
NYT COOKING
EXPLORING RECIPE POPULARITY

PROJECT CONTEXT

2014

NYT COOKING APP LAUNCHES

16,000 legacy recipes are migrated to the app and website and access is free to all.

2017

PAID ACCESS BEGINS

After building a strong and famously active user community, NYTC begins charging for subscriptions.

2018

CONTINUED GROWTH OF APP

By 2018, paid subscriber count had grown to 120,000, and the library recipe grew to 19,000.

2020

COMPETITION IS FIERCE

Heightened by the pandemic, Americans are cooking at home more now than any point in the past decade.

Sources: How NYT Cooking amassed 120,000 subscriptions in a year and a half; Survey: Cooking more at home could become the new normal post-pandemic

PROJECT CONTEXT

2014

NYT COOKING APP LAUNCHES

16,000 legacy recipes are migrated to the app and website and access is free to all.

2017

PAID ACCESS BEGINS

After building a strong and famously active user community, NYTC begins charging for subscriptions

2018

CONTINUED GROWTH OF APP

By 2018, paid subscriber count had grown to 120,000, and the library recipe grew to 19,000.

2020

COMPETITION IS FIERCE

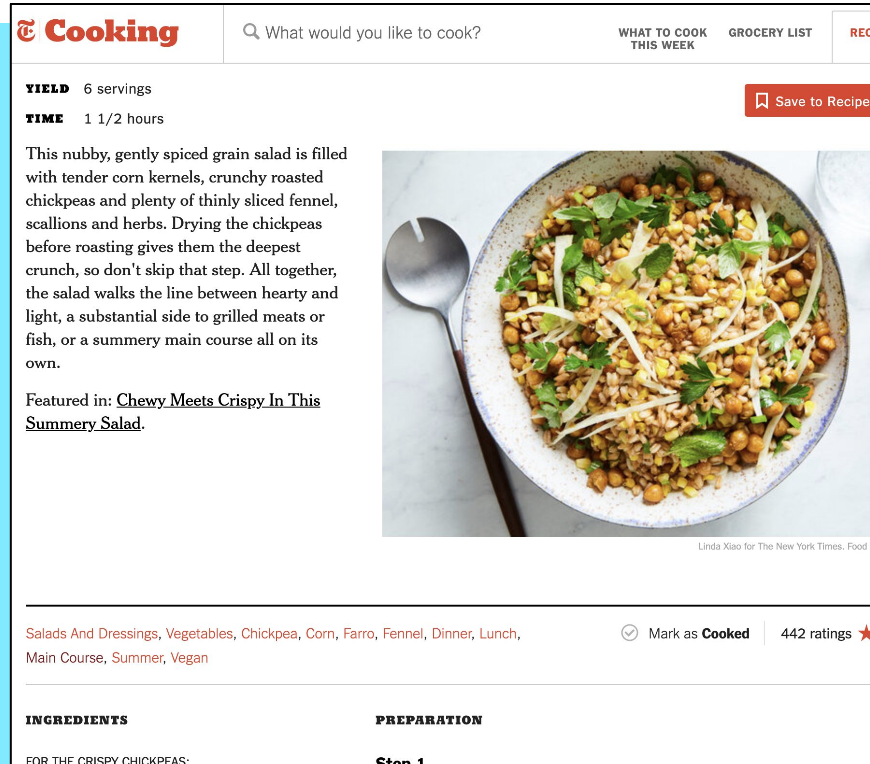
Heightened by the pandemic, Americans are cooking at home more frequently now than any point in the past decade.

Understanding user preferences is key to success in an increasingly competitive food media landscape.

Sources: How NYT Cooking amassed 120,000 subscriptions in a year and a half; Survey: Cooking more at home could become the new normal post-pandemic

