# **Analisys Datawarehouse Menggunakan Star Schema Untuk Mengetahui Rating Film Berdasarkan Genre Dan Tahun Release**

Oleh

Fitria Risqina

*Jurusan Teknik Informatika, Fakultas Teknik Elektro Dan Teknologi Informasi, Institut Teknologi Adhi Tama Surabaya*

**Abstrak**

Banyaknya film yang hadir dipasaran perfilman mengakibatkan persaingan yang sangat ketat. Oleh karena itu analisa ini dilakukan dengan tujuan memberikan informasi lebih lanjut mengenai aspek-aspek yang membuat film sukses dari sudut pandang pengguna berdasarkan data yang didapatkan dari IMDB atau sebuah basis data online yang berkaitan dengan film. Kumpulan data mencakup 85.855 film dengan atribut seperti deskripsi film, peringkat rata-rata, jumlah suara, genre, dan lain-lain. Analisa yang tepat, benar dan lengkap menghasilkan informasi yang sangat berharga.

***Kata Kunci :*** *Database, Datawarehouse, Star Schema, Datasets, Analisys, ETL (Extract, Transform, Load)*

1. **Pendahuluan**

Sumber data pada datawarehouse berasal dari berbagai macam format, software, platform dan jaringan yang beda. Data warehouse juga merupakan salah satu sistem pendukung keputusan, yaitu dengan menyimpan data dari berbagai sumber, mengorganisasikannya dan dianalisa oleh para pengambil kebijakan. Akan tetapi datawarehouse tidak dapat memberikan keputusan secara langsung. Namun ia dapat memberikan informasi yang dapat membuat user menjadi lebih paham dalam membuat kebijakan strategis. Data digital di tampung dalam sebuah basisdata (database) digunakan untuk membuat kebijakan, melakukan strategi manajemen, atau mengambil keputusan. Tumpukan data di database tidak akan berarti jika tidak diolah untuk menghasilkan sebuah informasi yang dapat menunjang proses bisnis atau untuk menghasilkan sebuah keputusan. Berdasarkan permasalahan yang telah dipaparkan diatas dapat dirumuskan permasalahan sebagai berikut: Bagaimana merancang Data Warehouse untuk mengetahui informasi lebih lanjut mengenai aspek yang membuat film sukses dari sudut pandang pengguna berdasarkan data yg di dapat dari imdb dengan star skema sebagai metode perancangan untuk menghasilkan sebuah informasi. permasalahan akan dibatasi pada hal- hal yang berkaitan dengan perancangan data warehouse ini sebagai berikut:

1. Data ditransformasikan dalam format yang berbeda- beda dari sumber database yang berbeda-beda ke database data warehouse.
2. Data diproses melalui proses Extract Transform Load (ETL) untuk disimpan ke database data warehouse.
3. Analisa data direpresentasikan menggunakan teknologi Online Analytical Processing (OLAP) dalam bentuk laporan analisa
4. Skema Perancangan Data Warehouse adalah Star schema

Adapun tujuan yang ingin dicapai dalam peneltian ini adalah:

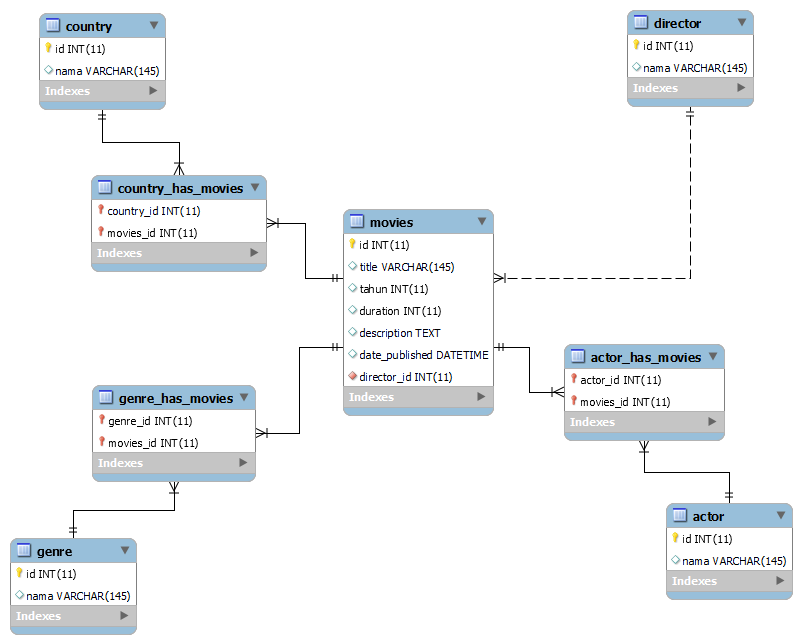
1. Mengetahui proses transformasi data dari sumber yang berbeda melalui mekanisme Extract Transform Load (ETL) dalam merepresentasikan analisa data menggunakan teknologi Online Analytical Processing (OLAP).
2. Menambah pengetahuan dan wawasan tentang Database, Data Warehouse dan proses yang terlibat di dalamnya.

**2. Pembahasan**

Pada bagian ini akan dibahas mengenai beberapa bagian dalam proses analisa diantaranya adalah sebagai berikut,

**2.1 Rancangan Skema Star Schema**

Datawarehouse memiliki sebuah struktur yang terspesialisasi untuk itu yang disebut star schema. Struktur ini mendefinisikan sebuah query dalam datawarehouse. Star schema berupa sebuah fact table yang dikelilingi oleh dimension tables. Setiap star secara teknis adalah sebuah datamart dan setiap datawarehouse bisa memiliki banyak stars atau Data Marts. Kegunaan dari star schema adalah memodifikasi tabel OLTP yang telah ternormalisasi menjadi sebuah struktur yang sederhana dan terdernomalisasi dalam datawarehouse. Dalam skema ini, ada fact table di tengah dan di sekelilingnya terdapat dimension table. Sehingga jika digambarkan akan terlihat seperti gambar 1.



