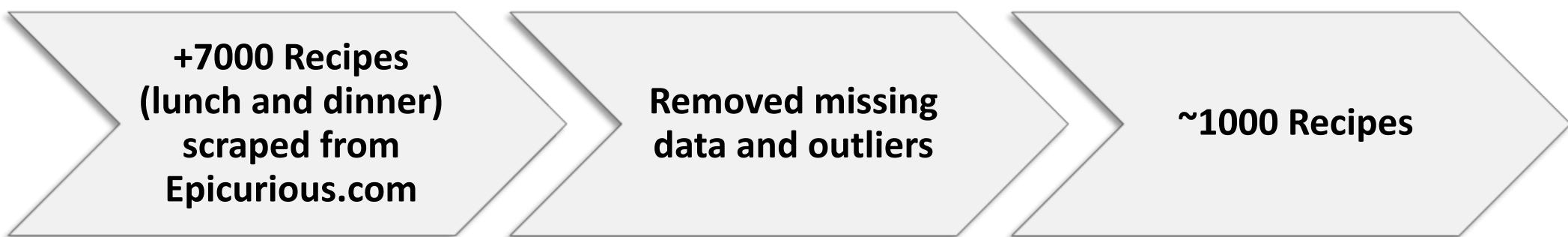




Predicting the Success of a Recipe

What can we conclude from the ratings in Epicurious.com?

Web Scraping and Data Cleaning



The screenshot shows a recipe page from Epicurious. At the top left is the 'epicurious' logo. To its right is a 'Follow' button with a red 'f' icon. Below the logo is a photo of a dish. To the right of the photo are the ratings '3/4 ⌂' and '67% MAKE IT AGAIN'. To the right of these are sections for 'NUTRITIONAL INFO' and 'MENUS & TAGS'. The 'NUTRITIONAL INFO' section contains a table of values per serving (4 servings). The 'MENUS & TAGS' section includes a 'THIS RECIPE IS FEATURED IN:' section with a thumbnail image of a dish and the text '4 New Recipes for Rice Lovers', and a 'TAGS:' section with various cooking terms.

NUTRITIONAL INFO

Calories	817	Carbohydrates	88 g(29%)
Fat	29 g(44%)	Protein	47 g(95%)
Saturated Fat	8 g(40%)	Sodium	175 mg(7%)
Polyunsaturated Fat	6 g	Fiber	3 g(12%)
Monounsaturated Fat	12 g	Cholesterol	172 mg(57%)

per serving (4 servings)

MENUS & TAGS

THIS RECIPE IS FEATURED IN:

4 New Recipes for Rice Lovers

TAGS:

BON APPÉTIT CHICKEN RICE AVOCADO POACH DINNER LUNCH
GINGER BASIL GREEN ONION/SCALLION

Choosing the Target Variable

SHRIMP WITH HERBY WHITE BEANS AND TOMATOES

BY DAVID TAMARKIN | EPICURIOUS | JANUARY 2017



Success Score = Rating * "Make it again"



