Team : Maria Jose Siles Navarro, Carl Tondo

Executive Summary:

The objective of this project is to gather datasets of user interactions and recipe data from Food.com (made up with 18 years worth of data). Using the datasets of recipes and user interaction, we are able to develop queries that show different user and recipe interactions. As an example, we are able to display the highest rated recipe, or the recipe with the most user interaction (rating/review).

Data Sources :

We gathered our data from Kaggle.com, specifically data sources made by Bodhisattwa Prasad Majumder and Shuyang Li. Using these datasets, they provided us with 18 years worth of data collected from Food.com. However, we didn’t need to analyze that huge amount of data, so we cleaned the data on Python using the pandas package. We cleaned the data by only receiving rows that fit within a three-month range that we chose to analyze.

[Food.com Data Source Link](https://www.kaggle.com/shuyangli94/food-com-recipes-and-user-interactions)

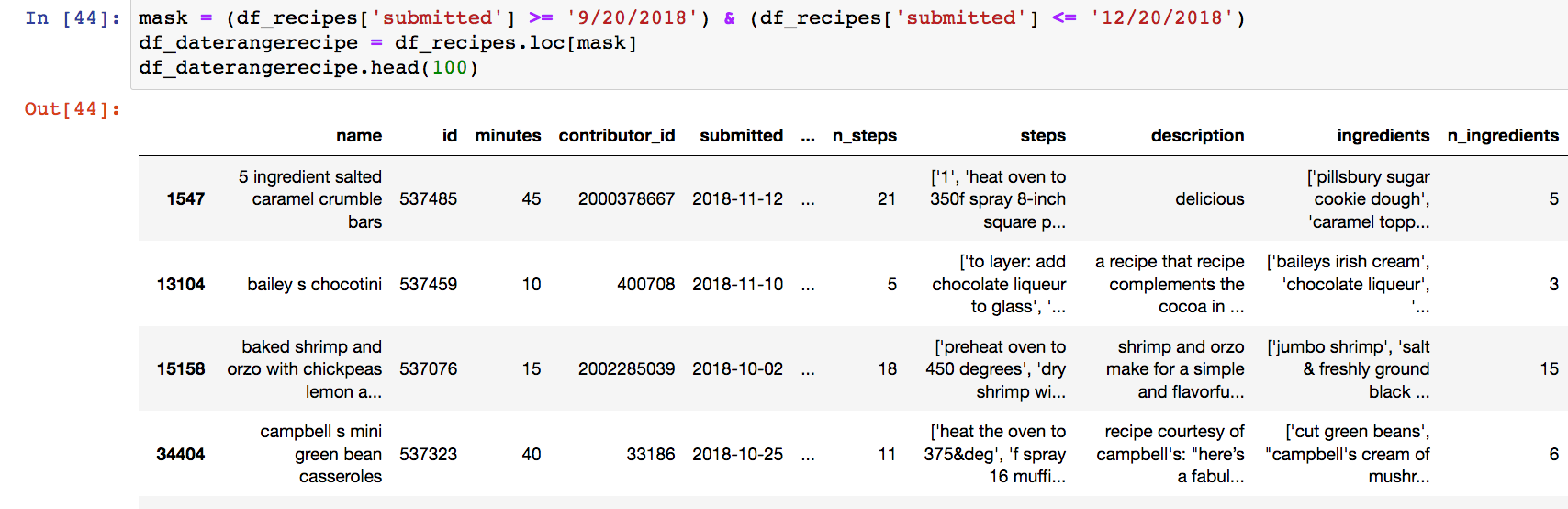
Data Dictionary:

