# To predict the average rating value of BBC Good Food recipes

-To improve quality of BBC Good Food recipes

#6

Claire May Paula Eva

# Agenda

- Problem Definition
- Data Preparation
- Data Exploration
- Modeling
- Evaluation
- Recommendation

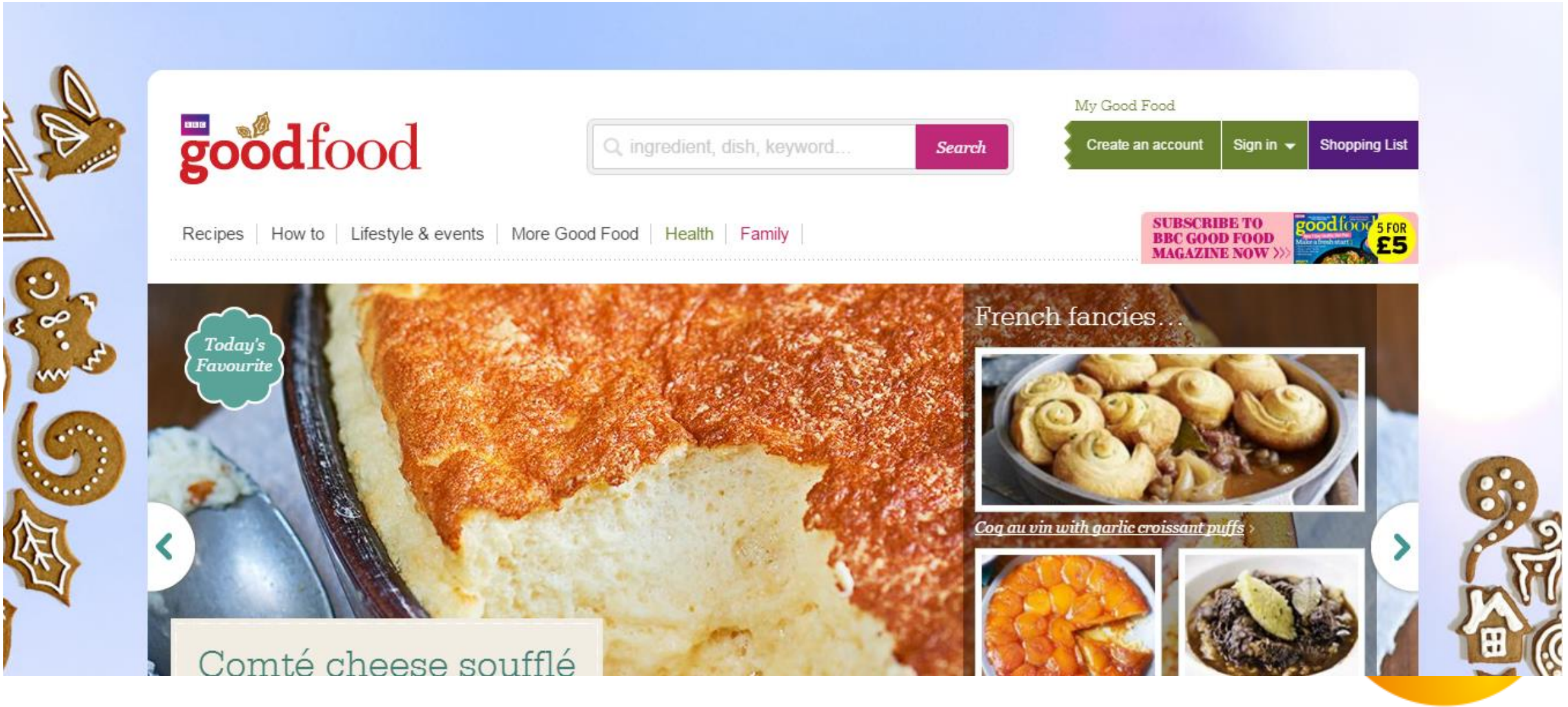
# Business Problem

- **Goal**
  - To improve the quality of BBC Good Food recipes
- **Stakeholder**
  - Client: BBC Good Food
  - Website visitor
- **Challenge**
  - Not the most popular recipe website
- **Opportunity**
  - To make the website more attractive(WOM)

# Analytic problem

- **Goal: What are you predicting?**
  - Average rating value of each new recipe
- **Why?**
  - Website Visitors tend to read the high rating recipe
- **What is a success?**
  - The predicted rating value is closed to the real value.

# Data Preparation


goodfood
ingredient, dish, keyword...
Search
My Good Food
Create an account
Sign in
Shopping List
Recipes
How to
Lifestyle & events
More Good Food
Health
Family
SUBSCRIBE TO BBC GOOD FOOD MAGAZINE NOW
5 FOR £5
Today's Favourite
Comté cheese soufflé
French fancies...
Coq au vin with garlic croissant puffs

# Roasted ratatouille pasta

(26 ratings) *Magazine subscription – 5 issues for £5*

**Prep:** 15 mins
**Cook:** 30 mins

Easy

Serves 2

This veg packed pasta dish provides all of your 5-a-day - and it's delicious!

*Share:* Facebook | Pinterest | Twitter | g+ | ✉

Nutrition *per serving*

Easily halved | Healthy | Vegetarian

| kcalories | protein | carbs | fat | saturates | fibre | sugar | salt |
|---|---|---|---|---|---|---|---|
| 450 | 15g | 83g | 9g | 1g | 9g | 16g | 0.07g |

*Save to My Good Food* | *Print*


## Ingredients

- 1 small aubergine, trimmed and cut into chunks
- 1 courgette trimmed and cut into chunks
- 1 red onion, thinly sliced
- 2 garlic cloves, unpeeled and left whole
- 1 tbsp olive oil
- 200g tomatoes
- 175g penne pasta
- good handful basil leaves

## Method

1. Heat oven to 200C/fan 180C/gas 6. Tip the veg and garlic into a roasting tin. Drizzle over the oil, then season and toss together. Roast for 20 mins, add the tomatoes, then roast for a further 10 mins.

2. Cook the pasta, drain and reserve 4 tbsp of water. Tip pasta, water and basil into the veg and toss. Squeeze over the soft garlic, to serve.

*Recipe from Good Food magazine, July 2007*

## Alternative recipes

Roasted ratatouille pasta
(0 ratings)

Roasted ratatouille chicken
(44 ratings)

Rosemary chicken with oven-roasted ratatouille
(33 ratings)

katysimpson | *9th Sep, 2012*

I stir pesto through the pasta before adding and add some fresh salmon in with the veg to roast and its lovely.

*Sign in or register to post comments*

sbloom | *4th Sep, 2012*

I added a little pesto and instead of using aubergine I used mushrooms which meant it wasn't really ratatouille but good all the same!

*Sign in or register to post comments*


Reader offer: £10 off craft beer

*Get a crate of eight hand-crafted beers for just £14 with free p&p.*


Reader Offer: Delicious

Dealing with NA; Derive Variables; delete those predictors when predicting

| id | url | recipe | country | comment | ratingnumber | average_ratingvalue | kcalories | protein | carbs | fat | saturates | fibre | sugar |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | http://www | 'Panforte' | europe | 3 | 2 | 4.375 | 294 | 18 | 7 | 28 | 16 | 2 | 6 |
| 2 | http://www | 'Butter p | europe | 3 | 2 | 4.375 | 429 | 9 | 39 | 26 | 12 | 4 | 5 |
| 3 | http://www | 'Doved' | america | 0 | 0 | 0 | 245 | 14 | 16 | 12 | 6 | 6 | 6 |
| 4 | http://www | 10-minute | other | 72 | 76 | 4.75 | 327 | 13 | 33 | 17 | 5 | 2 | 7 |
| 5 | http://www | 10-minute | asia | 21 | 22 | 3.977275 | 494 | 37 | 69 | 10 | 2 | 4 | 9 |
| 6 | http://www | 10-minute | america | 1 | 1 | 3 | 686 | 48 | 66 | 28 | 10 | 5 | 12 |
| 7 | http://www | 10-minute | asia | 4 | 2 | 2.5 | 546 | 28 | 68 | 20 | 3 | 3 | 9 |
| 8 | http://www | 10-minute | europe | 5 | 4 | 3.75 | 482 | 20 | 8 | 62 | 0 | 4 | 18 |
| 9 | http://www | 10-minute | europe | 2 | 2 | 5 | 109 | 4 | 1 | 13 | 0 | 2 | 6 |
| 10 | http://www | 10-minute | europe | 2 | | | | 16 | 18 | 9 | 1 | 5 | 0 |