**Gambar 1** Skema Star Schema Table Analisys Film

**2.2 Proses ETL (Extraction, Transformation dan Loading)**

1. Extraction

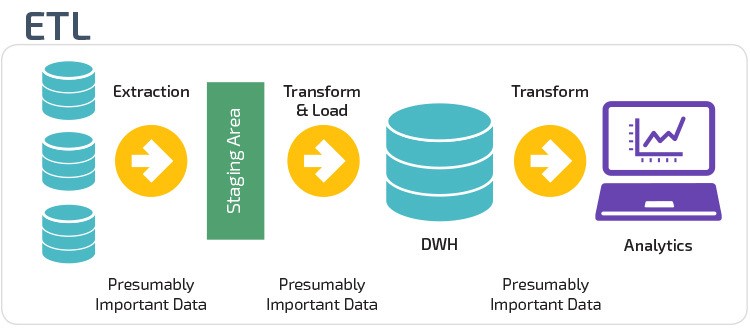
Extraction adalah suatu proses yang mengidentifikasikan seluruh sumber data yang relevan dan kemudian mengambil data dari sumbersumber data tersebut Extraction adalah suatu proses yang mengidentifikasikan seluruh sumber data yang relevan dan kemudian mengambil data dari sumbersumber data tersebut.

1. Transform

Transformation adalah suatu proses yang memiliki peran dalam melakukan perubahan dan integrasi skema data serta struktur yang telah didefinisikan sebelumnya oleh data warehouse.

1. Loading

Loading adalah suatu proses pemindahan data secara fisik dari sistem operasional ke dalam data warehouse. Seluruh proses diatas dapat dilihat pada gambar 2.



**Gambar 2** Proses ETL (Extract, Transform, Load)

* 1. **Proses Analisa**

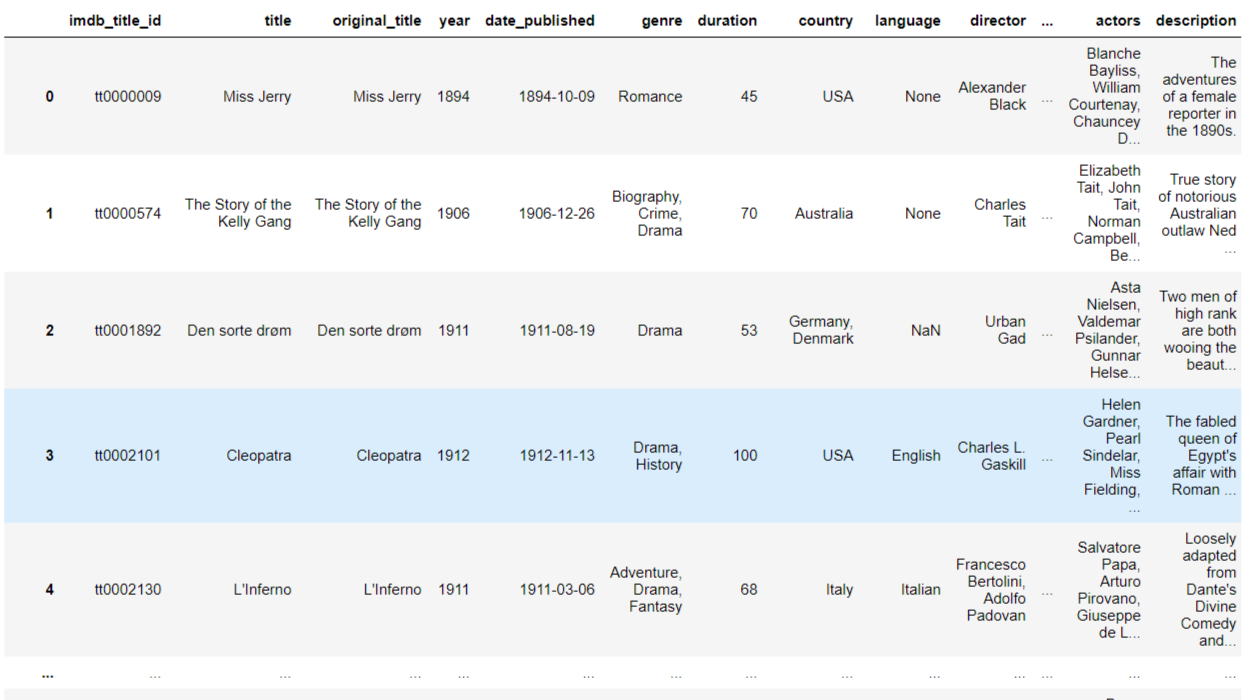
Pada bagian proses analisa ini akan terdapat beberapa tahapan, diantaranya adalah sebagai berikut,

* + 1. **Identifikasi Data Film**

Identifikasi Data melibatkan beberapa dataset yang telah didapatkan dari IMDB, dimana terdapat beberapa dataset dengan data yang berbeda-beda, diantaranya adalah

1. Movies

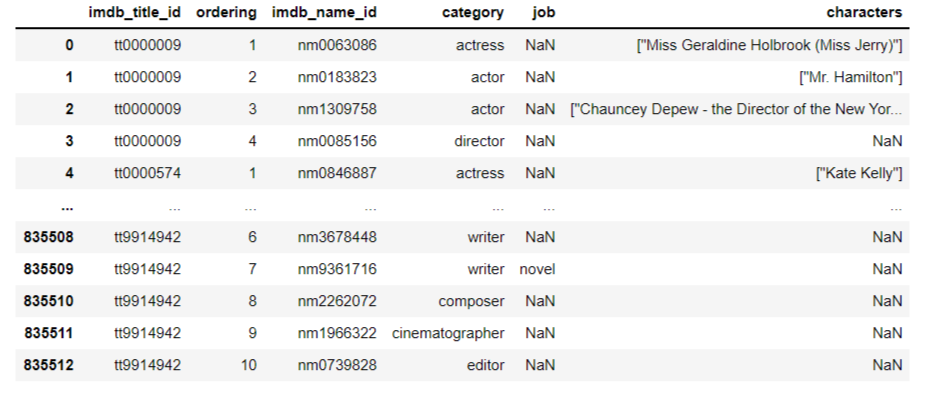
Dataset Movies berisikan data movie yang dapat dilihat pada gambar berikut,