|  |  |  |  |
| --- | --- | --- | --- |
| Field Name | Data Type | Description | Example |
| user\_id | bigint | The user’s number ID | 2002372706 |
| recipe\_id | bigint | The recipe’s number ID | 63786 |
| date | date | The date of user interaction with recipes. | 2018-12-20 |
| rating | int | The user rating of a recipe. | 5 |
| review | text | A user review of a recipe. | Finally, I found… |
| name | varchar | name of the recipe | baked shrimp |
| minutes | int | how long a recipe takes to make | 45 |
| contributor\_id | bigint | ID of the user that posted the recipe | 33186 |
| submitted | date | date of when recipe was submitted | 2018-11-12 |
| tags | varchar | recipe tags to improve search and filters | [’60-min-or-less’} |
| nutrition | varchar | basic nutrition facts of recipe | [52.8, 60.9, …] |
| n\_steps | int | amount of steps a recipe needs to be made | 21 |
| steps | varchar | Actual descriptions of steps that a person needs to follow | [‘1’ preheat, …] |
| description | varchar | quick description of what the recipe is | delicious |
| ingredients | varchar | the ingredients needed for the recipe | [‘pillsbury sugar’, …] |
| n\_ingredients | int | The amount of ingredients needed for a recipe | 5 |

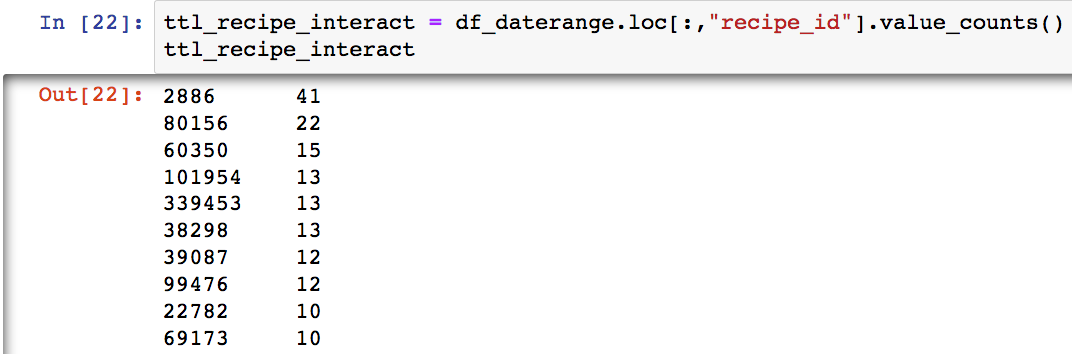
Data Cleaning:



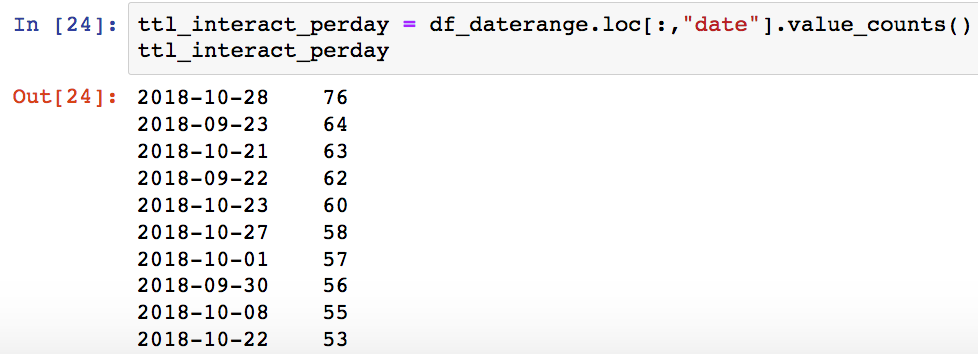
This is the first part of our data cleaning where we needed to create a date range so we didn’t need to handle an incredible amount of data. This allows us to analyze data within a three-month period for user interactions.



This is the second part of our data cleaning where we do the same thing of creating a date range, but for a different table that includes more descriptions about the recipe.