Data we collect

| salt | serving | level | step | ngredient | cooktime | preptime | datePublis | cookmethod | cate_Dinner | cate_Main.course | cate_Side.dish |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.2 | 12 | Easy | 174 | 15 | 30 | 15 | 2013/12/1 | NA | 0 | 0 | 0 |
| 0.9 | 10 | Moderately easy | 286 | 11 | 90 | 20 | 2013/10/1 | Baked | 0 | 1 | 0 |
| 2.1 | 2 | Easy | 93 | 8 | 10 | 5 | 2015/8/1 | Pan fried | 0 | 0 | 1 |
| 0.88 | 2 | Easy | NA | 8 | 0 | 10 | 2009/8/1 | NA | 0 | 1 | 0 |
| 2.91 | 2 | Easy | 95 | 9 | 5 | 5 | 2009/5/1 | NA | 1 | 1 | 0 |
| 1.43 | 1 | Easy | 92 | 8 | 10 | 0 | 2007/1/1 | NA | 1 | 1 | 1 |
| 2.31 | 2 | Easy | 98 | 6 | 0 | 0 | 2003/10/1 | NA | 0 | 0 | 0 |
| 1.5 | 2 | Easy | 78 | 5 | 0 | 0 | 2005/3/1 | NA | 1 | 1 | 0 |
| 0.52 | 12 | Easy | 111 | 7 | 0 | 10 | 2004/8/1 | NA | 1 | 1 | 0 |
| 1.2 | 4 | Easy | NA | 7 | 0 | 10 | 2005/8/1 | NA | 1 | 1 | 0 |

| id | url | |
|---|---|---|
| 1 | nforte-pies | |
| 2 | les-cheese | ‘Butter pie’ |
| 3 | loved-peas | |
| 4 | cous-salad | 10- |
| 5 | e-pad-thai | |
| 6 | eese-wrap | )-minute ste |
| 7 | dle-supper | 10-minute |
| 8 | e-tortellini | |
| 9 | ean-toasts | 10-m |
| 10 | bean-salad | 10-mi |

| salt | binned_serving |
|---|---|
| 0.2 | 4 |
| 0.9 | 3 |
| 2.1 | 2 |
| 0.88 | 2 |
| 2.91 | 2 |
| 1.43 | 1 |
| 2.31 | 2 |
| 1.5 | 2 |
| 0.52 | 4 |
| 1.2 | 3 |

serving vs ratingValue

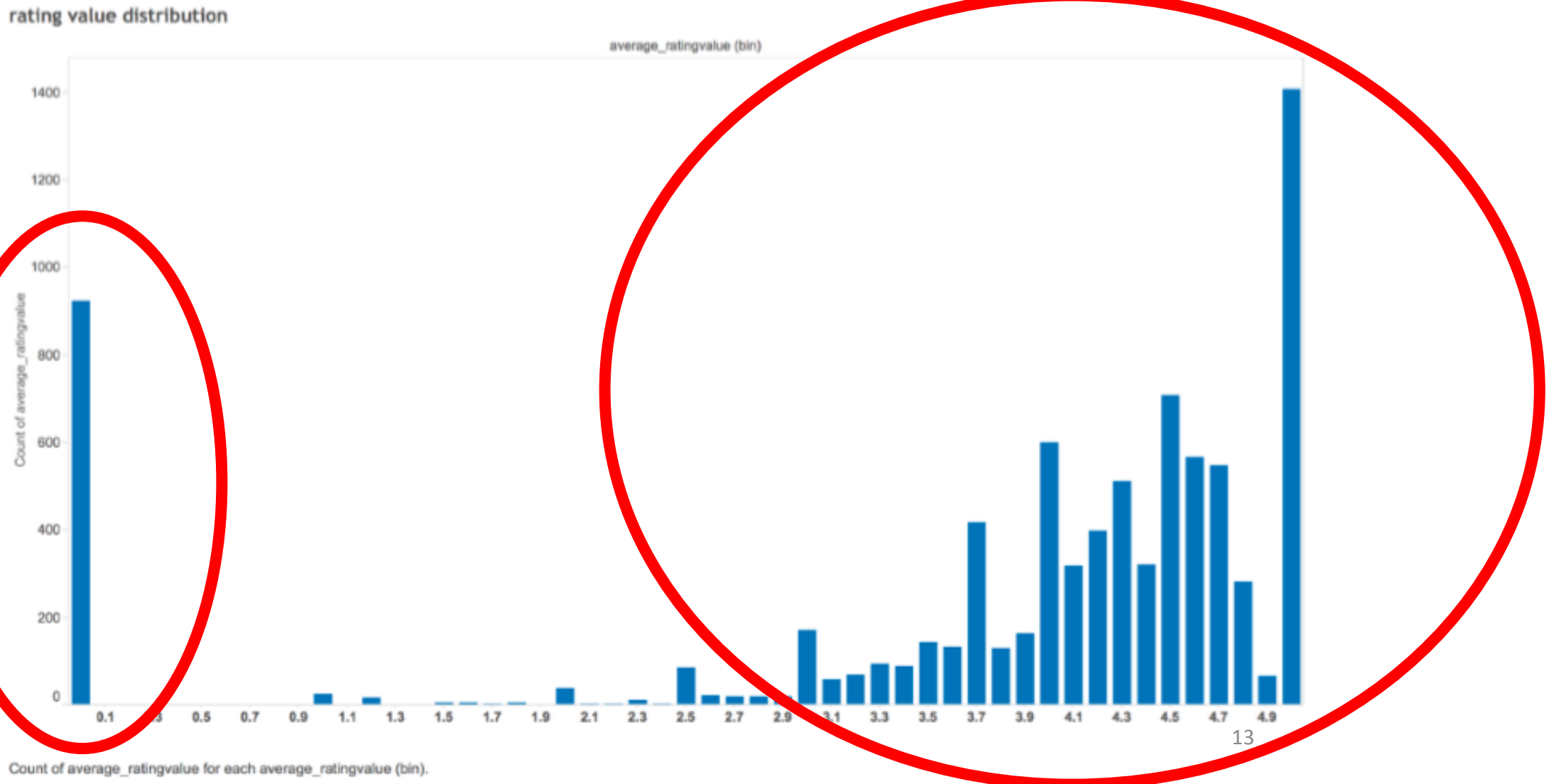

| sugar |
|---|
| 6 |
| 5 |
| 6 |
| 7 |
| 9 |
| 12 |
| 9 |
| 18 |
| 6 |
| 0 |

| Side.dish |
|---|
| 0 |
| 0 |
| 1 |
| 0 |
| 0 |
| 1 |
| 0 |
| 0 |
| 0 |
| 0 |

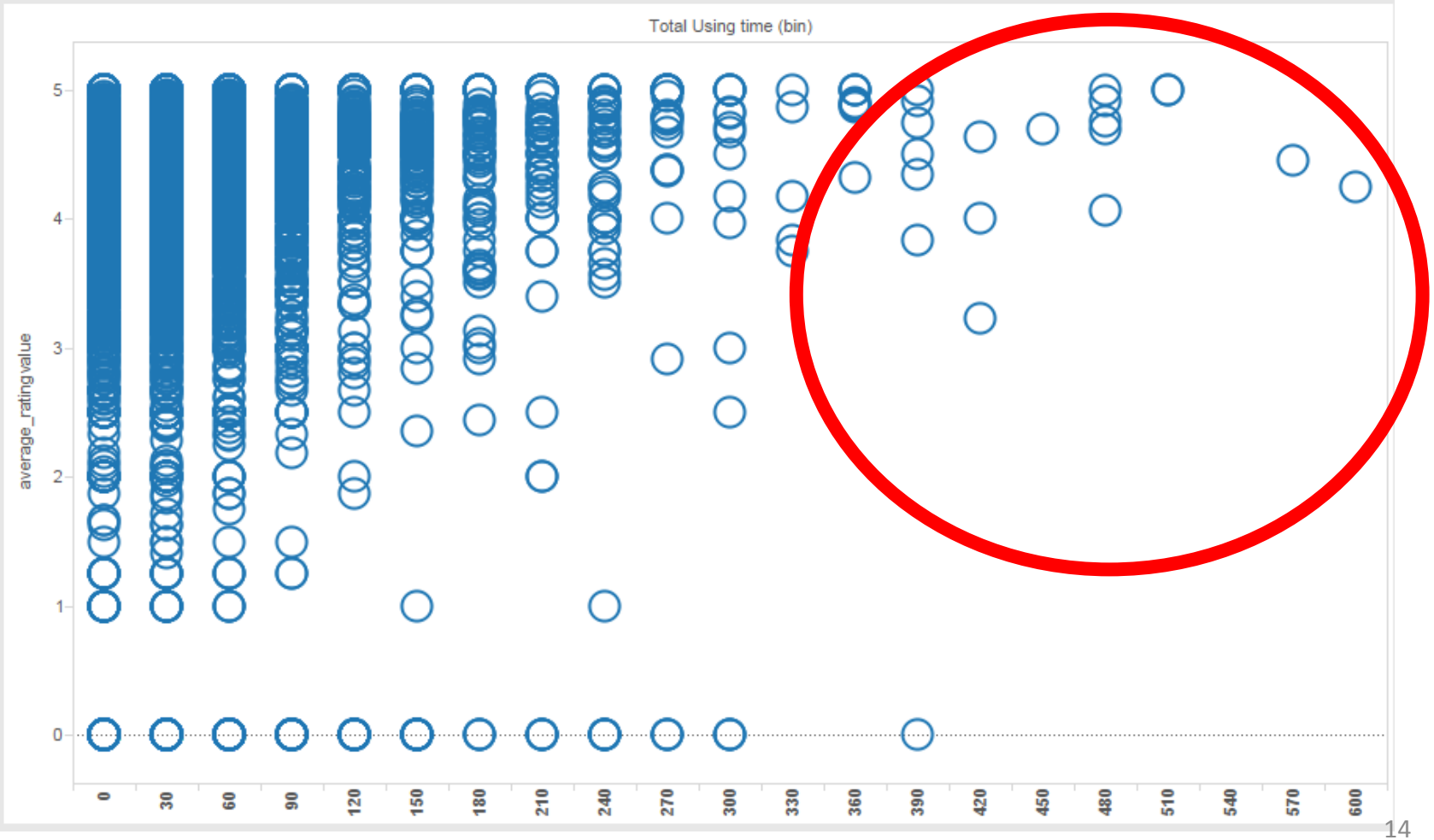
Average_ratingvalue vs. serving. The view is filtered on serving, which ranges from 1.00 to 20.00.

