

epicurious

Webscrape, EDA, & Hypothesis Test

Eric Wang

Project Description

Web scrape and Hypothesis test on Epicurious recipes. I want to analyze ingredients to see if they are an indicator for nutritional content. Also I would like to see if the tags are reliable for determining nutrition.

Motivation

Our nutrition is a very important aspect of our lives. Also cooking is a an enjoyable experience. It's fun to make delicious meals but even better when you can find balance by eating healthy. Looking at the types of recipes on the popular website epicurious.com can give insight on what is delicious and healthy.

EDA

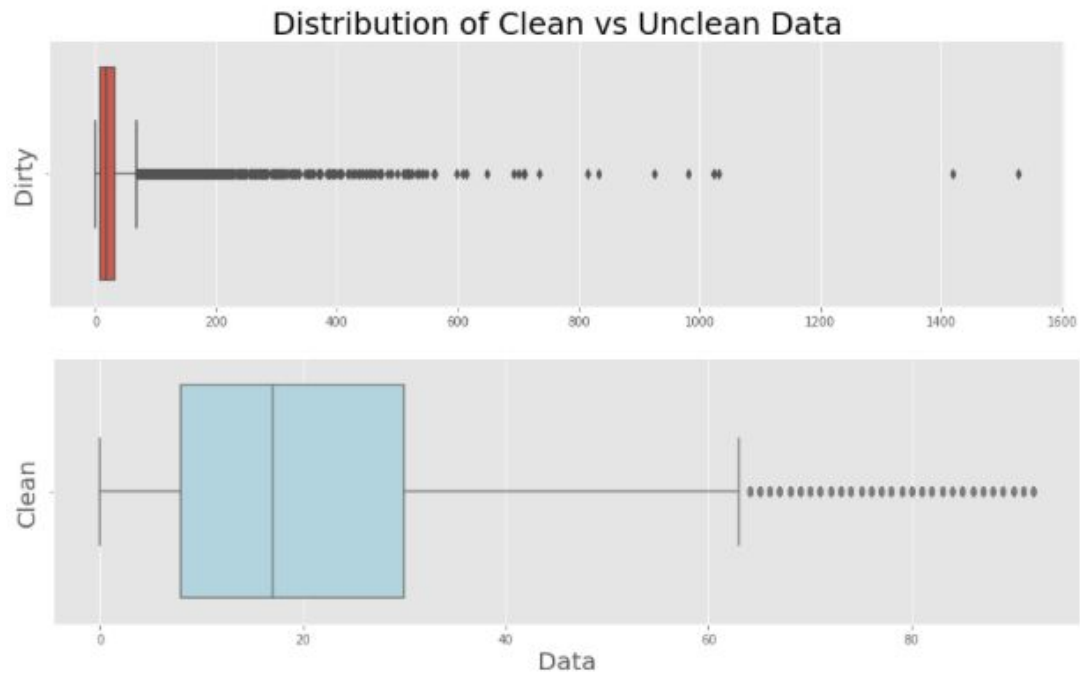
One of the issues was servings per recipes. There were instances when a recipe with high calorie count was notated as one serving. This was unrealistic so I referenced upper limit recommendation to set a threshold. (Tolerable Upper Intake Level (UL): maximum daily intake unlikely to cause adverse health effects.)

WHAT IS YOUR UPPER LIMIT ON FAT FOR THE CALORIES YOU CONSUME?

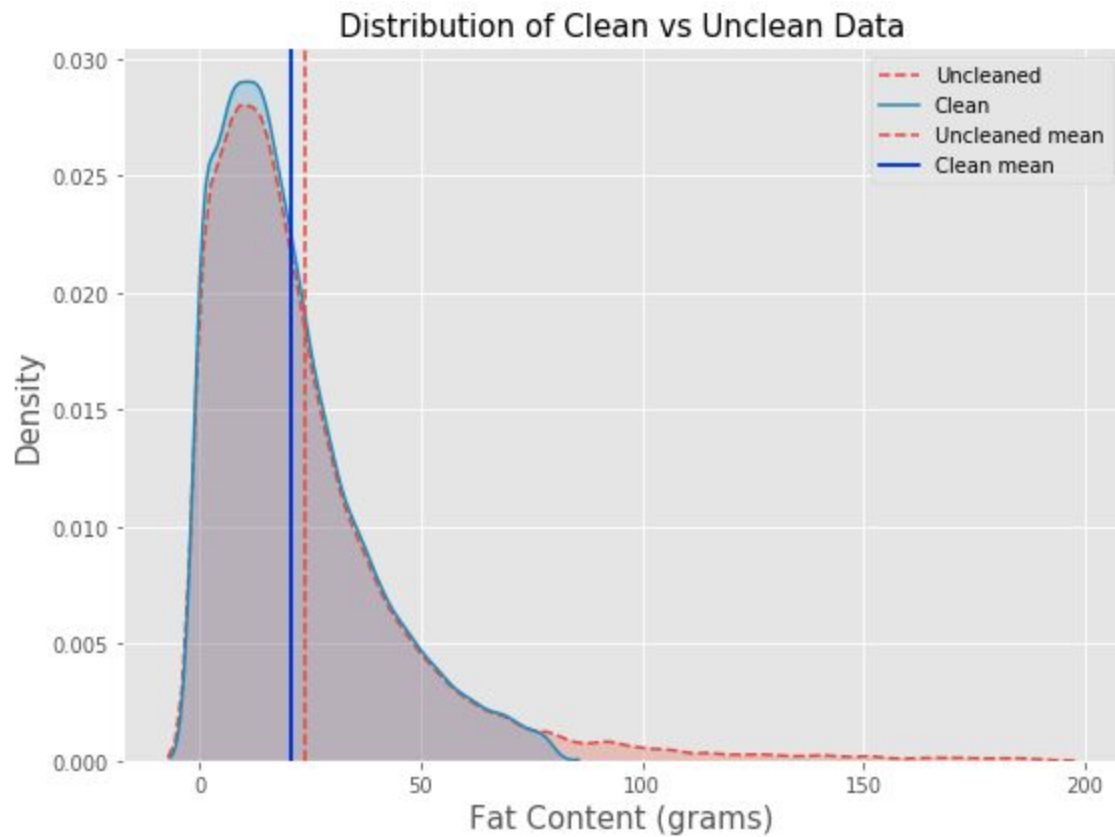
Total Calories per Day	Saturated Fat in Grams	Total Fat in Grams
1,600	18 or less	53
2,000*	20 or less	65
2,200	24 or less	73
2,500*	25 or less	80
2,800	31 or less	93

