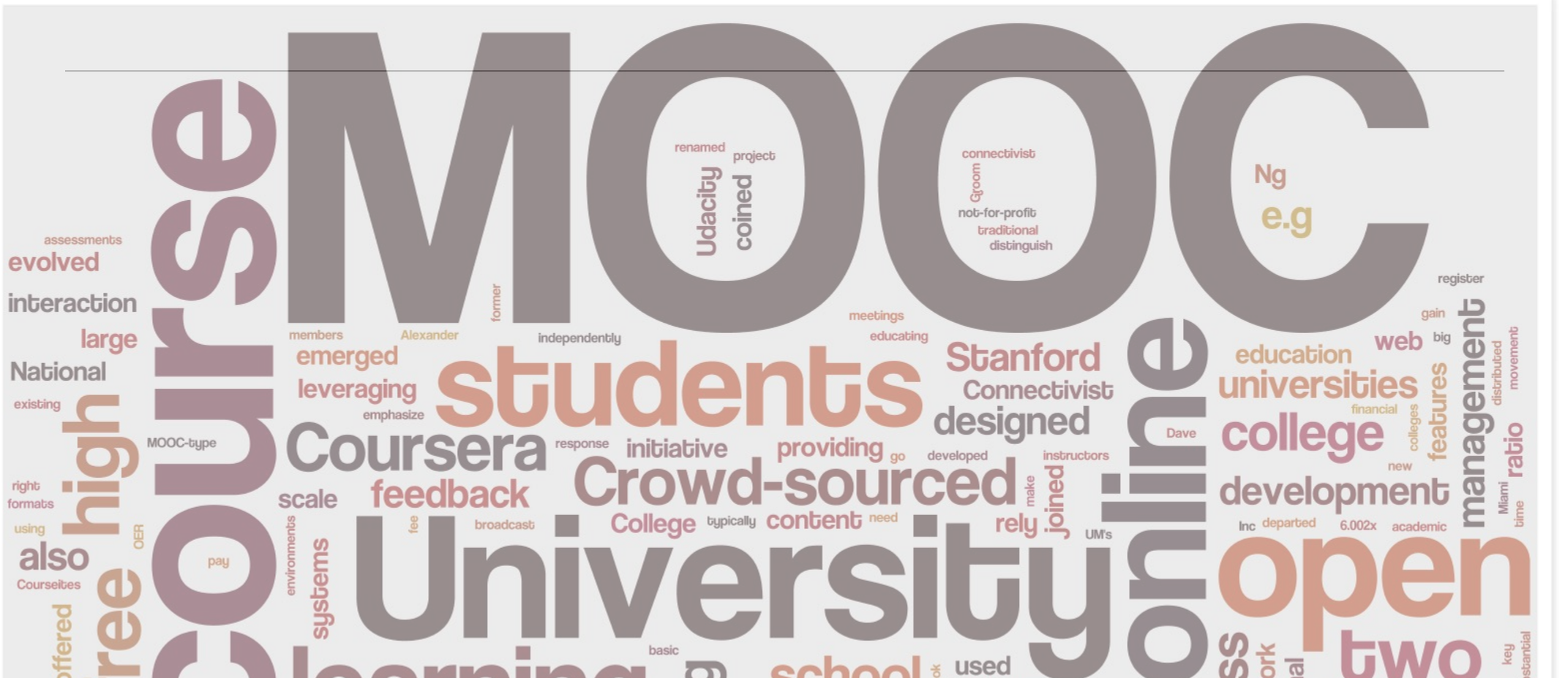
What drives up enrollment in MOOCs?

A multivariate linear regression analysis

Paru Jayaprakash | METIS SF 2018



JISC WEBINAR
moderated by David Kernohan
@dkernohan

#JiscWebinar

WHAT IS A MOOC?

NO ONE MOOC - MANY MANY VARIETIES (just like Learners)

Martin Weller HISTORY

@opencontent David Wiley

George Siemens
Stephen Downes
Dave Cormier

#DS106
Jim Groom
Martha Burtis
Alan Levine

benefits
- low cost
- employee recognition

Concerns
- not so open
- us-centric disciplines?