# Data Exploration

# Average Rating Value distribution

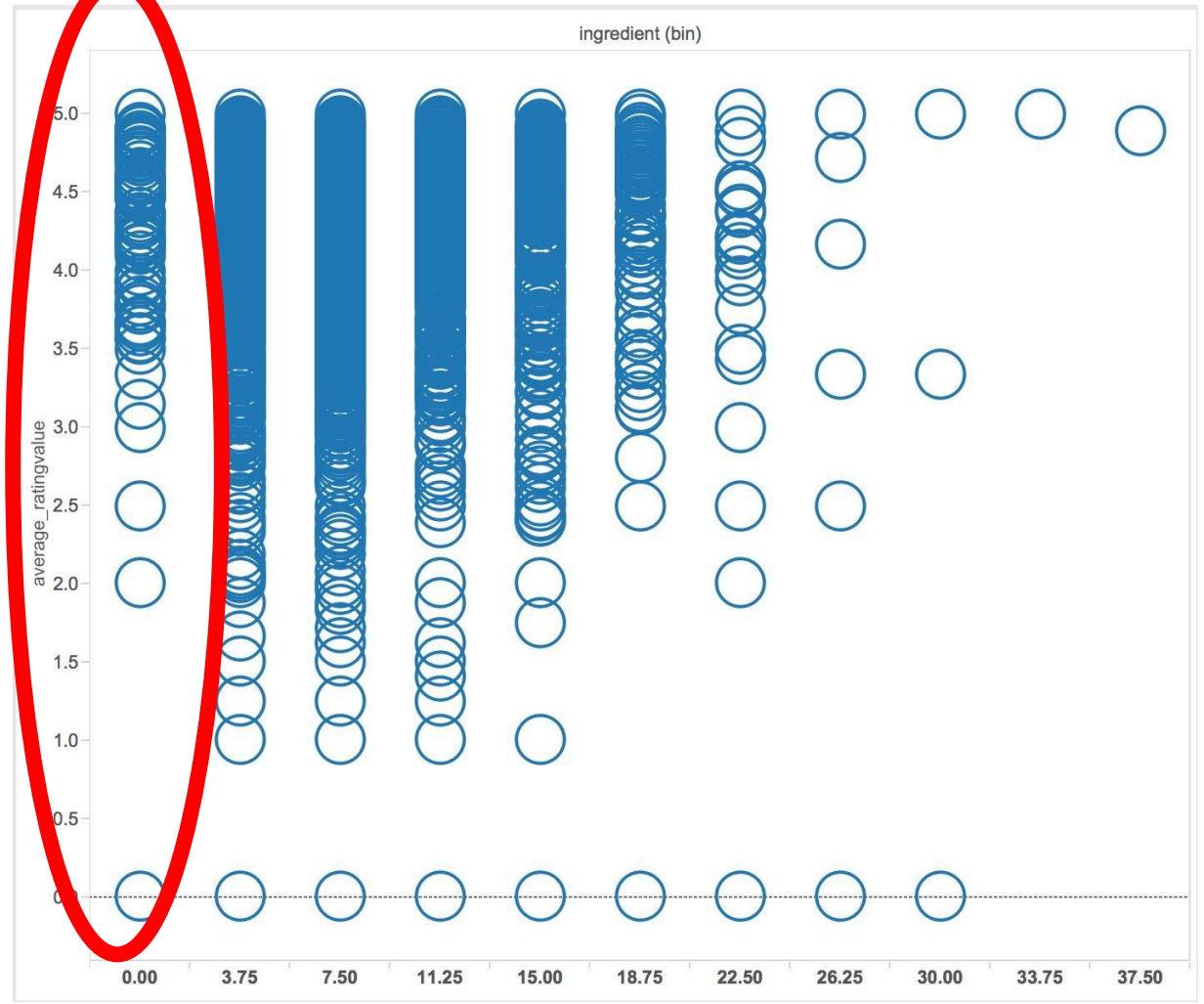


# Total Using time (Cook time +Prepare time ) vs AVG Rating value



# Cooking time vs Country

# Ingredient vs Average Rating Value



# Modeling

# Method

- Which methods (Supervised/Numerical)
  - **PCA**
  - **Decision Tree**
  - **Regression**
  - **KNN**

# PCA

PC1 Score > |0.18|

1. kcalories
2. carbs
3. fat
4. saturates
5. sugar
6. Step
7. ingredient
8. Total_time
9. cate_Side.dish
10. cate_Dessert
11. level_Easy
12. level_Moderately easy

