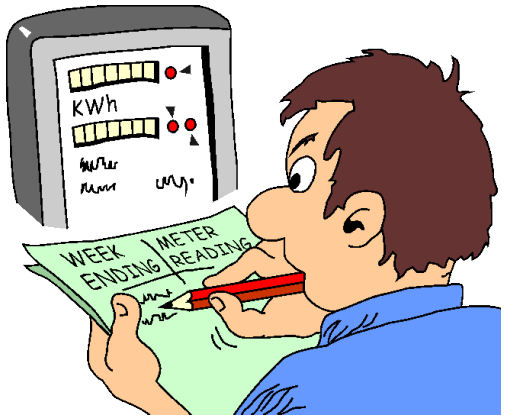
Predicting IPO returns

- Goal of Model: Predicting first day returns for an IPO using public information prior to offer
- Purpose: To use this model to pick IPOs to invest in as an Individual investor
- Task type: Supervised Learning
  - Prediction

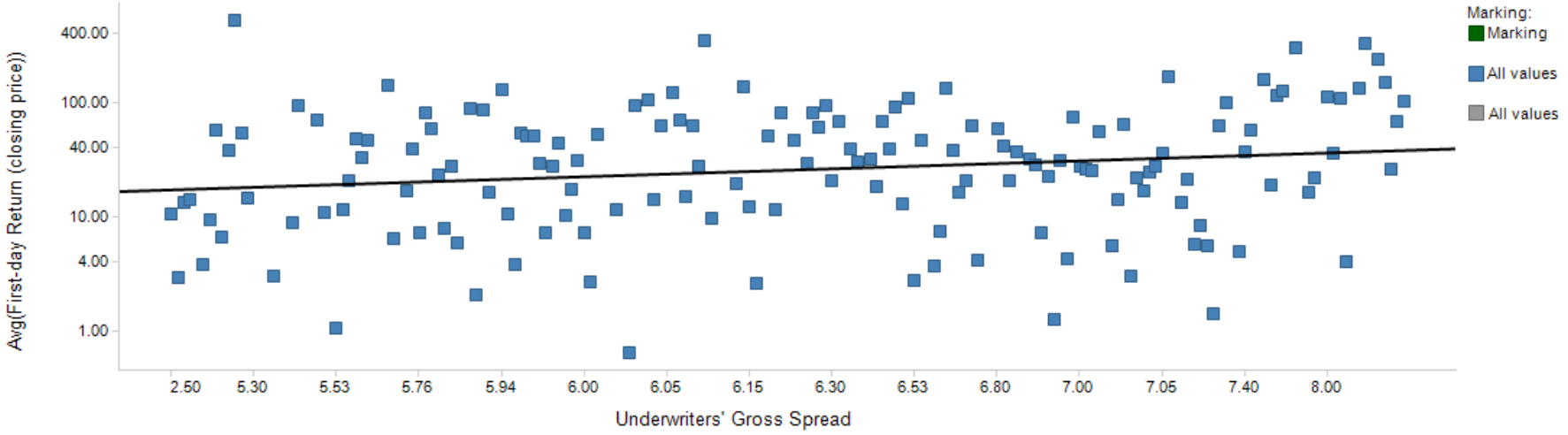
Data Collection

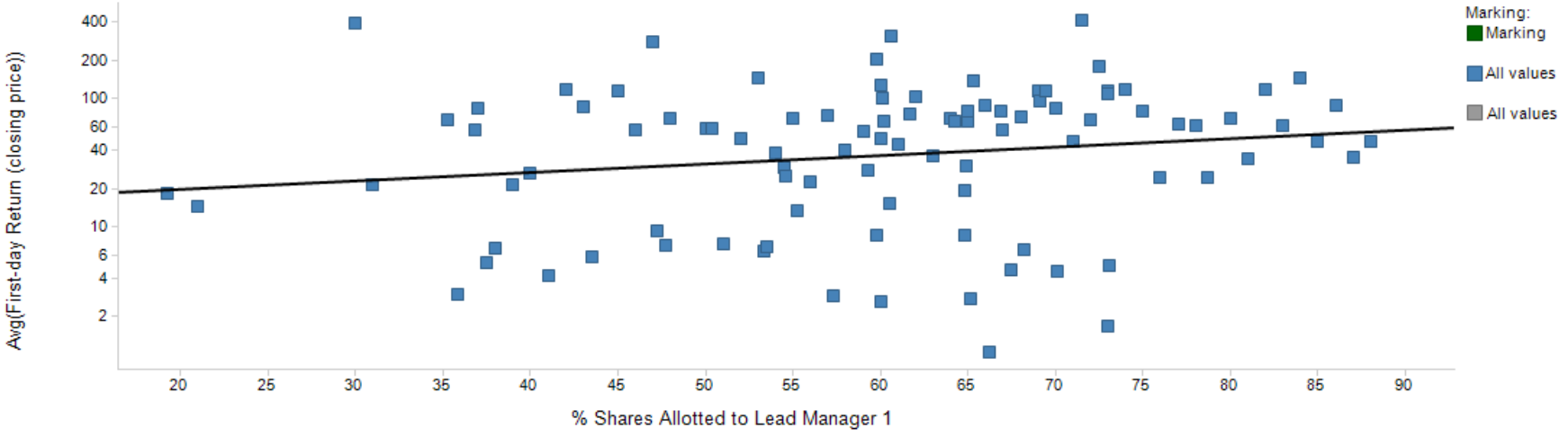


- Japanese IPO data from 1997-2009*
- 1561 IPOs
- Data Pre-processing
  - Industry(categorical) : 35 industries
    - 3 were spelling errors, corrected
    - Remove Air Trans (1), Fishery & Forestry (2) industries
  - Removed first 128 entries (1997-1999) as they had no data for 2 columns : Underwriter's fees & Allocation to BRLM
  - New Columns
    - Minimum bid size
    - Secondary Offering %age
  - Creation of Dummy Variables
    - BRLMs – 3, on the basis of Gross proceeds of IPO
    - Industry – 4, binned by average return
    - Market – whether the IPO was OTC or not

*Source: Kaneko and Pettway's Japanese IPO Database (KP-JIPO) : http://www.fbc.keio.ac.jp/~kaneko/KP-JIPO/top.htm

Exploration





Methods Used

- Linear Regression

Validation Data scoring - Summary Report

	Total sum of squared errors	RMS Error	Average Error
	3541081.888	90.8530139	-9.49800911

Test Data scoring - Summary Report

	Total sum of squared errors	RMS Error	Average Error
	3463687.293	110.049056	0.282298672

- K-Nearest Neighbours

Validation Data scoring - Summary Report (for k=5)

Total sum of squared errors	RMS Error	Average Error
4017396.95	96.77066527	-12.9477102

Test Data scoring - Summary Report (for k=5)

Total sum of squared errors	RMS Error	Average Error
3779848.996	114.961973	-1.04298592

- Regression Trees

Total Sum of Squares	2909471
Mean Squared Error	9063.774
RMSE	95.20385
Average Error	-4.82212

Conclusion

- Prediction algorithms do not give a reasonable prediction of IPO returns from public information.
  - High RMSE: 90%+ in all 3 methods
- Not Surprising, since if this was so easily possible using public data, then people would already have made money off it.