

# Recipe Categories with the Smallest Amount of Calories and Points on Skinnytaste.com

## Introduction

Many struggle with finding information on nutritious and diverse recipes daily. This paper examines the Skinnytaste website for healthier recipes. Skinnytaste provides the user with information about calories and points and has divided all recipes into key categories.

This paper examines the distribution of calories and Weight Watcher's points in different recipe categories on the first 50 pages of Skinnytaste. To achieve this, recipe data will be scraped with BeautifulSoup and processed in Google Colab with Python.

## Data Collection

For this paper, the page will be scraped to extract the following variables:

- recipe title
- calories
- points
- images
- categories (keys)
- summaries
- URL links to recipes

The URL link to the recipe and the image is retrieved from the index page while scraping. The rest of the data is scraped from the recipe title page. The image is the smaller thumbnail

provided on the index page since the larger image was wrapped with varying HTML tags on the title page.

In this scraping, we exclude the recipe entries that are not single recipes since they do not provide easily accessible information about points and calories, which are essential for the study. To accommodate this, the scraping ignores entries that do not provide information about Weight Watcher's points.

After the scraping, we are left with 938 recipes from the current content of the 50 first pages of recipes.

## Data Analysis

Let us first examine the overall statistics for the extracted data. The mean for calories is 237.30 with a median of 225.5, the mean for points is 4.97, and the median is 5.

	Calories	Points
count	938.000000	938.000000
mean	237.293497	4.969083
std	116.445717	2.797717
min	8.000000	0.000000
25%	146.625000	3.000000
50%	225.500000	5.000000
75%	308.000000	7.000000
max	608.000000	13.000000

Figure 1: Statistics on Calories and Points

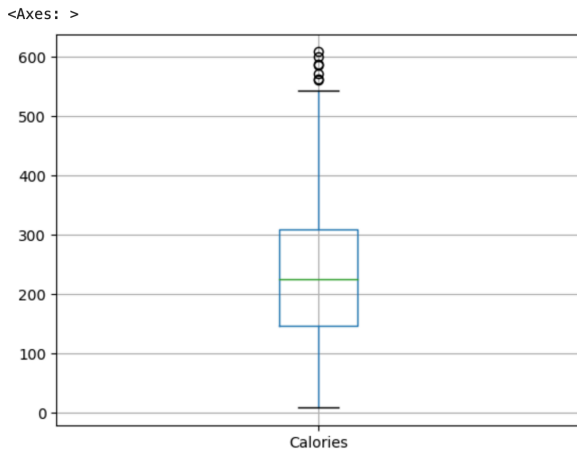


Figure 2: Box plot of calorie distribution

Here is the box plot of the distribution of the calories. Most recipes are under 300 calories, and only a few are outliers with up to 600 calories.

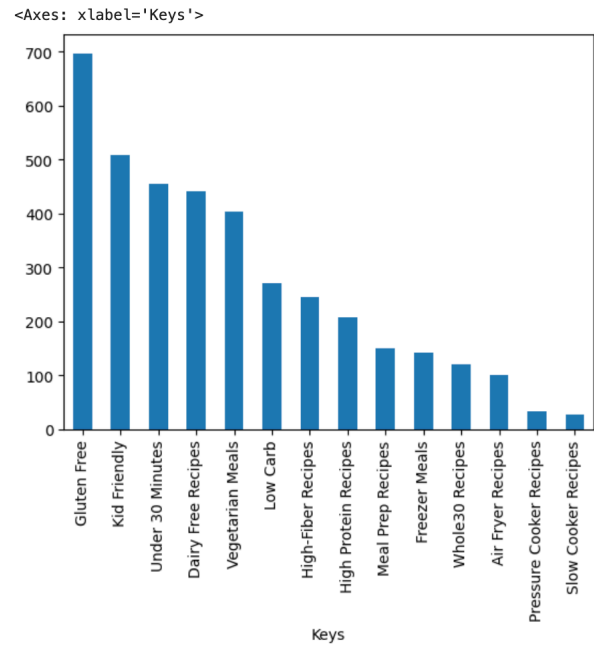


Figure 4: Bar chart of distribution of categories

This bar chart shows the distributions of recipe categories. Of the 938 extracted recipes, almost 700 are gluten-free. It is noteworthy that a recipe often has more than one category. The site offers several options for people with dietary restrictions.

The following bar charts examine the relationship of categories, calories, and points. As one recipe might have many categories, there is some overlap. Yet the analysis points to some trends on the Skinnytaste recipe site.

