What should I cook tonight?

Md Kaf Shahrier

Introduction:

Making selections about what to prepare every day might be difficult in the fast-paced world of today, particularly if one is trying to keep diet balanced. Finding meals that satisfy particular dietary requirements, such as calorie intake and personal health points, can be especially difficult given the deluge of recipe options accessible online. Many people find it difficult to sort among the various recipes, which frequently results in dissatisfaction and unhealthy eating choices.

It is imperative to address this issue because dietary decisions have a direct impact on one's general health and wellbeing. Eating a balanced diet and consuming the appropriate number of calories can help prevent heart disease, diabetes, and obesity, among other health issues. Furthermore, tailored recipe recommendations can support people in following their dietary objectives and simplify the process of maintaining a healthy lifestyle. Our quality of life can be enhanced and healthier eating habits promoted by streamlining the decision-making process.

The goal of this project is to create an approachable tool that uses data analysis, web scraping and visualization methods to offer customized recipe recommendations. The system is able to provide recommendations that are in line with the user's nutritional objectives by gathering and evaluating data from a well-known recipe website. With this method, recipe information such as names, images, calorie counts and summaries are extracted and the data is filtered and shown according to user-specified nutritional preferences. The ultimate objective is to develop a user-friendly interface that makes meal planning easier and encourages better eating practices.

Data Collection:

The data for this project has been collected from **Skinnytaste** through web scraping of the first 50 pages of recipe index. From each recipe page, both **quantitative** and **categorical** data were extracted as part of this process. The scrapping of the data form the web has been done through **Beautiful Soup** which is a python library that can pull out data form HTML and XML files. For every recipe, the following intriguing data points were scraped.

The **categorical** features are:

- **Name of the Food:** The recipe's title, which describes the kind of food.
- **Image of the Food:** The URL of the original source for the picture of the dish.
- **Summary:** A brief description or overview of the recipe.
- **Recipe Key:** A unique identifier tag for each recipe. In total 14 of these type keys are present in SkinnyTaste.

The **quantitative** features are:

- **Calories:** The recipe's overall calorie count.
- **Personal Points:** A numerical rating system (1-16) used to evaluate recipes based on specific dietary preferences.

The scraped data has been saved locally as both **Excel** and **CSV** files to aid in future analysis and utilization.

After finishing the scrapping the unique values of each column has been found. The table below shows the unique value count of each feature.

Features	No. of unique entries
Name of the Food	1000
Image of the Food	1000
Summary	998
Recipe Key	435
Calories	435
Personal Points	14

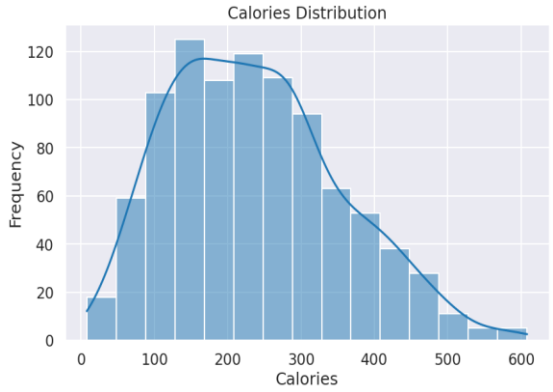
One of the primary challenges faced during the data collection process was the need to scrape the links of every individual recipe page. After gathering the links, navigating to each page and extracting data for each recipe proved to be time-consuming. Additionally, the data was not consistent across all pages which were also super challenging to extract the data.

Data Analysis:

We will examine the data visualization in this section in order to derive important conclusions and observations. Understanding various facets of the gathered data is aided by the distinct viewpoints that each visualization offers on the dataset.

Calories Distribution:

The distribution of calories among different recipes from the SkinnyTaste website is shown in the following visualization. The recipes are clearly concentrated in the calorie range of about 50 to 600, as indicated by the histogram, with the maximum frequency found in the 150 to 300 calorie range. Given the website's emphasis on healthy and weight-conscious cooking, this suggests that the majority of the recipes are created with low to moderate calorie counts.



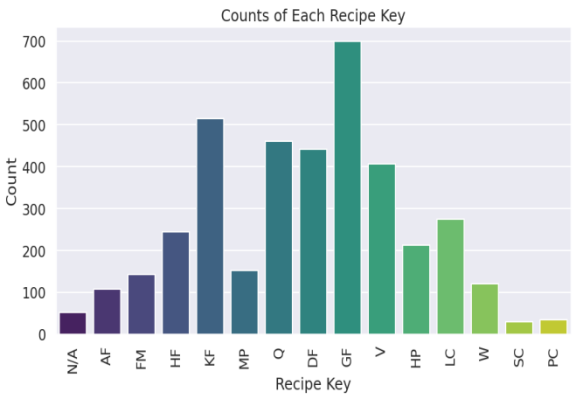
Since there are fewer dishes with greater calorie counts, the distribution is right-skewed. It appears from this skewness that although there are certain dishes with more calories, they are not as prevalent. The concept that most of the dishes are meant to be reasonably light and appropriate for those trying to maintain or reduce their caloric consumption is reinforced by the fact that most recipes are well below the 500-calorie threshold.

