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