Project Submission Requirements

- **A note specifying which dataset you analyzed**

I investigated TMDb movie data (cleaned from original data on Kaggle).

- **A statement of the question(s) you posed**

I posted 6 questions in the Jupyter Notebook and investigated data respectively. Most questions required bivariate analysis.

Questions:
I investigated the movie dataset. I would like to answer the questions below:

1. In which years the Comedy genre has been produced?
2. Is budget and profit correlated?
3. Do longer movies need more budget?
4. Do high profit movies have high popularity?
5. Do the movies' popularity increase year by year?
6. How do the length of movies change overtime?

- **A description of what you did to investigate those questions**

Firstly, I downloaded the data from the provided link, uploaded it to the Jupyter Notebook, imported required libraries for the investigation and analysis. Then I started to wrangle the data (assessed, cleaned). Then I saved the edited file. After that I started to look for the answers for my questions. I used univariate and bivariate analysis.

- **Documentation of any data wrangling you did**

I separated the data cleaning process to quality and tidiness issues and cleaned respectively.
Quality issues
- Erroneous data type in one variable (release_date)
- One duplicated record
- 0 record in several rows for revenue and budget variables
- Missing values/null values (cast(10790), homepage(2936), director(10822), tagline(8042), keywords(9373), overview(10862), genres(10843), production_companies(9836))
Tidiness issues
- Some variables are not needed for my investigation ('imdb_id', 'homepage', 'tagline', 'keywords', 'overview', 'budget_adj', 'revenue_adj')

- **Summary statistics and plots communicating your final results**

I conducted univariate and bivariate analysis.