

The [movie dataset on which this case study is based](#) is a database of 5000 movies catalogued by [The Movie Database \(TMDb\)](#). The information available about each movie is its budget, revenue, rating, actors and actresses, etc. In this case study, we will use this dataset to determine whether any information about a movie can predict the total revenue of a movie. We will also attempt to predict whether a movie's revenue will exceed its budget.

In Part 1 of this case study, we will inspect, clean, and transform the data.

You can download the [Jupyter Notebook for Case Study 7, Part 1 at this link](#).