The project report should clearly describe:

1. what data sources you used
2. why you chose those data sources
3. what target you chose (i.e., what is the intended use of the data, ...)
4. what difficulties you had to overcome to wrangle the data sources into the target data model
5. what techniques you did use
6. what you managed to achieve and what you failed to do

**Table of Contents**

**Abstract:**

This report describes an end to end flow for the Data Wrangling project. The primary objective is to identify interesting data sources to work on this project, to wrangle the data sources into a suitable target data model and to present the data to a larger audience.

**Data Sources:**

The data sources for this project is taken as IMDB and Rotten tomatoes.

**Why you chose those data sources:**

Data analysis for movie is an important way to analyse movie revenue and to understand user choice. Through data analysis the most popular genres, performance in recent years, earned revenue by the moves can be discovered. As any movie production involves huge cost and efforts, this an interesting topic to be analysed. The purpose is to perform data analysis on movie ratings by critics, the primary factors that influence a movie’s success that is measured by movie rating, budgets and revenue generated for last 10 years. Additionally, sentiment analysis is performed in order to understand the user review and sentiments accordingly. Our team created this database and obtained the information in this dataset from web scraping, using rvest package.

**What target you chose (i.e., what is the intended use of the data, ...):**

To perform the data analysis multiple approaches are taken. Firstly, data sources are identified.

**What difficulties you had to overcome to wrangle the data sources into the target data model**

Certain websites have anti-scraping policies. If someone overdo it, they will be blocked, and you will begin to see captchas instead of product details. Of course, you can also learn to work your way around the captchas using different services available. However, you do need to understand the legality of scraping data and whatever you are doing with the scraped data.

**What techniques you did use**

**What you managed to achieve and what you failed to do**