# Summary

Along with bitcoin and ostrich farms, movies are among the riskiest places to invest money. What if we could reduce risk by making better decisions to improve the probability of generating higher revenue? How correlated are genre, director, casting, and other dimensions with revenue? Is the viewer rating of the movie correlated with revenue? Taking into account competition, what combination of actor, genre, and director will likely produce the greatest revenue?

# Data

|  |  |  |
| --- | --- | --- |
| Data Set | Description | Purpose |
| [Kaggle “The Movies” Dataset](https://www.kaggle.com/rounakbanik/the-movies-dataset) | The main Movies Metadata file. Contains information on 45,000 movies featured in the Full MovieLens dataset. Features include posters, backdrops, budget, revenue, release dates, languages, production countries and companies.  Cast and Crew Information for all our movies. Available in the form of a stringified JSON Object. | Provides the following elements:   1. Movie title 2. Revenues 3. IMDB ID 4. Cast & crew |
| [IMDB](http://www.imdb.com/) | Joined with IMDB ID from Kaggle data set:  The IMDb rating and votes information for titles, weighted average of all the individual user ratings, number of votes the title has received | Provides the following elements:   1. IMDB rating 2. Weighted average rating 3. Number of votes received 4. Birth year of actor (to determine if actor is still available) |
| [CPI (Consumer Price Index)](https://fred.stlouisfed.org/series/CUUR0000SA0R) | Consumer price index since 1892 | Provides the following element to adjust revenue data for inflation:   1. CPI Index |



# Design Mock-Up

The scatter plot control will allow the user to select x and y dimensions to dynamically produce the above charts. Additionally, a dynamic bar charting control will allow the user to select dimensions for the histograms. The point is to allow the user to combine different dimensions to home in finding the right combination of genre, actor, director, quality etc. to derive the greatest revenue.

The mock-up of the site will look something like the following with dynamic controls which allow selection and filtering of dimensions.

