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.

Data Dictionary:

Field Name	Data Type	Description	Example
user_id	bigint		
recipe_id	bigint		
date	date		
rating	int		
review	text		

<u>Project Summary</u>: We determine the range and data cleaning of the food recipes and interactions.

Methodology:

- Find users interactions of the food and receipts
- Find two graphs on that lead to cleaning
- -ERD DIAGRAM
- -Table Schema
- -Notebook
- -SQL
- -Five queries
- -Execute Summary

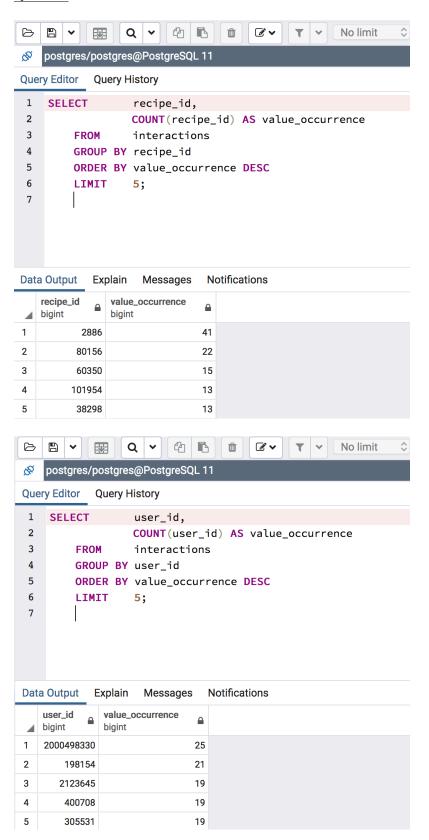
Data Collection:

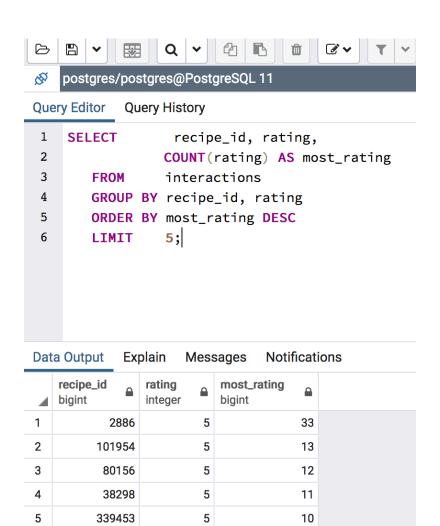
- Raw Interactions
- Raw Recipes

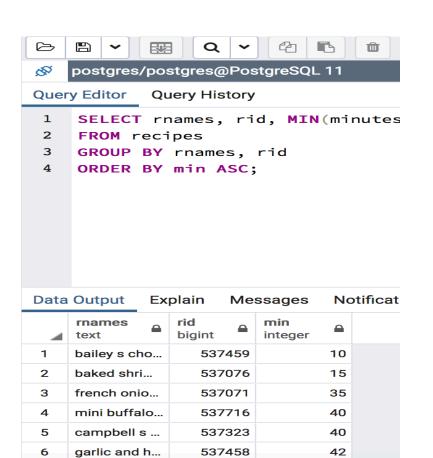
Table Schema:

```
1
      SELECT
                   recipe_id,
                 COUNT(recipe_id) AS value_occurrence
 2
 3
        FROM
                 interactions
 4
        GROUP BY recipe_id
 5
        ORDER BY value_occurrence DESC
 6
        LIMIT
                 5;
 7
 8
      SELECT
                   user id,
                 COUNT(user_id) AS value_occurrence
 9
10
        FROM
                 interactions
11
        GROUP BY user_id
        ORDER BY value_occurrence DESC
12
13
        LIMIT
                 5;
14
                   recipe_id, rating,
15
      SELECT
                COUNT(rating) AS most_rating
16
17
       FROM
                interactions
18
       GROUP BY recipe_id, rating
       ORDER BY most_rating DESC
19
20
       LIMIT
                5;
21
22
     SELECT rnames, rid, MIN(minutes)
23
     FROM recipes
      GROUP BY rnames, rid
24
25
      ORDER BY min ASC;
26
    SELECT rnames, rid, MIN(n_steps)
27
28
     FROM recipes
29 GROUP BY rnames, rid
30
      ORDER BY min ASC;
```

Queries:







5 ingredient...



- 1 SELECT rnames, rid, MIN(n_steps)
- 2 FROM recipes
- 3 **GROUP BY** rnames, rid
- 4 ORDER BY min ASC;

Data Output		Explain Messages		otifications
4	rnames text	rid bigint	min integer	
1	bailey s cho	537459	5	
2	one pan no	537351	7	
3	moist ginge	537543	8	
4	pumpkin sp	537319	10	
5	thanksgivin	537039	10	
6	campbell s	537323	11	
7	mini buffalo	537716	12	