Stanford
EdX
Udacity

Lou McGill LEARNER PERSPECTIVE

- AUTHENTIC
- BUILD COMMUNITY
- portfolio
- #DS106
- #CHANGE11
- SUBGROUPS
- SUPPORT
- CHALLENGES
- CULTURAL LANGUAGE
- EXPERIMENT
- ACCESS TO HIGH PROFILE ACADEMICS + INSTITUTIONS
- SELF REGULATION
- MOTIVATIONS

@loumcgill

EXPERIENCED LEARNERS LOW RETENTION

INSPIRATIONAL

Jonathan Worth EXPERIENCE

INSTRUCTOR

CURATE a journey £Context

#picbod

CONNECT EXPERTISE

CONNECT WITH FABRIC of INDUSTRY

@jonathan_worth

DAVID WHITE CREDIBILITY

BRAND

See Me

certificate MASTERY

ACADEMIA OR Learning?

@daveowhite

OPEN

NOT ALWAYS INCLUSIVE

GOOD TEACHING STILL IMPORTANT!

NETWORK LEARNING FOCUS

REQUIRES DIGITAL LITERACY

into familiar SPACES

play learning social

COMFORT

FACILITATE

@giuliaforsythe July 2012

Source of Information

Scraping Level 1Features

Machine Learning A-Z™: Hands-On Python & R In Data Science

BEST SELLER | 285 lectures • 41 hours • All Levels

Learn to create **Machine Learning** Algorithms in Python and R from two Data Science experts. Code templates included. | By Kirill Eremenko

$10.99
~~$199.99~~
4.4
(48,991 ratings)

Python for Data Science and Machine Learning Bootcamp

143 lectures • 21.5 hours • All Levels

Learn how to use NumPy, Pandas, Seaborn , Matplotlib , Plotly , Scikit-Learn , **Machine Learning**, Tensorflow , and more! | By Jose Portilla

$10.99
~~$194.99~~
4.5
(24,322 ratings)

Data Science, Deep Learning, & Machine Learning with Python

BEST SELLER 90 lectures • 12 hours • Beginner

Go hands-on with the neural network, artificial intelligence, and **machine learning** techniques employers are seeking! | By Sundog Education by Frank Kane

$10.99
~~$159.99~~
4.4
(9,499 ratings)

Features: Title, No of Lectures, No of Hours, Student Level, Sale Price, Real Price, Stars, Ratings

Scraping Level 2 Feature

Gift This Course

Machine Learning A-Z™: Hands-On Python & R In Data Science

Learn to create Machine Learning Algorithms in Python and R from two Data Science experts. Code templates included.

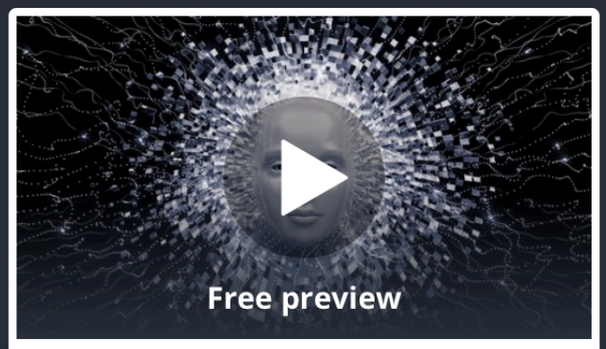
BEST SELLER ★★★★½ 4.4 (48,991 ratings) 278,156 students enrolled

Created by Kirill Eremenko, Hadelin de Ponteves, SuperDataScience Team, SuperDataScience Support

Last updated 6/2018

English CC English [Auto-generated], Italian [Auto-generated], 4 more

Enrollment



$10.99 ~~$199.99~~

94% off

1 day left at this price!

Add To Cart

Buy Now

30-Day Money-Back Guarantee

Includes:

- 41 hours on-demand video
- 24 Articles
- 2 Supplemental Resources
- Full lifetime access
- Access on mobile and TV
- Certificate of Completion

What Will I Learn?

- Master Machine Learning on Python & R
- Make accurate predictions
- Make robust Machine Learning models
- Use Machine Learning for personal purpose
- Have a great intuition of many Machine Learning models
- Make powerful analysis
- Create strong added value to your business
- Handle specific topics like Reinforcement Learning, NLP and Deep Learning

