# IT00CH95-3004 Data Science

# Mini Project 1: What Should I Cook Tonight?

Md Hasibul Haque Zahid (2302302)

## Introduction

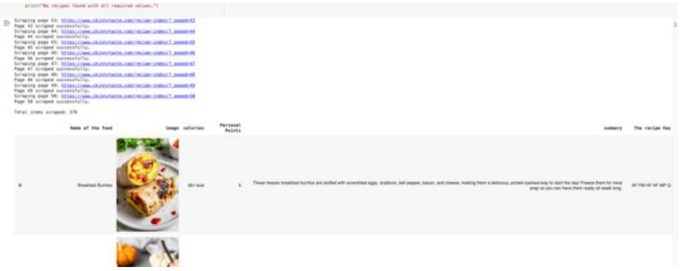
The purpose of this project is to provide recommendations based on calorie and point ranges to help users decide "What should I cook tonight?" This was achieved by scraping and analyzing data from the food recipe website Skinnytaste. The data that was scraped included a variety of details, such as the dish's name, picture, nutritional value, user-submitted points, and a recipe synopsis. Users can choose a meal that best suits their dietary requirements by using this information to sort food choices by calorie count. Simply enter a calorie and point range.

## Data Collection

To collect data, I used **BeautifulSoup**, a popular Python library for web scraping. The first 50 pages of the website **Skinnytaste** were scraped, which involved parsing the HTML structure to extract relevant details. The data collected included:

- **Food Name**: The title of the dish.
- **Image**: A URL link to the image of the dish.
- **Calories**: The calorie counts of each dish.
- **Personal Points**: Weight Watchers points associated with each recipe.
- **Summary**: A brief description of the dish.
- **Recipe Key**: A special feature available on the website for filtering recipes.

To store the scraped data, I used Python dictionaries and converted them into a **Pandas DataFrame** for easy manipulation and analysis. The data was then exported to a CSV file for persistent storage and future use.



```
    print("No recipes found with all required values.")

Scraping page 43: https://www.skinnytaste.com/recipe-index/?_paged=43
Page 43 scraped successfully.
Scraping page 44: https://www.skinnytaste.com/recipe-index/?_paged=44
Page 44 scraped successfully.
Scraping page 45: https://www.skinnytaste.com/recipe-index/?_paged=45
Page 45 scraped successfully.
Scraping page 46: https://www.skinnytaste.com/recipe-index/?_paged=46
Page 46 scraped successfully.
Scraping page 47: https://www.skinnytaste.com/recipe-index/?_paged=47
Page 47 scraped successfully.
Scraping page 48: https://www.skinnytaste.com/recipe-index/?_paged=48
Page 48 scraped successfully.
Scraping page 49: https://www.skinnytaste.com/recipe-index/?_paged=49
Page 49 scraped successfully.
Scraping page 50: https://www.skinnytaste.com/recipe-index/?_paged=50
Page 50 scraped successfully.

Total items scraped: 378
```

|  | Name of the food | image | calories | Personal Points | summary | The recipe Key |
|---|---|---|---|---|---|---|
| 0 | Breakfast Burritos |  | 261 kcal | 1 | These freezer breakfast burritos are stuffed with scrambled eggs, scallions, bell pepper, bacon, and cheese, making them a delicious, protein-packed way to start the day! Freeze them for meal prep so you can have them ready all week long. | [illegible] |

**Fig: Scraping the Data**

## Data Analysis

After collecting the data, I performed exploratory data analysis (EDA) to uncover interesting insights about the food recipes available on the website.

### Calories Distribution

I visualized the distribution of calories across the recipes using a histogram. Most recipes had a calorie count below 500 kcal, with a few outliers having significantly higher calorie content. This is useful for users looking to find low-calorie recipes.

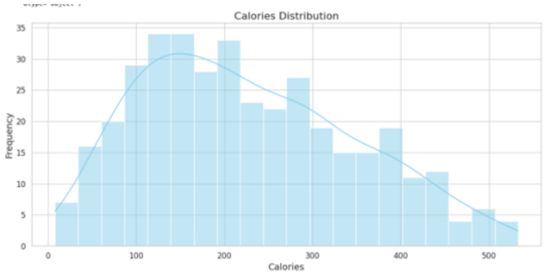


**Fig: Calories Distribution**

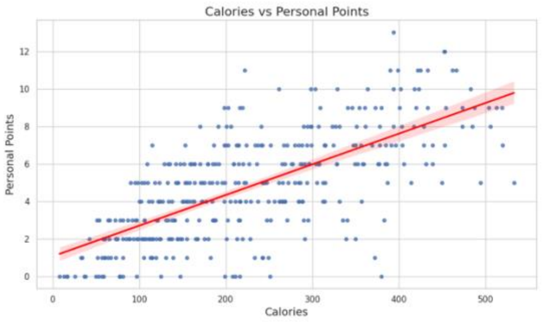


**Fig: Calories Vs Personal Points**

### Personal Points Distribution

Weight Watchers' personal points were also analyzed, showing a wide distribution across recipes. The majority of recipes had points between 2 and 10, making them suitable for people following the Weight Watchers program.

