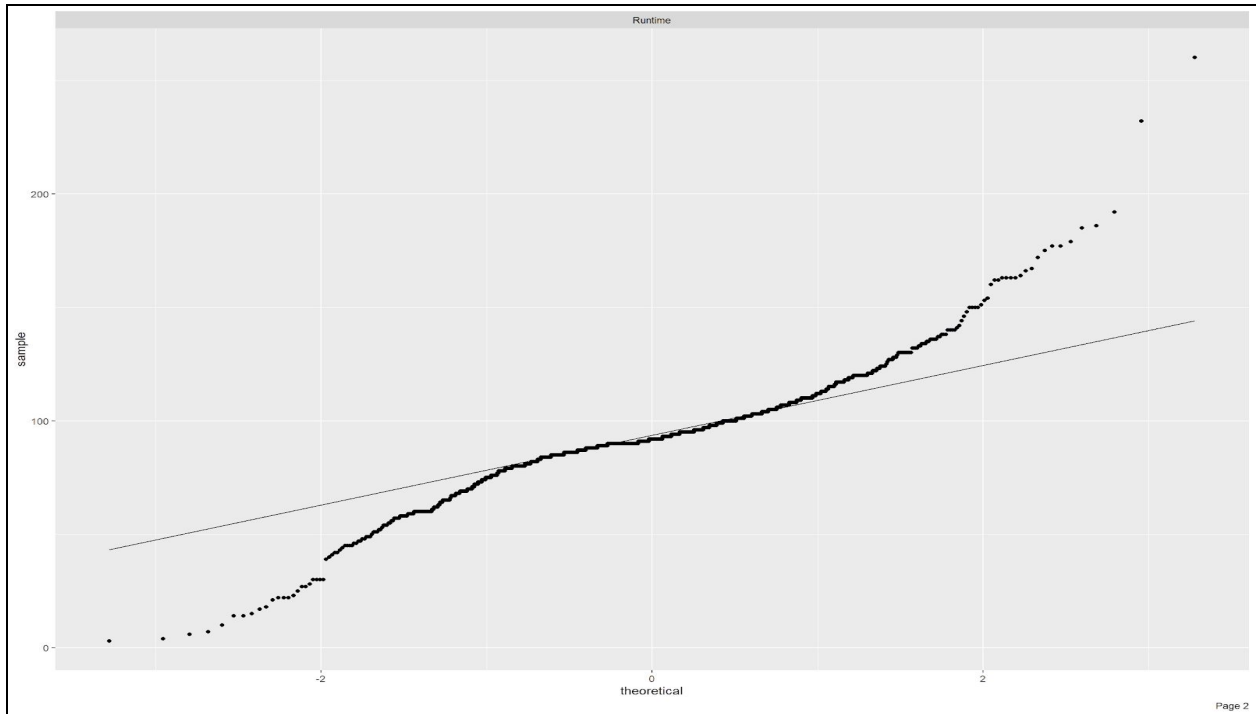
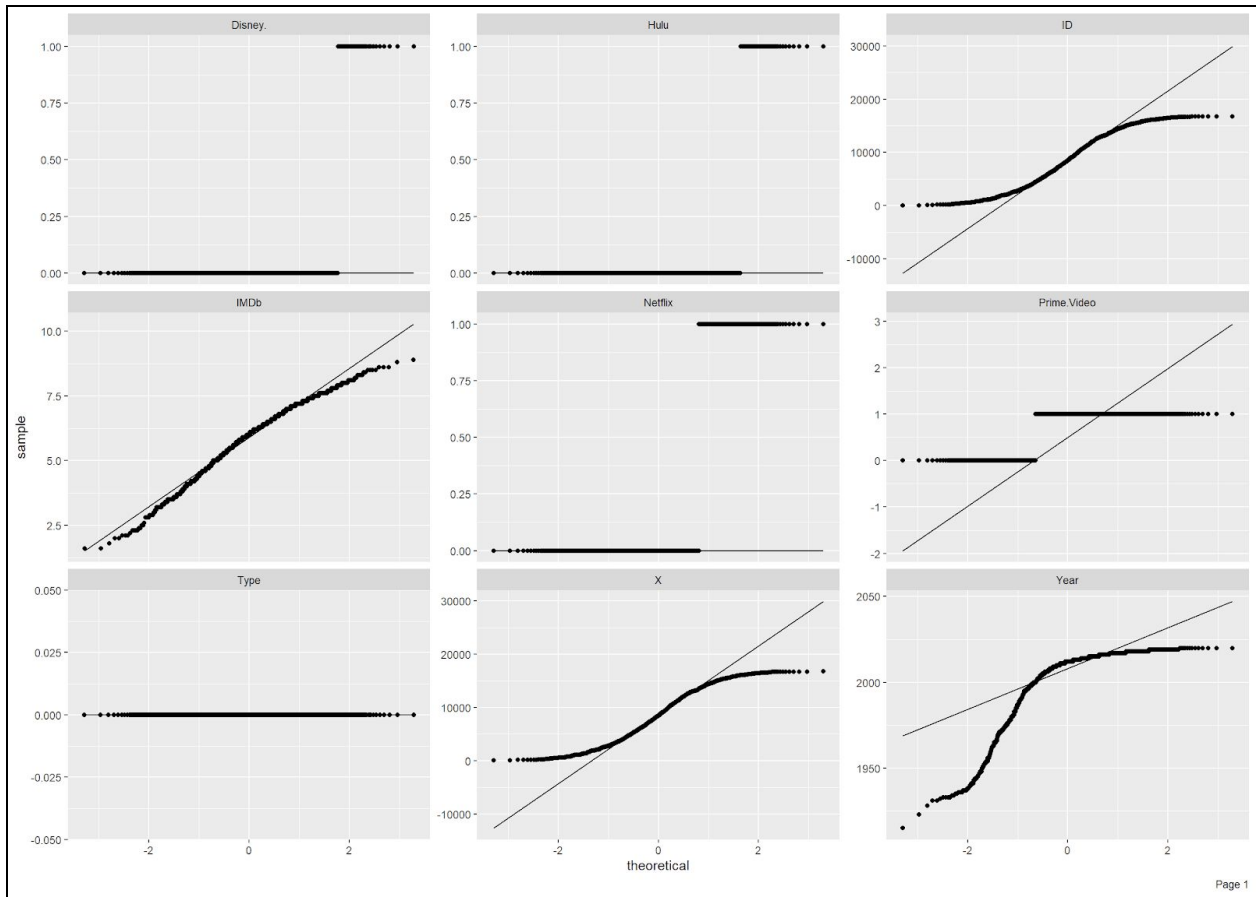
Data pada kolom Prime Video berisi informasi mengenai data film yang tersedia di Prime Video. Pada Bar Chart terlihat film yang tersedia pada Prime Video berjumlah lebih dari 4000 film.

Chart Bar Disney+



Data pada kolom Disney+ berisi informasi mengenai data film yang tersedia di Disney+. Pada Bar Chart terlihat film yang tersedia pada Disney+ hanya berjumlah beberapa ratus film saja.

QQPLOT



End Note :

Beberapa data pada kolom seperti Netflix , Hulu , Prime Video atau Disney+ sewaktu waktu dapat mengalami perubahan. Ketidaksesuaian dalam kolom Age , Genres , Director , atau runtime bisa terjadi , kolom - kolom tersebut membutuhkan kejelasan sumber yang lebih dalam dalam penentuan informasi yang valid.

Source : <https://www.kaggle.com/ruchi798/movies-on-netflix-prime-video-hulu-and-disney>