

Cooking

Predicting the Number of Recipe Ratings

Madeline Vossbrinck



Outline

- Background and Data
- Features and Methodology
- Results
- Conclusions and Future Work



Background

A recipe's number of ratings can be used as a measure of user engagement. The ratings themselves offer valuable feedback to both the recipe's author and NYT Cooking.



Data

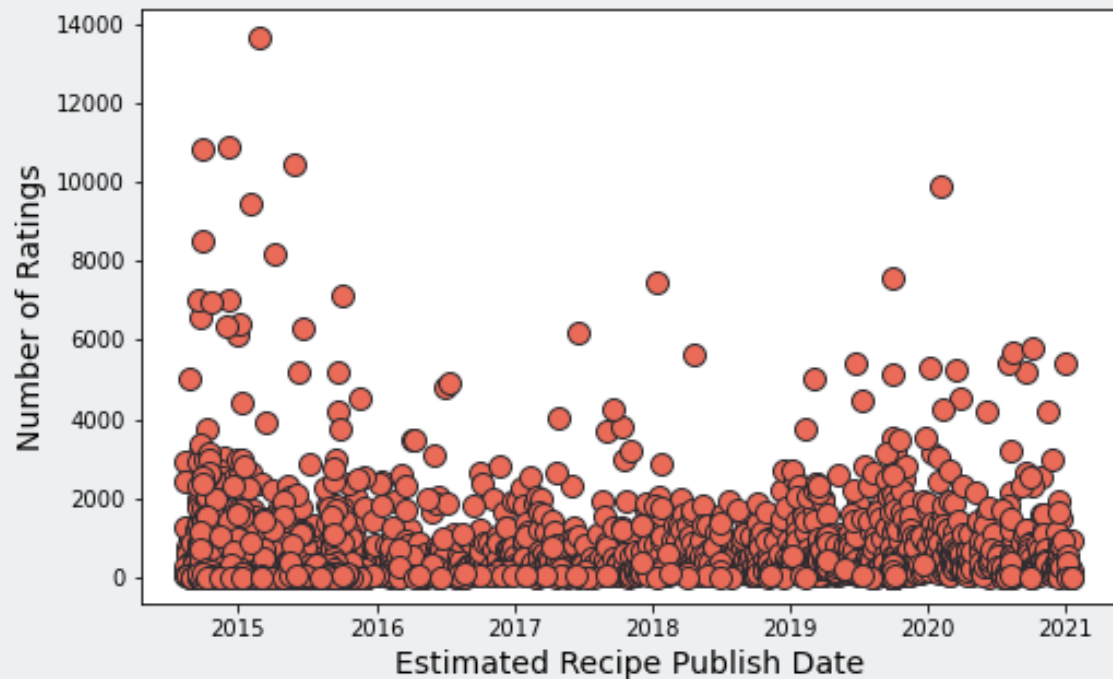
- NYT Cooking - <https://cooking.nytimes.com>
- Wayback Machine - <https://archive.org>

Tools

- BeautifulSoup, Selenium, Extruct
- statsmodels, sklearn



Number of Ratings by Estimated Publish Date



Value Summary:

Min	0
25th	15
50th	103
75th	402
Max	13,705
Mean	395


Features Considered

- Number of Preparation Steps
- Aggregate Rating Value (1-5)
- Preparation Time (minutes)
- Number of Ingredients
- Days Since Estimated Publish Date
- **Keyword Tags** (main course, easy, vegetarian, etc.)
- **Author** (Melissa Clark, Mark Bittman, etc.)



Categorizing Authors

Author	Total Ratings	Recipe Count	Average #
Melissa Clark	183448	271	676.9
Mark Bittman	116898	249	469.5
Sam Sifton	74158	76	975.8
David Tanis	63287	147	430.5
Alison Roman	61910	43	1439.8
Ali Slagle	61533	41	1500.8
Julia Moskin	61022	84	726.5
Martha Rose Schulman	36808	192	191.7
Colu Henry	28090	27	1040.4
The New York Times	27596	25	1103.8



There is a drop in total ratings after Julia Moskin

Selected Model

Predicted Number of Ratings =

343.8 * Aggregated Rating Value + 0.083 * Recipe Time In Minutes
- 0.225 * Days Since Estimated Publish Date
+ 975.9 * Author Is Ali Slagle + 916.5 * Author Is Alison Roman
+ 4.08 * Author is David Tanis + 419.1 * Author is Julia Moskin
+ 241.9 * Author is Mark Bittman + 278.2 * Author is Melissa Clark
+ 646.1 * Author is Sam Sifton

$R^2 = 15\%$

RMSE = 690

Conclusions

- Linear regression does not predict number of ratings well using the current variables and categories
- The author's impact on the number of ratings needs to be further explored
- There are thousands of possible keyword tags for recipes, making it difficult to find meaningful ones

Future Work

- Further investigate author's effect on number of ratings
- Scrape remaining recipes
- Find better measure for number of steps



Future Work Given Access To All Data

- Pull metrics for number of times a recipe site was shown in a search, visited, saved, commented on, and shared
- Analyze user activity to find out if there are patterns in recipe views and ratings based on author and/or keyword tags



Questions?

Appendix

Top 10 Recipes By Rating

Recipe Name	# of Ratings	Agg Rating	Author	Est. Publish Date
No-Knead Bread	13,705	5	Mark Bittman	2/23/15
Old-Fashioned Beef Stew	10,920	5	Molly O'Neill	12/8/14
Creamy Macaroni and Cheese	10,852	5	Julia Moskin	9/28/14
Marcella Hazan's Bolognese Sauce	10,474	5	The New York Times	5/31/15
Caramelized Shallot Pasta	9,910	5	Alison Roman	2/2/20
Oven-Roasted Chicken Shawarma	9,448	5	Sam Sifton	1/31/15
Chocolate Chip Cookies	8,506	5	David Leite	9/28/14
Roasted Chicken Provençal	8,173	5	Sam Sifton	4/4/15
Vinegar Chicken With Crushed Olive Dressing	7,576	4	Alison Roman	9/30/19
Salted Chocolate Chunk Shortbread Cookies	7,475	5	Alison Roman	1/14/18