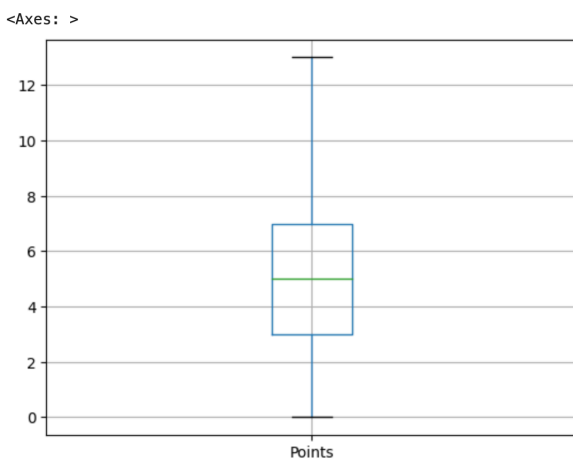


Figure 3: Box plot of points distribution

This boxplot shows the distribution of the Weight Watcher's points. 75 % of the recipes have a maximum of 7 points, up to 13 points.

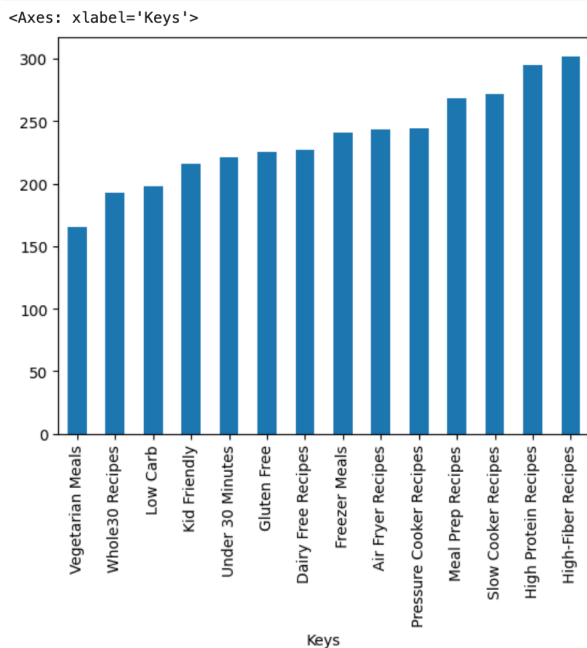


Figure 5: Bar chart of categories' median calories in ascending order

This bar chart ranks the categories in ascending order based on median calorie values. Vegetarian meals have the lowest calories together with Whole30 recipes and low-carb recipes.

High-protein recipes and high-fiber recipes generally have higher calorie values.

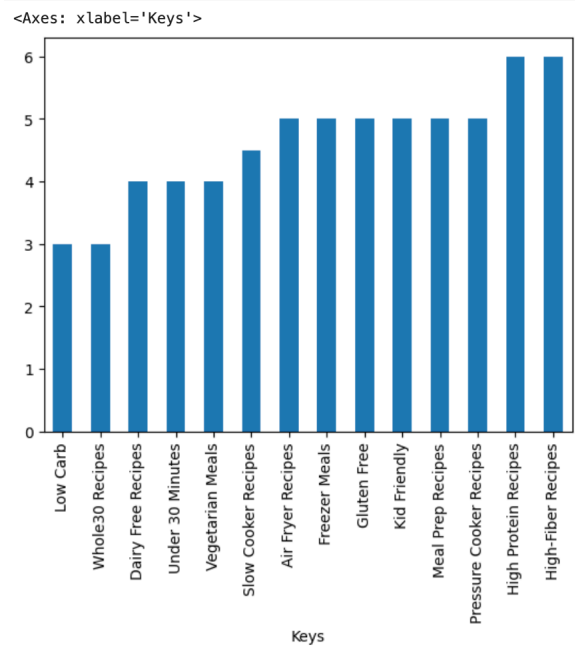


Figure 6: Bar chart of categories' median points in ascending order

In this bar chart, we examine the same phenomena for the distribution of points in different recipe categories.

The lowest median groups are low-carb and Whole30 recipes, with a median of 3 points. Vegetarian meals have a median of 4 points, ranking somewhat higher than in the comparison with the calorie distribution analysis in Figure 5.

High-protein and high-fiber recipes also have the highest point values in this comparison.

## Conclusions

In this paper, we have scraped recipes to compare from the Skinnytaste recipe website.

We have extracted titles, calories, points, summaries, images, and links from the site for analysis. Images were scraped from the index page since there were inconsistencies in the HTML on the title pages on the site regarding the wrapping tags of the image tag.

To handle the issue of some entries on the site being collections of recipes, they were filtered out by excluding entries that lacked information about points.

The distribution of recipe categories based on calorie and point medians was explored, ranking them in ascending order. On Skinnytaste, vegetarian recipes often have fewer calories.

Many interesting aspects could be analyzed with the data frame. Further explorations could examine the relationship between calories and points, as the vegetarian category had the lowest calories, but not the lowest points.