What Makes for a Popular Book Review?

Michael Burnam-Fink
Metis Project Luther

Business Case

- Popular book reviews are engaging

goodreads

METIS

Business Case

- Engagement on Goodreads.com impacts book sales on Amazon

amazon

METIS

Tools

Scraping

Selenium

Beautifulsoup

Modelling

SM

StatsModels

scikit learn

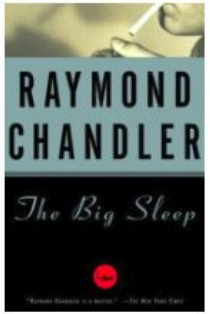
Visualizing

matplotlib

seaborn

Data Source

Michael Burnam-Fink's Reviews > The Big Sleep



The Big Sleep (Philip Marlowe, #1)
by Raymond Chandler

Michael Burnam-Fink's review — May 05, 2012 · edit

★★★★★

bookshelves: mystery, fiction, 2012

7 highlights (Private)

What can you say about "The Big Sleep" that hasn't already been said? This is a classic of hard boiled noir. The language is as glamorous as the dames, as punchy as private-eye Marlowe, and as dark as the souls of the criminals, dissolute rich, and corrupt cops who inhabit the world. If you haven't read it, you're missing out.

1 like

READING PROGRESS

- Started Reading Add a date
- May 5, 2012 – Shelved
- May 5, 2012 – Shelved as: mystery
- May 5, 2012 – Shelved as: fiction
- May 5, 2012 – Shelved as: 2012
- May 5, 2012 – Finished Reading

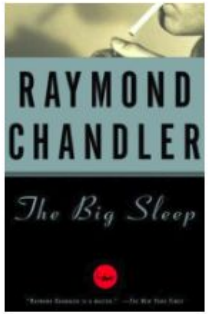
METIS

Data Source

- Word Count
- Image Count
- Book info
- Rating

- User Friends
- Average Rating
- Total Times Reviewed

Michael Burnam-Fink's Reviews > The Big Sleep



The Big Sleep (Philip Marlowe, #1)
by Raymond Chandler

Michael Burnam-Fink's review — May 05, 2012 · edit

★★★★★

bookshelves: mystery, fiction, 2012

7 highlights (Private)

What can you say about "The Big Sleep" that hasn't already been said? This is a classic of hard boiled noir. The language is as glamorous as the dames, as punchy as private-eye Marlowe, and as dark as the souls of the criminals, dissolute rich, and corrupt cops who inhabit the world. If you haven't read it, you're missing out.