* Percent Daily Values on Nutrition Facts Labels are based on a 2,000 calorie diet. Values for 2,000 and 2,500 calories are rounded to the nearest 5 grams to be consistent with the Nutrition Facts Label.

EDA



EDA



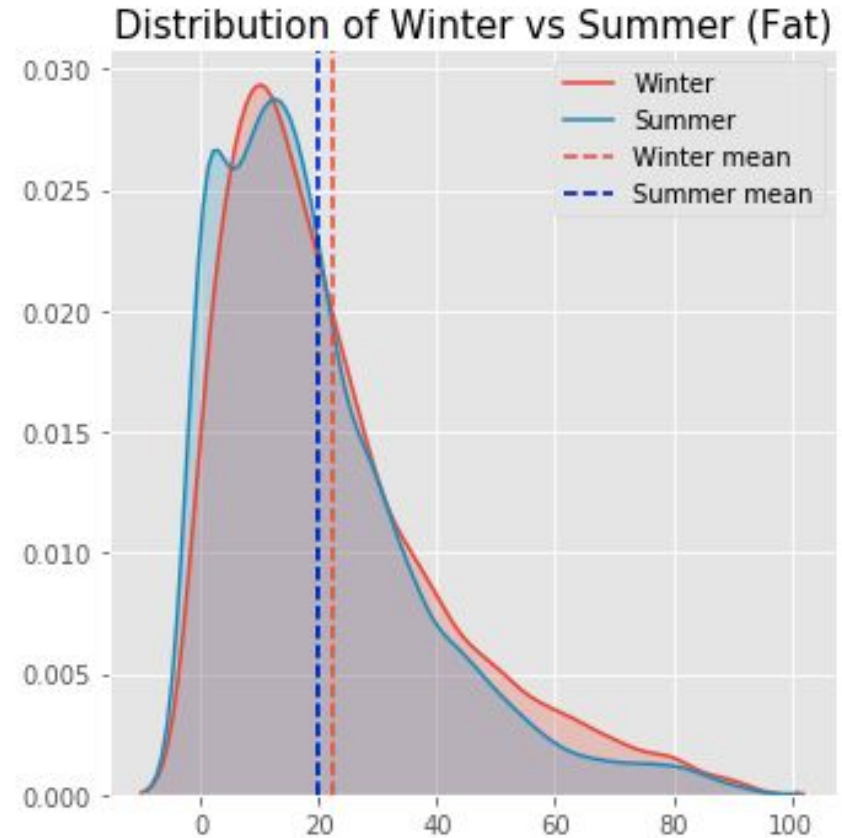
Hypothesis Testing

Hypothesis Test 1: Trying to lose weight is common in the summer. I chose to look at data with summer and winter tags and compare the fat content. I performed a U-test on my hypothesis

H₀: summer fat content is higher than winter.

H₁: winter fat is higher.

P<.02, Reject the null hypothesis (0.000398)



Hypothesis Testing

Hypothesis Test 2: There are a lot of recipes for sweets and desserts on Epicurious. I extracted all the recipes with sugar as an ingredient and compared with those that did not. Before examining the data I would have assumed calorie count on recipes would be higher.

H0: sugar recipe calorie count is same or higher than sugarless.

H1: sugarless recipe calorie count is higher.

P <.02, Reject the null hypothesis

Hypothesis Test 3: I wanted to test something on the healthier side of the spectrum. I filtered the recipes for tags that included the word Vegetarian and compared it with those that did not. This was one of the most popular tags.

H0: Non-vegetarian fiber content is the same or higher

H1: Vegetarian recipes have higher fiber content

P <.02, Reject the null hypothesis

Conclusion

Although some of the data needed to be wrangled, I believe my hypothesis tests shows that the tags can be a reliable source when determining what kind of meals you want to cook. Whether it is a fatty winter dessert or a healthy summer/vegetarian dish, Epicurious is a great resource for delicious meals.

Future considerations:

1. I would look into more features to scrape in the future. I feel like I missed out by not looking into the publish date of the recipes as that would have given a lot of insight in patterns over the years.
2. Scraping for serving size might give a little more insight into the meals, however many of them do not follow a formal scale.
3. There is a lot more data to be cleaned. Specifically the ingredients. I feel that there would have been much more interesting insight.
4. There may be a better scale of caloric and fat intake to use. This will require further research on nutritional values.