**Principal Components**

| Feature\Cc | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| kcalories | -0.360382 | 0.2135661 | -0.055749 | 0.0740914 | -0.126255 | 0.0831524 | -0.062651 | 0.0984474 | -0.072925 | 0.084075 |
| saturates | -0.314791 | 0.0187841 | -0.094939 | 0.0583895 | -0.122726 | -0.02801 | -0.072783 | 0.0333066 | -0.109549 | 0.1778462 |
| step_fix | -0.305228 | -0.088956 | -0.042564 | -0.200891 | 0.1683691 | -0.009394 | 0.1042846 | 0.0144485 | 0.0077521 | -0.189111 |
| fat | -0.28242 | 0.1703323 | -0.177788 | 0.0503788 | -0.227387 | -0.041854 | -0.067245 | 0.0270962 | -0.142331 | 0.1789507 |
| ingredient | -0.258852 | 0.0883467 | 0.1152498 | -0.131407 | 0.2334604 | 0.0862629 | -0.005411 | 0.0196304 | -0.046383 | -0.300141 |
| total_time | -0.249291 | 0.003518 | 0.0498074 | -0.177978 | 0.1159491 | -0.068563 | 0.0370586 | -0.093095 | -0.011838 | -0.225029 |
| level_Modera | -0.247228 | -0.099153 | -0.09018 | -0.26252 | 0.1826946 | -0.108702 | 0.1688936 | 0.0912196 | 0.2265115 | 0.2987548 |
| sugar | -0.243675 | -0.138516 | 0.0963765 | 0.1958007 | -0.111182 | 0.1178058 | -0.178049 | 0.1867852 | 0.1127815 | -0.015503 |
| cate_Dessert | -0.190699 | -0.248172 | 0.029359 | 0.1736357 | -0.037069 | 0.0171582 | -0.185734 | 0.0868247 | 0.1015246 | 0.1375141 |
| carbs | -0.182505 | -0.017415 | 0.1014272 | 0.159428 | 0.0420721 | 0.2477789 | -0.161892 | 0.2210445 | 0.2050901 | -0.048757 |
| protein | -0.143459 | 0.3486759 | -0.002257 | -0.107515 | -0.014775 | -0.005696 | 0.0990458 | -0.092066 | -0.180468 | -0.009703 |
| cate_Afternoc | -0.125445 | -0.267551 | -0.021438 | 0.1227785 | 0.0498123 | 0.1320839 | -0.13043 | -0.015142 | -0.199684 | -0.105271 |
| level_For the | -0.124919 | -0.060851 | -0.03963 | -0.116279 | 0.0864124 | 0.0002175 | 0.0507817 | 0.0232107 | -0.022784 | -0.243319 |
| cate_Treat | -0.119939 | -0.27929 | -0.007661 | 0.1749271 | 0.0459901 | 0.14005 | -0.104136 | -0.007899 | -0.189143 | -0.053416 |
| country_euro | -0.088086 | -0.164461 | -0.35417 | 0.0138476 | -0.38684 | -0.25305 | 0.0969212 | -0.164346 | 0.0927294 | -0.223883 |
| salt | -0.060362 | 0.2185025 | -0.064813 | -0.02614 | 0.0037885 | 0.1501453 | 0.0642446 | -0.034522 | -0.120893 | -0.068326 |
| cate_Main.cou | -0.053878 | 0.3716351 | 0.0308265 | -0.083075 | -0.012066 | -0.006782 | 0.039511 | -0.063325 | -0.176862 | -0.028931 |
| fibre | -0.038945 | 0.2339494 | 0.0236637 | 0.1010814 | -0.027359 | 0.1345353 | -0.09345 | 0.0793135 | 0.2656919 | -0.191285 |
| binned_servin | -0.034204 | -0.24556 | -0.005365 | -0.011747 | 0.1739924 | -0.103632 | -0.067899 | -0.143029 | -0.184873 | -0.1214 |
| cate_Pasta.cou | -0.014453 | 0.0811323 | -0.147946 | 0.0482313 | -0.205003 | -0.029909 | -0.065546 | 0.0648844 | -0.061313 | 0.0945475 |
| cate_Fish.Cou | -0.005674 | 0.036961 | -0.008188 | -0.025242 | -0.005287 | -0.008981 | 0.0440944 | -0.017487 | 0.0134574 | -0.096078 |
| cate_Dinner | 0.0022581 | 0.2737705 | -0.045219 | -0.039816 | 0.0156415 | -0.136022 | -0.127384 | 0.0360485 | 0.0993242 | 0.0802942 |
| country_amer | 0.0091749 | 0.0240558 | 0.1553586 | 0.0539158 | 0.1423804 | 0.2500807 | 0.0116115 | 0.186673 | -0.195326 | 0.1261968 |
| cate_Breakfas | 0.0217191 | -0.065923 | -0.118381 | -0.019459 | -0.068423 | 0.4936909 | 0.3543561 | -0.198952 | 0.1100265 | 0.0596869 |
| cate_Brunch | 0.0282028 | -0.039196 | -0.197018 | -0.024637 | -0.055925 | 0.4876091 | 0.3280159 | -0.139344 | 0.0804853 | 0.0713291 |
| cate_Cocktails | 0.0383541 | -0.041902 | 0.1266348 | 0.0912854 | -0.08662 | -0.131672 | 0.4146803 | 0.4674997 | -0.120815 | -0.076117 |
| cate_Soup.cou | 0.0392375 | 0.0243168 | -0.016004 | 0.0275374 | 0.0105505 | 0.005038 | -0.02797 | 0.0289154 | 0.1911242 | -0.463175 |
| country_othe | 0.0392433 | 0.0780326 | 0.1235398 | -0.006094 | 0.1300898 | 0.0434801 | -0.070759 | 0.0612546 | 0.1118553 | 0.3039909 |
| cate_Drink | 0.0544289 | -0.05503 | 0.1419064 | 0.1100826 | -0.107271 | -0.141416 | 0.4283064 | 0.442154 | -0.100157 | -0.057542 |
| cate_Canapes | 0.0676781 | -0.064152 | -0.109352 | -0.046555 | 0.1652882 | -0.103857 | 0.0162526 | -0.066077 | -0.33237 | 0.0908161 |
| cate_Vegetabl | 0.072078 | 0.017461 | -0.16549 | -4.94E-05 | 0.1276891 | -0.0506 | -0.090376 | 0.1671435 | 0.2583436 | 0.0129239 |
| country_asia | 0.0783915 | 0.1378548 | 0.2493437 | -0.054161 | 0.2962924 | 0.1040023 | -0.083654 | 0.0265094 | -0.047815 | -0.019012 |
| cate_Buffet | 0.081429 | -0.074931 | -0.302087 | -0.068882 | 0.2244705 | -0.044871 | -0.107059 | 0.1392175 | -0.269418 | 0.0648211 |
| cate_Supper | 0.0844199 | 0.1390441 | -0.374174 | -0.009688 | 0.0544977 | 0.1062992 | -0.123184 | 0.2921002 | -0.037435 | -0.104309 |
| cate_Lunch | 0.0923965 | 0.1246644 | -0.337138 | -0.036813 | 0.0755624 | 0.0533722 | -0.096597 | 0.2662113 | 0.0868296 | -0.163396 |
| cate_Starter | 0.0997025 | -0.120701 | 0.1281712 | -0.509927 | -0.306646 | 0.1379335 | -0.199507 | 0.1574011 | -0.045317 | -0.007539 |
| cate_Condime | 0.0997025 | -0.120701 | 0.1281712 | -0.509927 | -0.306646 | 0.1379335 | -0.199507 | 0.1574011 | -0.045317 | -0.007539 |
| cate_Snack | 0.1022751 | -0.077858 | -0.327111 | -0.027999 | 0.1268665 | 0.1966481 | -0.048834 | 0.1607868 | -0.220693 | -0.00222 |
| cate_Side.dish | 0.1960632 | -0.043801 | -0.166746 | -0.000141 | 0.1885313 | -0.089835 | -0.044347 | 0.0020576 | 0.2750637 | 0.1038185 |
| level_Easy | 0.2817352 | 0.1171655 | 0.1004627 | 0.2928089 | -0.205905 | 0.1024718 | -0.179054 | -0.095071 | -0.204861 | -0.187423 |

# PCA

## KNN

**Training Data Scoring - Summary Report (for k = 19)**

| Total sum of squared errors | RMS Error | Average Error |
|---|---|---|
| 1.2745E-27 | 6.77702E-16 | -1.07222E-17 |

**Validation Data Scoring - Summary Report (for k = 19)**

| Total sum of squared errors | RMS Error | Average Error |
|---|---|---|
| 6263.10433 | 1.502323046 | -0.035436414 |

**Test Data Scoring - Summary Report (for k = 19)**

| Total sum of squared errors | RMS Error | Average Error |
|---|---|---|
| 6153.5106 | 1.467338675 | 0.039820815 |

## TREES

**Training Data scoring - Summary Report (Using Full-Grown Tree)**

| Total sum of squared errors | RMS Error | Average Error |
|---|---|---|
| 5146.7402 | 1.3618668 | 5.13703E-16 |

**Validation Data scoring - Summary Report (Using Full-Grown Tree)**

| Total sum of squared errors | RMS Error | Average Error |
|---|---|---|
| 6982.6545 | 1.5862761 | -0.04575178 |

**Test Data scoring - Summary Report (Using Full-Grown Tree)**

| Total sum of squared errors | RMS Error | Average Error |
|---|---|---|
| 6590.3266 | 1.5185264 | -0.01962091 |

## Regression

**Training Data Scoring - Summary Report**

| Total sum of squared errors | RMS Error | Average Error |
|---|---|---|
| 1019.54 | 0.64247 | 1.77096E-15 |

