

Question/need:

- What is the framing question of your analysis, or the purpose of the model/system you plan to build?

Client is a vegetarian recipe developer who has worked on some recipes and is curious to find out if these recipes would receive good ratings before publishing them. He has approached me to predict ratings based on all the vegetarian recipes present on [allrecipes.com](https://www.allrecipes.com).

- Who benefits from exploring this question or building this model/system?

Client can make changes to the new recipes based on the predictions if needed and understand what are the key factors for best rated recipes.

Data Description:

- What dataset(s) do you plan to use, and how will you obtain the data?

Data will be obtained through web scraping from [allrecipes.com](https://www.allrecipes.com) and loaded to a pandas dataframe containing 1000 recipe names, ingredients and other features.

- What is an individual sample/unit of analysis in this project? What characteristics/features do you expect to work with?

Each row is a recipe with its details and features include-

- 1) Recipe Name
- 2) Rating
- 3) Prep Time
- 4) Cook Time
- 5) Servings
- 6) Yield
- 7) Calories
- 8) Protein

9) Carbohydrates

10) Fat

11) Sodium

- If modeling, what will you predict as your target?

Aim of the project is to predict ratings of the users who tried the recipes using linear regression. User can rate a recipe on a scale of 5 stars, and no. of stars mean -

Number of Stars	Interpretation
5	Loved it!
4	Liked it
3	It was okay
2	Didn't like it
1	Couldn't eat it

Tools:

- How do you intend to meet the tools requirement of the project?
 - ☐ Data will be scraped using a combination of Selenium and Beautiful soup (or just beautiful soup, will need to try)
 - ☐ Scikit learn will be used for importing linear regression
 - ☐ Visualizations are performed using Seaborn and Matplotlib
- Are you planning in advance to need or use additional tools beyond those required?

None

MVP Goal:

- What would a [minimum viable product \(MVP\)](#) look like for this project?

MVP will have scraped data from [allrecipes.com](https://www.allrecipes.com) into a pandas dataframe with some visualizations as part of EDA such as pairplot to understand pairwise relationships in the dataset and an initial linear regression model.