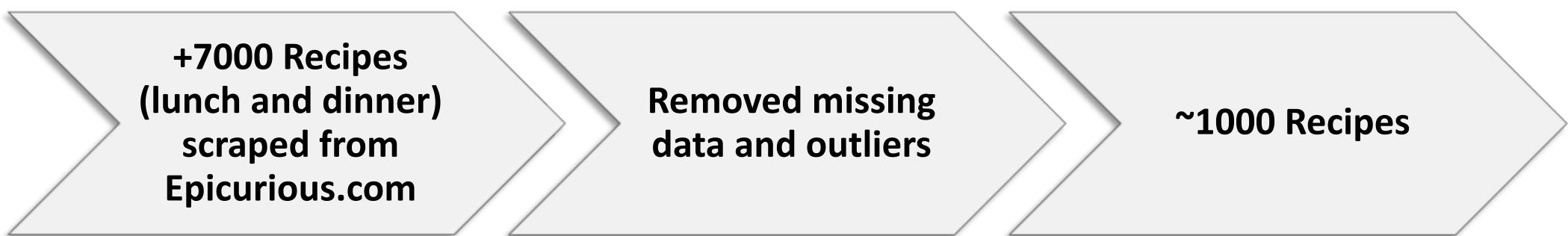




# Predicting the Success of a Recipe

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

# Web Scraping and Data Cleaning



**epicurious**

Follow

**3/4** **67%**

REVIEWS (3)

MAKE IT AGAIN

**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)

Powered by EDAMAM

**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**

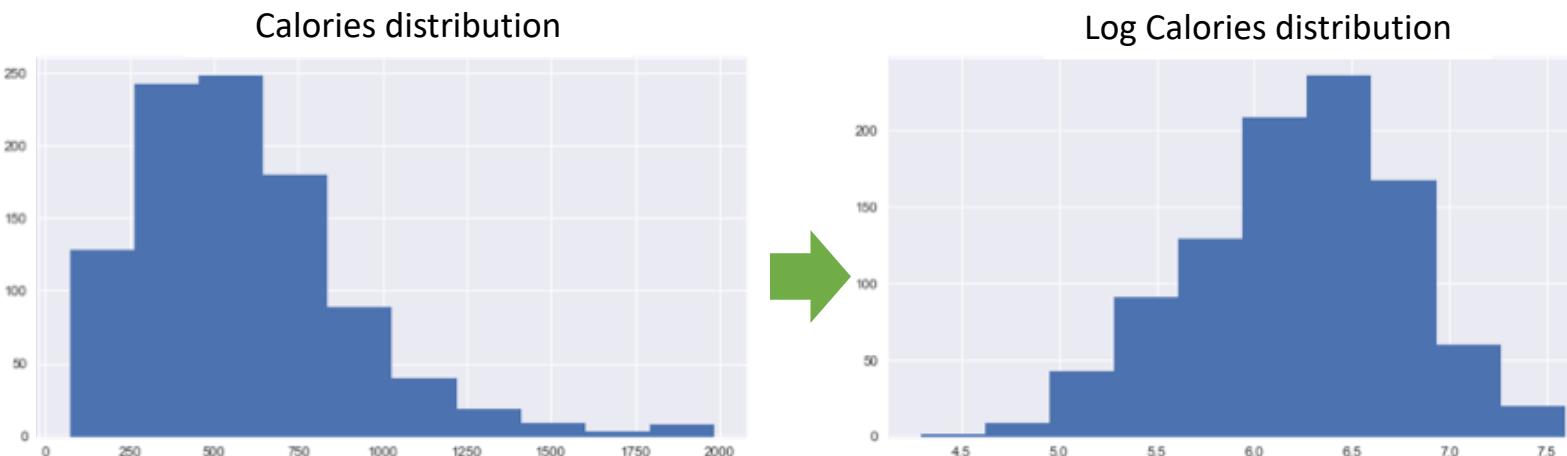
- Calories
- Protein
- Cholesterol
- Etc.