DATA COLLECTION

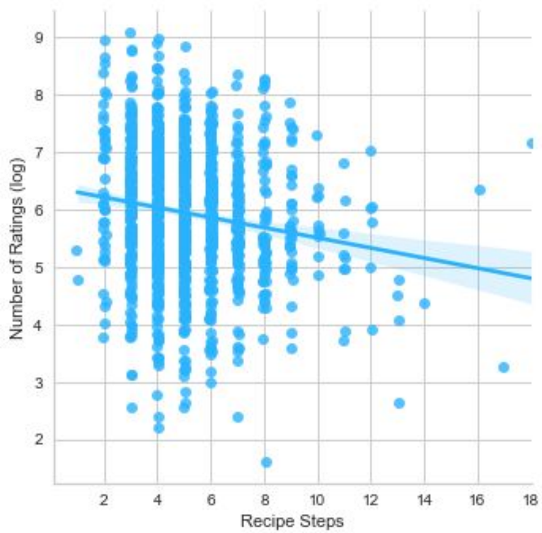
- Scraped recipe data for 1,920 dinner recipes from NYTCooking.com in late September 2020
- Target variable:
  - **Ratings Count**, or the number of users who have reported marking as recipe as “cooked”
- Features include:
  - Ingredients (and count)
  - Recipe steps (and count)
  - Total time to cook
  - Author
  - Recipe tags
  - Date published (or re-released)



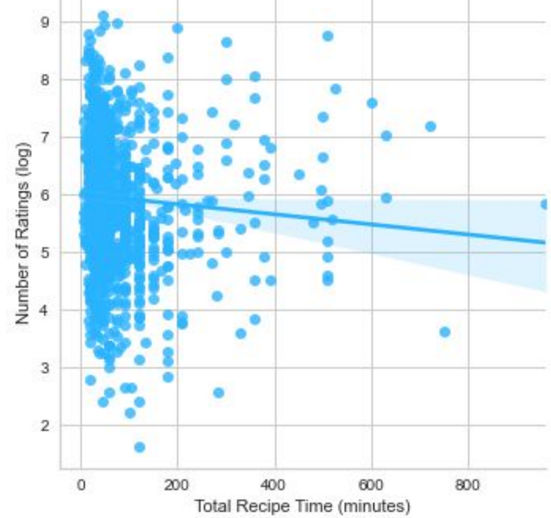
WHICH FEATURES ARE ASSOCIATED WITH TOTAL RATINGS?

EFFORT VS POPULARITY: MILD NEGATIVE CORRELATION

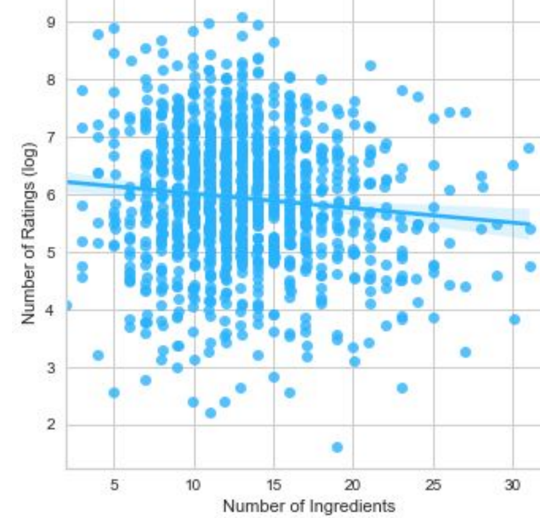
RECIPE STEPS



TOTAL TIME



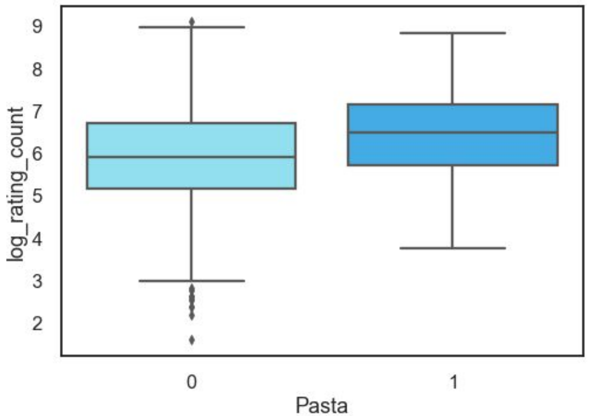
NO. OF INGREDIENTS

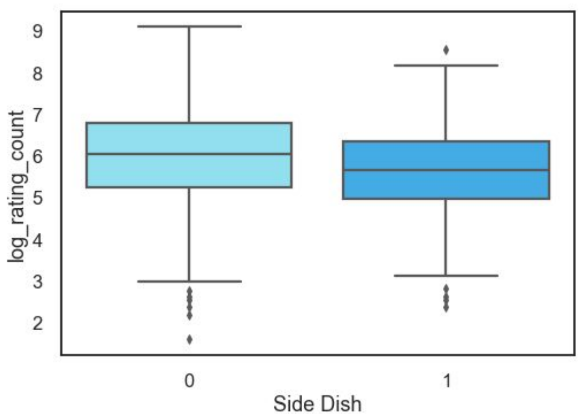


All measures of recipe effort have a small negative correlation with the number of ratings.