Choosing the Features

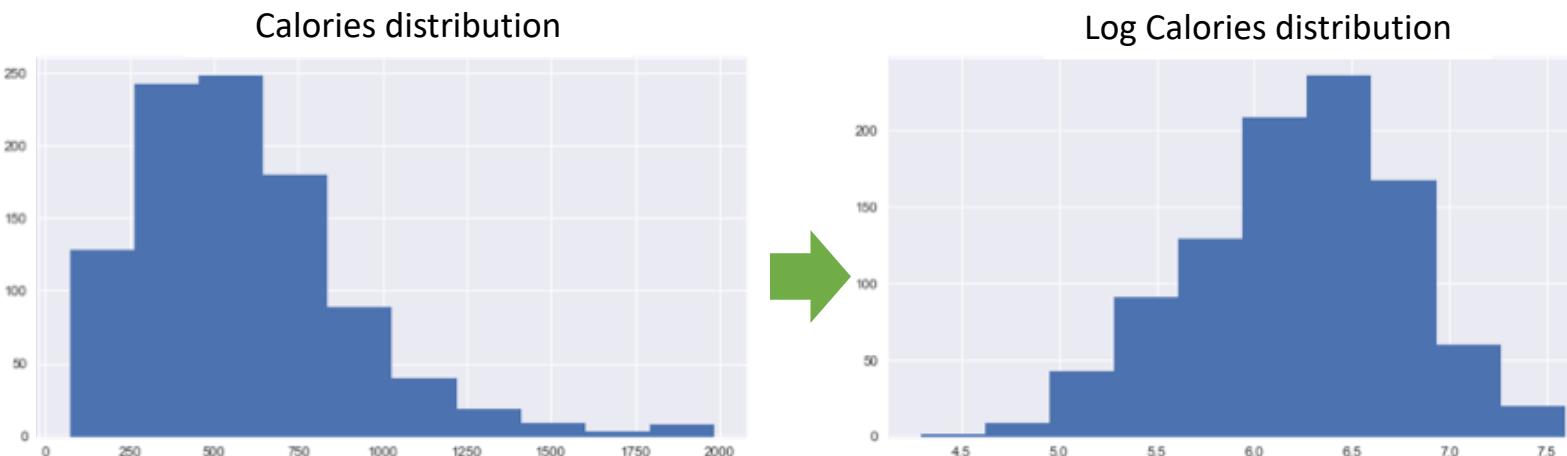
Independent Variables

- 1. Nutrition Data
 - Calories
 - Protein
 - Cholesterol
 - Etc.
- 2. Ingredients (dummy)



Data Wrangling

- **Log transformation in Nutrition Data**



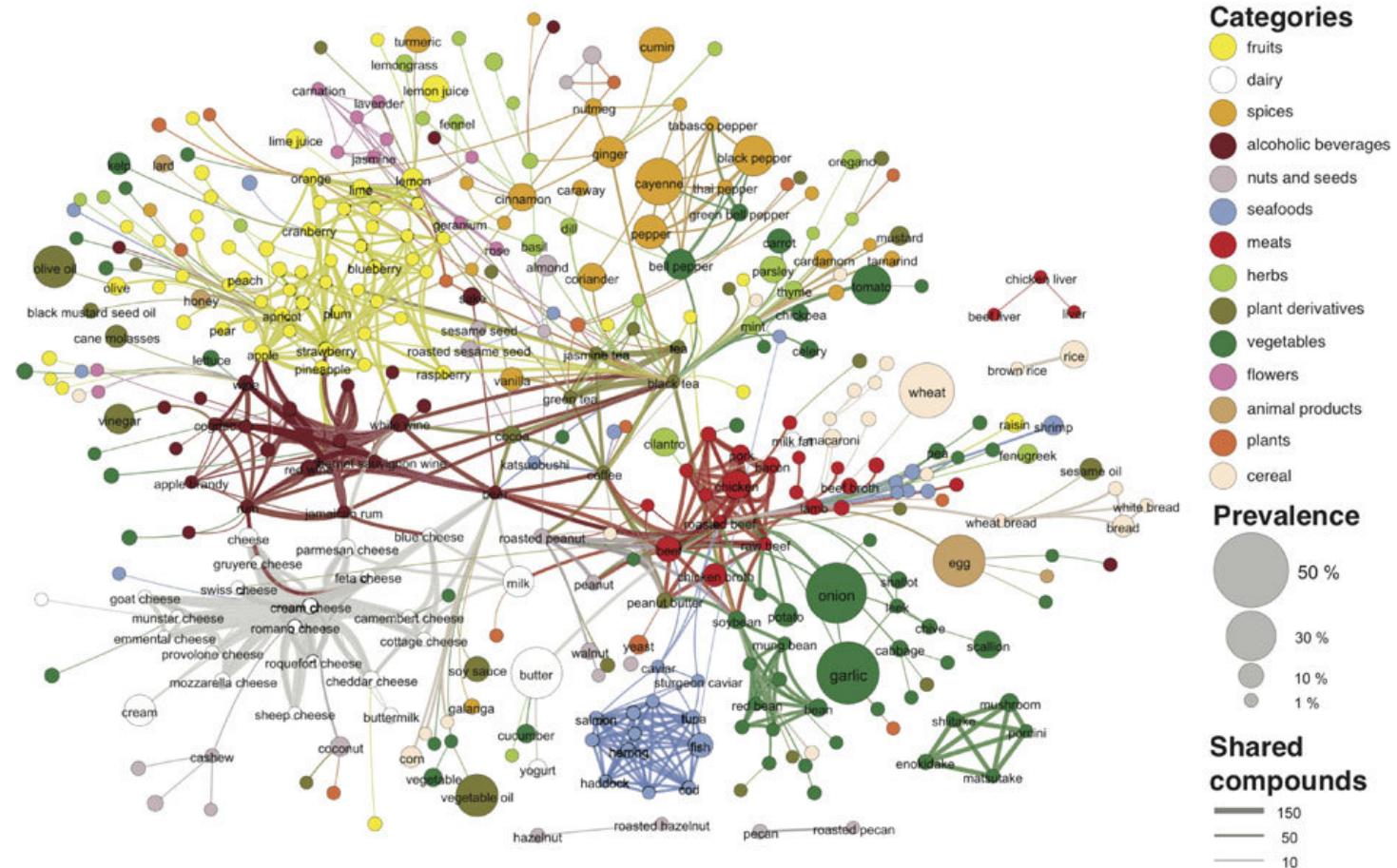
- **Feature Selection for choosing Ingredients**
 - Recursive Feature Elimination (`sklearn.feature_selection`)
 - Ranks the predictors based on their contribution to the model.
 - Selected 50 most important features

