**1. Airplane Crashes Since 1908 - Full history of airplane crashes throughout the world, from 1908-present**

A complete data set of all the Airplane Crashes since 1908. It includes information such as date, time, location, flight type, operator, fatalities etc. Also includes a Summary describing the event in brief. Potential data users include passengers and frequent fliers. Passengers can use this to find and analyze trends and find locations that are prone to crashes. It can also generate trends to analyze the safest and the least safe airliners and routes etc.

**URL:** <https://www.kaggle.com/saurograndi/airplane-crashes-since-1908>  
<https://opendata.socrata.com/profile/Aiden/aide-nnnn>  
  
**License:** Open Database License (<http://opendatacommons.org/licenses/odbl/1.0/>)

**2. IMDB 5000 Movie Dataset - 5000+ movies’ data scraped from IMDB website**

This dataset contains the data of 5000+ movies scraped from IMDB.com. There are more than 2000 unique director names, 30,000 actors including data such as duration, aspect ratio, genres etc. Chuan San at NYC Data Science has used it build models predicting the IMDB rating of the movies using the parameters from the dataset. This dataset can prove valuable to movie studios to find the most bankable actors, directors etc. Also the actors, screenwriters and directors who make movies that are most likely to yield a higher score i.e. critical and popular acclaim. Also it can view the trends using the historical data to predict the box-office success of future titles.

**URL:** <https://blog.nycdatascience.com/student-works/machine-learning/movie-rating-prediction/><https://www.kaggle.com/deepmatrix/imdb-5000-movie-dataset>

**License:** Open Database License (<http://opendatacommons.org/licenses/odbl/1.0/>)

**3. Y Combinator Companies - Publicly launched YC companies funded from summer 2005 to summer 2016**

This data set contains the vertical, year of launch, and descriptions for publicly launched YC companies between the summer of 2005 and the summer of 2016. It includes very popular companies like reddit etc.

Potential data users include investors, entrepreneurs and crowd-sourcing companies. They can use analyses from this dataset to find patterns and trends that are predictive of success or failure of a startup. Even economists can use this data. The data contains companies categorized by industry. This can help economists analyze the relative growth rates of startups from different industries and other interesting information pertaining to startups.

**URL:** <https://www.kaggle.com/benhamner/y-combinator-companies>

**License:** Open Database License (<http://opendatacommons.org/licenses/odbl/1.0/>)

**Bibliography:**

Aiden. (2009, June 24). Airplane crashes and fatalities since 1908. Retrieved September 14, 2016, from Socrata, <https://opendata.socrata.com/Government/Airplane-Crashes-and-Fatalities-Since-1908/q2te-8cvq>

Sun, C., & Academy, N. D. S. (2016, August 22). NYC data science academy. Retrieved September 14, 2016, from NYC Data Science, <https://blog.nycdatascience.com/student-works/machine-learning/movie-rating-prediction/>

Hammer, B. (2016). *Y Combinator companies - Kaggle*. Retrieved September 14, 2016, from Kaggle, <https://www.kaggle.com/benhamner/y-combinator-companies>