1 like

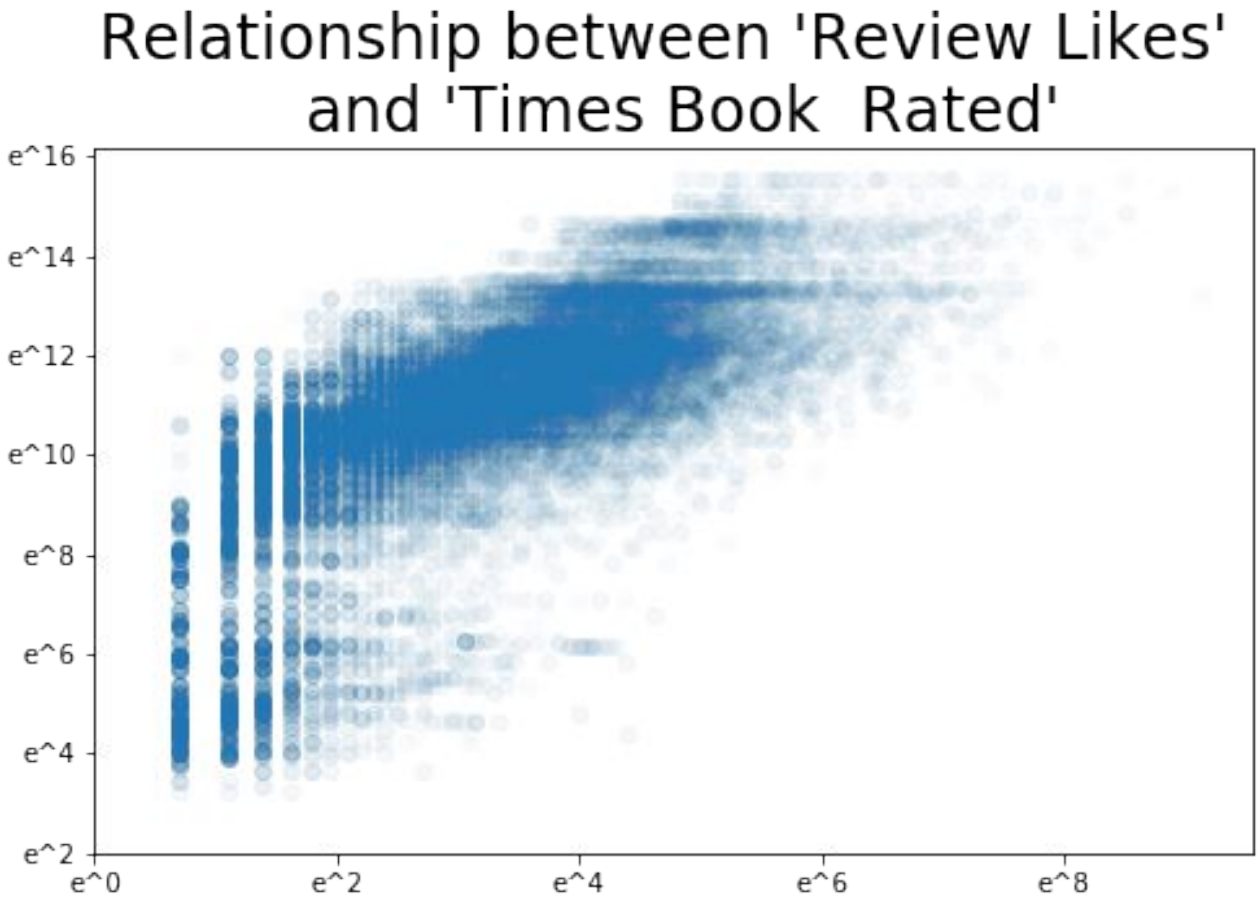
Target

READING PROGRESS

- Started Reading Add a date
- May 5, 2012 – Shelved
- May 5, 2012 – Shelved as: mystery
- May 5, 2012 – Shelved as: fiction
- May 5, 2012 – Shelved as: 2012
- May 5, 2012 – Finished Reading