**Validation Data Scoring - Summary Report**

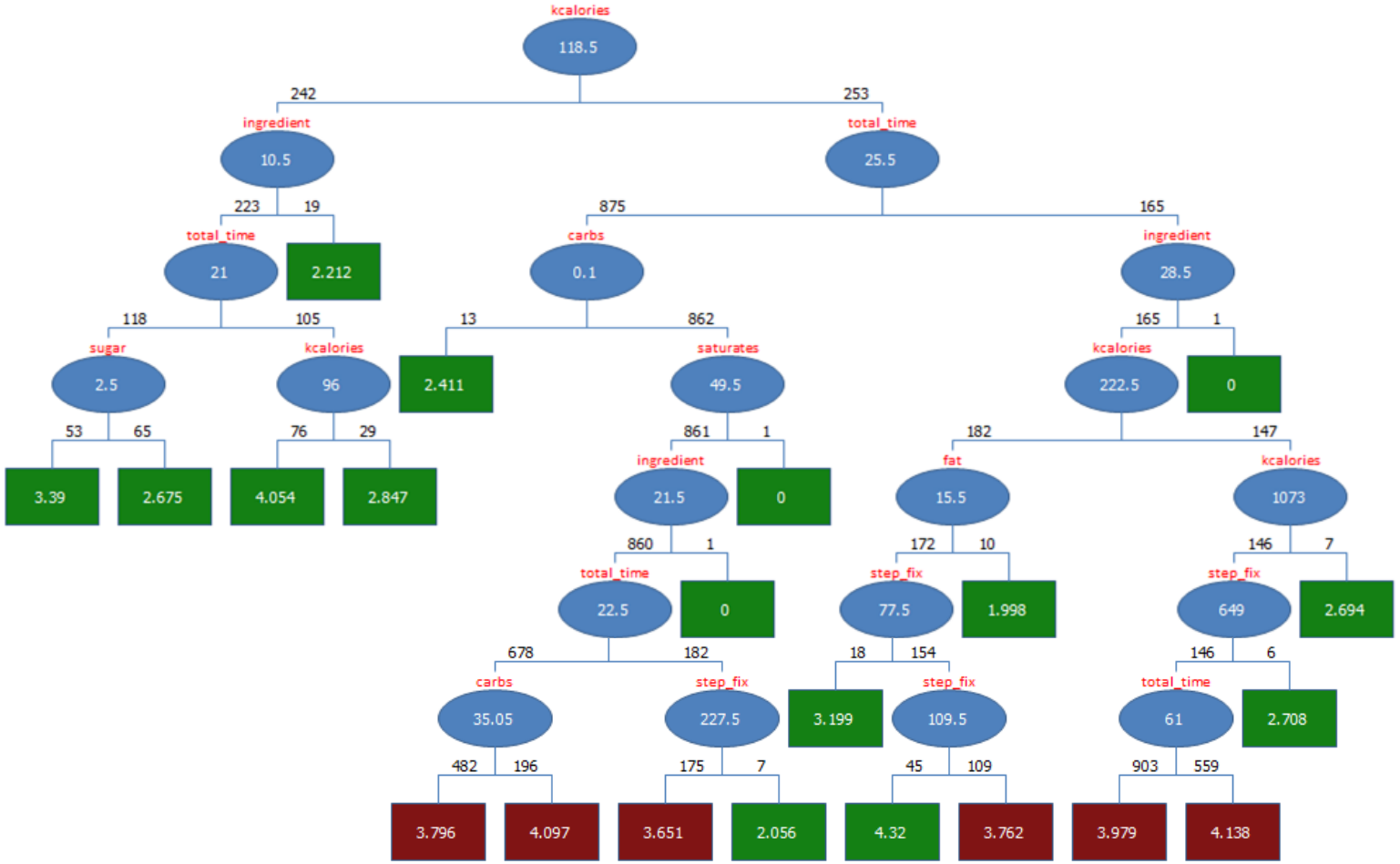
| Total sum of squared errors | RMS Error | Average Error |
|---|---|---|
| 1181.97 | 0.69176 | -0.00297033 |

**Test Data Scoring - Summary Report**

| Total sum of squared errors | RMS Error | Average Error |
|---|---|---|
| 1087.89 | 0.65368 | 0.024141097 |

# Decision Tree

**PCA**



Explore

Run

With PCA Variable selection

| Validation | Total SSE | RMS Error | Aver. Error |
|---|---|---|---|
| KNN | 6263.10 | 1.50 | -0.036 |
| Trees | 6982.65 | 1.59 | -0.046 |
| Regression | 1181.97 | 0.69 | -0.003 |

# STEPWISE

1. saturates
2. salt
3. total_time
4. cate_Dinner
5. cate_Side.dish
6. cate_Afternoon.tea
7. cate_Supper
8. cate_Starter
9. country_america
10. country_asia
11. country_europe
12. country_other
13. level_Moderately easy

**Variable Selection**

| Subset Link | #Coeffs | RSS | Cp | R? | Adjusted R? | Probabili |
|---|---|---|---|---|---|---|
| Choose Subset | 1 | 1197.5498 | 34552.5425 | -13.5782 | -13.5782 | |
| Choose Subset | 2 | 219.672 | 4324.8472 | -1.6741 | -1.6752 | |
| Choose Subset | 3 | 181.1578 | 3136.2362 | -1.2053 | -1.2071 | |
| Choose Subset | 4 | 123.4563 | 1354.4759 | -0.5029 | -0.5047 | |
| Choose Subset | 5 | 116.8644 | 1152.6964 | -0.4226 | -0.4249 | |
| Choose Subset | 6 | 115.8343 | 1122.8533 | -0.4101 | -0.413 | |
| Choose Subset | 7 | 114.7643 | 1091.7748 | -0.3971 | -0.4005 | |
| Choose Subset | 8 | 113.5367 | 1055.8242 | -0.3821 | -0.386 | |
| Choose Subset | 9 | 112.561 | 1027.6633 | -0.3702 | -0.3747 | |
| Choose Subset | 10 | 111.8447 | 1007.5194 | -0.3615 | -0.3665 | |
| Choose Subset | 11 | 111.3198 | 993.2928 | -0.3551 | -0.3606 | |
| Choose Subset | 12 | 110.8495 | 980.7547 | -0.3494 | -0.3554 | |
| Choose Subset | 13 | 109.2946 | 934.6865 | -0.3305 | -0.337 | |
| Choose Subset | 14 | 80.5617 | 48.4495 | 0.0193 | 0.0141 | 0.0 |
| Choose Subset | 15 | 80.2122 | 39.6447 | 0.0236 | 0.018 | 0.0 |
| Choose Subset | 16 | 79.911 | 32.334 | 0.0272 | 0.0213 | 0.0 |
| Choose Subset | 17 | 79.6862 | 27.3861 | 0.03 | 0.0236 | 0.0 |
| Choose Subset | 18 | 79.5456 | 25.0401 | 0.0317 | 0.025 | 0.1 |
| Choose Subset | 17 | 79.5533 | 23.2768 | 0.0316 | 0.0253 | 0.1 |
| Choose Subset | 16 | 79.5639 | 21.603 | 0.0314 | 0.0255 | 0.1 |
| Choose Subset | 15 | 79.5812 | 20.1407 | 0.0312 | 0.0257 | 0.2 |
| Choose Subset | 14 | 79.6591 | 20.5461 | 0.0303 | 0.0252 | 0.1 |

# Stepwise

## KNN

**Training Data Scoring - Summary Report (for k = 20)**

| Total sum of squared errors | RMS Error | Average Error |
|---|---|---|
| 198.329 | 0.26734 | 0.00015 |

**Validation Data Scoring - Summary Report (for k = 20)**

| Total sum of squared errors | RMS Error | Average Error |
|---|---|---|
| 6413.21 | 1.52022 | -0.03916 |

**Test Data Scoring - Summary Report (for k = 20)**

| Total sum of squared errors | RMS Error | Average Error |
|---|---|---|
| 6364.69 | 1.4923 | 0.02215 |

## TREES

**Training Data scoring - Summary Report (Using Full-Grown Tree)**

| Total sum of squared errors | RMS Error | Average Error |
|---|---|---|
| 5146.4285 | 1.3618256 | -5.56752E-16 |

**Validation Data scoring - Summary Report (Using Full-Grown Tree)**

| Total sum of squared errors | RMS Error | Average Error |
|---|---|---|
| 6815.2486 | 1.5671456 | -0.026397183 |

**Test Data scoring - Summary Report (Using Full-Grown Tree)**

| Total sum of squared errors | RMS Error | Average Error |
|---|---|---|
| 6392.5745 | 1.4955701 | -0.002265808 |

## Regression

**Training Data Scoring - Summary Report**

| Total sum of squared errors | RMS Error | Average Error |
|---|---|---|
| 1008.9 | 0.63911 | 4.09245E-15 |