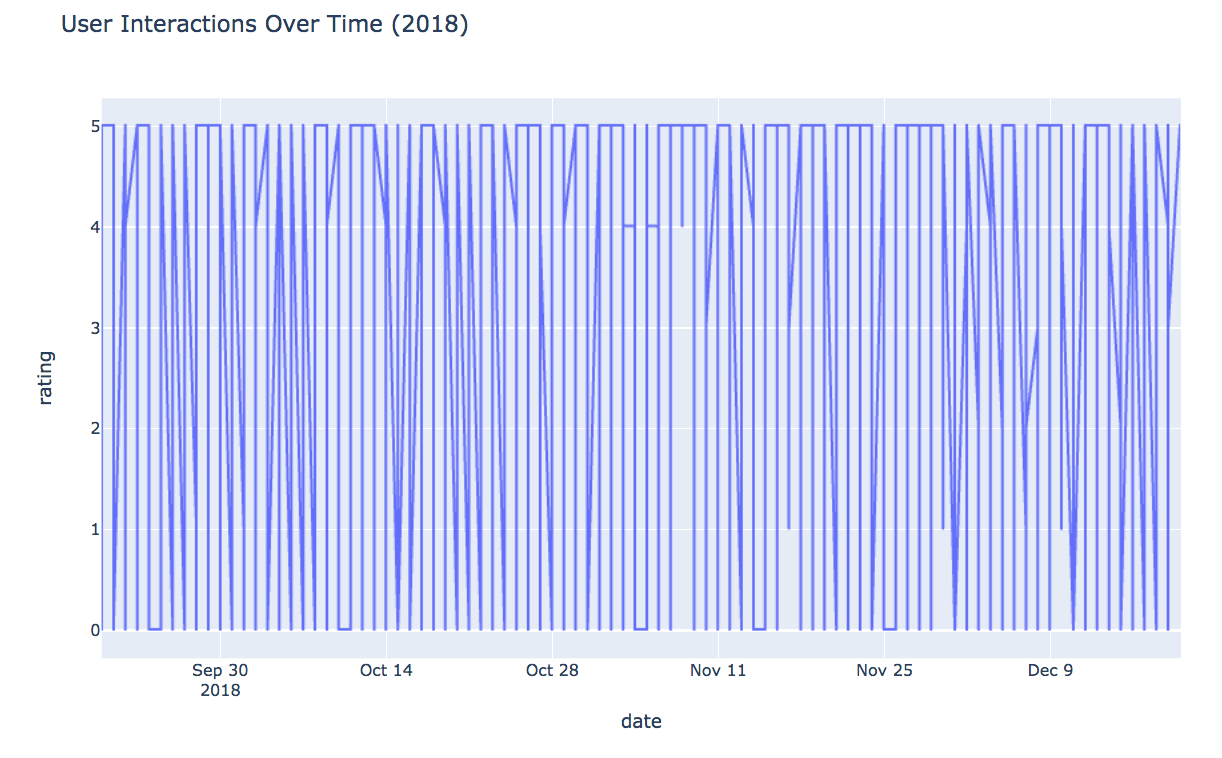


This python script gives us a general overview of recipes with the most user interactions within that three-month period.