METIS

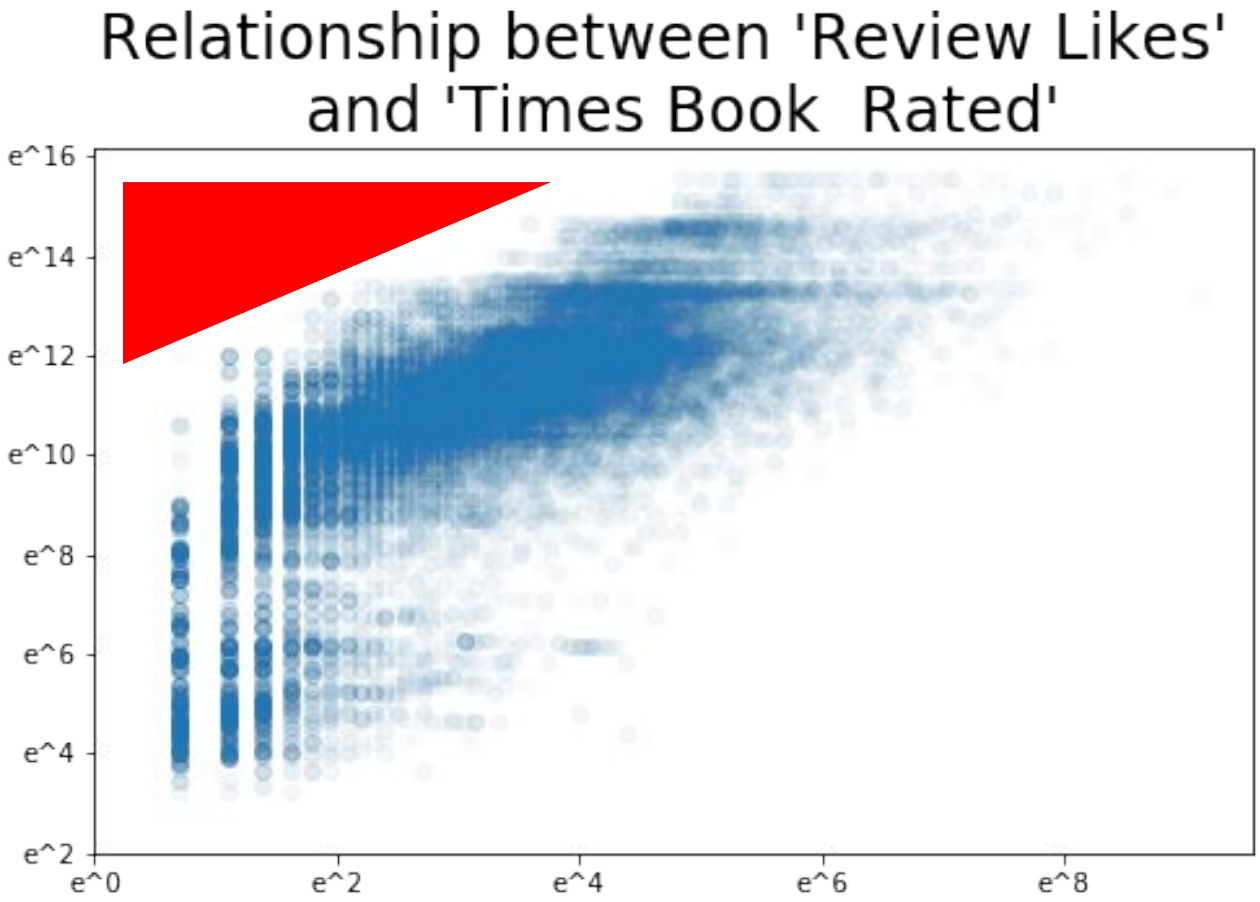
EDA


Relationship between 'Review Likes' and 'Times Book Rated'
e^16
e^14
e^12
e^10
e^8
e^6
e^4
e^2
e^0
e^2
e^4
e^6
e^8

