Ratemakers

(Making it rate)

Movie Consultation Services

Co-Founders: Olivia and Jill

Project Overview

Ratemakers

1. Defining **Business Case** and **hypotheses**
2. Finding **datasets of interest**
3. (Treating and cleaning the **data)**
4. (Mapping and querying the **data)**
5. **Obstacles**
6. Gathering **insights**

Defining Business Case and hypotheses

Business case:

- Movies are expensive enterprises that don’t always pay dividends. Therefore, our business helps clients determine which type of movies they should make.

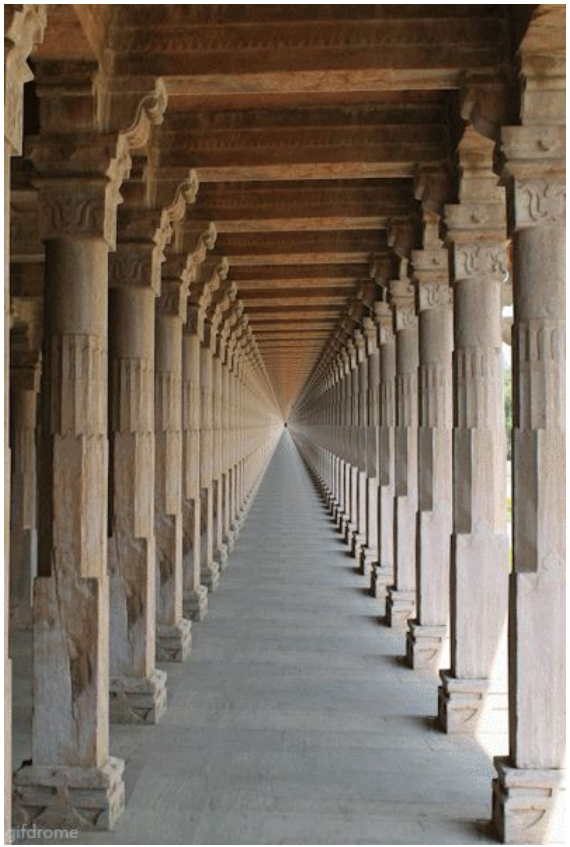
Hypothesis:

- Ratings are a predictor of gross profit
- Drama yields higher gross profits than comedy
- Budgets are directly correlated with gross profit

Finding datasets of interest

We were looking for datasets with:

- Movie name
- Movie genre
- Ratings
- Budget
- Profit



And we found:

- IMDB movie dataset
  - 'names', 'date_x', 'score', 'genre', 'overview', 'crew', 'orig_title', 'status', 'orig_lang', 'budget_x', 'revenue', 'country', 'index',
- Film Genre Statistics
  - 'Genre', 'Year', 'Movies Released', 'Gross', 'Tickets Sold', 'Inflation-Adjusted Gross', 'Top Movie', 'Top Movie Gross (That Year)','Top Movie Inflation-Adjusted Gross (That Year)'

Treating and cleaning the data

Merging the tables:

- We wanted a workable dataset that only included information about movies that were listed in both datasets

Dropping irrelevant columns:

- 'names', 'date_x', 'genre', 'overview', 'crew', 'orig_title', 'status', 'orig_lang', 'revenue', 'country', 'index', 'Movies Released', 'Top Movie Gross (That Year)', 'Top Movie Inflation-Adjusted Gross (That Year)'

Renaming, reordering, and formatting the columns:

- ‘top_movie’, 'year', 'genre', 'score', 'tickets_sold', 'budget', 'gross', 'inflation_adjusted_gross', 'unique_name'

Dropping rows with null values:

- There weren’t any!

Mapping and querying the data

Querying the data to find:

- The number of movies per genre
- The genre has the highest average rating
- The average rating of movies per genre
- The most popular genre per year
- The genres with the highest gross profit
- the budgets, and how they affected profits
- The correlation between tickets sold and gross profit
- The average number of tickets sold and gross profit for all movies per year
- The top 3 highest rated movies, their genre and year
- The 3 lowest rated movies, their genre and year
- The unique number of genres per year
- The average gross profit per genre with information on data points used to acquire the average
- count everything where score(ratings) is > 80 then sort and order by genre
- same as above but creating a column for multiple score thresholds

Obstacles

Asynchronous work, mainly:

- Immovable appointments and the Deutsche Bahn keeping team members apart resulted in:
  - Unilateral decisions about hypotheses, wrangling, etc.
  - Only one night when the team members were able to proximally, simultaneously work together

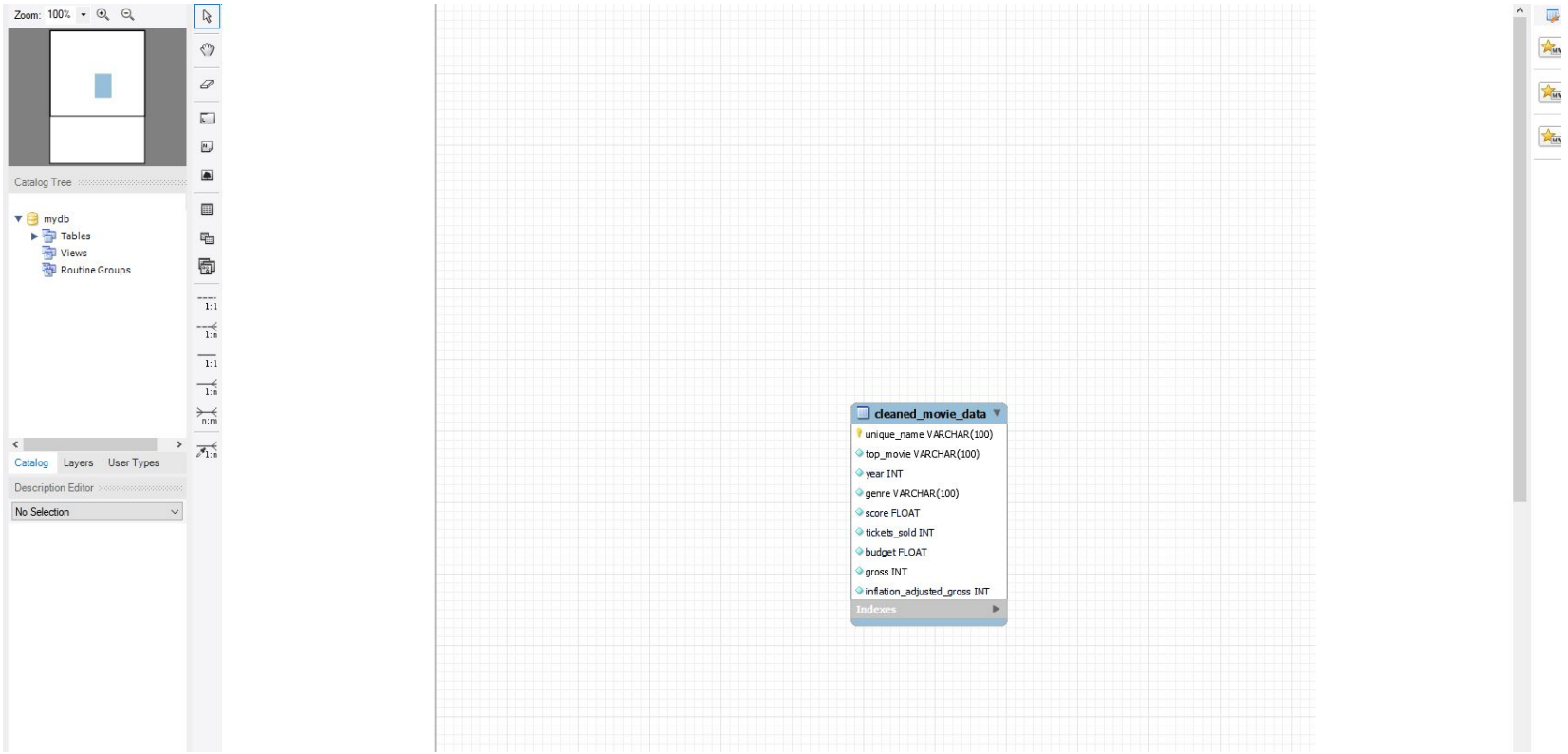
SQL being a pain to install:

- Made for a slow start and required playing catch up

Everything typical involved with figuring out how to code:

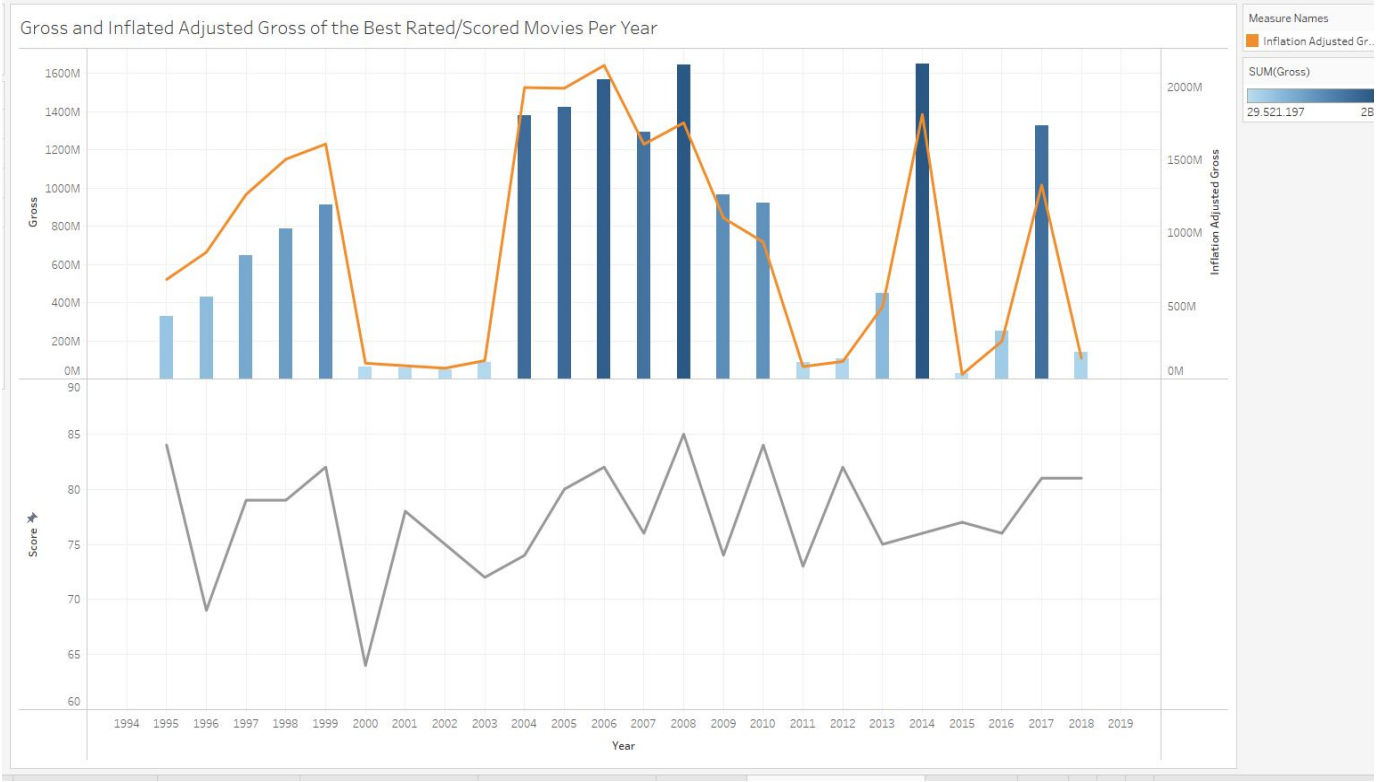
- Self-explanatory

Schema


Zoom: 100%
Catalog Tree
mydb
Tables
Views
Routine Groups
Catalog
Layers
User Types
Description Editor
No Selection
cleaned_movie_data
unique_name VARCHAR(100)
top_movie VARCHAR(100)
year INT
genre VARCHAR(100)
score FLOAT
tickets_sold INT
budget FLOAT
gross INT
inflation_adjusted_gross INT
Indexes

