Mini Project 1 – What should I cook tonight?

Hektor Dahlberg (41685)

September 24, 2024

Contents

1 Introduction

The problem for the mini project is “what should I cook tonight?”.

2 Data Collection

2.1 How are you collecting data?

The data is scraped from the web site via looping through 50 pages of recipes and while doing that also going through all the articles of recipes and finding the data for step 2 “filter interesting data”.

2.2 What are the challenges in scraping the data? How do you store the scraped data for a further easy use?

Challenges for scraping the data was that there were also cookbooks in the data, but this problem was solved by using an if statement to check if the article had a recipe in it, if not the article was skipped. The data was stored in a pandas data frame for easier handling for steps 3 and 4 in the mini project.

3 Data analysis: Visualize the data, what are some interesting observations you could report?

For visualisation I only used one histogram to represent the calories distribution and for the recipe key distribution and points distribution bar plots where used. Interesting observations from the data is that there are not many recipes with an review under 4.00 points and a lot of recipes with a perfect 5.

4 Conclusion: What were the bottlenecks? How did you overcome them?

Bottlenecks for me was to try to understand the task as there is a lot of possible ways to interpret the tasks, this problem was solved by just choosing one interpretation and do the project according to that interpretation.