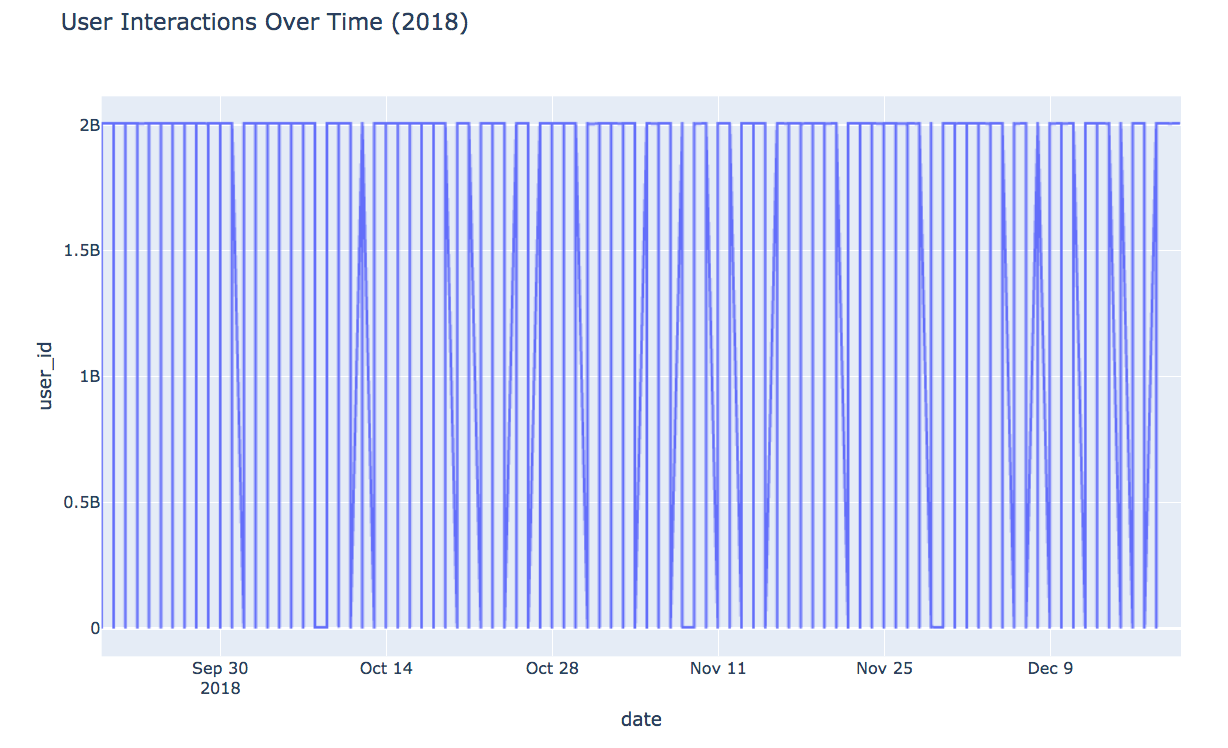


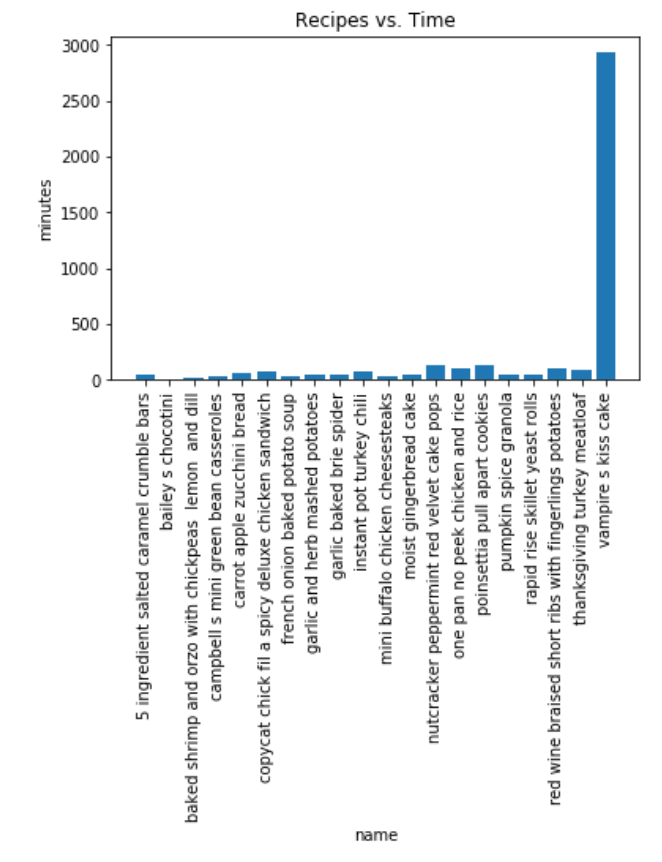
This allows us to see which date within the three-month period has the most user interactions.

Graphs:



This graph demonstrates the number of ratings submitted by users within a three-month period. We can see that between October 28 and November 11 has the least amount of user interaction.

This shows us the activity of users, by analyzing when each user interacts with a recipe through either a review or a rating.



This displays the submitted recipes and their respective cooking times. This allows us to see which recipes take the longest or the shortest.