**Gambar 3** Datasets Movies

1. Title

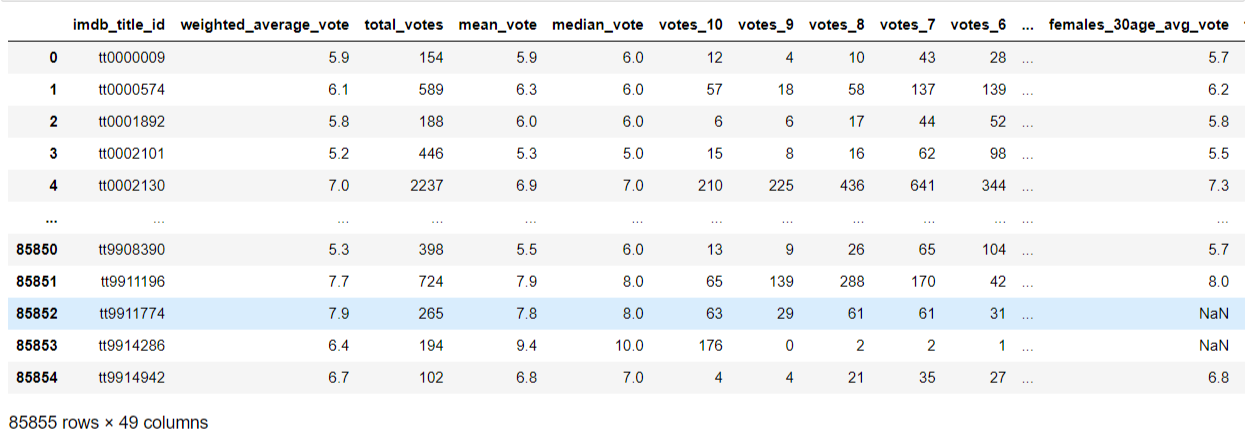
Dataset Titles berisikan data category dan character yang dapat dilihat pada gambar berikut,



**Gambar 4** Datasets Titles

1. Rating

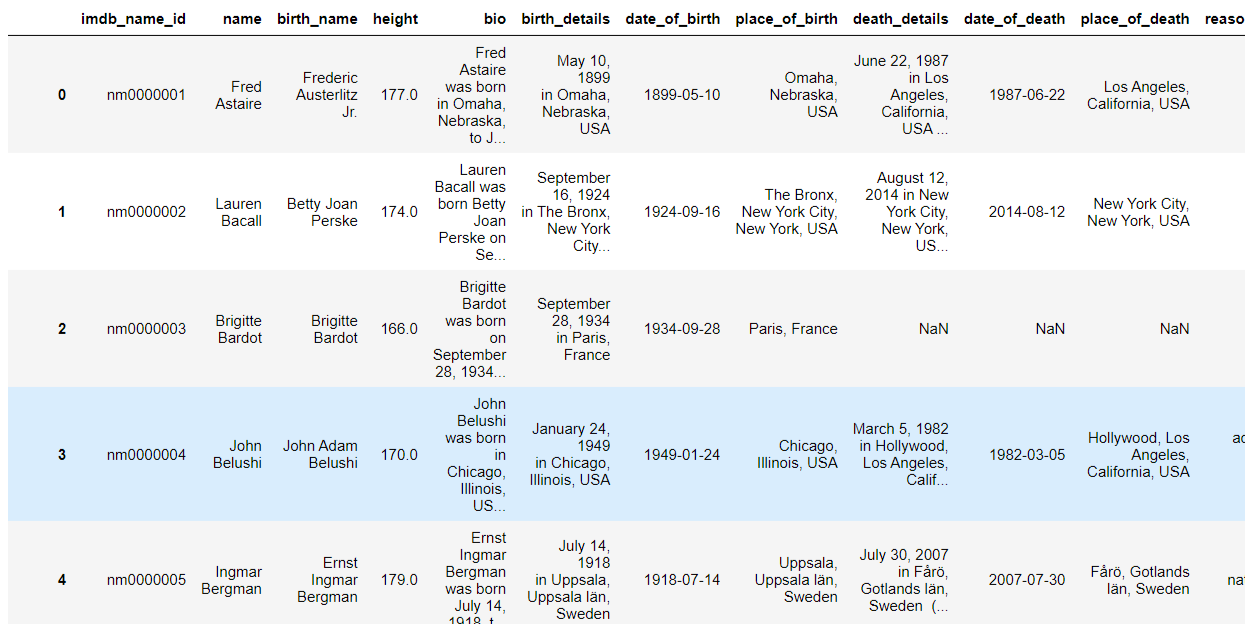
Dataset Rating berisikan data vote dengan berbagai macam tipe vote yang dapat dilihat pada gambar berikut,