It's interesting to note that the modal calorie content of the recipes falls between 150 and 200 calories. This suggests that SkinnyTaste favors dishes that strike a compromise between flavor and calorie content. The wide range of calorie values—from almost 0 to more than 600 calories—also suggests that there are many different kinds of dishes that may be made to accommodate different dietary requirements and tastes.

Recipe Key Distribution:

A bar plot displaying the counts of each recipe key from the SkinnyTaste website is the second visualization. This plot sheds light on the range and regularity of the many recipe kinds that are posted on the website.

With almost 700 dishes, the most popular category is Gluten-Free (GF). This large figure suggests that there is a high priority placed on accommodating people who are gluten intolerant or who choose to follow a gluten-free diet. The large number of gluten-free recipes on SkinnyTaste indicates that it's a useful tool for people who must or choose to eliminate gluten in their diets.



With 500–600 recipes apiece, Kid-Friendly (KF) and Quick (Q) recipes likewise exhibit high frequency. These categories' popularity is a reflection of the demands of time-pressed families and individuals searching for quick and easy meal options. Recipes that are kid-friendly are designed with families with young diners in mind, making the food tasty and appropriate for smaller tummies.

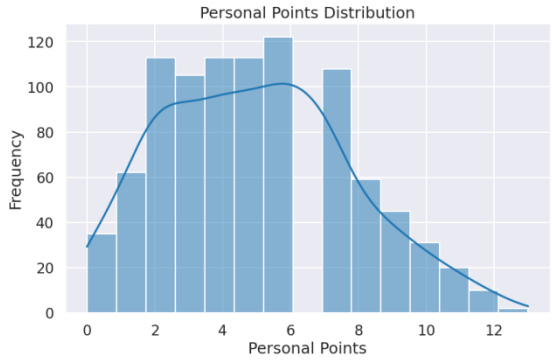
Lower counts are found in less popular categories such Meal Prep (MP), Whole 30 (W), Slow Cooker (SC), and Pressure Cooker (PC). These categories are nonetheless significant for users who are interested in particular diet plans or cooking techniques even though they are less represented. The existence of these categories indicates that, despite being less frequently requested,

SkinnyTaste provides customized recipes to satisfy specific demands.

The recipes that don't fit into any particular key or weren't categorized are probably found in the N/A category, which has the lowest count. This tiny figure indicates that the majority of the recipes on the website are clearly organized, making it simple for consumers to navigate and search.

Personal Point Distribution:

The distribution of personal points histogram makes it evident how SkinnyTaste recipes are grouped according to users' personal points. This distribution is especially important to people who follow diet plans based on points, like Weight Watchers, where managing dietary consumption is facilitated by the accumulation of personal points.



Most recipes fall between two and six points, indicating a right-skewed distribution of personal points. Most recipes appear to be modest in points, as indicated by the histogram's peak, which is located at about six points. That means that SkinnyTaste wants to offer recipes that are appropriate for regular meals as part of a balanced diet and don't include a lot of personal touches.

Throughout this section, the distribution of data of various features has been discussed. An examination of SkinnyTaste's recipe data shows that moderate-calorie meals are highly valued; the majority of recipes have 150–300 calories, with a peak at 150 calories. A wide variety of dietary choices are featured in the recipe key distribution, with a focus on vegetarian, dairy-free, and gluten-free options. Furthermore, the site's emphasis on offering wholesome dishes that complement point-based dietary plans is highlighted by the personal point's distribution, which leans toward the 2 to 6 point range. All things considered, SkinnyTaste provides a varied assortment of wholesome dishes that suit different dietary requirements and tastes.

Conclusion:

Many bottlenecks occurred during the data collection and analysis process. One major obstacle was navigating the inconsistent structure of recipe, which made it challenging to consistently scrape data. Additionally, pagination and link accessibility made it difficult to initially scrape the links from the recipe index page. In order to remedy this, a methodical process was put into place to remove links from each page before compiling specific recipe information. Besides this, it was also difficult to extract the recipe keys since they were extracted altogether as a string. Then those were separated based on space ultimately counting the occurrence of each recipe key. Moreover, to handle the missing values 'N/A' tags has been used. Apart from this calorie distribution & personal points were extracted as a string but those were converted into float and integer respectively before the visualization.

Overall, by considering all the things, this task has given me a thorough understanding of the web scraping procedure. The experience was quite enjoyable for me because it improved my technical abilities and gave me the chance to learn about fresh ideas and approaches. My understanding of the complexities of data-driven projects has grown as a result of the gratifying and difficult process of gathering and evaluating data.