ERD Diagram:

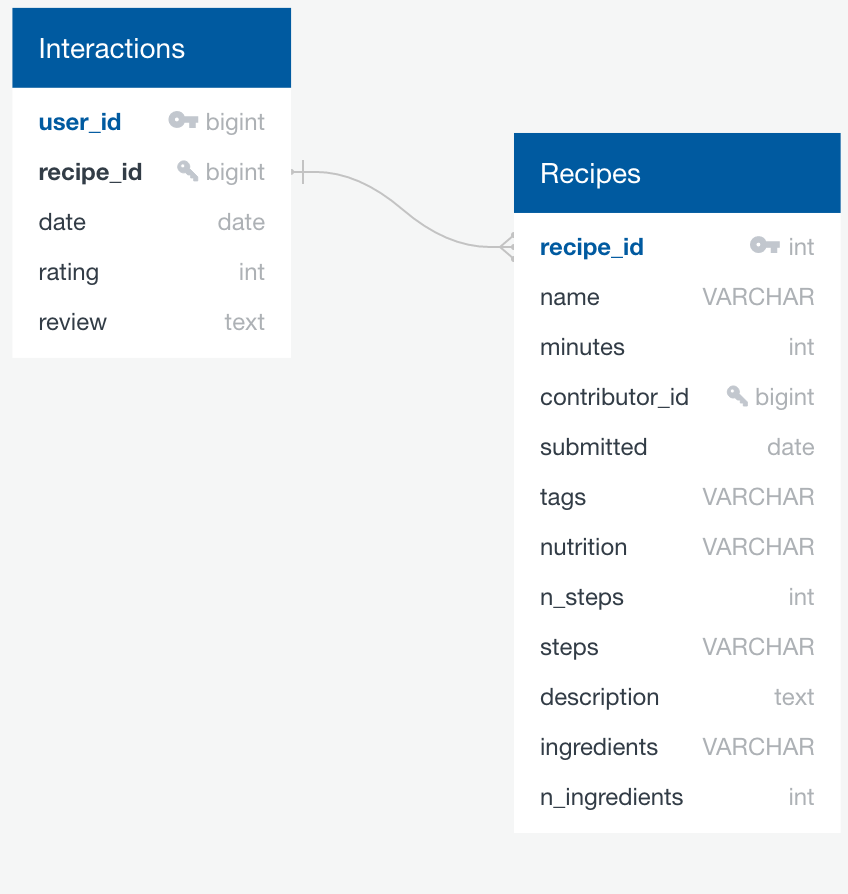
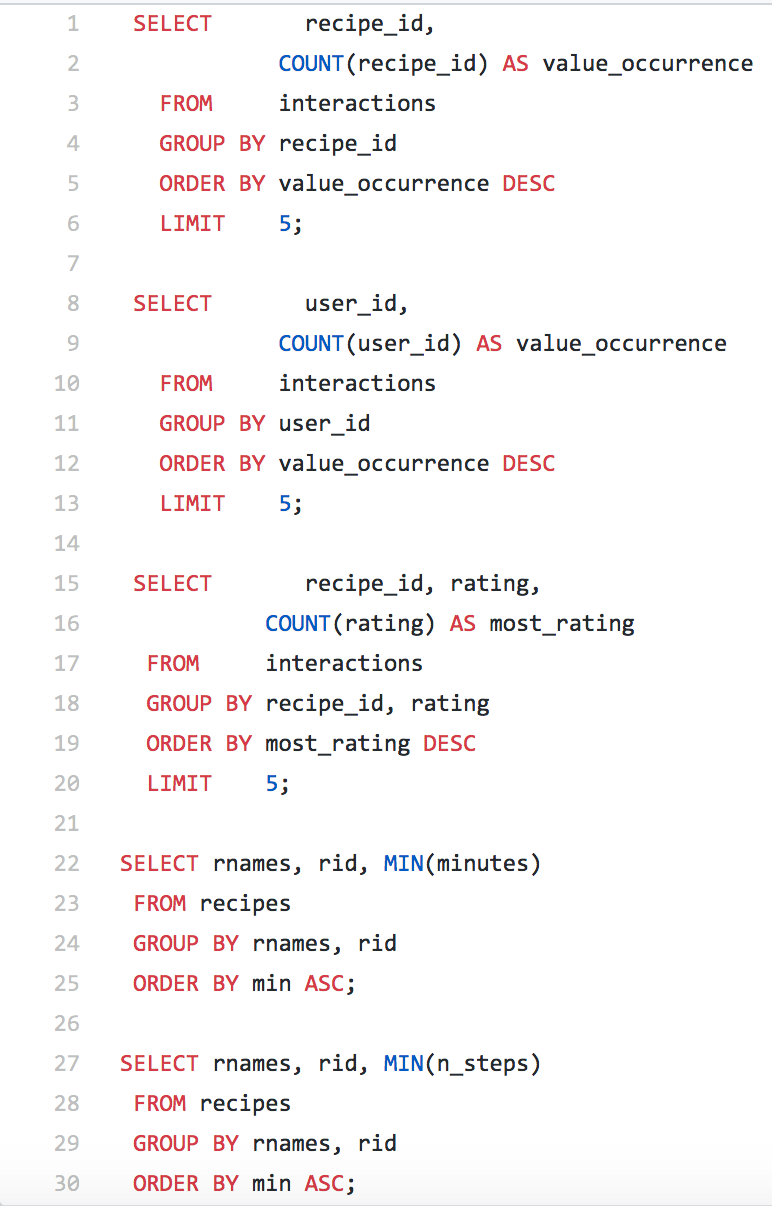