TRENDS ACROSS TAG CATEGORIES

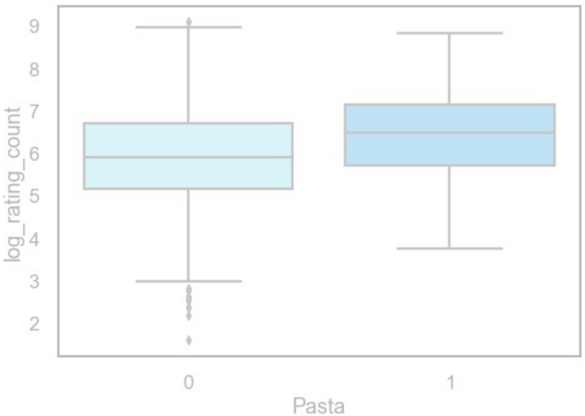
DISH TYPE

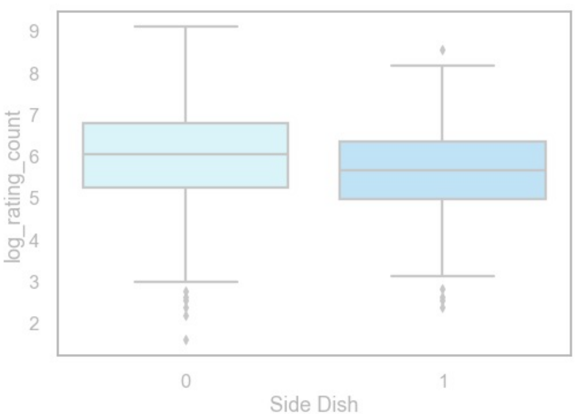




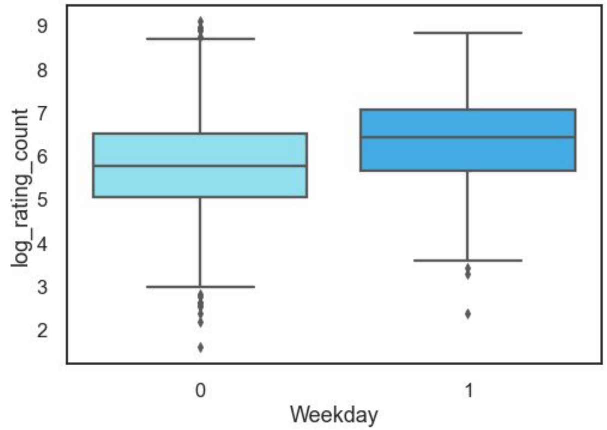
TRENDS ACROSS TAG CATEGORIES

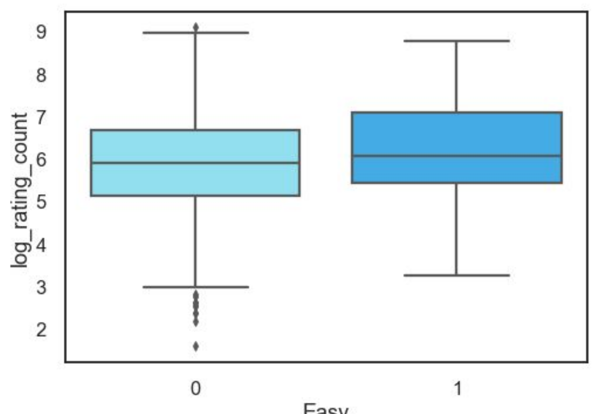
DISH TYPE





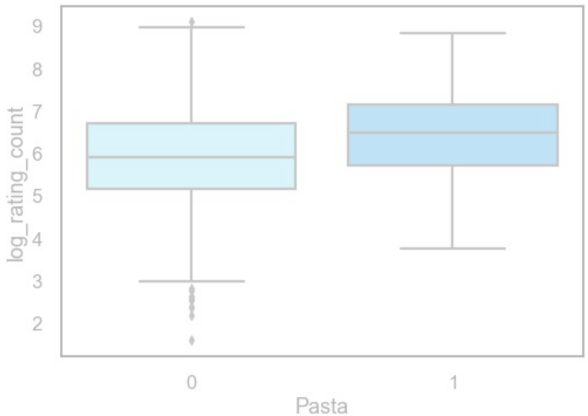
DIFFICULTY SIGNAL

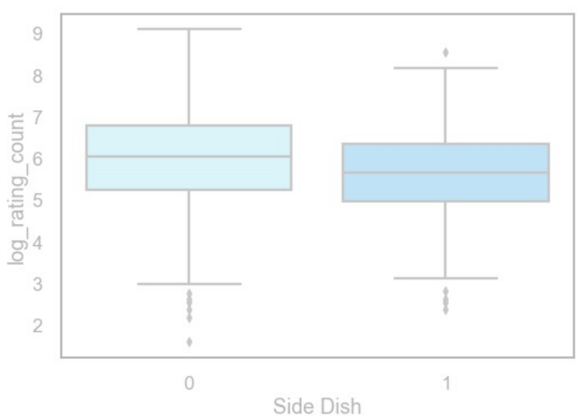




