

MCU box office

Analysing data using SQL

By Abdurrahman Jakda

April 2022

Purpose

- The aim of this project is to find interesting information using data from all 27 released MCU film since April 2022. This will include focusing on individual movies and then the phases.

Dataset

- The data used for this project has come from David Dong via [Kaggle](#), ranging from 05/02/2008 to 17/12/2021.
- The dataset first starts from I have cleaned the data so it can be imported to and suitable for PostgreSQL.
- The columns in this dataset are: ID, movie title, MCU phase, release date, tomato meter, audience score, movie duration, production budget, opening weekend, domestic box office, worldwide box office.

Simple queries used

Stats	Highest grossing	Lowest grossing	Most expensive	Highest rated
Movie	Avengers: Endgame	The Incredible Hulk	Avengers: Endgame	Spiderman: No Way Home



Data on the different phases in the MCU

Phase	1	2	3	4
Movies per phase	6	6	11	5

Stats	Highest grossing	Lowest grossing	Most expensive	Highest rated
Phase	3	1	3	4



Overall trends and conclusions

- Phase 3 is the most successful phase with the movies it outperforming others in terms of gross and ratings.
- Avengers: Endgame is the leading factor to phase 3 success as it has brought in the most money but also the higher number of movies in this phase could be said to be a contributing factor.
- Phase 1 has the lowest rating but with phase 4 only having 5 movies out so far a fair conclusion cannot be made right now.