**Validation Data Scoring - Summary Report**

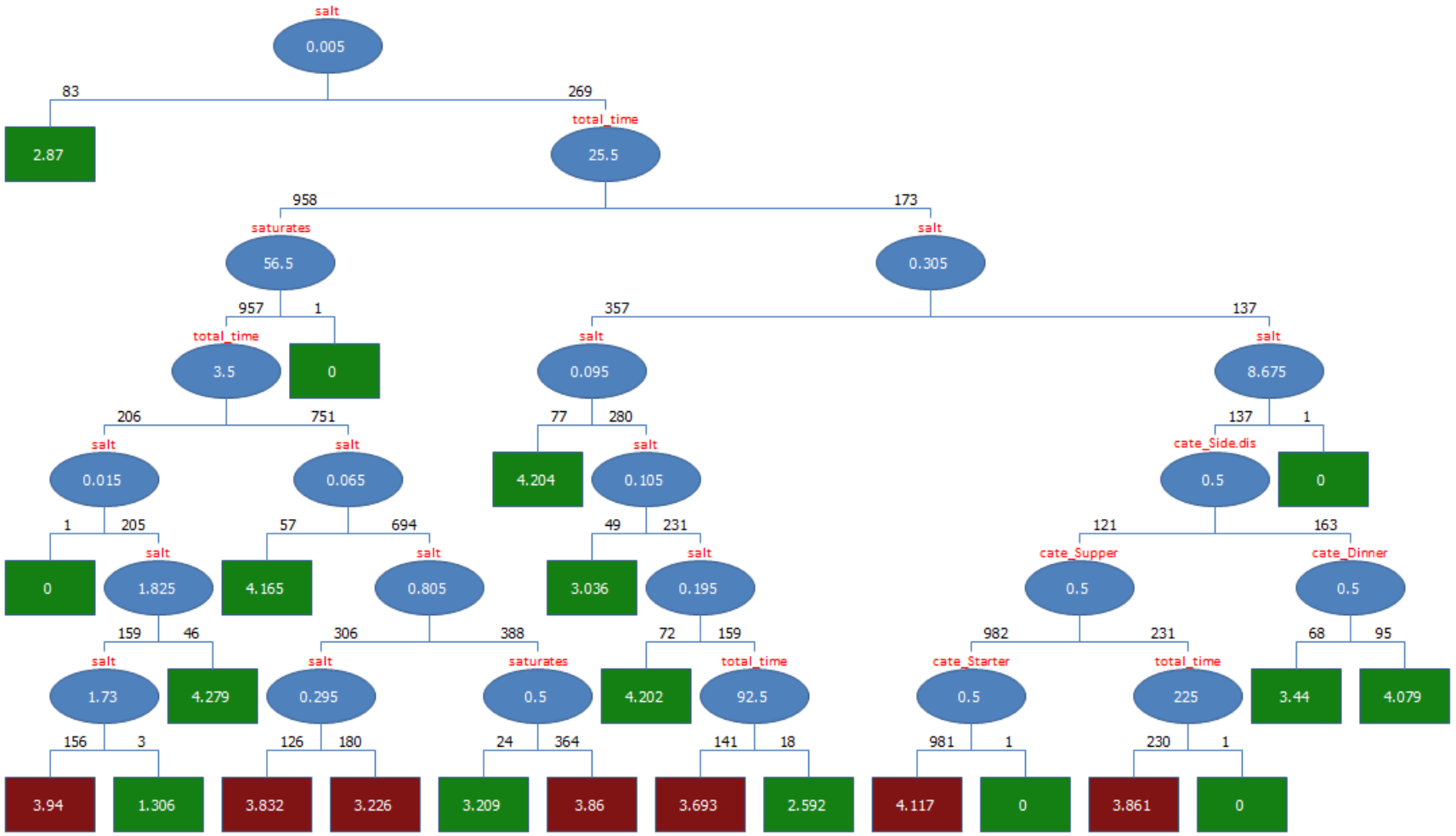
| Total sum of squared errors | RMS Error | Average Error |
|---|---|---|
| 1176.87 | 0.69026 | -0.00173737 |

**Test Data Scoring - Summary Report**

| Total sum of squared errors | RMS Error | Average Error |
|---|---|---|
| 1076.17 | 0.65015 | 0.021656965 |

# Decision Tree

## Stepwise



**Explore**

**Run**

With PCA Variable selection

| Validation | Total SSE | RMS Error | Aver. Error |
|---|---|---|---|
| KNN | 6263.10 | 1.50 | -0.036 |
| Trees | 6982.65 | 1.59 | -0.046 |
| Regression | 1181.97 | 0.69 | -0.003 |

**2nd-Run**

With stepwise Variable selection

| Validation | Total SSE | RMS Error | Aver. Error |
|---|---|---|---|
| KNN | 6413.21 | 1.52 | -0.04 |
| Trees | 6815.25 | 1.57 | -0.026 |
| Regression | 1176.87 | 0.69 | -0.002 |

# Performance Evaluation

Explore

# Comparison

We choose Regression (Stepwise)!

2nd-Run

With stepwise Variable selection

| Validation | Total SSE | RMS Error | Aver. Error |
|---|---|---|---|
| KNN | 6413.21 | 1.52 | -0.04 |
| Trees | 6815.25 | 1.57 | -0.026 |
| Regression | 1176.87 | 0.69 | -0.002 |

# Why Regression?

**1.Error is the smallest**

**2.Specific Value**

-Client want to know the exact value to see the improvement

**3.Coefficient**

-good for explain, not black box

# Detail of Regression

**Regression Model**

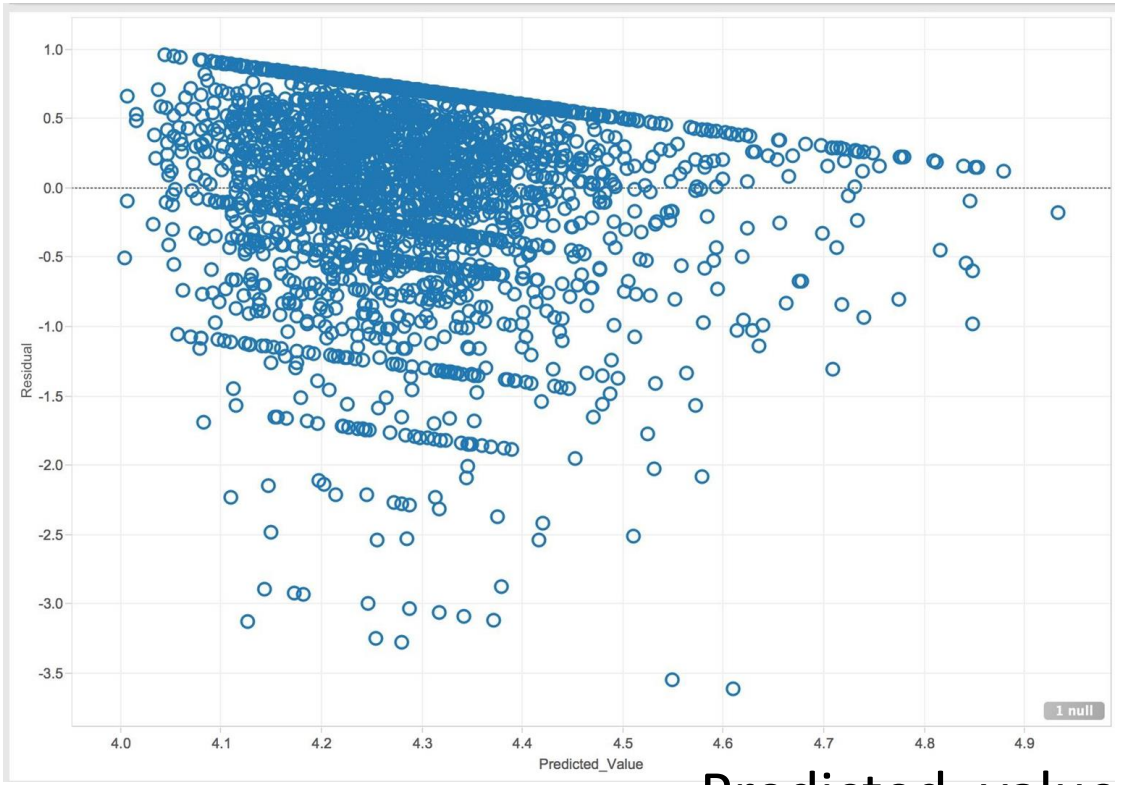
| Input Variables | Coefficient | Std. Error | t-Statistic | P-Value | CI Lower | CI Upper | RSS Reduction |
|---|---|---|---|---|---|---|---|
| Intercept | 4.23268 | 0.057102655 | 74.12409883 | 0 | 4.12071 | 4.34466 | 45480.6 |
| saturates | 0.00289 | 0.001749085 | 1.653735067 | 0.09831 | -0.00054 | 0.00632 | 3.00608 |
| salt | 0.02035 | 0.011977373 | 1.699427094 | 0.08937 | -0.00313 | 0.04384 | 0.00546 |
| total_time | 0.00086 | 0.000276594 | 3.113441923 | 0.00187 | 0.00032 | 0.0014 | 6.51486 |
| cate_Dinner | -0.10653 | 0.028400099 | -3.75097945 | 0.00018 | -0.16222 | -0.05084 | 7.90184 |
| cate_Side.di | 0.12083 | 0.037147827 | 3.252546606 | 0.00116 | 0.04798 | 0.19367 | 2.97768 |
| cate_Aftern | -0.0977 | 0.045347988 | -2.15443874 | 0.0313 | -0.18662 | -0.00878 | 1.61348 |
| cate_Supper | -0.08404 | 0.032720637 | -2.56850645 | 0.01027 | -0.14821 | -0.01988 | 2.6203 |
| cate_Starter | 0.40145 | 0.103865309 | 3.865148356 | 0.00011 | 0.19778 | 0.60513 | 6.31266 |
| country_am | 0.06988 | 0.068381604 | 1.021843048 | 0.30696 | -0.06422 | 0.20397 | 0.41633 |
| country_asi | -0.05537 | 0.060479828 | -0.9155069 | 0.36002 | -0.17397 | 0.06323 | 2.28417 |
| country_eur | 0.03465 | 0.052604697 | 0.658735035 | 0.51013 | -0.0685 | 0.13781 | 0.18954 |
| level_Mode | 0.0933 | 0.040392016 | 2.309759909 | 0.02098 | 0.01409 | 0.1725 | 2.19066 |

