

CSC343 Project - Design Changes

1. Changes in Investigative Questions

- Remove the third investigative question that uses rating data from different rating websites

We made this change since we decided to only keep ratings from IMDb. We could not find a dataset that links movie ids that each movie rating website uses for the same movie. We also considered linking by name and other attributes, but it is risky to do it this way since we cannot affirm that those attributes can form a superkey.

2. Changes in Schemas and Specifications

- Remove Mdate from Movie schema

We removed Mdate since we will not study questions that would require a detailed date.

- Add usaRevenue to Finance schema; change type for figure related attributes in Finance schema

We added usaRevenue, so we can compare its USA market performance and global market performance.

We changed the type for budget, usaRevenue and revenue from int to bigint since we found out int may not be large enough to hold some of the values.

- Remove rating data from FilmTV and Rotten Tomatoes from Rating schema; Add IMDb movie average ratings from people in different age ranges to Rating schema

We removed rating data from websites other than IMDb for the reason we mentioned above.

We added average rating data for various age ranges, including 'all ages', 'under 18 years old', 'from 18 to 30 years old', 'from 30 to 45 years old', and 'above 45 years old' since we want to answer more detailed questions later like "what affects teen's rating?"

- Change primary key for Award schema

We changed the primary key to all the attributes since situations like the following may exist: in the same year and for the same award, a person did not win the award for one category, but won the award for another category.