METIS

EDA

THERE SHOULD BE DATA HERE

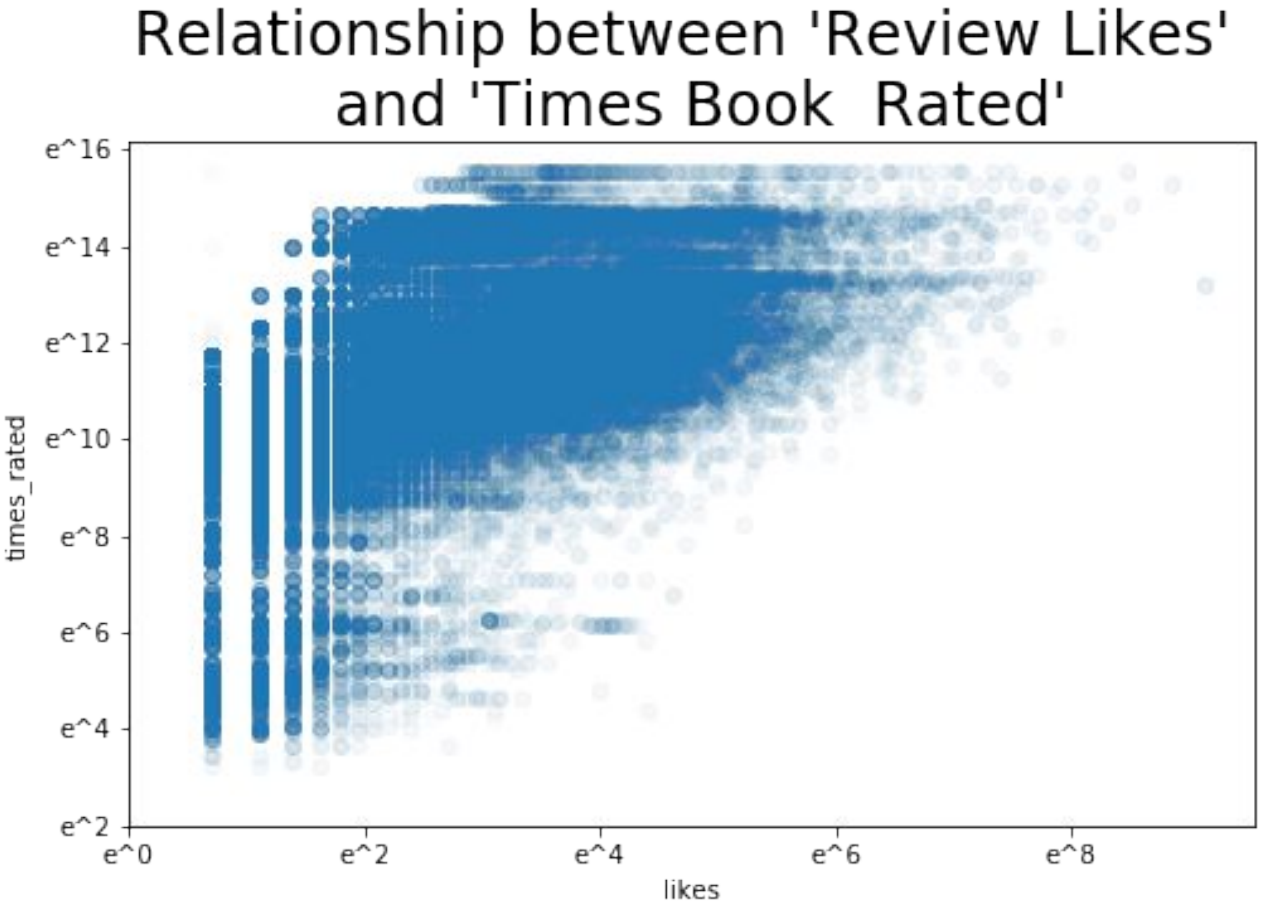


METIS

METIS

EDA

75,000 records
200 MB of data



METIS

Code: Multiprocessing

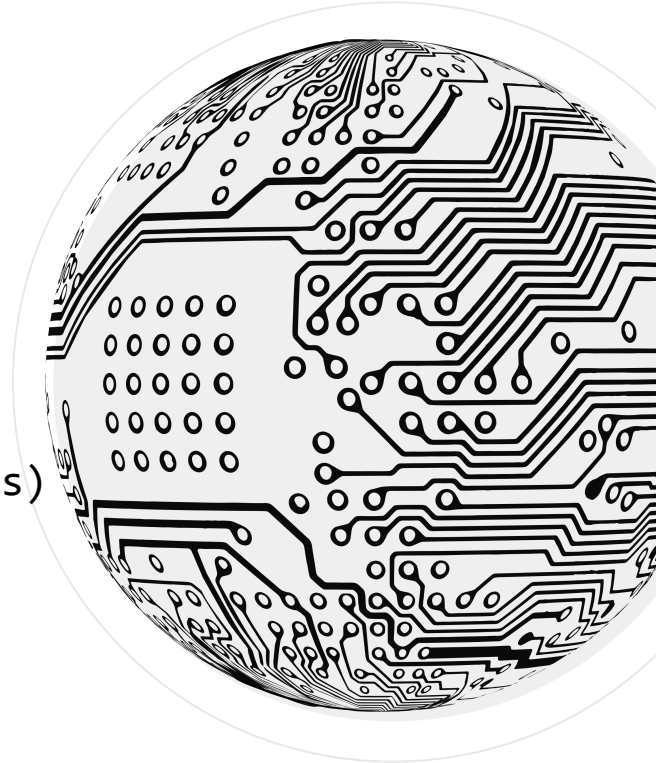
```
import multiprocessing

def reviewscraper(url)
review_urls = [url_1, url_2... url_300]

pool = multiprocessing.Pool()
reviews = pool.map(review_scraper, review_urls)

df = pd.DataFrame.from_records(reviews)

outfile = open(‘title book.pkl’,'wb')
pickle.dump(df,outfile)
```