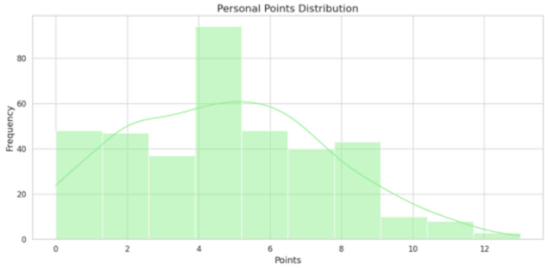


Fig: **Personal Points Distribution**

### Recipe Key Distribution

The **Recipe Key** is a unique feature on the website that categorizes recipes (e.g., vegan, gluten-free, etc.). This categorization helps users easily find recipes based on their dietary preferences. The distribution of these categories was visualized using a bar chart to give an overview of the most common recipe types.

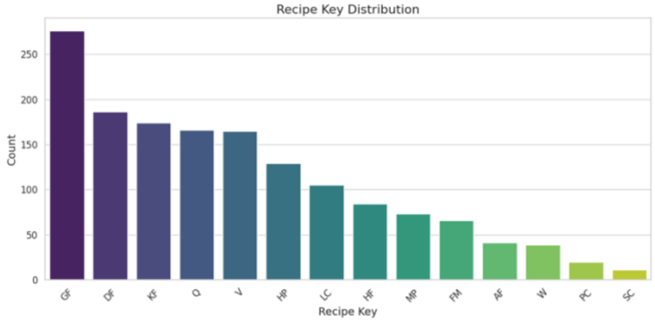


**Fig: Recipe Key distribution**

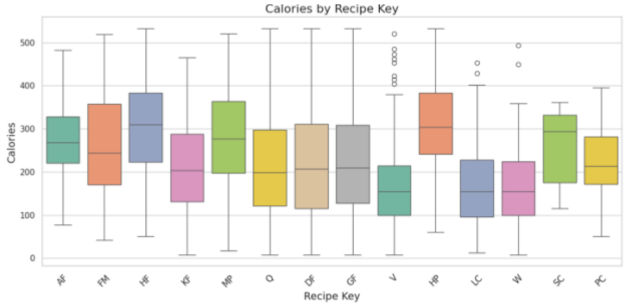
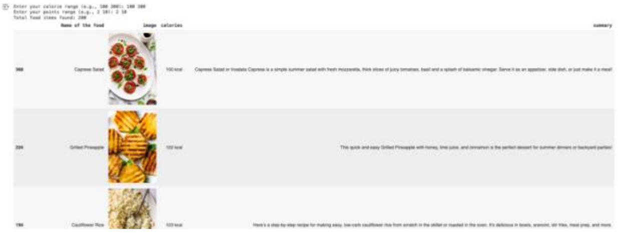


**Fig: Calories Vs Recipe Key**

### User Interaction

One of the primary requirements of the project was to allow user interaction. The program allows the user to input a calorie range and a point range, and it outputs the first 10 recipes that meet these criteria. The output includes:

- The name of the dish
- An image of the dish & Caloreis
- The summary of the recipe



```
Enter your calorie range (e.g., 100 300): 100 300
Enter your points range (e.g., 2 10): 2 10
Total food items found: 200
```

| | Name of the food | image | calories | summary |
|---|---|---|---|---|
| 368 | Caprese Salad | | 100 kcal | Caprese Salad or Insalata Caprese is a simple summer salad with fresh mozzarella, thick slices of juicy tomatoes, basil and a splash of balsamic vinegar. Serve it as an appetizer, side dish, or just make it a meal! |
| 224 | Grilled Pineapple | | 102 kcal | This quick and easy Grilled Pineapple with honey, lime juice, and cinnamon is the perfect dessert for summer dinners or backyard parties! |
| 184 | Cauliflower Rice | | 103 kcal | Here's a step-by-step recipe for making easy, low-carb cauliflower rice from scratch in the skillet or roasted in the oven. It's delicious in bowls, arancini, stir fries, meal prep, and more. |

Fig: **How User will interact**

This interaction is useful for users who have specific dietary goals and are looking for personalized recipe suggestions based on calorie and point ranges.

**Conclusion**

**Bottlenecks and Solutions**

During the project, I encountered several bottlenecks that significantly slowed down the progress of scraping data from the website. The challenges came from both technical and structural aspects of the website. Here's a breakdown of the key bottlenecks and how I overcame them:

1. **Pagination Issues:** Another bottleneck involved navigating through multiple pages to scrape all 50 pages of data. The website uses a complex pagination system, and initially, I had difficulty properly identifying the URLs and handling page changes automatically.

**Solution**: After trying a few different methods, I was able to programmatically construct URLs for subsequent pages using BeautifulSoup by identifying the patterns in the pagination system. I then iterated through each page while ensuring the data was being captured correctly.

2. **Inconsistent Data Structure:** I faced issues with missing or inconsistent data fields across different recipes. For example, some pages were missing calorie information or personal points, causing errors during the scraping process.

**Solution**: I had to implement error handling and conditional logic to account for these inconsistencies. By checking for the presence of required fields and skipping or marking incomplete records, I ensured that the scraping process could continue without crashing.

This project demonstrates the use of web scraping and data analysis techniques to solve the problem of meal planning based on nutritional preferences. However, the process was not without its challenges. Web scraping using BeautifulSoup required handling dynamic content and inconsistencies in the data structure. Additionally, managing pagination and large datasets was time-consuming.

Despite these challenges, the project was successful in gathering, analyzing, and presenting the data in a way that provides valuable insights to users.