**Gambar 5** Datasets Rating

1. Names

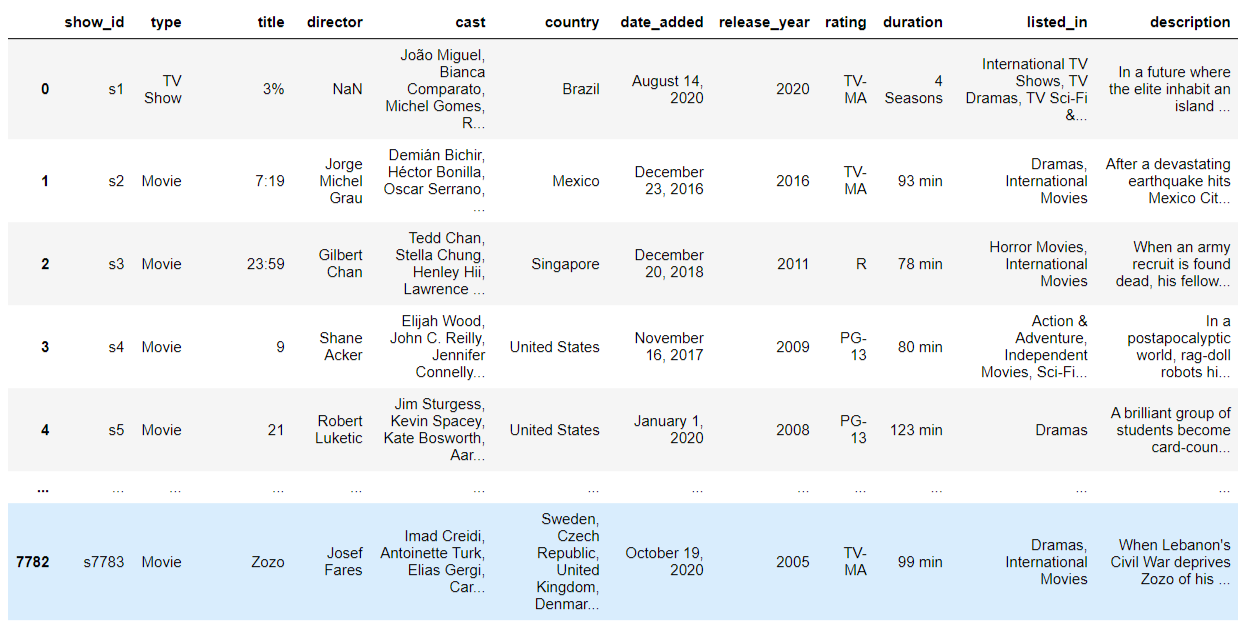
Dataset names berisikan data user yang melakukan vote yang dapat dilihat pada gambar berikut,



**Gambar 6** Datasets Names

1. Netflix Titles

Dataset Netflix Titles berisikan data movie yang dapat dilihat pada gambar berikut,

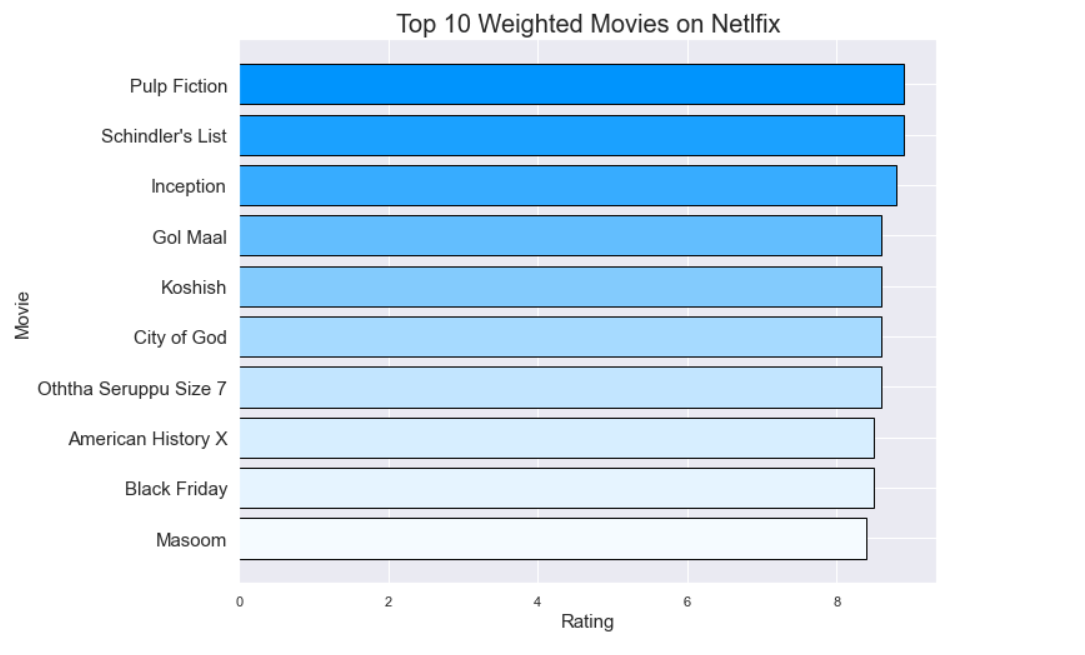


**Gambar 7** Datasets Netflix Titles

**2.3.2 Hasil Analisa**

Dari hasil identifikasi dan analisa terdapat beberapa hasil diantaranya adalah sebagai berikut,