METIS

First Linear Regression

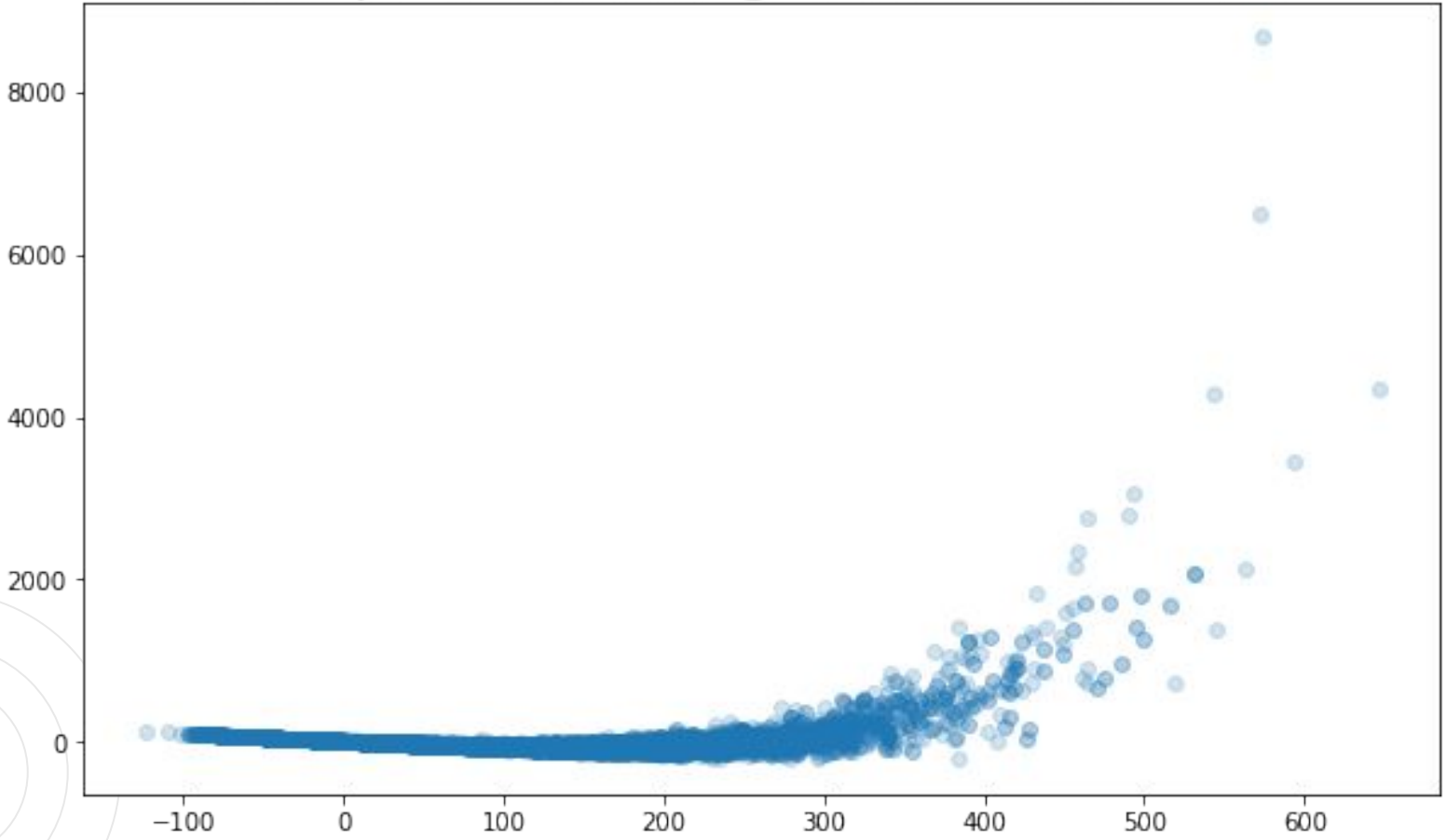
Dep. Variable:	y	**R-squared:**	0.256
Model:	OLS	**Adj. R-squared:**	0.255
Method:	Least Squares	**F-statistic:**	1226.
Date:	Thu, 11 Oct 2018	**Prob (F-statistic):**	0.00
Time:	14:28:31	**Log-Likelihood:**	-3.1845e+05
No. Observations:	53576	**AIC:**	6.369e+05
Df Residuals:	53560	**BIC:**	6.371e+05
Df Model:	15		
Covariance Type:	nonrobust		