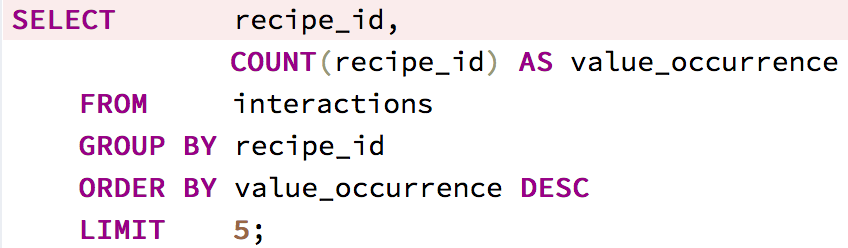
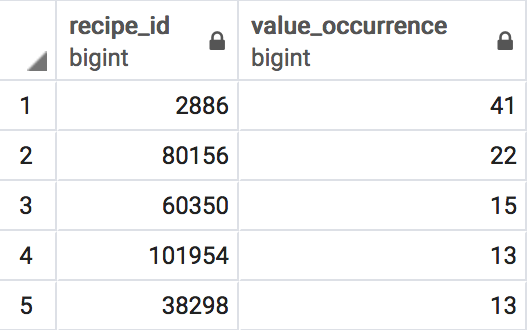