| | |
|---|---|
| Residual DF | 2457 |
| R? | 0.03448 |
| Adjusted R? | 0.02977 |
| Std. Error Estimate | 0.6408 |
| RSS | 1008.9 |

P-value < 0.05

# Regression Assumption

Residual




Predicted_value

So, we adjust to logY

# Original

**Training Data Scoring - Summary Report**

| Total sum of squared errors | RMS Error | Average Error |
|---|---|---|
| 1008.9 | 0.63911 | 4.09245E-15 |

**Validation Data Scoring - Summary Report**

| Total sum of squared errors | RMS Error | Average Error |
|---|---|---|
| 1176.87 | 0.69026 | -0.00173737 |

**Test Data Scoring - Summary Report**

| Total sum of squared errors | RMS Error | Average Error |
|---|---|---|
| 1076.17 | 0.65015 | 0.021656965 |

# LN_Regression

**Training Data Scoring - Summary Report**

| Total sum of squared errors | RMS Error | Average Error |
|---|---|---|
| 79.9631 | 0.17993 | 1.68907E-15 |

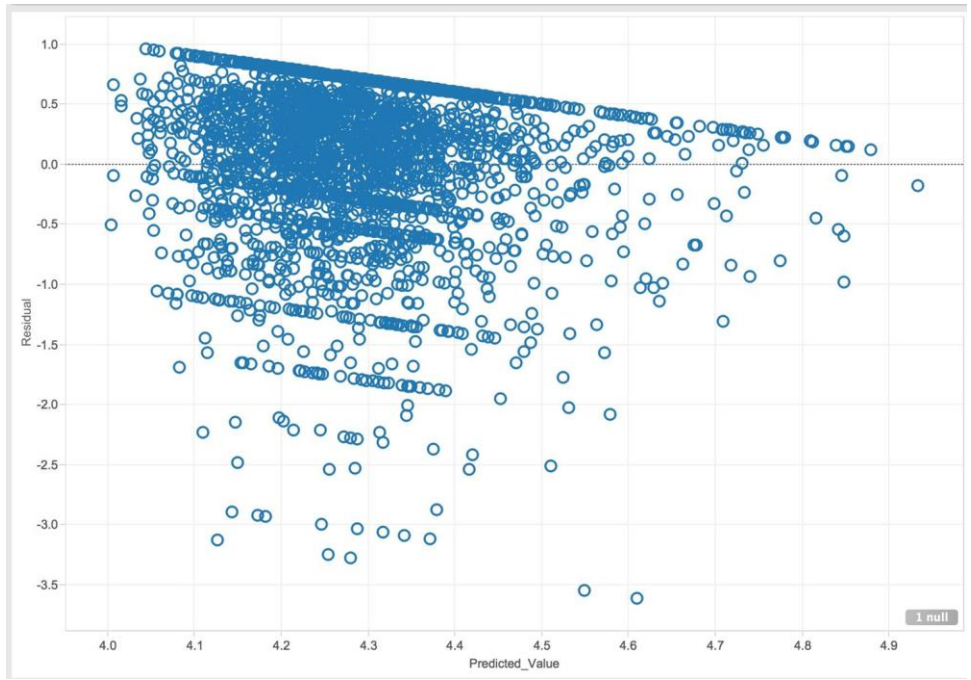
**Validation Data Scoring - Summary Report**

| Total sum of squared errors | RMS Error | Average Error |
|---|---|---|
| 109.218 | 0.21028 | -0.00397528 |

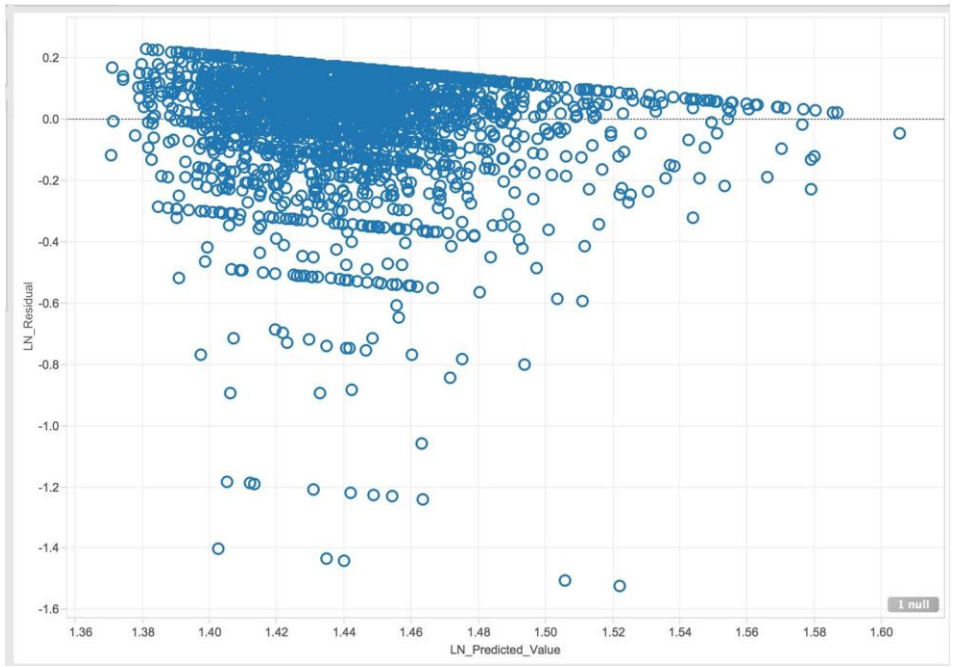
**Test Data Scoring - Summary Report**

| Total sum of squared errors | RMS Error | Average Error |
|---|---|---|
| 88.8314 | 0.18679 | 0.004613984 |

