

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		

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

Queries:

postgres/postgres@PostgreSQL 11

Query Editor Query History

```
1 SELECT recipe_id,
2         COUNT(recipe_id) AS value_occurrence
3 FROM interactions
4 GROUP BY recipe_id
5 ORDER BY value_occurrence DESC
6 LIMIT 5;
7
```

Data Output Explain Messages Notifications

	recipe_id bigint	value_occurrence bigint	
1	2886	41	
2	80156	22	
3	60350	15	
4	101954	13	
5	38298	13	

postgres/postgres@PostgreSQL 11

Query Editor Query History

```
1 SELECT user_id,
2         COUNT(user_id) AS value_occurrence
3 FROM interactions
4 GROUP BY user_id
5 ORDER BY value_occurrence DESC
6 LIMIT 5;
7
```

Data Output Explain Messages Notifications

	user_id bigint	value_occurrence bigint	
1	2000498330	25	
2	198154	21	
3	2123645	19	
4	400708	19	
5	305531	19	



postgres/postgres@PostgreSQL 11

Query Editor Query History

```
1  SELECT      recipe_id, rating,
2              COUNT(rating) AS most_rating
3  FROM        interactions
4  GROUP BY    recipe_id, rating
5  ORDER BY    most_rating DESC
6  LIMIT       5;
```

Data Output Explain Messages Notifications

	recipe_id bigint	rating integer	most_rating bigint	
1	2886	5	33	
2	101954	5	13	
3	80156	5	12	
4	38298	5	11	
5	339453	5	10	

postgres/postgres@PostgreSQL 11

Query Editor

Query History

1

2

3

4

SELECT

rnames,

rid,

MIN(minutes

FROM

recipes

GROUP BY

rnames,

rid

ORDER BY

min

ASC;

Data Output

Explain

Messages

Notificat

	<div>rnames</div> <div>text</div> <div></div>	<div>rid</div> <div>bigint</div> <div></div>	<div>min</div> <div>integer</div> <div></div>
1	bailey s cho...	537459	10
2	baked shri...	537076	15
3	french onio...	537071	35
4	mini buffalo...	537716	40
5	campbell s ...	537323	40
6	garlic and h...	537458	42
7	5 ingredient...	537485	45