Incentives

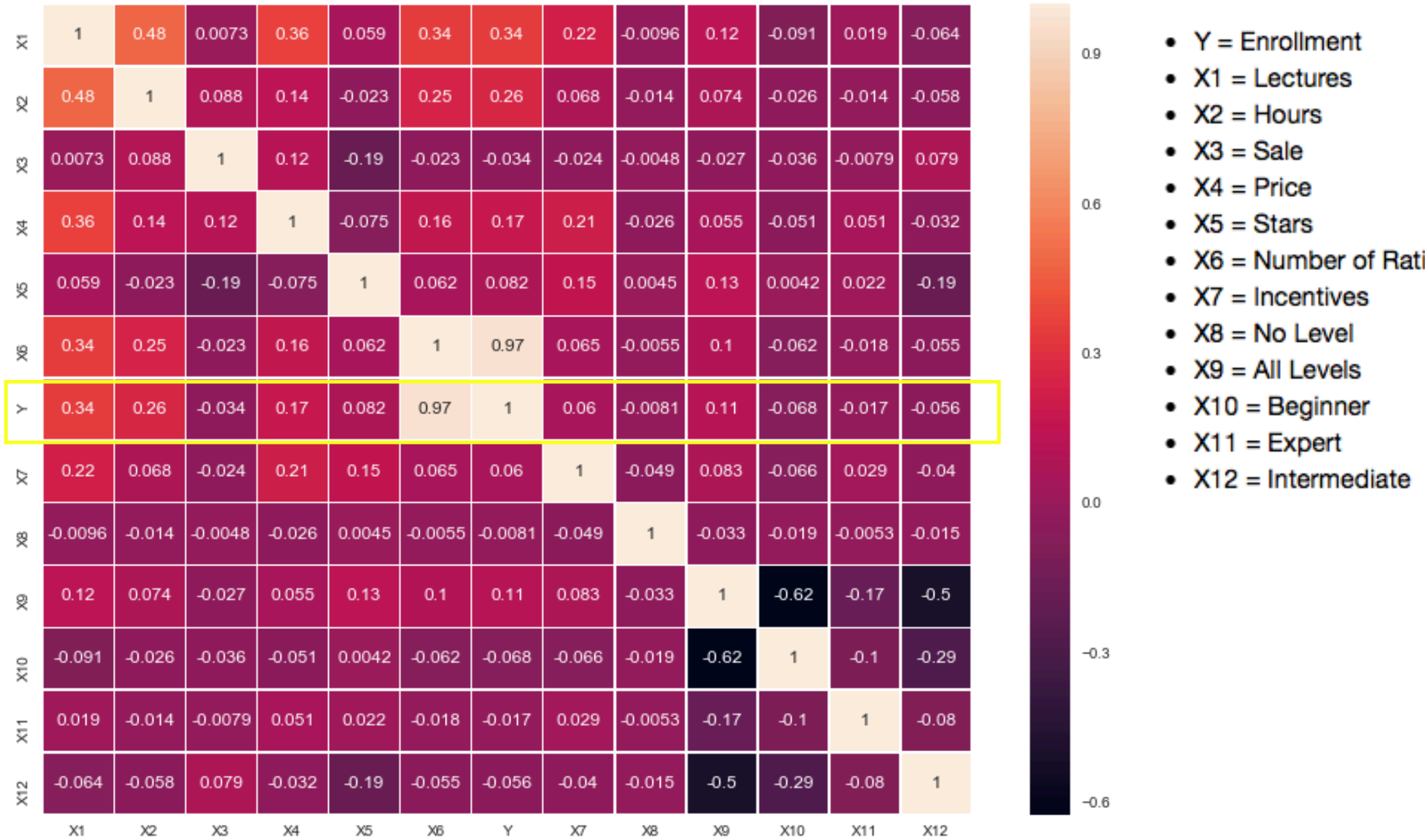
Scraping Summary

No of pages scraped	160
Records	1286
Used records	986
Features	12

Minimization metric

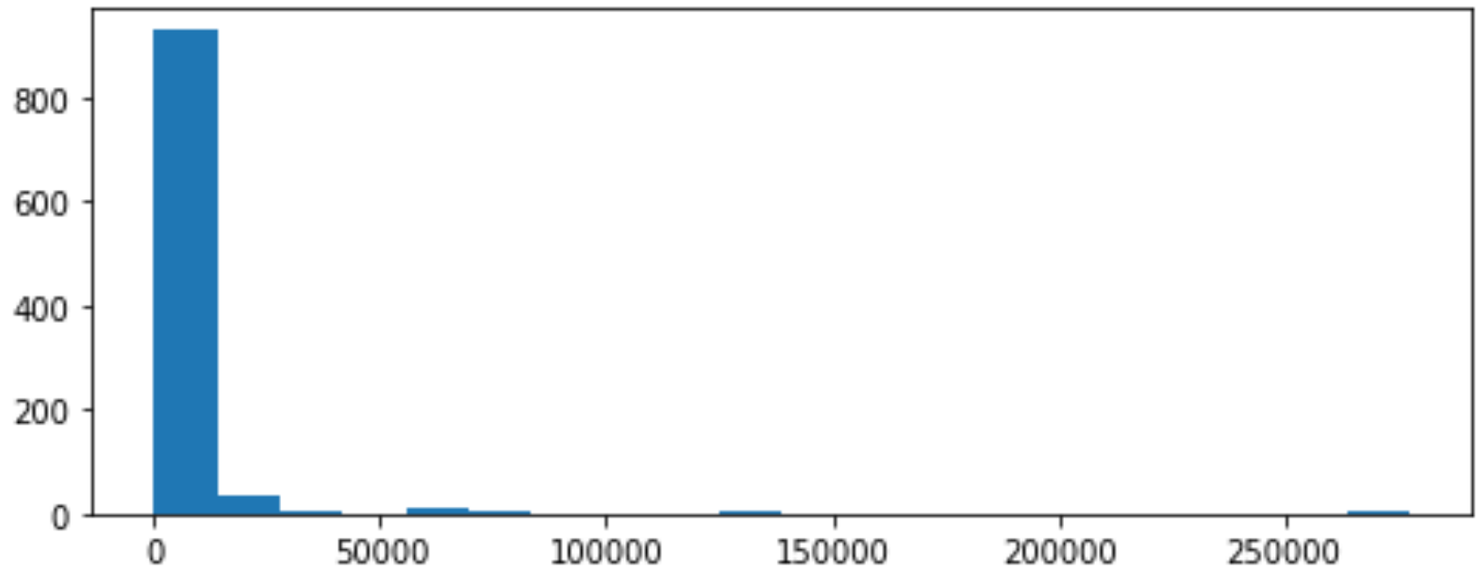
Mean squared error	$\mathrm{MSE} = \frac{1}{n}\sum_{t=1}^{n} e_t^2$
Root mean squared error	$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2}$
Mean absolute error	$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n} \lvert e_t \rvert$
Mean absolute percentage error	$\mathrm{MAPE} = \frac{100\%}{n}\sum_{t=1}^{n} \left\lvert \frac{e_t}{y_t} \right\rvert$