Preliminary Analysis

	index	success_score	rating	calories	carbohydrates	protein	cholesterol	sodium
Top Recipes		4.0000	4.00	703.30	44.96	39.31	149.61	978.71
Bottom Recipes		1.4108	2.64	607.45	53.51	26.64	116.66	712.30
Comparison		2.8400	1.52	1.16	0.84	1.48	1.28	1.37

On average, the 100 best rated recipes have more calories, protein, cholesterol and sodium, and less carbohydrates than the 100 worse rated recipes.

Preliminary Analysis



Source: Article "What A Global Flavor Map Can Tell Us About How We Pair Foods (Nature.com)

The success of a recipe is a result of the combination of flavors and preparation methods

Exploratory Data Analysis

Linear Regression

- OLS Regression on training set (70% of data) beginning with the selected features.

OLS Regression Results

Dep. Variable:	success_score	R-squared:	0.950
Model:	OLS	Adj. R-squared:	0.946
Method:	Least Squares	F-statistic:	216.8
Date:	Thu, 01 Feb 2018	Prob (F-statistic):	0.00
Time:	18:44:32	Log-Likelihood:	-693.53
No. Observations:	676	AIC:	1497.
Df Residuals:	621	BIC:	1745.
Df Model:	55		
Covariance Type:	nonrobust		