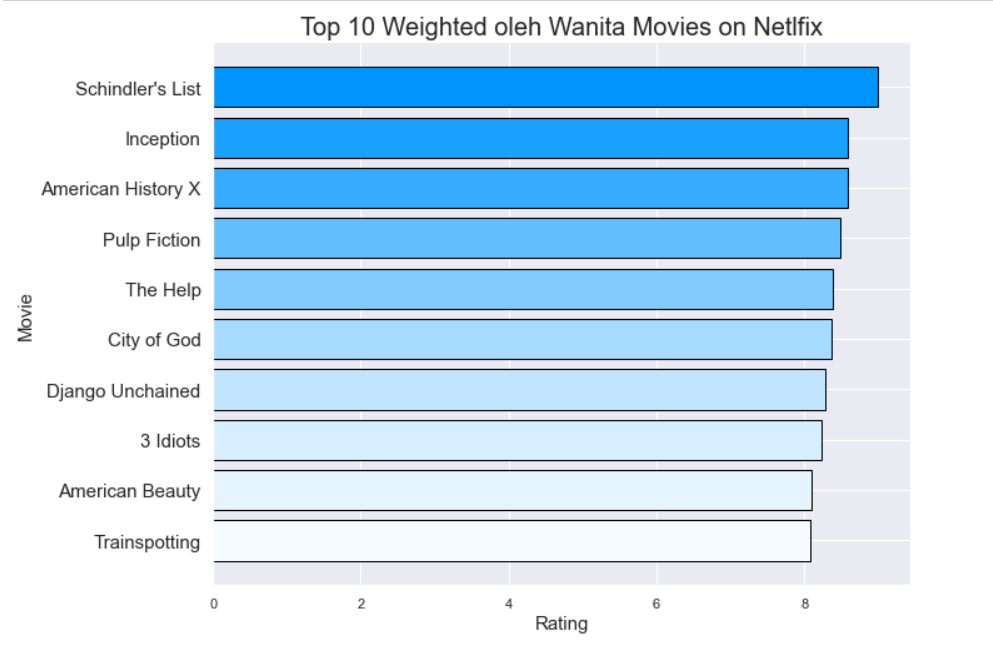
**2.3.2.1 Data Peringkat Teratas**

****

**Gambar 8** Grafik Peringkat Teratas

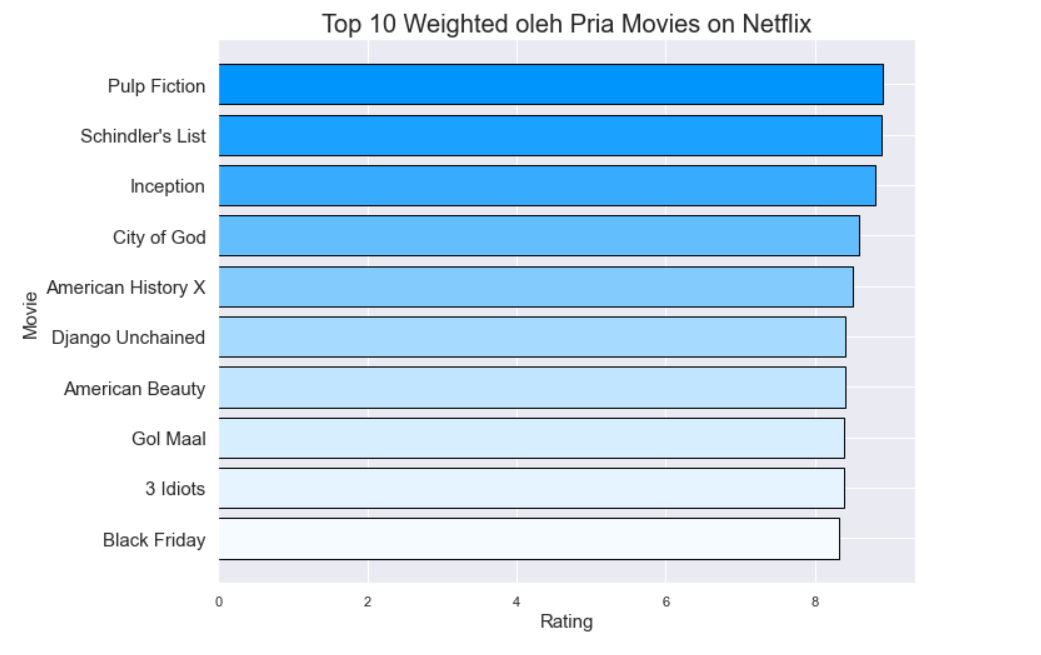
Dari gambar diatas dapat diambil kesimpulan bahwa terdapat data dengan 10 peringkat teratas berdasarkan total rata-rata voting yang dilakukan oleh user.

**2.3.2.2 Data Peringkat Teratas Berdasarkan Hasil Vote Pria & Wanita**

****

**Gambar 9** Grafik Peringkat Teratas Dari Vote Wanita

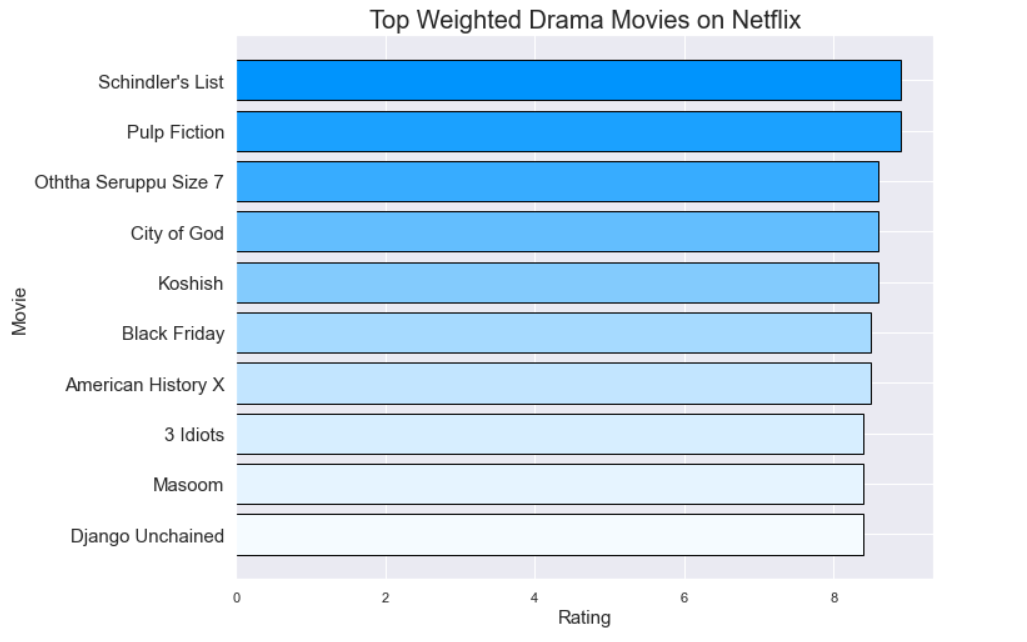
Dari gambar diatas dapat dilihat bahwa hasil vote Wanita menghasilkan 10 peringkat film teratas dimana peringkat utama juga masuk ke dalam hasil rata-rata voting secara keseluruhan.

