Machine Learning

We first train-test split to split our dataset into train dataset, which is used to train our predicting model. The test dataset is used to test out our model and see how accurately it can predict the values of the response variables, which are ratings, revenue, and profit.

For our model of choice, we will mainly be using Bayesian Model.

So why Bayesian model instead of a linear regression or other models?

The main reason is that Bayesian analysis is usually more accurate with smaller samples. Due to the limitation of our API keys for some of our datasets, we were not able to extract large amounts of data and had to work with a smaller dataset. Additionally, Bayesian Models are more flexible and can handle more complex models.

The difference between Bayesian Model and other frequentist models such as linear regression is in the Bayesian view, a probability is assigned to a hypothesis as the degree of belief on the occurrence of an event, while in the frequentist view, a hypothesis is tested without being assigned a probability.

With our limited dataset size due to the limits of our API keys, we figured that using Bayesian Model would be more suited to us.

Firstly, we wanted to see if a single variable alone could accurately predict our 3 response variables. However, the results of the univariate models were not very promising, so we decided to run a multi-variate Bayesian model.

We decided to run multi-variate Bayesian Models as well to see if these models would help with our results.