**2. Ingredients (dummy)**



# Data Wrangling

- **Log transformation in Nutrition Data**



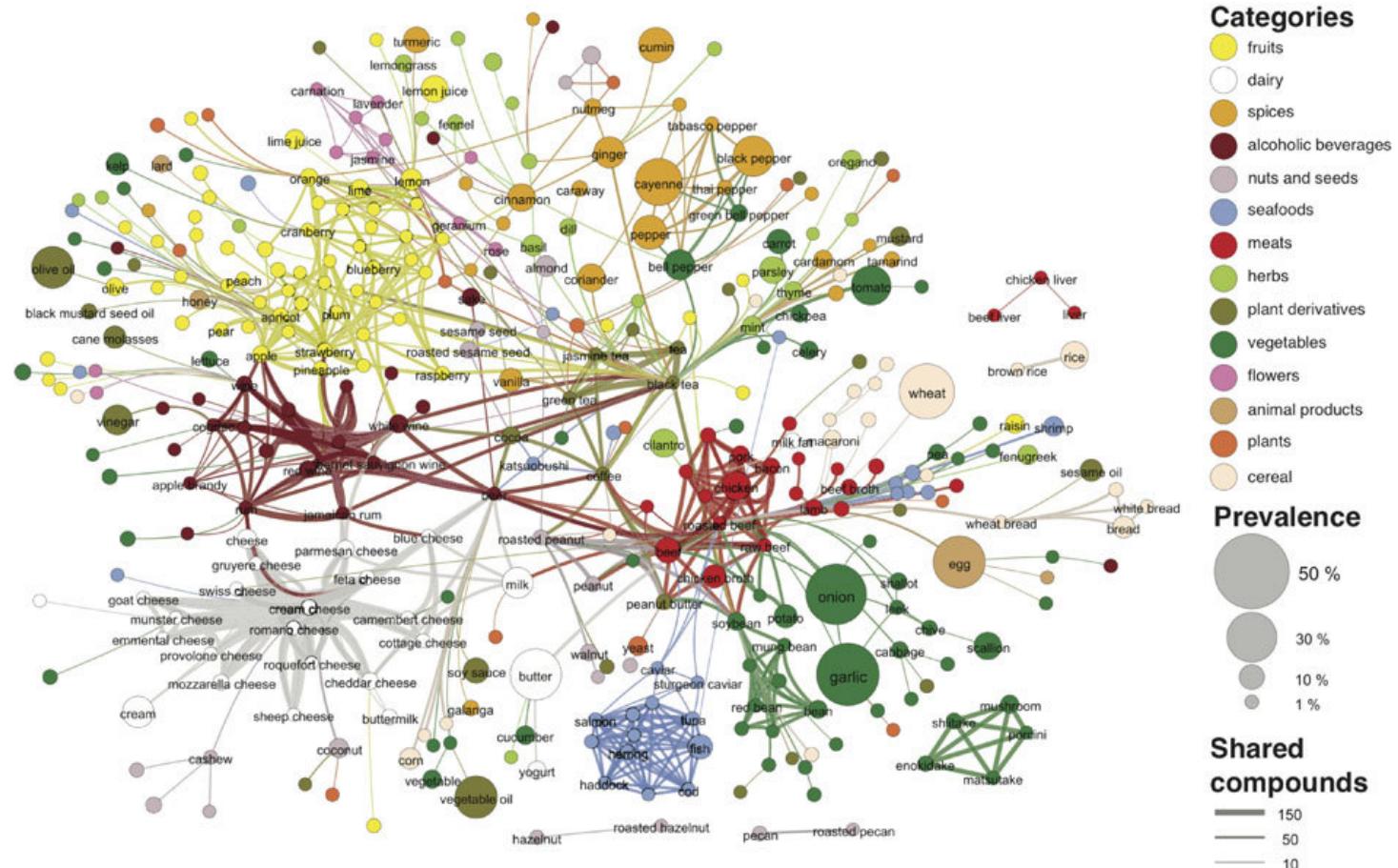
- **Feature Selection for choosing Ingredients**
  - Recursive Feature Elimination (`sklearn.feature_selection`)
    - Ranks the predictors based on their contribution to the model.
  - Selected 50 most important features

# Preliminary Analysis

	<b>index</b>	<b>success_score</b>	<b>rating</b>	<b>calories</b>	<b>carbohydrates</b>	<b>protein</b>	<b>cholesterol</b>	<b>sodium</b>
<b>Top Recipes</b>		4.0000	4.00	703.30	44.96	39.31	149.61	978.71
<b>Bottom Recipes</b>		1.4108	2.64	607.45	53.51	26.64	116.66	712.30
<b>Comparison</b>		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

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

MSE: 0.589

Removed features  
when p-value high

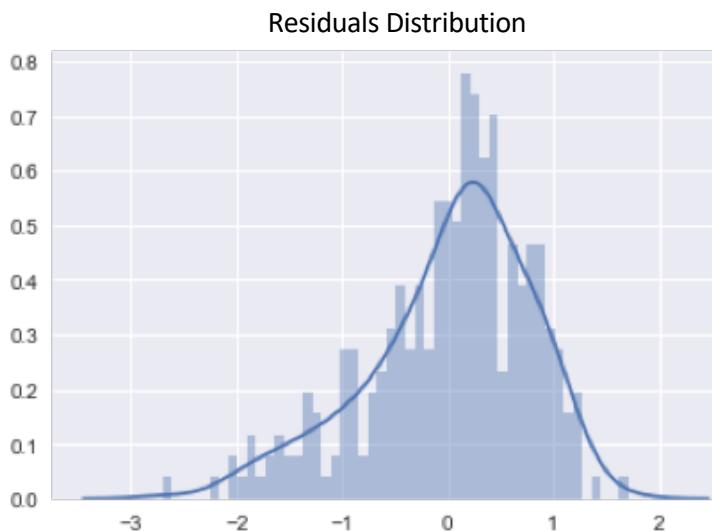
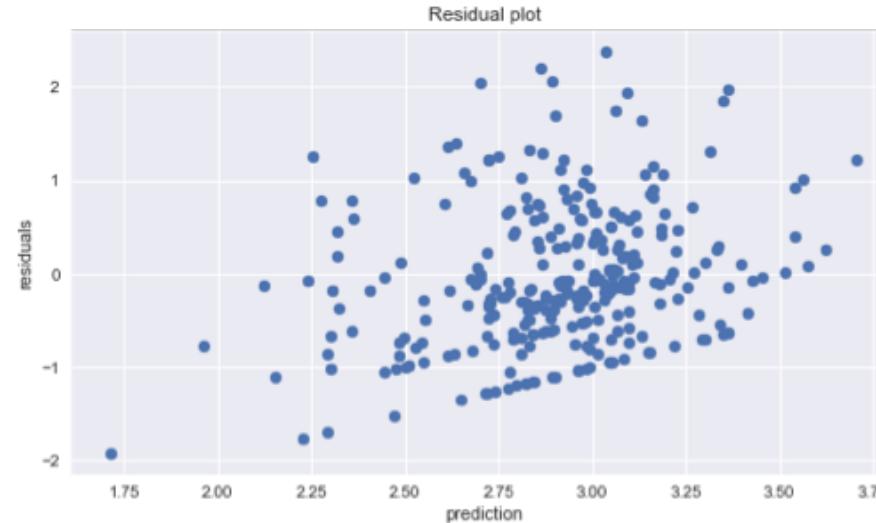