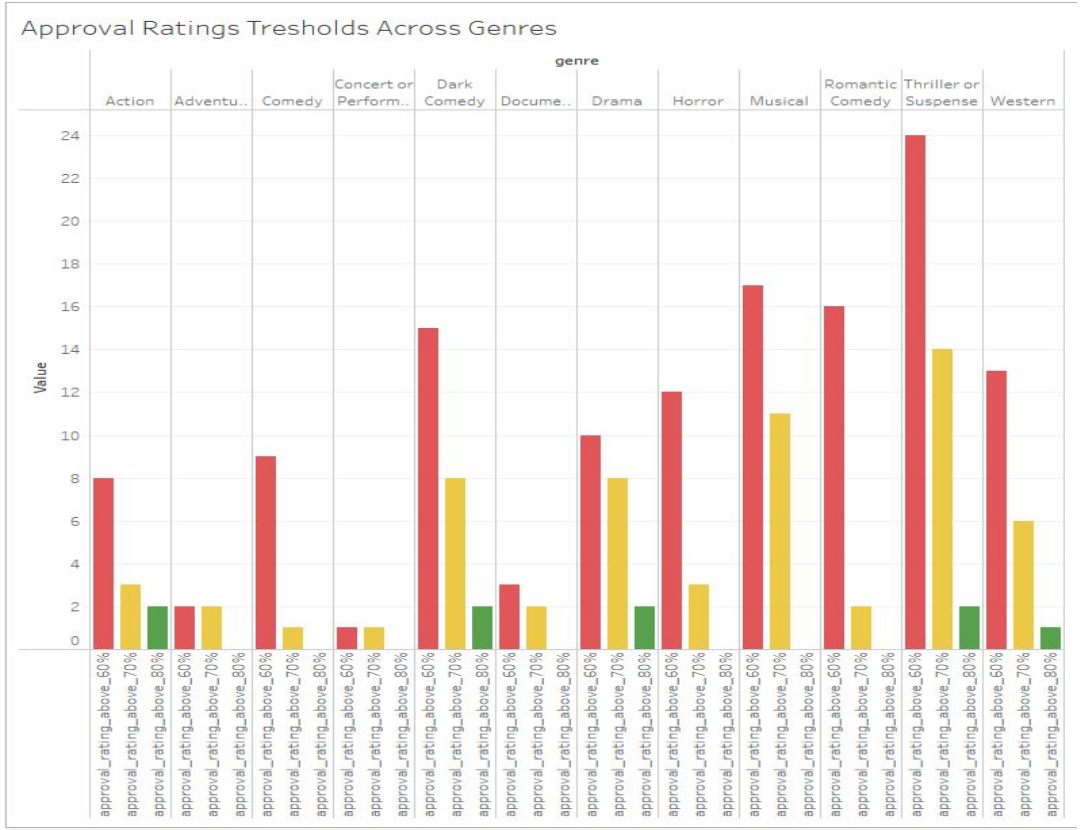
Gathering insights - **Gross and gross profit (adjusted for inflation) of the best rated movies PER YEAR**

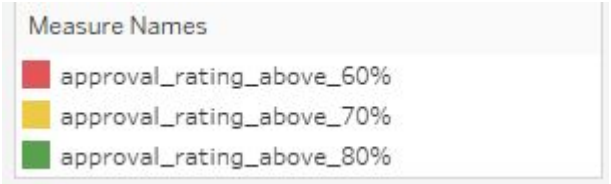
1. **Ratings are a predictor of gross profit**



Gathering insights - Approval rating thresholds by genre

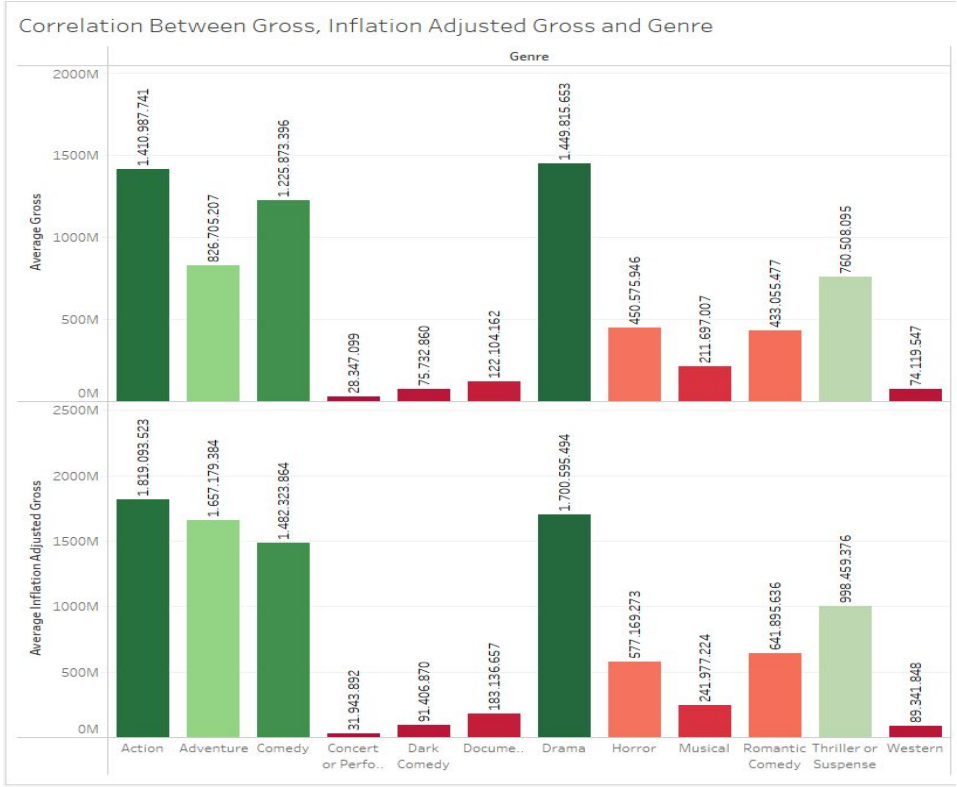
2. Drama yields higher gross profits than comedy

