Data 607 Assignment Week 2

Leticia Salazar

9/5/2021

Assignment - SQL and R

Choose six recent popular movies. Ask at least five people that you know (friends, family, classmates, imaginary friends if necessary) to rate each of these movies that they have seen on a scale of 1 to 5. Take the results (observations) and store them in a SQL database of your choosing. Load the information from the SQL database into an R dataframe.

For this assignment I used google forms [https://forms.gle/oWoD8iMbheFF69rd9] to conduct my survey. I asked my friends to rate the following movies:

- LA LA LAND [2016]
- COCO [2017]
- BIRD BOX [2018]
- AVENGERS: ENDGAME [2019]
- WONDER WOMAN 1984 [2020]
- TOM AND JERRY [2021]

For each movie I asked them to rate them using the following scale:

- 0 - Never seen it
- 1 - Poor
- 2 - Fair
- 3 - OK/Good
- 4 - Very Good
- 5 - Excellent

```
# Install package
#install.packages("RPostgreSQL")
require("RPostgreSQL")
```

```
## Loading required package: RPostgreSQL
```

```
## Loading required package: DBI
```

```
#Start creating connection with PostgreSQL
movie_ratings<-dbConnect(dbDriver("PostgreSQL"), dbname="Movie_Ratingsdb", host="localhost", port=5433,
```

```
#Get the tables from connection
dbListTables(movie_ratings);
```

```
## [1] "movie_ratings"  "movie_ratings2"
```

```
select * from public.movie_ratings
```

Show first table for movie ratings responses

Table 1: 5 records

id	participants	la_la_land	coco	bird_box	avengers_endgame	wonder_woman_1984	tom_and_jerry
1	SG	0	5	3	5	3	0
2	GF	3	5	0	0	0	0
3	CC	2	5	2	4	0	5
4	KP	1	5	3	5	2	1
5	BO	0	5	1	5	4	5

```
select * from public.movie_ratings2
```

Show second table for average movie ratings along with rotten tomatoes score and audience score.

Table 2: 6 records

movie_id	movie_name	avg_rating	tomatometer	audicence_score
1	LA_LA_LAND	1	9	4
2	COCO	5	8	5
3	BIRD_BOX	2	6	3
4	AVENGERS_ENDGAME	4	8	5
5	WONDER_WOMAN_1984	2	6	4
6	TOM_AND_JERRY	2	5	4

```
dbDisconnect(movie_ratings)
```

```
## [1] TRUE
```

Conclusion: When conducting this short survey one issue that came up was that not all of my friends had seen the movies I selected. This was something I expected and why I chose to add the rating of "0" - never seen it. By doing so, the movie ratings wouldn't all appear as "1" - Poor when in reality the participants do not know what the movie is about. In terms of the ratings, "COCO" was one of the movies enjoyed by everyone and rated "5"" all throughout followed by Avengers: Endgame with an average rating of "3.8". I was able to get ratings from Rotten Tomatoes {https://www.rottentomatoes.com/}, their audience scores

and created movie_ratings2 in postgresql to compare the averages from my friends alongside theirs. Based on Rotten Tomatoes, the most highly rated movie was "La La Land" but the audience score rated "Avengers: Endgame" and "COCO" higher. When I received the responses for my survey, google forms created charts for two additional questions I created:

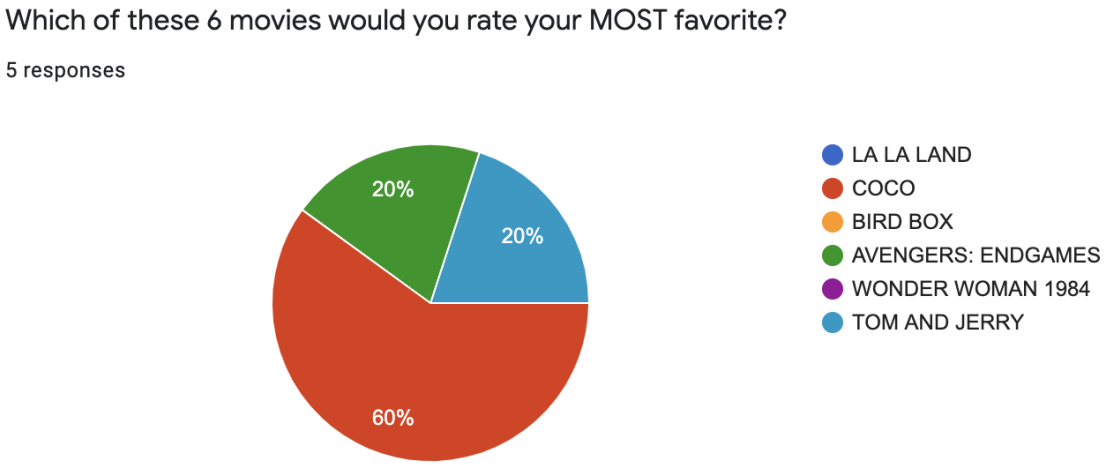

Figure 1: Most Favorite Movie

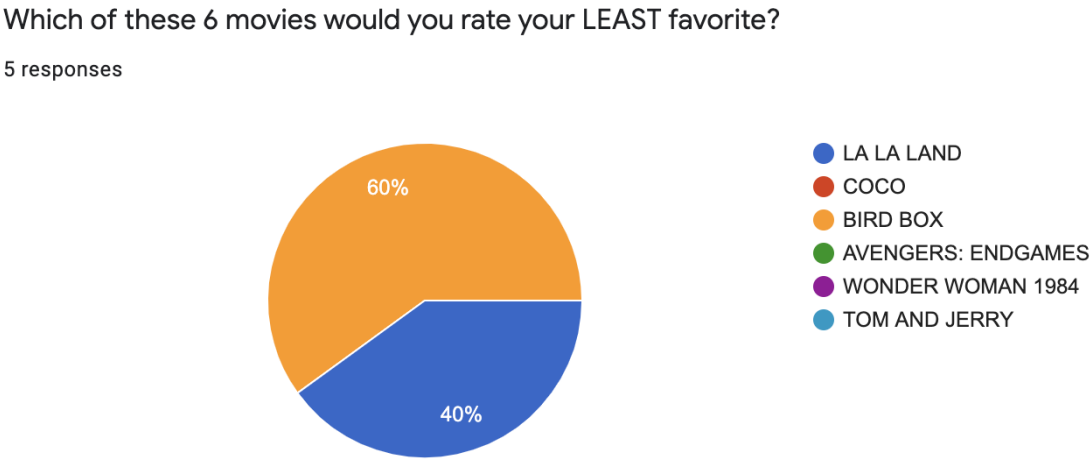

Figure 2: Least Favorite Movie