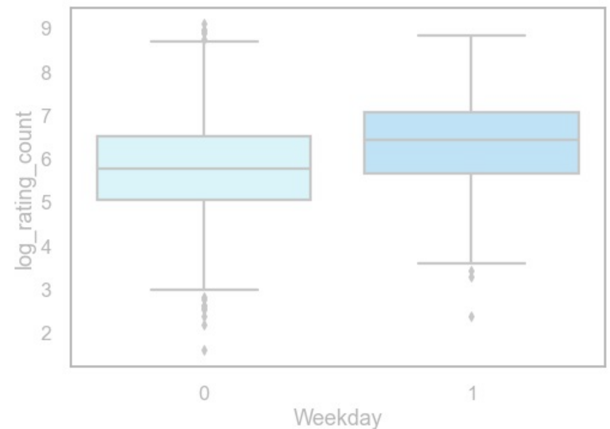
TRENDS ACROSS TAG CATEGORIES

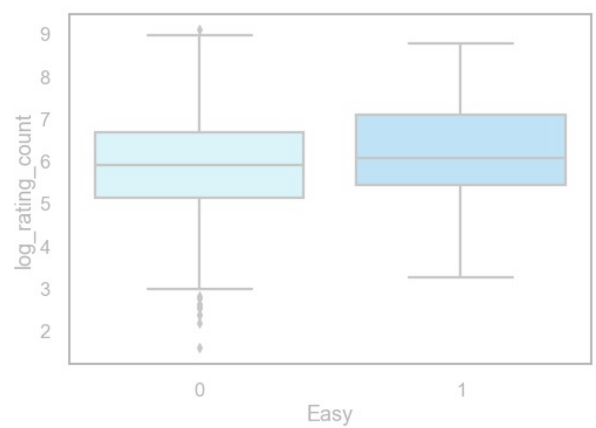
DISH TYPE



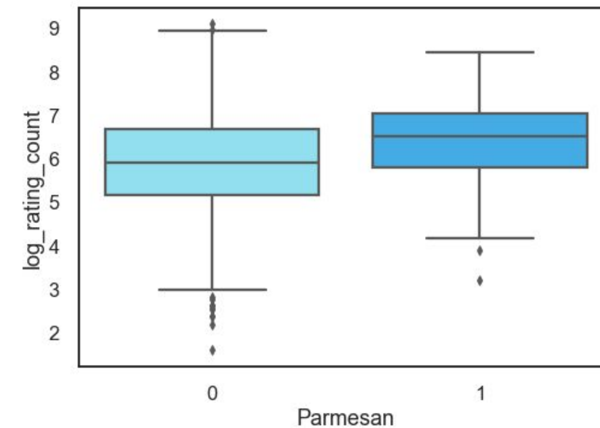


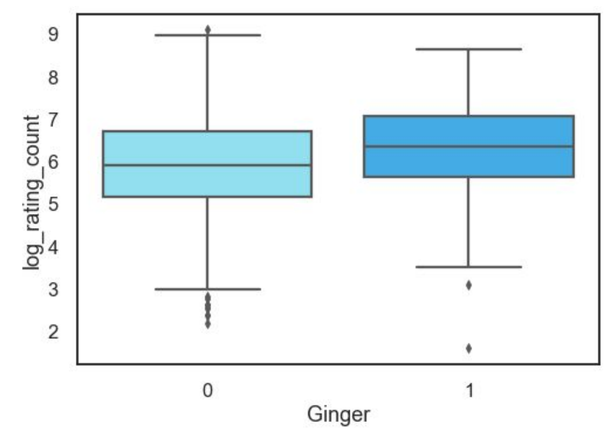
DIFFICULTY SIGNAL





INGREDIENTS





AUTHOR IMPACT

This chart shows the rating count distribution (log transformed) for the top 20 highest-volume authors.