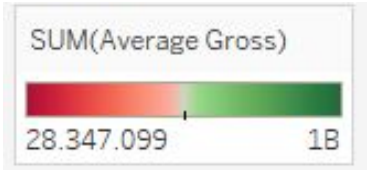




Gathering insights - **Profits by genre chart**

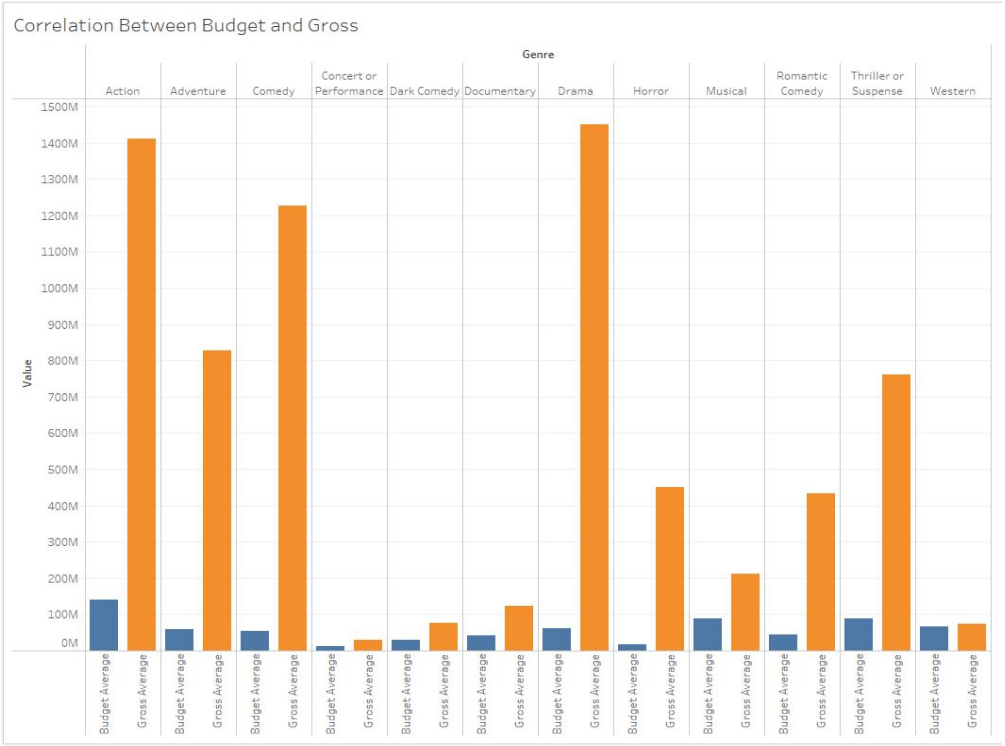
2. Drama yields higher gross profits than comedy

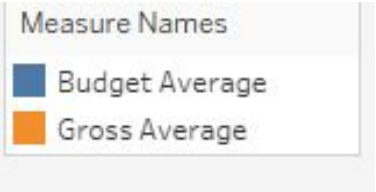




Gathering insights - **Correlation chart between budget and gross profit**

3. Budgets are directly correlated with gross profit





Summary

As per hypotheses:

- Ratings are a predictor of gross profit 🚫
- Drama yields higher gross profits than comedy ✅
- Budgets are directly correlated with gross profit 🚫

Thank you!

Olivia and Jill