OLS Regression Results

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

MSE: 0.576

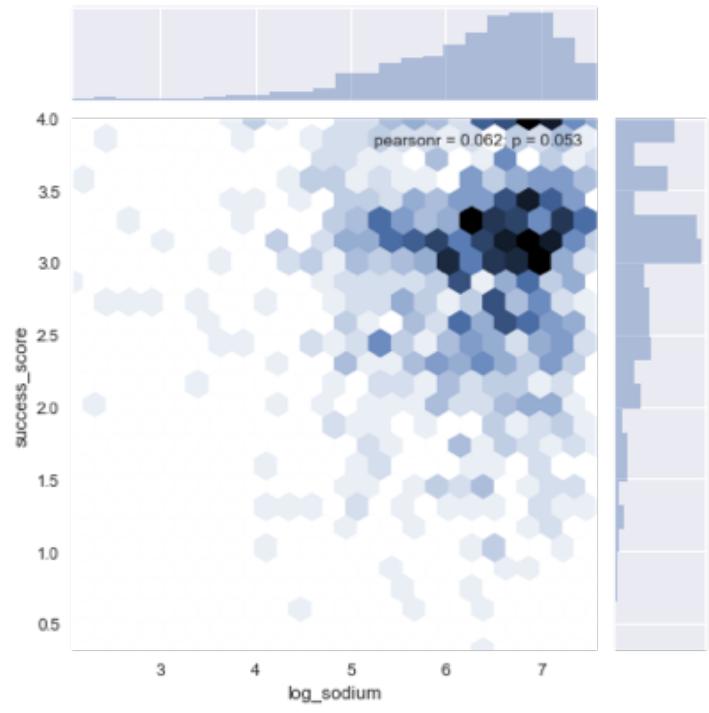
# Exploratory Data Analysis Results



# Exploratory Data Analysis Results

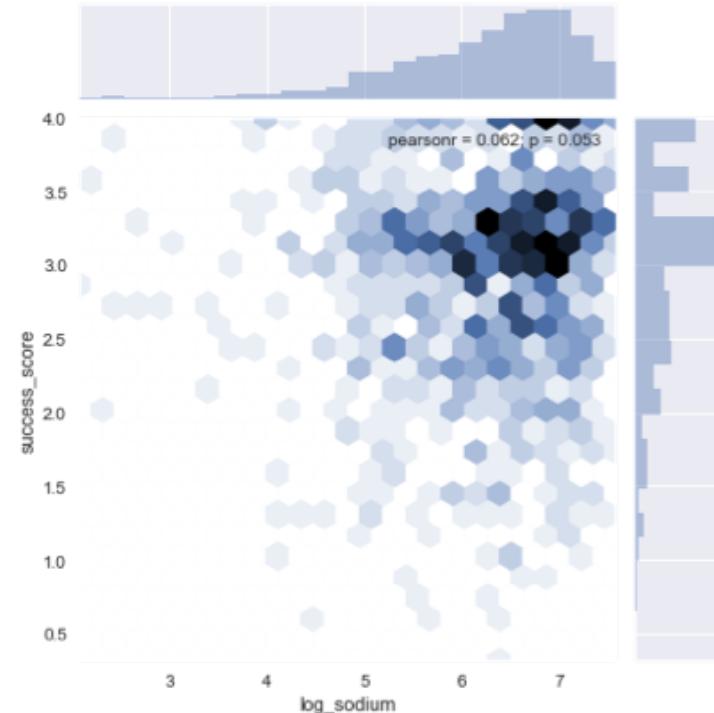
	coef	std err	t	P> t	[0.025	0.975]
log_sodium	0.1288	0.034	3.777	0.000	0.062	0.196

1% increase in log\_sodium, the success score increases 0.13



	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			



# Final Thoughts

---

- **Conclusions**

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

- **Next Steps**

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