There are clear differences in both volume and popularity across authors.

DIFFERENCES ACROSS PUBLICATION TIMING

RELEASE MONTH

Month	Total Recipes	Median Rating Count
Jan	128	528
Jun	80	474
Dec	59	465
Jul	82	435
Feb	94	427
Oct	102	380
Aug	118	370
Mar	95	364
May	96	343
Sep	122	341
Apr	105	331
Nov	110	301

RELEASE DAY OF WEEK

Weekday	Total Recipes	Median Rating Count
Sat	12	1082
Sun	130	629
Thu	109	564
Tue	125	557
Mon	92	553
Fri	99	363
Wed	624	299

MODEL RESULTS AND INTERPRETATION

FEATURES INCLUDED

After using LASSO regularization to narrow down the feature set and further iteration, the final model selected was a simple Linear Regression with the following features:

NINE AUTHORS	• (+) Alison Roman, Melissa Clark, Julia Moskin, Ali Slagle • (-) David Tanis, Sean Sherman, Gabrielle Hamilton, Florence Fabricant, “Other”
SEVEN TAGS	• (+) Chicken (aggregated), Easy (aggregated), Meatless (aggregated), Salmon, Italian (aggregated) • (-) Spring, Bacon
THREE PUBLICATION TIMING FEATURES	• (+) Sunday release • (-) Wednesday release, November release
ONE EFFORT FEATURE	• (+) • (-) Number of recipe steps (squared)

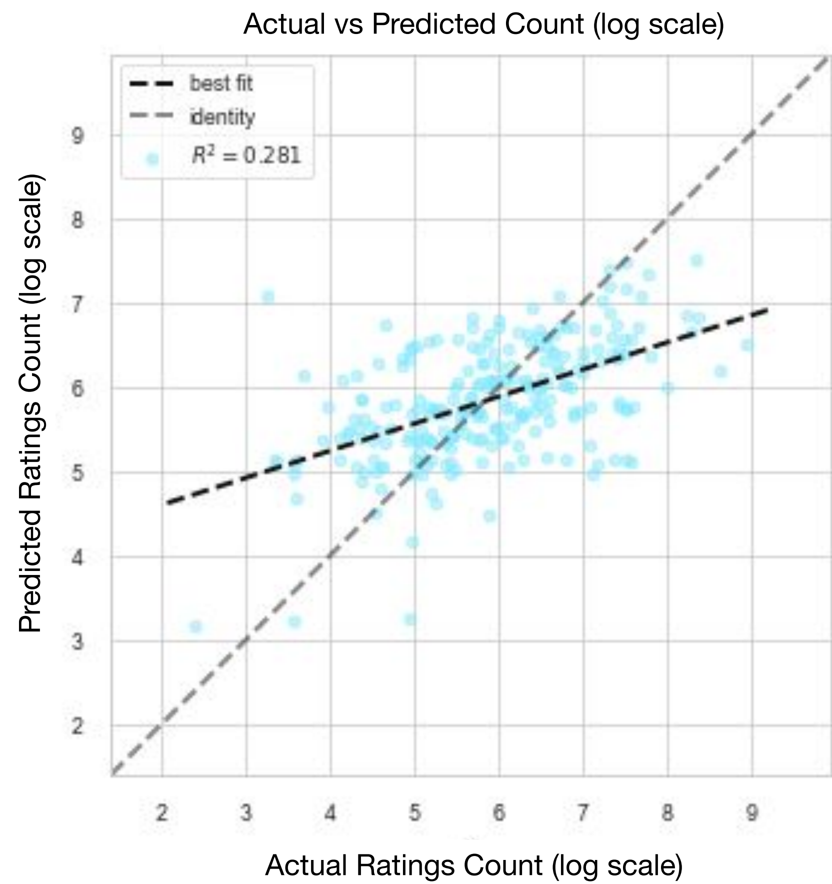
RESULTS

28%

R^2

856

RMSE*



*Represents error in actual number of ratings, not the log-transformed target.