R^2 : 0.256

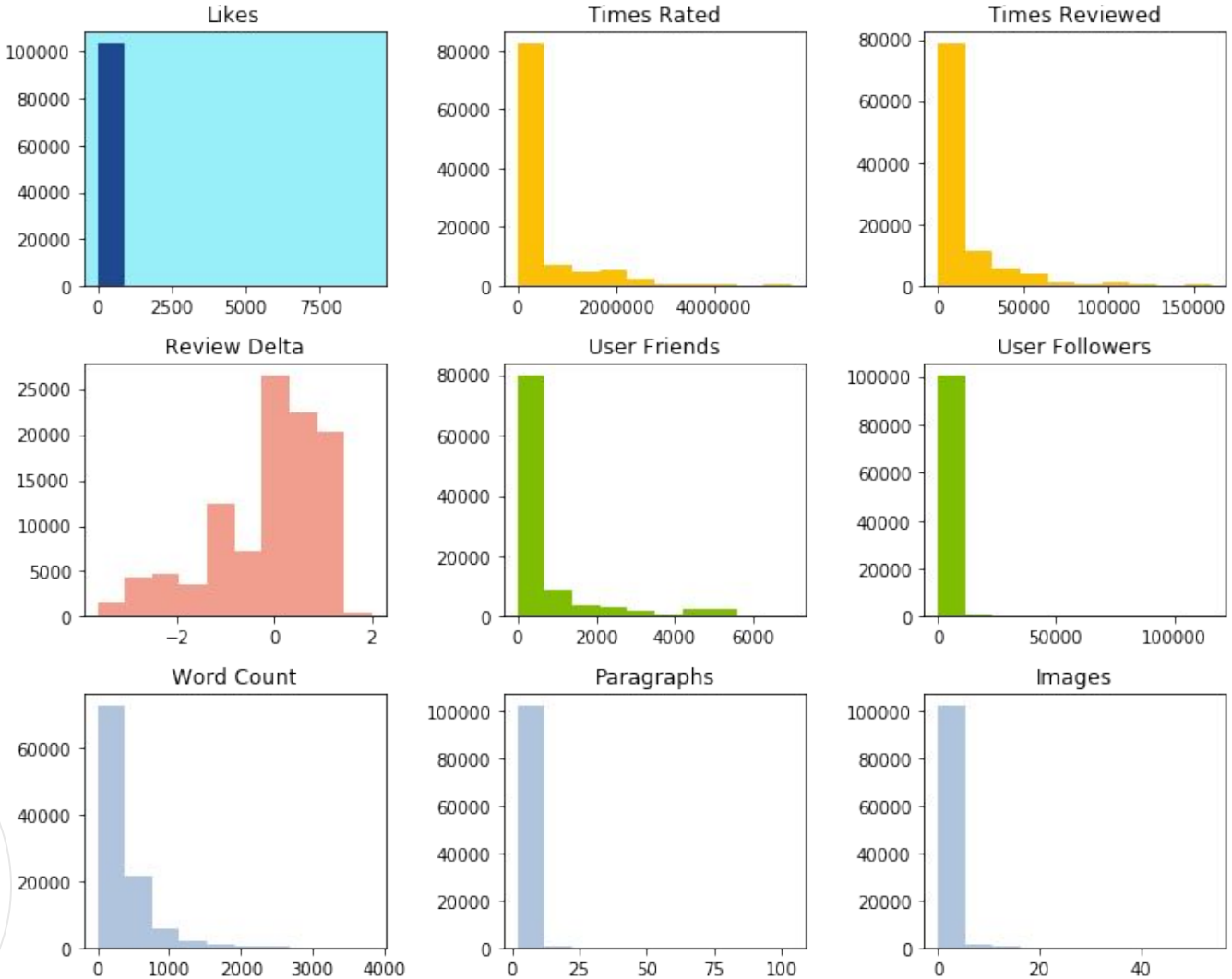
MSE: 5,883

METIS

Simple Linear Regression Residuals

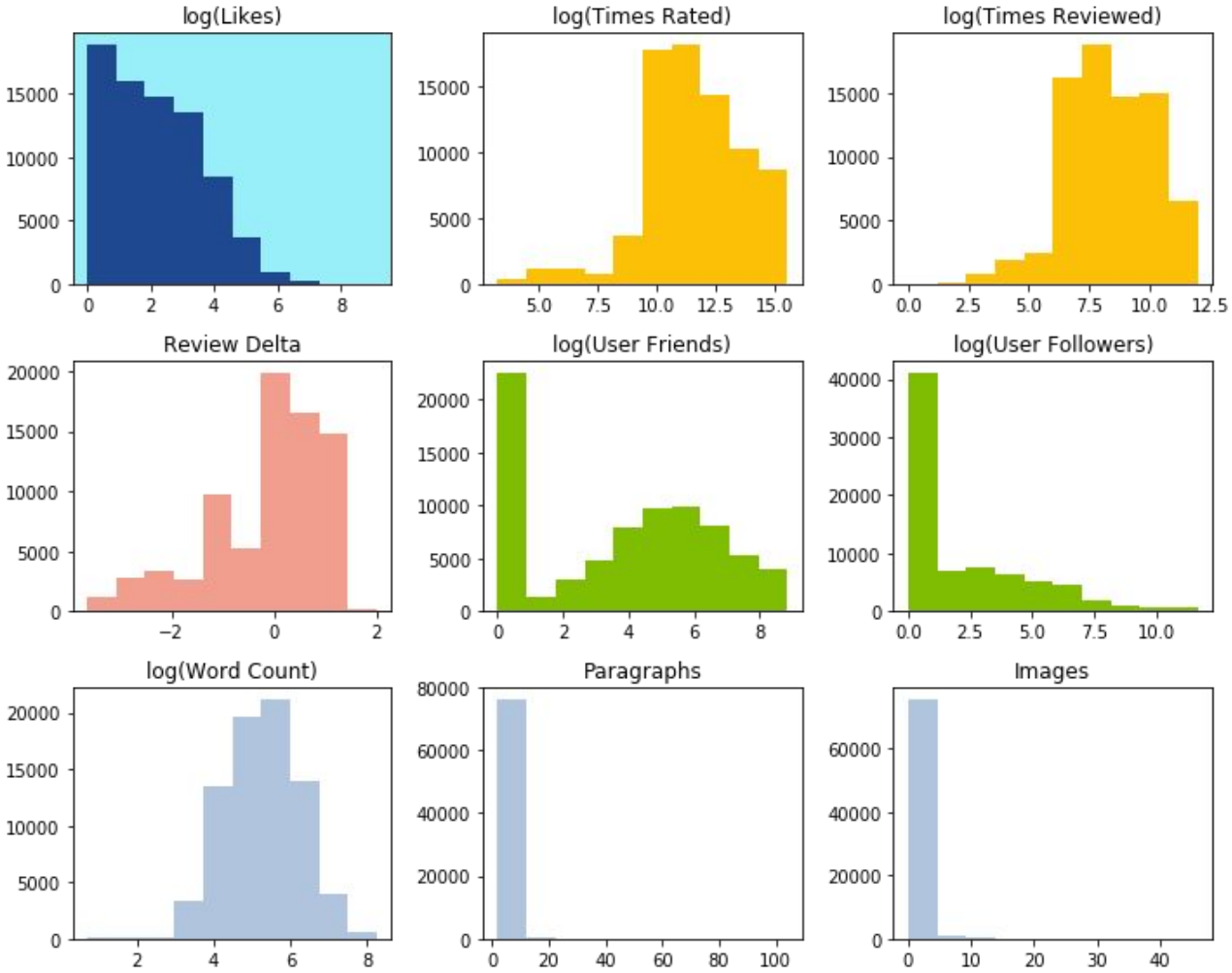


METIS

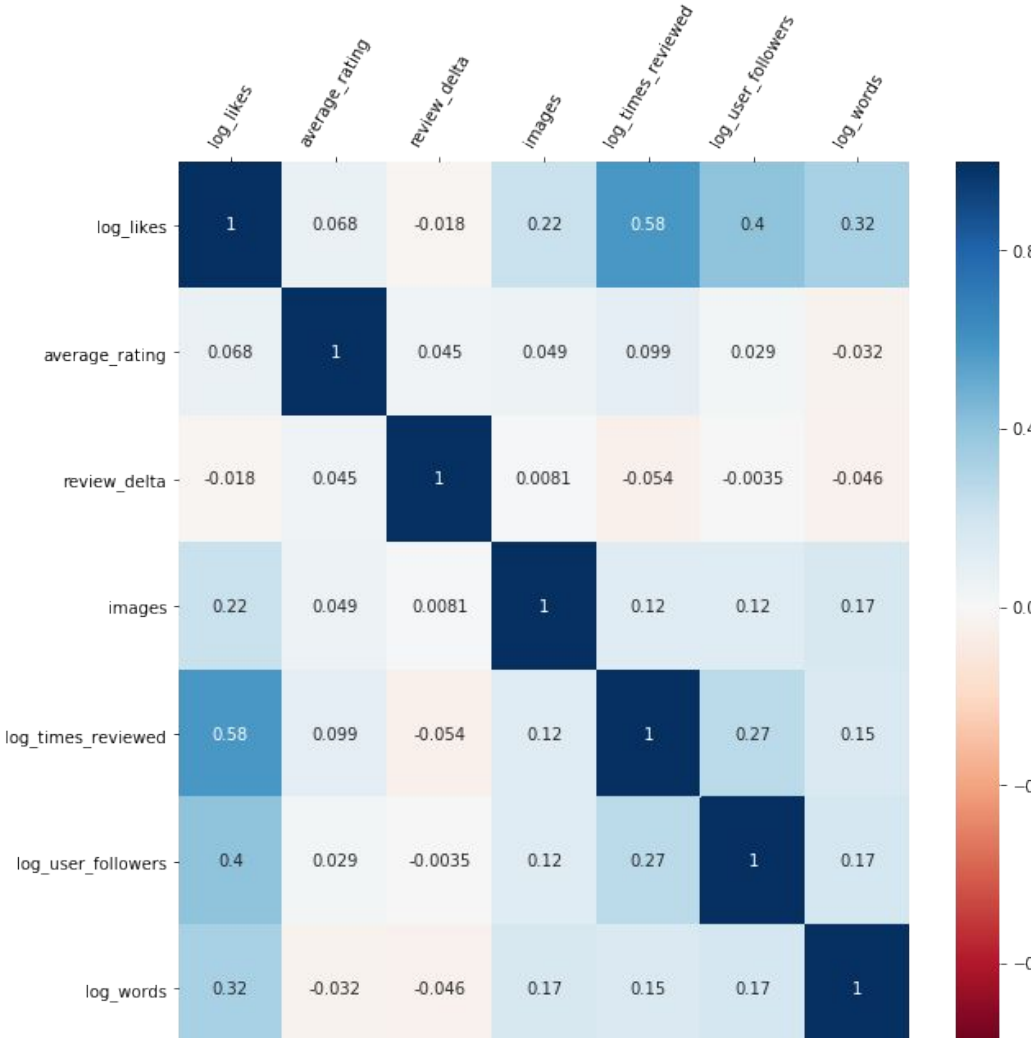

Likes
Times Rated
Times Reviewed
Review Delta
User Friends
User Followers
Word Count
Paragraphs
Images

METIS

Social networks
grow by
preferential attachment
and demonstrate
exponential distributions


log(Likes)
log(Times Rated)
log(Times Reviewed)
Review Delta
log(User Friends)
log(User Followers)
log(Word Count)
Paragraphs
Images

METIS


log_likes
average_rating
review_delta
images
log_times_reviewed
log_user_followers
log_words
log_likes 1 0.068 -0.018 0.22 0.58 0.4 0.32
average_rating 0.068 1 0.045 0.049 0.099 0.029 -0.032
review_delta -0.018 0.045 1 0.0081 -0.054 -0.0035 -0.046
images 0.22 0.049 0.0081 1 0.12 0.12 0.17
