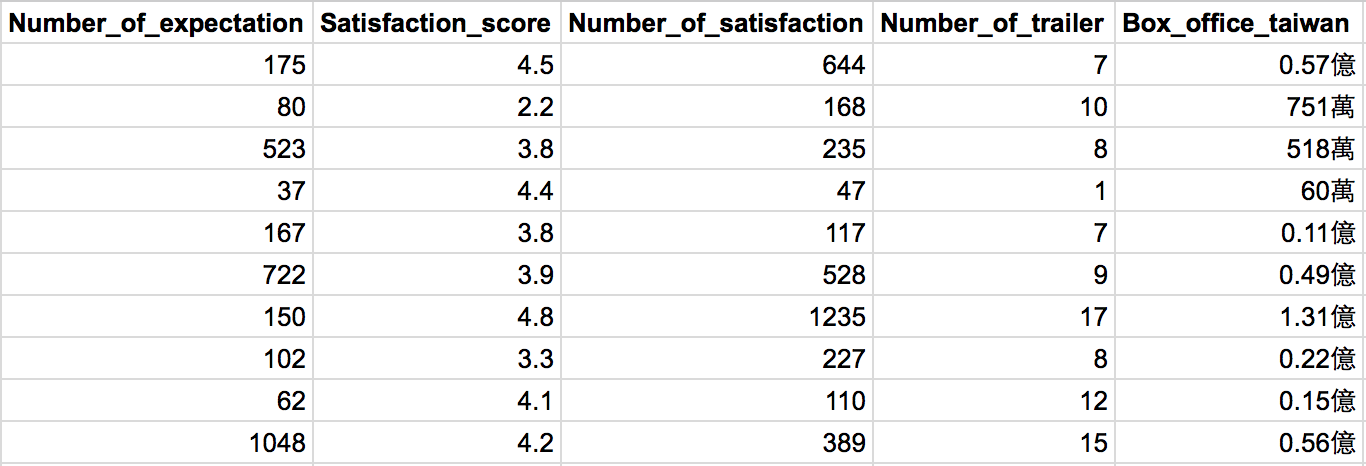
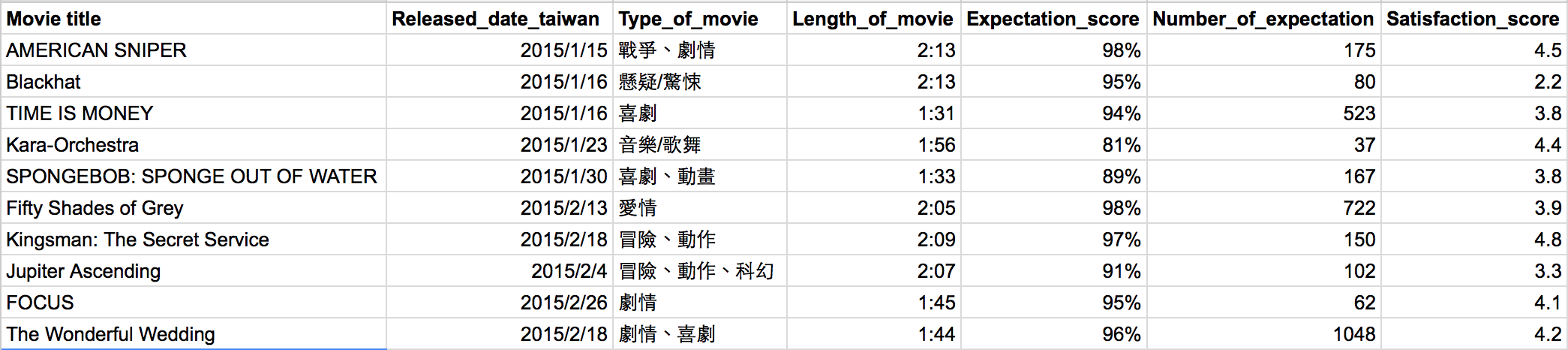
* **Suitable Title:** Box office revenue prediction in China based on Taiwan box office revenue to maximize their profits
* **Business Goal and Humanistic Evaluation**
* Who is the stakeholder or client? Movie theaters in China are our clients.
* Who will be affected by the solution? Film distribution corporations, theaters in China, Chinese customers of the theaters, advertisement companies.
* A description of the business goal. The business benefits relate to the theaters maximizing their profits by meeting the demands of their customers more efficiently. i.e. showing the movies that the customers want to see, arranging more times per day and seats. Customers will develop some form of loyalty to specific theaters when their demands are efficiently met.
* What opportunity is it creating? (a) To meet the demands of Chinese customers of movie theaters. (b) Advertisement companies can get more cases.
  + What shortcoming does it address? (a) We may encounter Sampling error, because some Taiwanese preference may different from Chinese. (b) We can’t predict movies that released at the same time or before Taiwan.
  + What are possible dangers of this application? Illegal company will copy the movies which have high revenue in Taiwan and sell it to customers before the movies are released in China.
  + Who might be harmed and how? Movie theaters, because customers may buy more copies instead of going to theaters.
  + What would be considered a success? Predicting box office revenue accurately so movie theaters can better prepare the logistics of showing movies in advance.
* **Analytics/Data Mining Goal**
* A description of the analytics objective. (1) Whether there is a strong relationship between Taiwanese box office revenue and China box office revenue. (2) To build the model using the Taiwan’s box office revenue to predict China’s box office revenue.
* Is this a supervised or unsupervised task? Is it predictive or descriptive? Is it retrospective or forward-looking? It is a supervised task, predictive and forward-looking.
* What is the main outcome variable(s) of interest? Box office revenue in China.
* **Data (Brief description of the available data.)**
* Some guidance on the data subset that will be used and the re-processing or preparation that might be needed based on your past experience. (1) Change categorical variables to numerical variables. (2) Handle missing data (3) Choose only the movies that released earlier in Taiwan than in China.
* Sample of ten rows (records) with ten columns (variables) that will be used.



And we also have released date in China and box office revenue in China.

* **Methods**
* What are some data mining methods to consider? (1) Visualization of box office revenue in China and box office revenue in Taiwan (2) Linear regression (3) KNN (4) Regression tree (5) C4.5 (J48)
* Which performance measures are appropriate? How do they map to the business goal? (a)To compare the error between box office revenue in Taiwan and box office revenue in China. So we can use RMSE, MAE, MAD. (b)Movie theaters will get more accurately predicted box office revenue so they can get higher income by doing better preparation and budgets arrangement in proper ways.
* **Implementation/Production**
* Operational requirements and/or constraints (for example, will the solution be run in real-time? will it require collecting new data? will it be a one-time analysis or ongoing?) We need to collect new data and it is a one-time analysis.