

MovieLens Project

By Shanti Jogi

Project Overview

The MovieLens project was aimed at analysing movie ratings data using Microsoft Power BI. The dataset included information about **movies**, **users**, and **ratings**. The goal was to clean, transform, and visualize the data to uncover insights about user preferences and rating trends.

Data Cleaning and Preparation

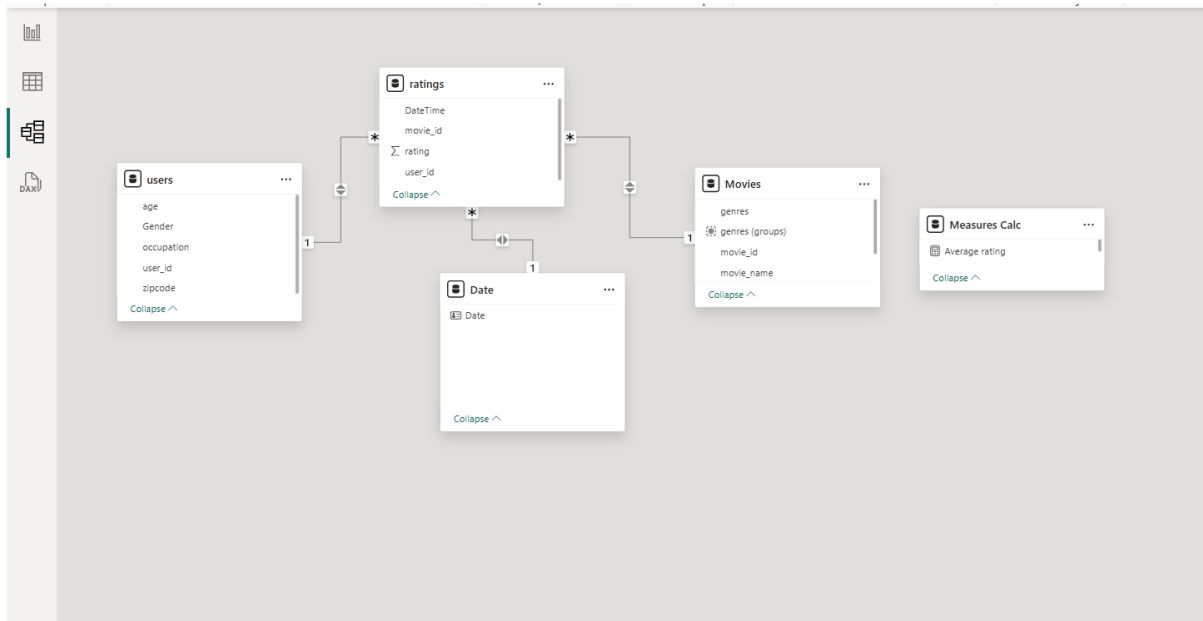
Performed in **Power Query**:

- Imported multiple datasets: movies, ratings, and users.
- Split the genres column using a delimiter (,) to handle multiple genres.
- Converted the rating timestamp (in seconds) to a standard **date and time** format.
- Removed unnecessary columns and renamed fields for clarity.

	1 user_id	2 movie_id	3 rating	4 DateTime
1	1	1193	5	31-12-2000 22:12:40
2	1	661	3	31-12-2000 22:35:09
3	1	914	3	31-12-2000 22:32:48
4	1	3408	4	31-12-2000 22:04:35
5	1	2355	5	06-01-2001 23:38:11
6	1	1197	3	31-12-2000 22:37:48
7	1	1287	5	31-12-2000 22:33:59
8	1	2804	5	31-12-2000 22:11:59
9	1	594	4	31-12-2000 22:37:48
10	1	919	4	31-12-2000 22:22:48
11	1	595	5	06-01-2001 23:37:48
12	1	938	4	31-12-2000 22:29:12
13	1	2398	4	31-12-2000 22:38:01
14	1	2918	4	31-12-2000 22:35:24
15	1	1035	5	31-12-2000 22:29:13
16	1	2791	4	31-12-2000 22:36:28
17	1	2687	3	06-01-2001 23:37:48
18	1	2018	4	31-12-2000 22:29:37
19	1	3105	5	31-12-2000 22:28:33
20	1	2797	4	31-12-2000 22:33:59
21	1	2321	3	31-12-2000 22:36:45

Data Modelling

- Built **entity relationships** between:
 - movies and ratings (via movieId)
 - ratings and users (via userId)
- Created a **Master Date Table** using DAX.
- Linked the Date Table with the Ratings table using the rating_date.
- Marked the Date Table as a **Date Table** to enable time intelligence functions.



Report View Customization

- Grouped columns into folders in the Report View for better organization.
- Created calculated columns and measures using DAX.

Visuals and Dashboard Elements

- **Cards** for key performance indicators (KPIs) like:
 - Total number of ratings
 - Number of users
 - Average rating
- **Bar/Column Charts** to show:
 - Top-rated movies
 - Most active users
- **Line Chart** to display rating trend.
- **Stacked Column Chart** to compare genre preferences by gender.
- **Slicers** for interactive filtering by genre, gender, and date range.

Analysis Highlights

- **Genre Preferences:** Notable variation in genre preference across gender.
- **Rating Trends:** Average ratings fluctuated across time, genres and the volume.
- **Top Movies:** Identified highest-rated and most-rated movies.
- **User Activity:** Highlighted the most active users based on the number of rated movies.

Tools Used

- Microsoft Power BI
- Power Query Editor
- DAX (Data Analysis Expressions)
- Data Modelling (Relationships, Date Table)

Outcome

- Developed an interactive Power BI dashboard.
- Cleaned and transformed MovieLens data for accurate analysis.
- Applied best practices in data modelling and visualization.
- Delivered meaningful insights about user behaviour and genre trends.