# Original

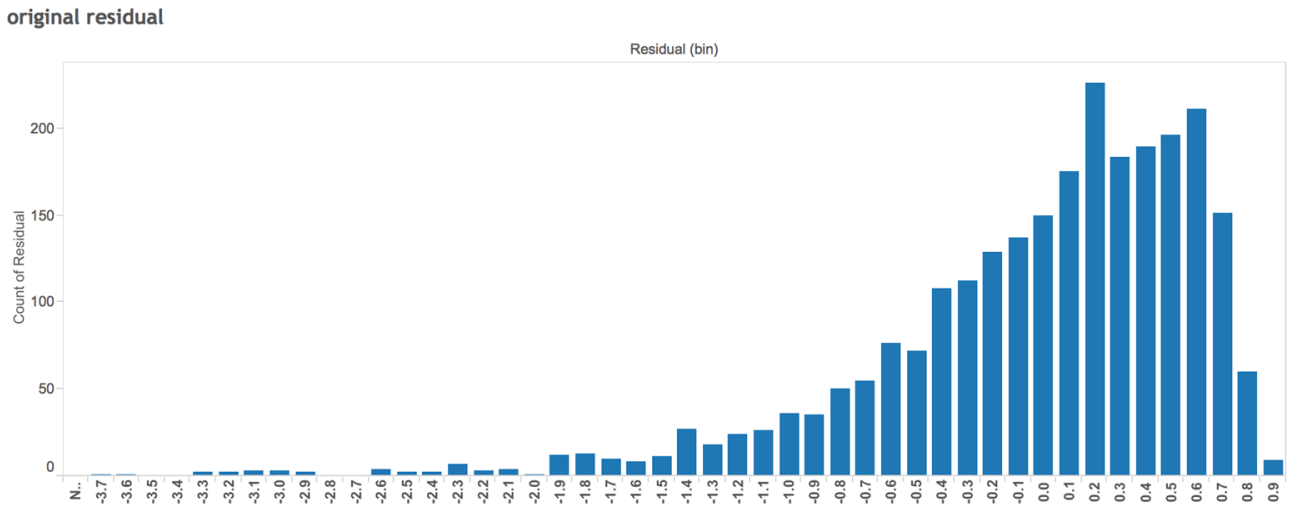


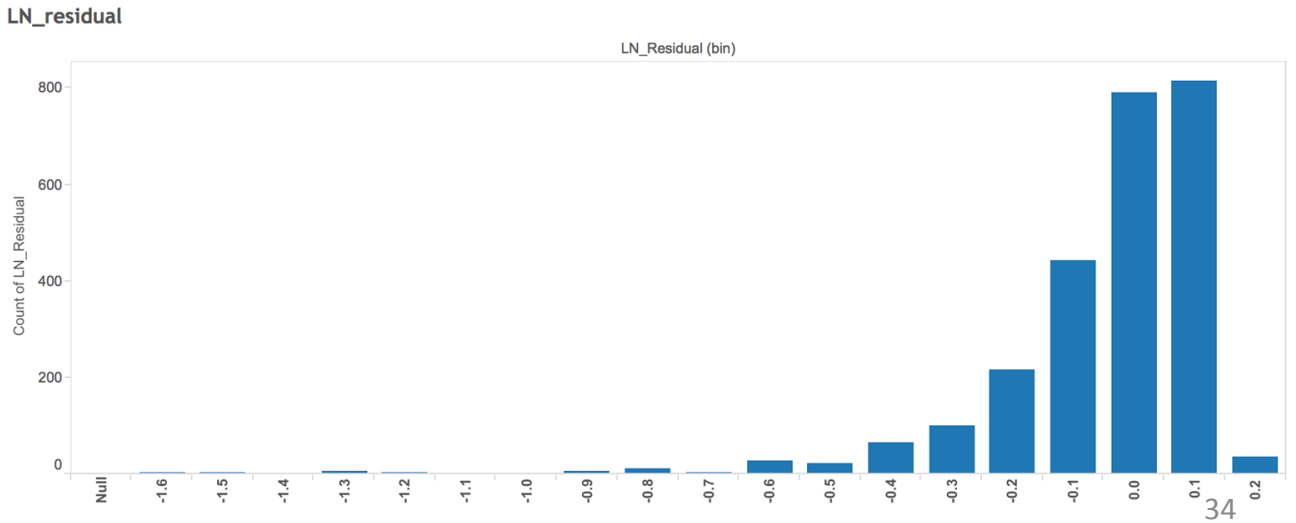

# LN_Regression



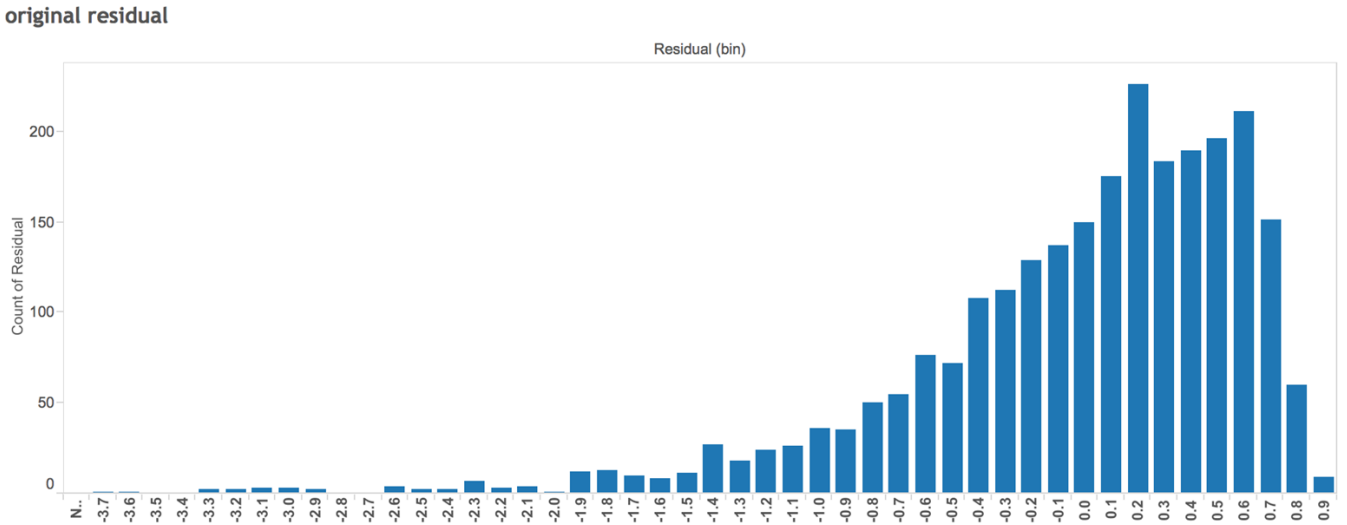
Original

LN_Regression

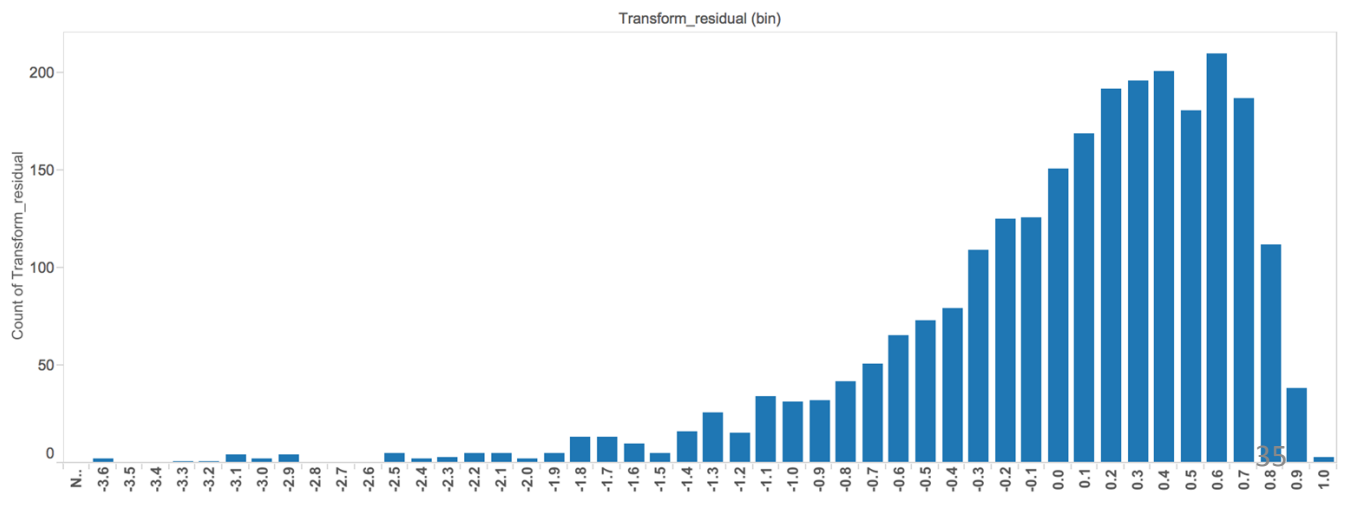


Original




Transformed



# Our Model!!!

log(Y)

= 0.000627*Saturates+0.0051*Salt+0.00022*Total_time

+(0.0254)*Dinner+0.02778*Side_dish+(0.0239)*Afternoon_tea

+(0.02198)*Supper+0.09807*Starter+0.02019*America

+(0.0109)*Asia+0.0117*Europe+0.02168*Level_Moderately easy

# Recommendation


# Recommendations

1. What should the client be aware of when deploying our model?

- Is our model only good for Europe recipe?
  (5,941 Europe recipe/ 8,408 Total Recipe= 70% )
- There are unit of measure in same variable
  Ex. serving:  500 ml, 700g,  4 x 75ml glasses

1. If we were to do this project again, we will.....

- Predict number of comment or people giving rating to recipe
- Forecast (considering time series)
- Build model for each cluster


# Recommendations

## 3. How could we improve our model?

- Get new data
- Find more efficient predictors
- Cross validation
- Ensemble

## 4. If you want to analyze data about "article"...

- Explore your data, know your data, play your data is very very important!!!
- Be aware of error, small might not be small in reality, because the range of rating_value is small.

# Thank you for listening!