David Blascyk

Michael Laska

Ingris O’Connor

Cindy Pendarvis

* The sources of data that you will extract from.

[https://www.kaggle.com/shuyangli94/food-com-recipes-and-user-interactions#RAW\_interactions.csv](https://www.kaggle.com/shuyangli94/food-com-recipes-and-user-interactions" \l "RAW_interactions.csv)

interactions\_test.csv

interactions\_train.csv

interactions\_validation.csv

PP\_recipes.csv

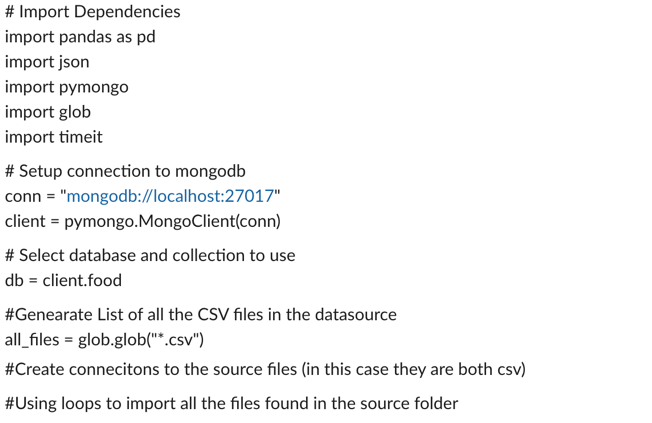
PP\_user.csv

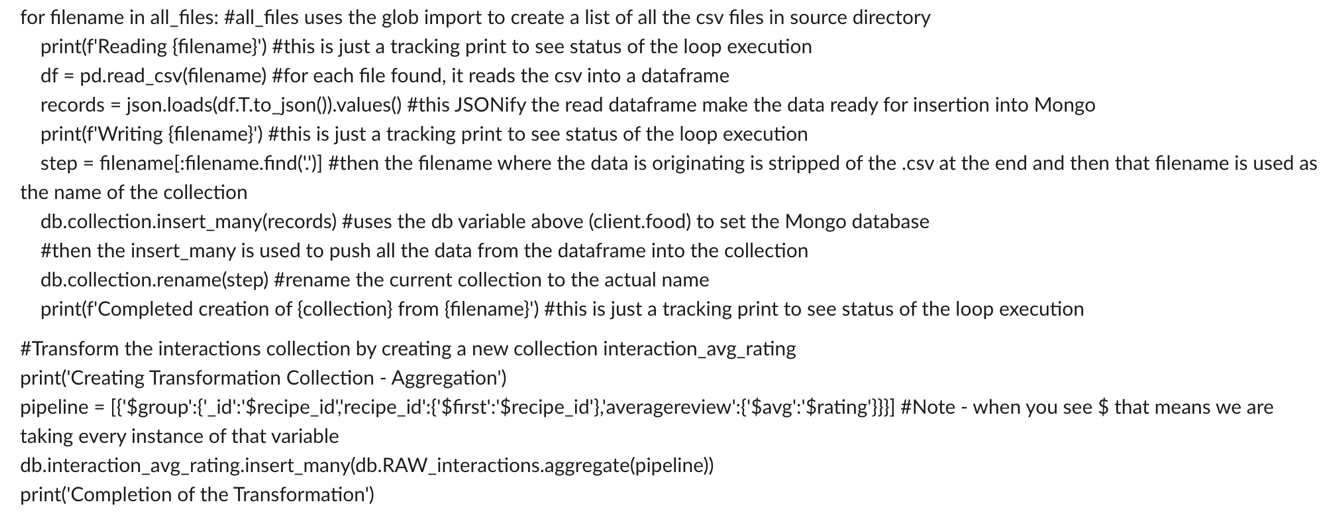
RAW\_interactions.csv

RAW\_recipes.csv

* The type of transformation needed for this data (cleaning, joining, filtering, aggregating, etc).

Created new collection grouped by recipe id to calculate average rating.





* The type of final production database to load the data into (relational or non-relational).

Used non-relational Mongo database.

* The final tables or collections that will be used in the production database.

Database: food

Collections: interactions\_test, interactions\_train , interactions\_validation, PP\_recipes , PP\_user, RAW\_interactions, RAW\_recipies, interaction\_avg\_rating