MSE: 0.589

Removed features
when p-value high

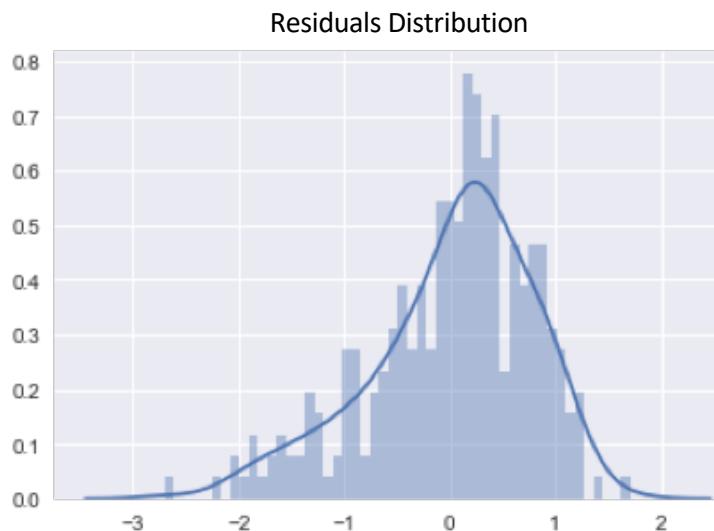
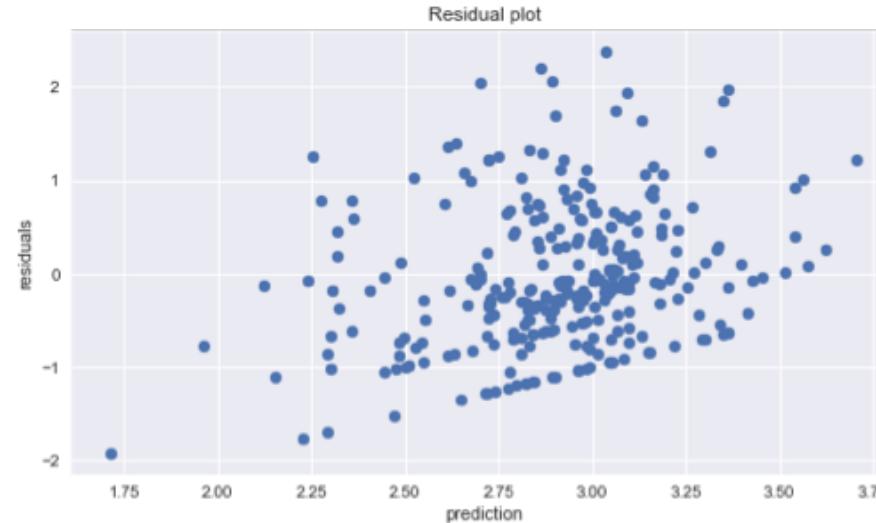


OLS Regression Results

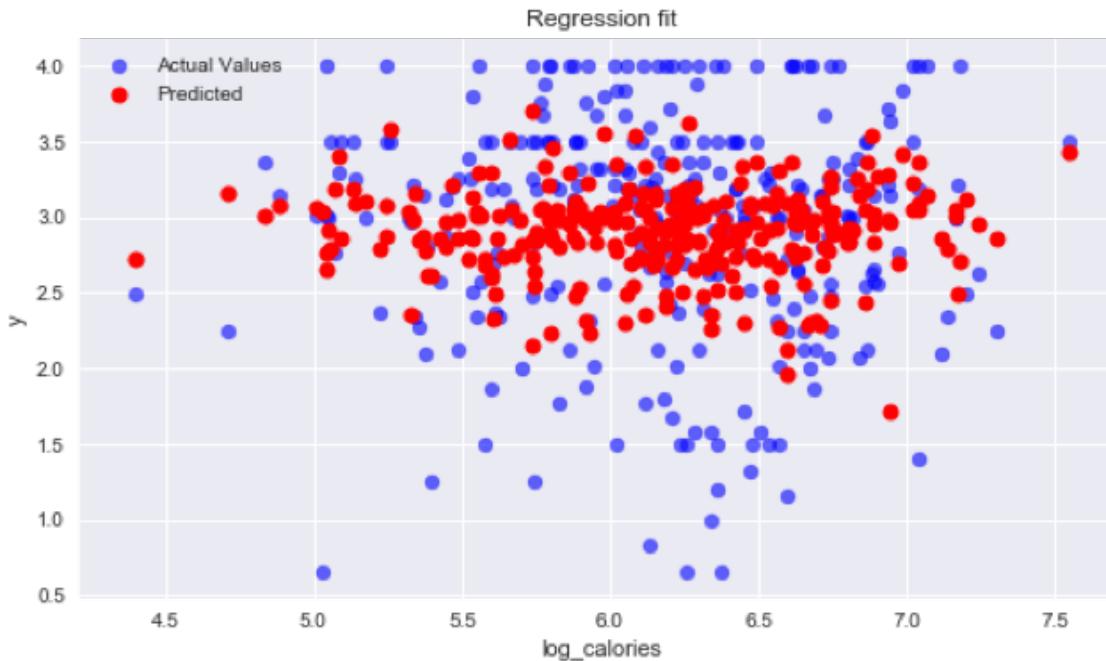
Dep. Variable:	success_score	R-squared:	0.951
Model:	OLS	Adj. R-squared:	0.949
Method:	Least Squares	F-statistic:	406.0
Date:	Thu, 01 Feb 2018	Prob (F-statistic):	0.00
Time:	18:47:23	Log-Likelihood:	-685.96
No. Observations:	676	AIC:	1434.
Df Residuals:	645	BIC:	1574.
Df Model:	31		
Covariance Type:	nonrobust		

