Proposal 1: Insight into startups and VC communities (data from Crunchbase)

The data set from crunchbase contains dataset for startups and venture capital communities. Data also contains funding, investment and acquisition data on over 40k companies

Questions to ask/answer

* What industries are hot?
* Are experienced founders better?
* Do VCs have an age bias?
* Are there characteristics of a company (industry, location) that differ by VC
* Do some VCs invest togetyher while others rarely do so?
* Are companies raising more money earier?
* Are we in a bubble?

Source: <https://blog.modeanalytics.com/five-public-dataset/>

Data Sets:

[crunchbase.acquisitions](https://modeanalytics.com/crunchbase/tables/acquisitions)

[crunchbase.companies](https://modeanalytics.com/crunchbase/tables/companies)

[crunchbase.investments](https://modeanalytics.com/crunchbase/tables/investments)

[crunchbase.rounds](https://modeanalytics.com/crunchbase/tables/rounds).

Proposal 2: House Price Prediction (Ongoing Kaggle)

With 79 explanatory variables describing (almost) every aspect of residential homes in Ames, Iowa, this competition challenges you to predict the final price of each home.

Source: <https://www.kaggle.com/c/house-prices-advanced-regression-techniques>

Data Sets @ <https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data>

* train.csv - the training set
* test.csv - the test set

Proposal 3: Integer Sequence Learning (Kaggle)

This competition challenges you create a machine learning algorithm capable of guessing the next number in an integer sequence. For example, next number in the sequence 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55 is 89.

Intrigued by the possibility of solving a puzzle via machine learning algorithms

Source: <https://www.kaggle.com/c/integer-sequence-learning>

Data Sets @ <https://www.kaggle.com/c/integer-sequence-learning/data>

* train.csv - the training set, contains full sequences
* test.csv - the test set, missing the last number in each sequence