****

**Gambar 10** Grafik Peringkat Teratas Dari Vote Pria

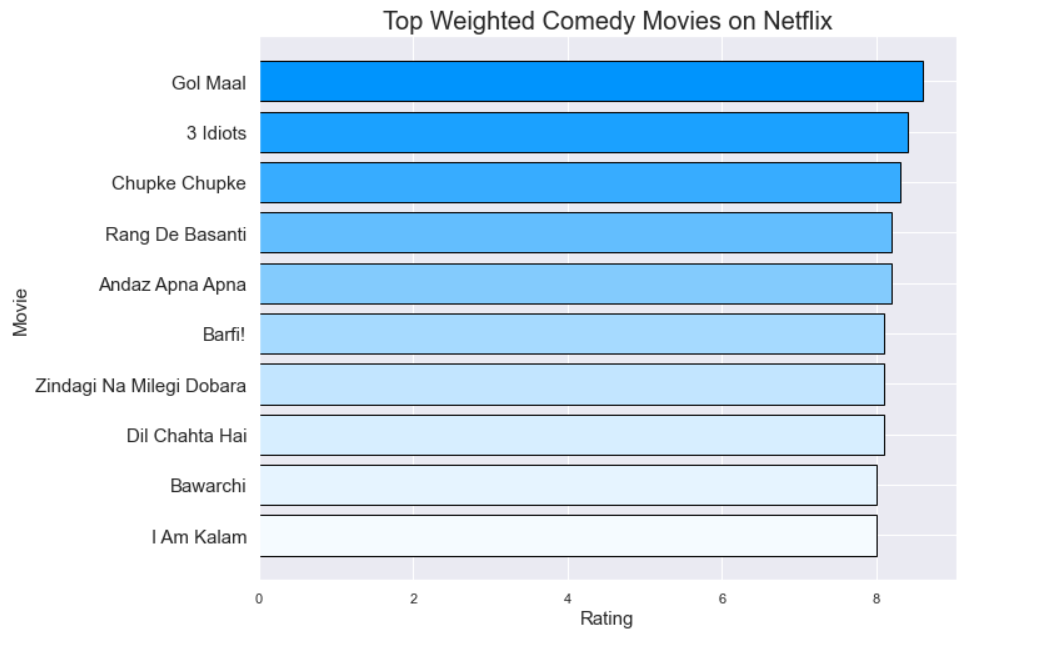
Dari gambar diatas dapat dilihat bahwa hasil vote Pria menghasilkan 10 peringkat film teratas dimana peringkat utama juga masuk ke dalam hasil rata-rata voting secara keseluruhan serta masuk juga dalam vote yang dilakukan oleh wanita.

**2.3.2.3 Data Peringkat Teratas Berdasarkan Genre**

****

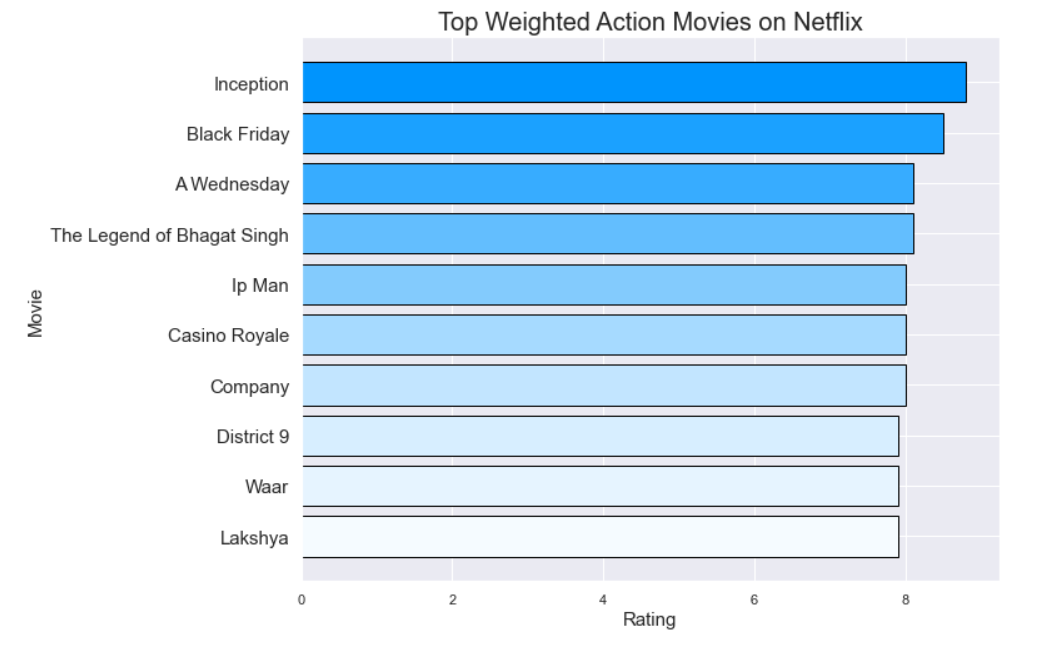
**Gambar 11** Grafik Peringkat Teratas Dari Genre Drama

Gambar diatas menunjukkan bahwa genre film drama menghasilkan 10 peringkat teratas dimana salah satu film bergenre drama tersebut terdapat juga pada total rata-rata voting dan juga yang digemari oleh Pria dan Wanita.

****

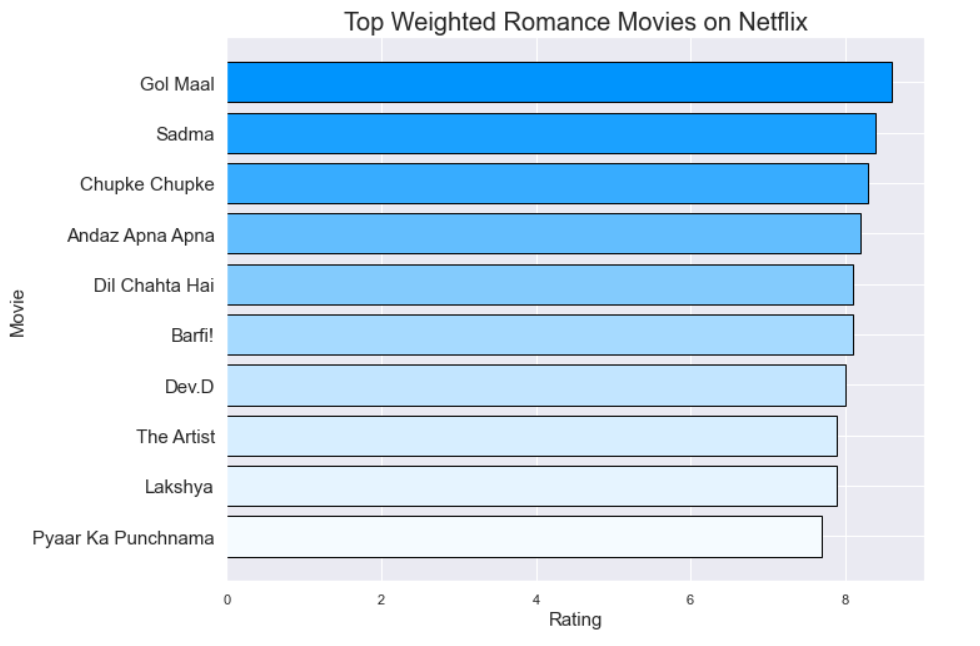
**Gambar 12** Grafik Peringkat Teratas Dari Genre Comedy