RECOMMENDATIONS

1. Invest in author acquisition and promotion: they can be a key driver of readership
2. Monitor tag popularity and consider creating more content in popular areas, especially if existing content count is low
3. Consider A/B testing recipe release timing to better understand the weekday release results seen here

NEXT STEPS

Augment the existing analysis:

- Dig into highest residuals and consider redefining outliers to improve model performance
- Implement time series sampling into this modeling, to better control for changes in readership and app UI that occurred over the data collection period
- Add additional data to the existing analysis, such as author social media following, and the cost or rarity of ingredients

NEXT STEPS

Augment the existing analysis:

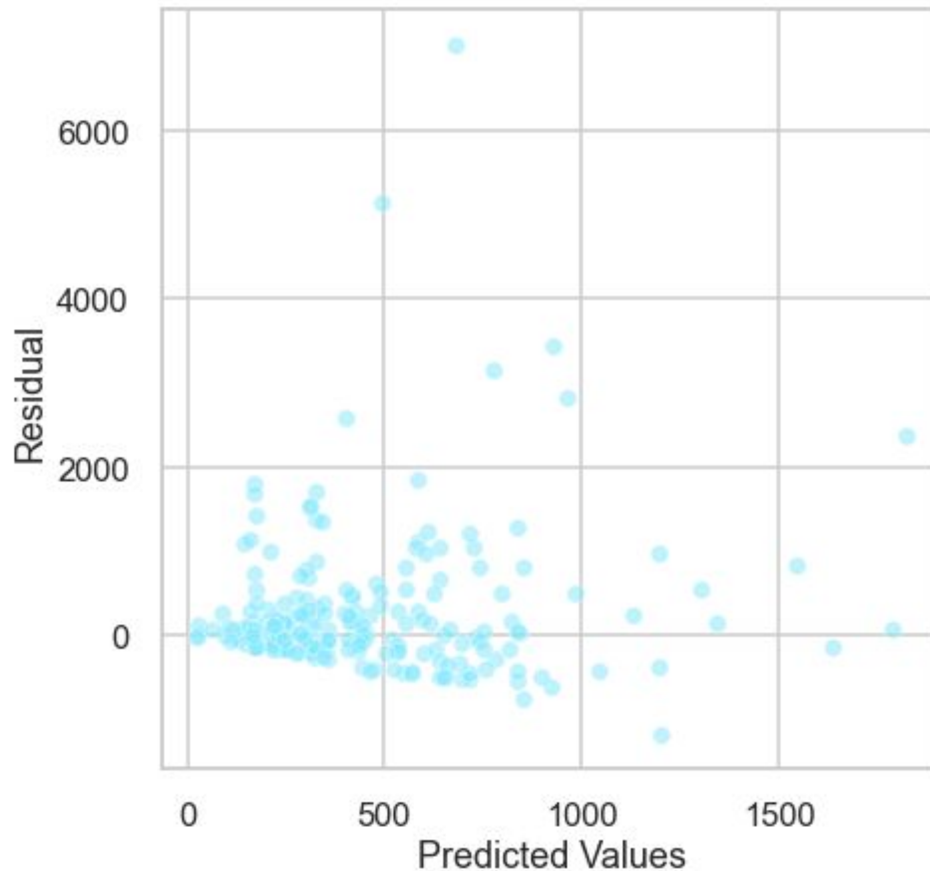
- Dig into highest residuals and consider redefining outliers to improve model performance
- Implement time series sampling into this modeling, to better control for changes in readership and app UI that occurred over the data collection period
- Add additional data to the existing analysis, such as author social media following, and the cost or rarity of ingredients

Implement new methodologies:

- Classification modeling techniques may be more appropriate than linear regression, and allow for a richer understanding of different recipe profiles
- Consider natural language processing to better take advantage of the text in the recipe topnote and steps, and identify other trends

APPENDIX

RESIDUALS PLOT



MODEL COEFFICIENTS

Note: coefficient values are scaled and correspond to log-transformed target variable, so further processing would be needed to apply directly to making predictions.

Feature	Coefficient
Chicken	0.25
Weekday	0.18
auth_Alison Roman	0.17
auth_Melissa Clark	0.13
auth_Julia Moskin	0.13
weekday_Sun	0.11
Vegan	0.11
auth_Ali Slagle	0.10
Salmon	0.08
Italian	0.08
Bacon	-0.06
Spring	-0.08
sqrd_num_steps	-0.08
month_Nov	-0.08
auth_David Tanis	-0.08
auth_Other	-0.10
auth_Sean Sherman	-0.12
auth_Gabrielle Hamilton	-0.12
auth_Florence Fabricant	-0.13
weekday_Wed	-0.16