Table Schema:

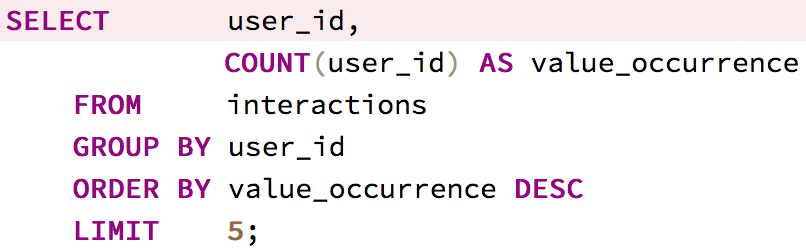


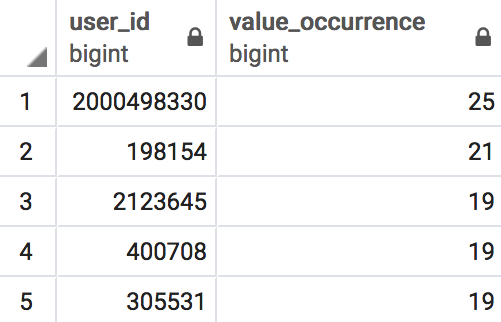
Queries:

INPUT: 

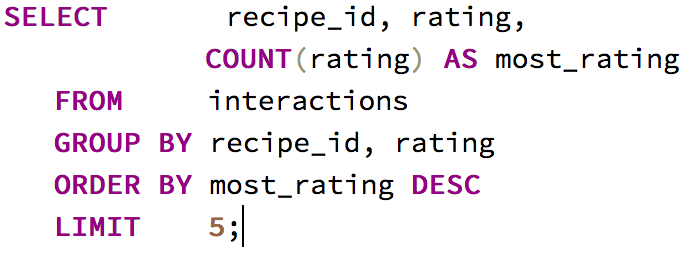
OUTPUT: 

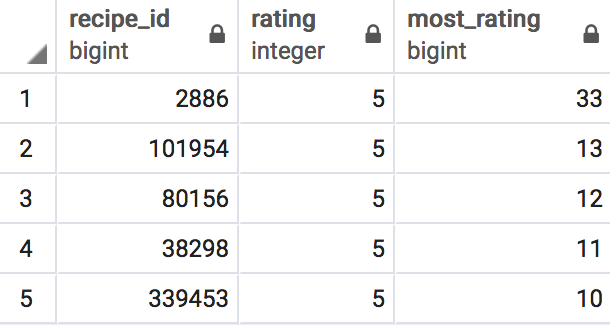
1. This shows us the top recipes with the most user interactions (that includes both reviews and ratings).

INPUT: 

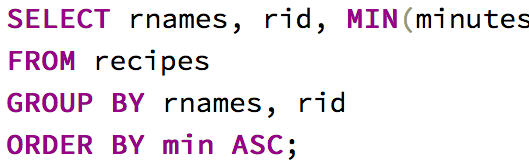
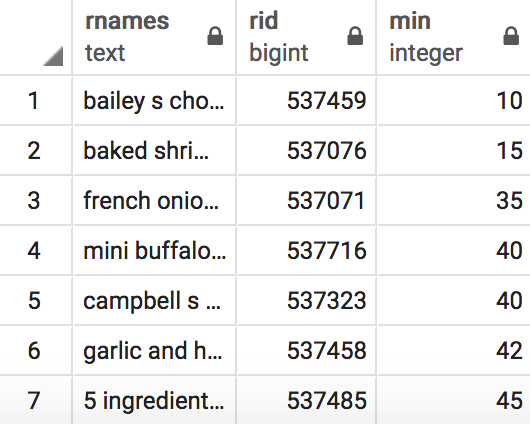
OUTPUT:

2. This shows us the most active users by tracking the users with the most interactions to a recipe.

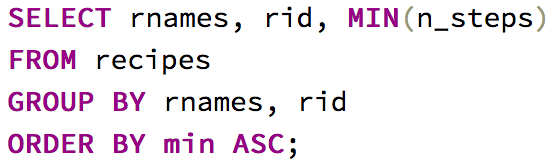
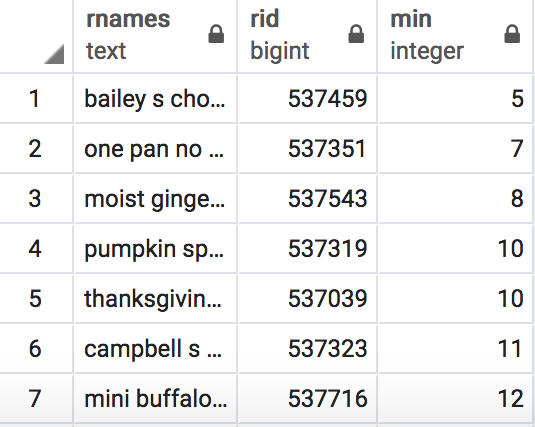
INPUT: 

OUTPUT: 

3. This shows us the recipes with the most ratings and the highest value ratings from users.

INPUT:   
OUTPUT: 

4. This shows us the recipes that can be made at the least amount of time for those that want quick under an hour of cooking.

INPUT:  
OUTPUT:

5. This shows us recipes that are simple to make for beginner due to their low number of steps.