Pada grafik film bergenre comedy agak berbeda dengan grafik sebelumnya, pada grafik ini rata-rata film yang muncul adalah film yang tidak ada pada grafik-grafik sebelumnya.

****

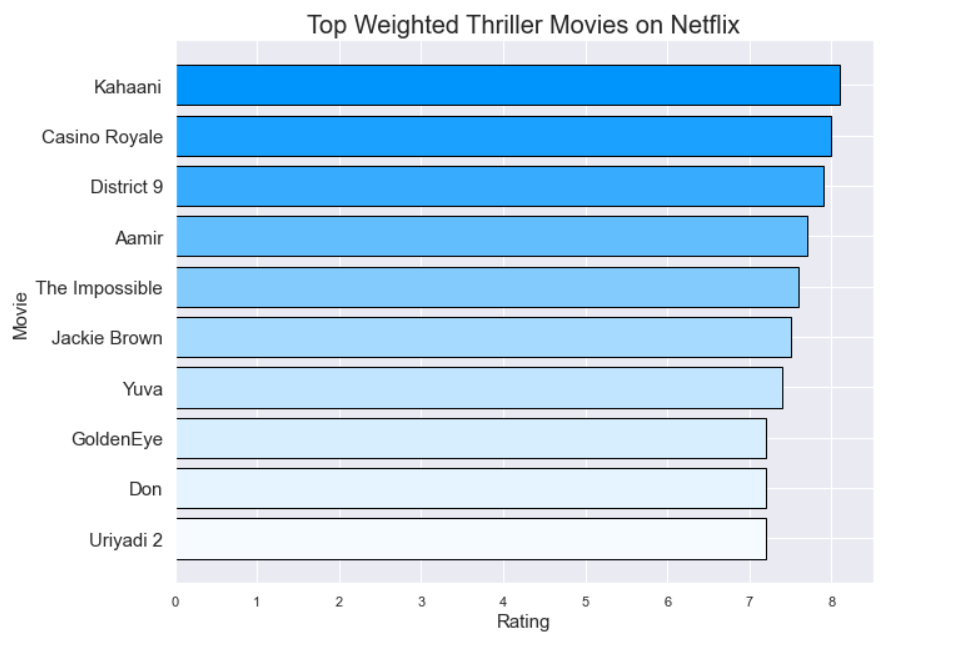
**Gambar 13** Grafik Peringkat Teratas Dari Genre Action

Pada grafik film bergenre action juga agak berbeda dengan grafik sebelumnya, pada grafik ini rata-rata film yang muncul adalah film yang tidak ada pada grafik-grafik sebelumnya.

****

**Gambar 14** Grafik Peringkat Teratas Dari Genre Romance

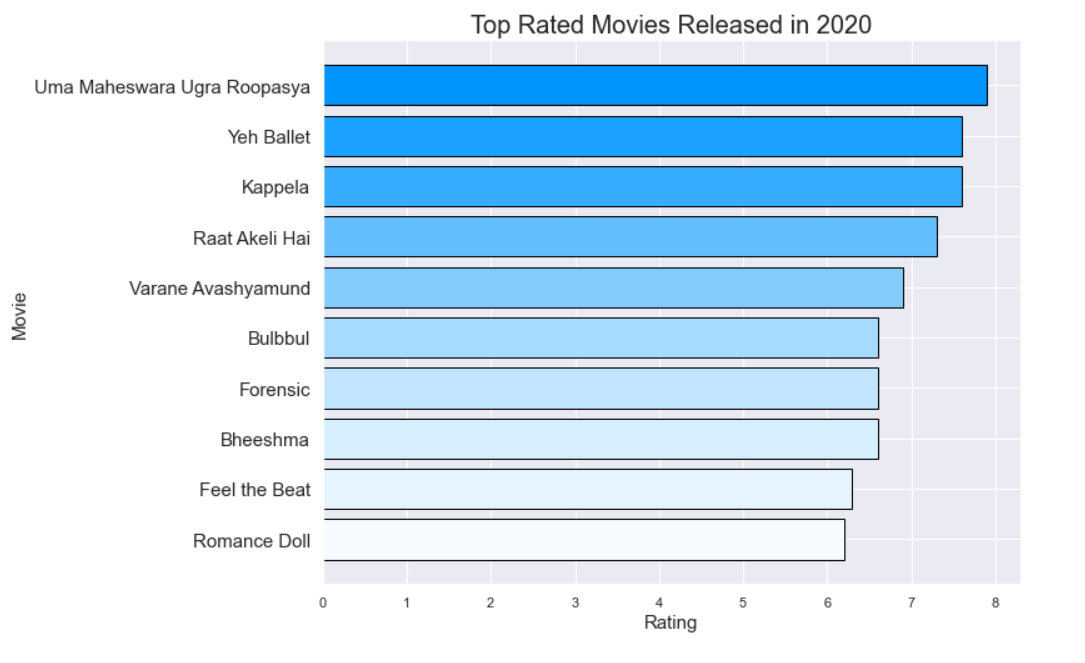
Pada grafik film bergenre Romance terdapat salah satu film yang muncul sebagai top teratas dimana film yang muncul pada grafik terdapat juga pada grafik sebelumnya yaitu grafik yang bergenre Comedy.

