Hardly Executive, Hardly Summary!

Success is measured in a myriad of ways and in the movie business, a studio needs either a financial success or a critical success to stay in business. All studios, of course, want both.

While still an industry highly dependent on creative genius, the movie industry in the United States, often referred to as Hollywood, regularly turns to the collected data of the past before beginning a new film project. Sometimes, a formula is found and becomes the starting point of a new film. It might seem like the end of an art form but this is also entertainment, where tried and true elements such as a certain genre, director, and of course, actors, can provide the vehicle to a good show.

Can predicting a future film’s success, either financially or critically, be reduced to an algorithm or two? Why not? The Academy Awards started in 1929 with a few years of modern filmmaking preceding it. A century of film comes with a century of film data, with exceptions. Cinema managers were still calling in the numbers every night before computerized ticketing, and may still do so in some locations.

We are using two established datasets. One, from Mojo, covering a film’s metadata and sales, and Rotten Tomatoes, the popular movie ratings site. Following extraction of each csv, we transformed the datasets into dataframes with the information we needed to explore the nature of the business and, using machine learning, build an application to determine if the elements of a prospective film will be successful with audiences.