

Chau (Pearl) Nguyen

Tug37553@temple.edu

Movie Lens data analyst

Movie Lens data contain all the rating of movie submitted by users. The information was contained by three separated tables. The first table contains all the information about the movies. The data field "movie_id" played as index on this table. With the users table, the data field "user_id" played as index on this table. The ratings table played as a join table. It's contain all the relationship between users and movies: all the movie rating that all the users had given from the history. This is exactly like a relational database.

One-third of the user's population, watched movies and rated are male. The male users are dominant on this data set. Therefore, we can't not conclude whether or not the certain occupations are male or female dominant. It's because the total survey is male dominant.

	user_id	age	sex	occupation	zip_code
291	292	35	F	programmer	94703
299	300	26	F	programmer	55106
351	352	37	F	programmer	55105
403	404	29	F	programmer	55108
420	421	38	F	programmer	55105
697	698	28	F	programmer	06906

t.

There are only 6 female programmers had given the rate. Which is proportionally less than male population, which is 60 male programmers. There were only 7 doctors gave out rating and none of them are female. We can only have concluded that more male programmers gave out movie ratings than female programmers. Only male doctors gave out movie ratings. However, this data set is small and don't have enough information about the date when this data set was recorded. The population of the people who in the data was unproportioned (from gender, occupation and age).