A 250-word paper summarizing your steps and any challenges you ran into during the project. Discuss the importance and relevance of this type of process if you were a data scientist. How often do you think you would have to do this to get the data you need?

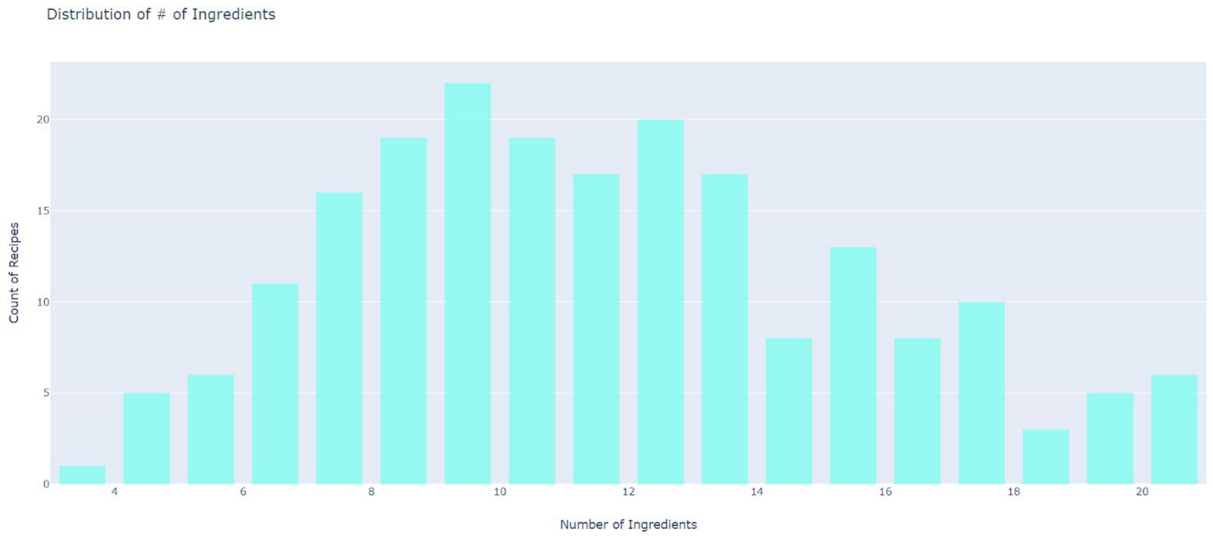
The first step was finding an API to pull the data from. I used the website https://apilist.fun/ and was very excited about all of the cool API's we can use. I wanted to find one that would provide the required number of minimum columns and ended up using The Meal DB.

This site allowed users to easily access the data for free if it is for educational purposes. I didn't even have to sign up to use the student key. I created a request that would pull 1000 random recipes from their website. I loaded the response into JSON and was able to create a dataframe from the dictionaries.

My transformations included making the Headers readable, replacing missing values and changing the strings data within the columns to title case. I replaced missing values with None, as the missing data fell into categorical columns so np.NaN and 0 did not make sense.

I created a new column that held the length of total ingredients in each recipe and plotted this distribution (see below).

The hardest part was honestly stopping myself here. I read the instructions again and it said to just do 2 transformations. I have many ideas in my head for ways I could continue analyzing the data, including looking at the most popular ingredients for each Meal Category or Area (cuisine). I could also look at the average number of ingredients by Meal Category or Area. I could apply a Tf-IDF transformation on the ingredients and use it to predict the cuisine, etc. So many ways to slice and dice data once you start getting in there.



Reference: https://www.themealdb.com/api.php?ref=apilist.fun