MSE: 0.576

Exploratory Data Analysis Results

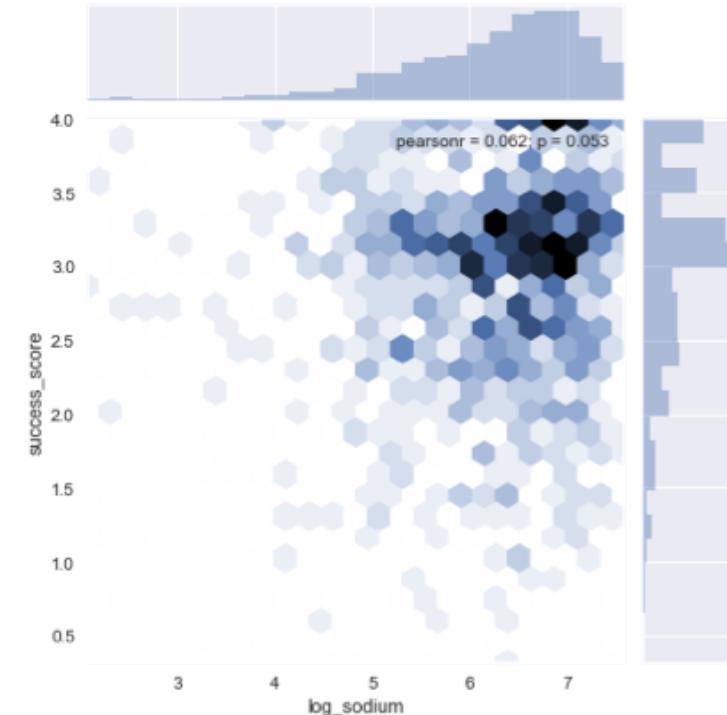


Exploratory Data Analysis Results



	coef	std err	t	P> t	[0.025	0.975]
log_calories	0.5571	0.051	10.922	0.000	0.457	0.657

1% increase in log_calories, the success score increases 0.56



Exploratory Data Analysis

Features selected are mostly not main ingredients, but flavor additions to the recipes

Prune			
Grains	Shallot	Lemon	
Ground Lamb	Pine Nut	Cheddar	
Nutmeg	Goat Cheese	Chile Pepper	
Pear	Coconut	Mint	
Orange Juice	Oat	Squash	
Okra	Celery	Walnut	
Trout	Cranberry	Feta	
log_calories	Monterey Jack	Lentil	
log_carbohydrates	Pistachio	Breadcrumbs	
log_protein	Pineapple		
log_sodium			



Conclusions

- People prefer more caloric and salty meal
- Main ingredients can not be individually analyzed in a recipe
- “Fancy” ingredients may increase (or decrease) the rating



Next Steps

- Use a classification method
- Extract proportions of ingredients
- Use NLP for analyzing the reviews