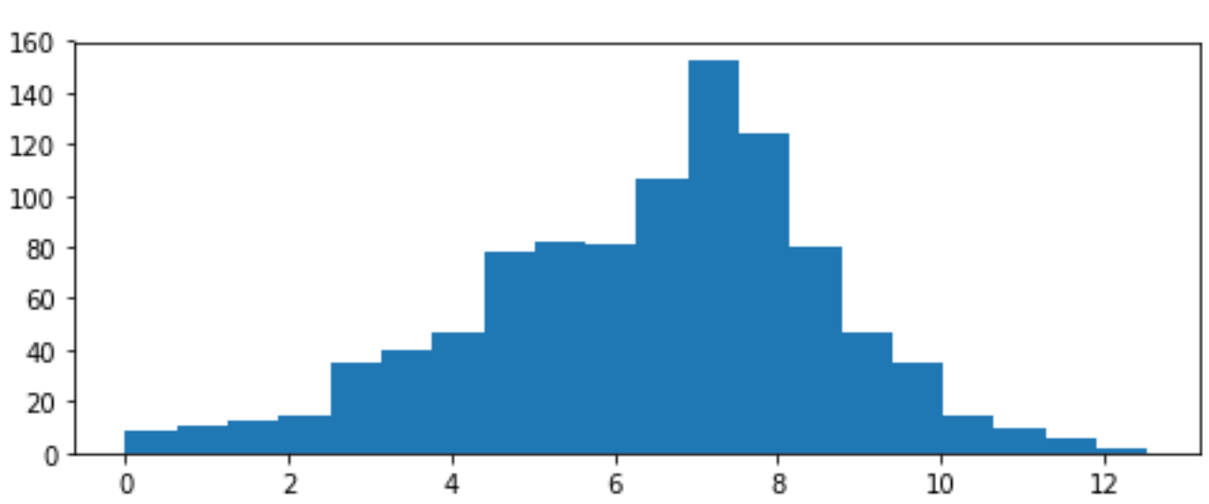
Exploratory Data Analysis



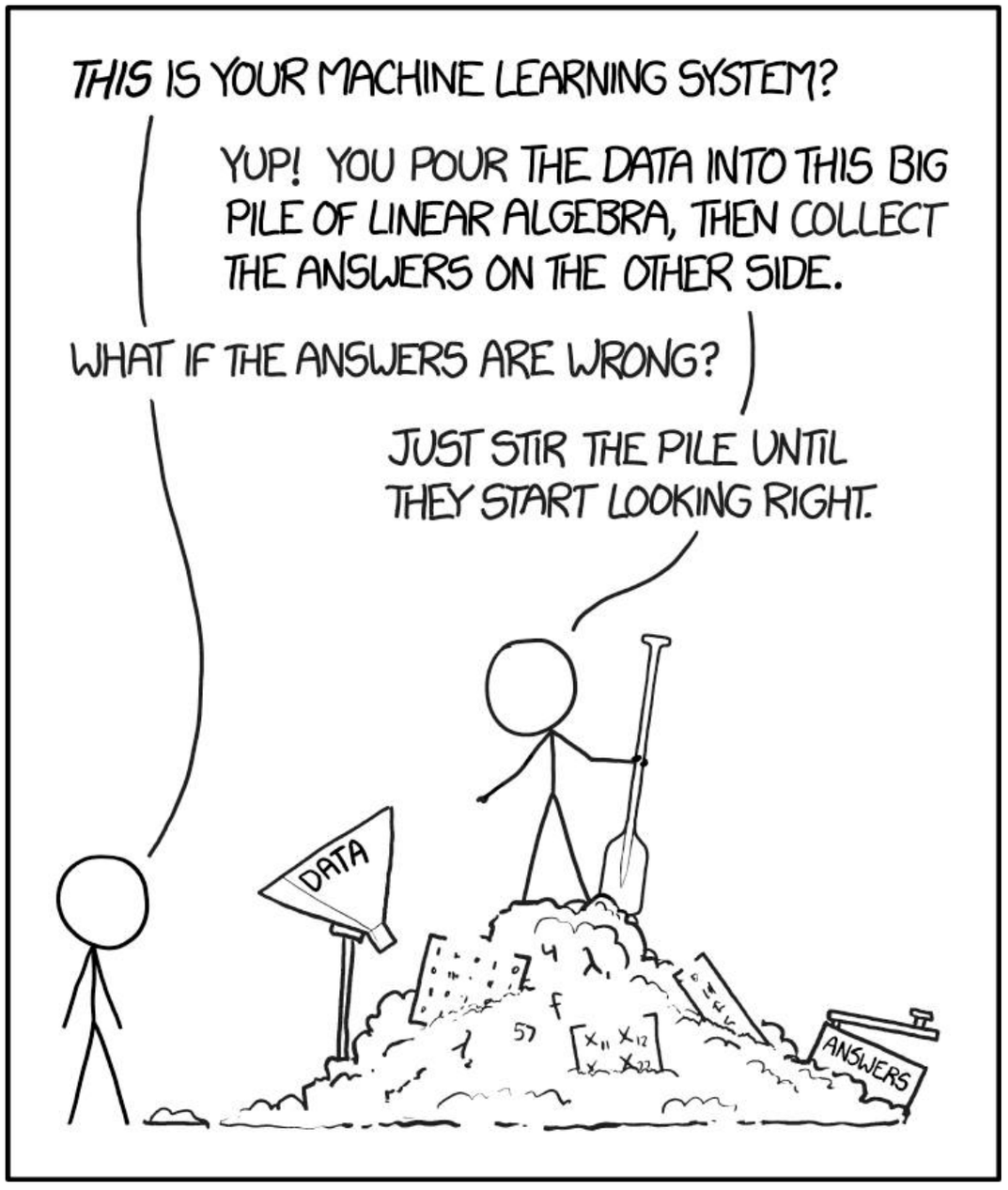
- Y = Enrollment
- X1 = Lectures
- X2 = Hours
- X3 = Sale
- X4 = Price
- X5 = Stars
- X6 = Number of Ratings
- X7 = Incentives
- X8 = No Level
- X9 = All Levels
- X10 = Beginner
- X11 = Expert
- X12 = Intermediate

A peak into the histogram of Enrollment data made me realize a log transformation was necessary





Using LASSO method to choose the most relevant features



The strongest features

Sale Price of Course
Total Price
Number of Stars
Ratings

Courses designed for all levels
Courses designed for Intermediate level
Courses designed for Expert level
Courses designed for Beginner level

Future Work

1. Tokenize course titles and descriptions to add more features
2. Scrape more data