log_times_reviewed 0.58 0.099 -0.054 0.12 1 0.27 0.15
log_user_followers 0.4 0.029 -0.0035 0.12 0.27 1 0.17
log_words 0.32 -0.032 -0.046 0.17 0.15 0.17 1

METIS

Final LR

Dep. Variable:	log_likes	**R-squared:**	0.461
Model:	OLS	**Adj. R-squared:**	0.461
Method:	Least Squares	**F-statistic:**	7642.
Date:	Thu, 11 Oct 2018	**Prob (F-statistic):**	0.00
Time:	14:29:55	**Log-Likelihood:**	-80973.
No. Observations:	53576	**AIC:**	1.620e+05
Df Residuals:	53569	**BIC:**	1.620e+05
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-3.6139	0.094	-38.551	0.000	-3.798	-3.430
average_rating	0.1185	0.022	5.427	0.000	0.076	0.161
review_delta	0.0416	0.004	9.747	0.000	0.033	0.050
images	0.1254	0.004	29.784	0.000	0.117	0.134
log_words	0.3031	0.005	59.478	0.000	0.293	0.313
log_times_reviewed	0.4092	0.003	145.136	0.000	0.404	0.415
log_user_followers	0.1319	0.002	70.752	0.000	0.128	0.136

Omnibus:	3352.761	**Durbin-Watson:**	2.005
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	4020.050
Skew:	0.649	**Prob(JB):**	0.00
Kurtosis:	3.340	**Cond. No.**	224.

R^2: 0.467

MSE: 34

METIS

METIS

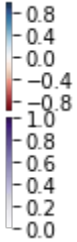
Model Improvement

- KFolds
  - Robust across test/train splits
- Lasso & Ridge
  - Decrease in performance
- Polynomial Features
  - Unstable improvements
- Polynomial + ElasticNetCV
  - Instability + Worse Performance

METIS

Analysis: What Increase Likes?

	avg. rating	review delta	image	log(words)	log(times_reviewed)	log(user_followers)
Corr:	0.087	0.014	0.21	0.31	0.58	0.41
Coef:	0.11	0.047	0.13	0.31	0.41	0.13



- Read popular books
- Write longer reviews
- Use more images
- Get more friends

METIS

Next Steps

Natural language processing

METIS