****

**Gambar 15** Grafik Peringkat Teratas Dari Genre Thriller

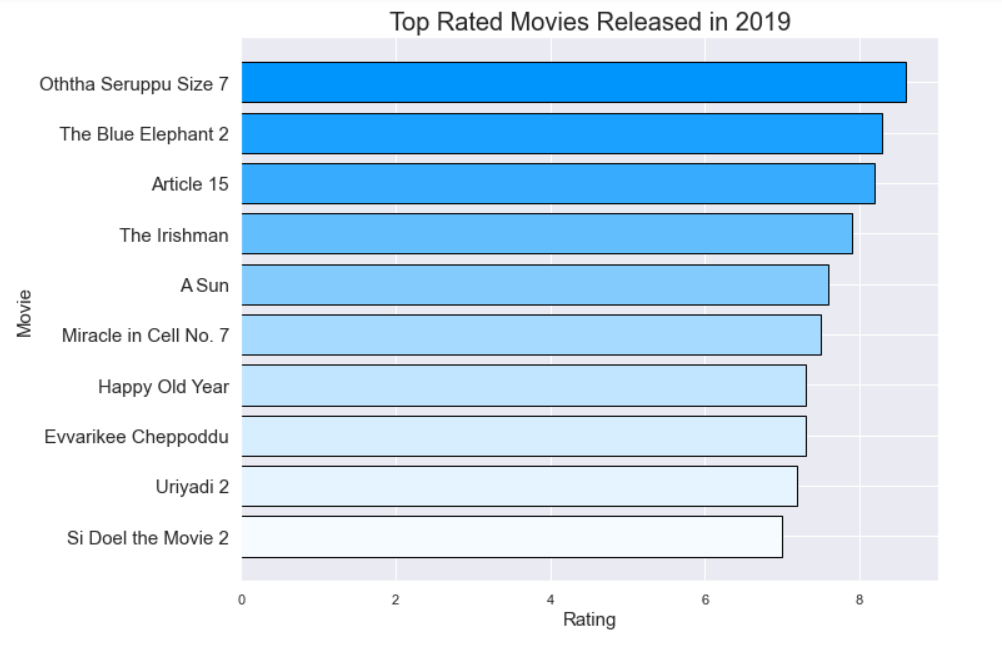
Pada grafik ini dapat disimpulkan bahwa film yang muncul adalah film yang tidak muncul pada grafik-grafik sebelumnya.

**2.3.2.4 Data Peringkat Teratas Berdasarkan Tahun Release**



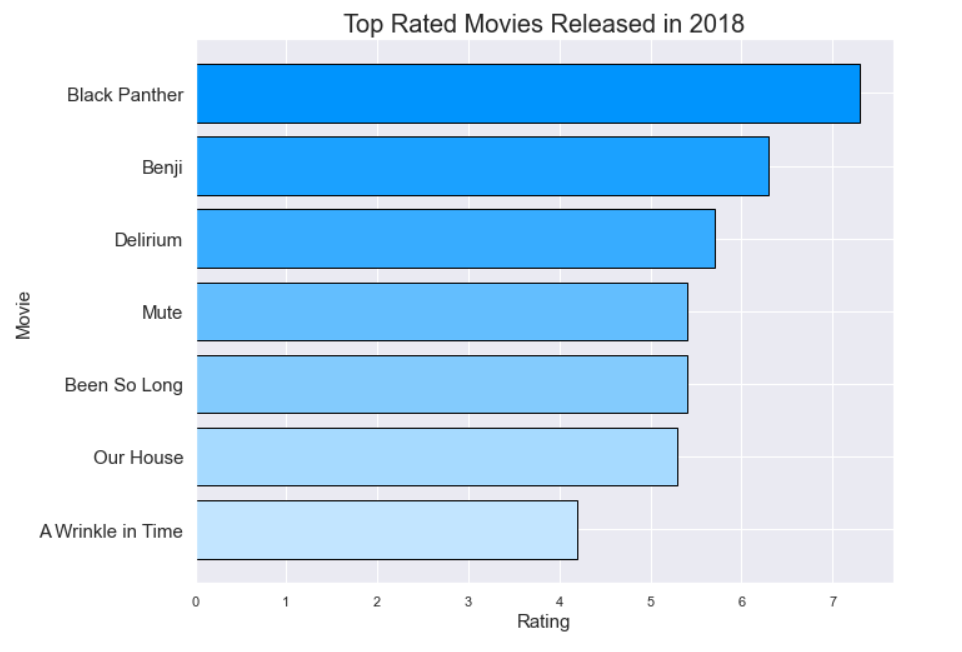
**Gambar 16** Grafik Peringkat Teratas Dari Tahun Release 2020

Pada grafik diatas digambarkan bahwa tahun release 2020 didominasi film-film bergenre Romance.



**Gambar 17** Grafik Peringkat Teratas Dari Tahun Release 2019

Pada grafik diatas menunjukkan bahwa tahun release 2019 peringkat teratas yang muncul terdapata beberapa genre dibandingkan dengan tahun release 2020 yang didominasi genre comedy



**Gambar 18** Grafik Peringkat Teratas Dari Tahun Release 2018

Pada tahun release 2018 menunjukkan bahwa yang mendominasi adalah film dengan genre action.

**3. Kesimpulan**

Kesimpulan yang didapatkan dari proses analisa dan implementasi datawarehouse pada penelitian ini adalah bahwa suatu film dapat sangat sukses dan digemari oleh semua kalangan adalah berdasarkan rating dan yang paling disukai oleh kalangan Pria dan Wanita berdasarkan genre, serta tahun release dari suatu film sangat berpengaruh juga. Kemudian implementasi dari datawarehouse sangatlah berguna untuk mendapatkan suatu hasil dari proses analisa yang dilakukan.