WMCS009-05.2023-2024.1B
Information Systems

Association Rule Analysis (Market Basket Analysis) Assignment

Dr. George Azzopardi

Task

- Use Jupyter notebook to implement the Apriori algorithm in Python

- **Input:**
  - CSV file: It must contain a binary representation of all concerned items. The first row must contain the header (i.e. the names of the items), and the remaining rows must contain only 1s and 0s separated by commas
  - Minimum support - the threshold used to select the frequent itemsets
  - Minimum confidence - the threshold used to determine the association rules

- **Output**
  - List of association rules that satisfy the given minimum support and confidence

- **Deliverable**
  - The python notebook and a short report describing how the main steps of the algorithm were implemented. These include the self-join and pruning (based on the Apriori principle) in determining the frequent itemsets, and the non-monotonicity property in determining the association rules.

Evaluation

- The evaluation of your assignment will be based on:
  - Use the provided myDataFile.csv as your test case. Test your algorithm with a minimum ***support of 0.005*** and ***confidence of 0.6***. List the number of frequent itemsets per layer and the association rules in the form “A -> B (confidence = ?” where A and B can be comma separated lists.
  - Correct implementation of the Apriori principle in determining the frequent itemsets
  - Correct implementation of the non-monotonicity property in the determination of the association rules
  - Readability of the algorithm and report

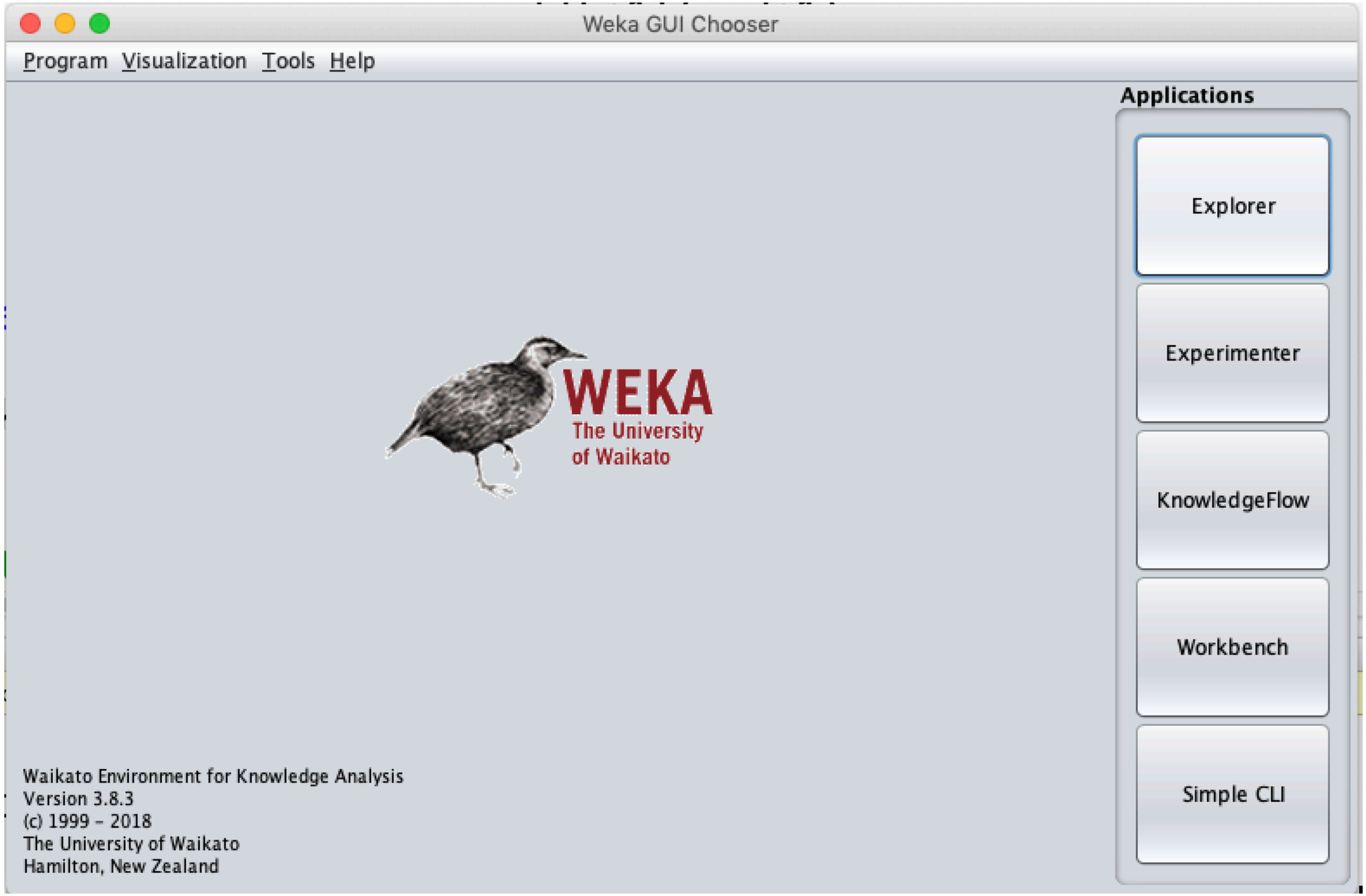
Groceries data set

- The file groceries.txt contains 9835 rows with comma separated items
- Each row can be considered as a receipt with a transaction of multiple grocery items

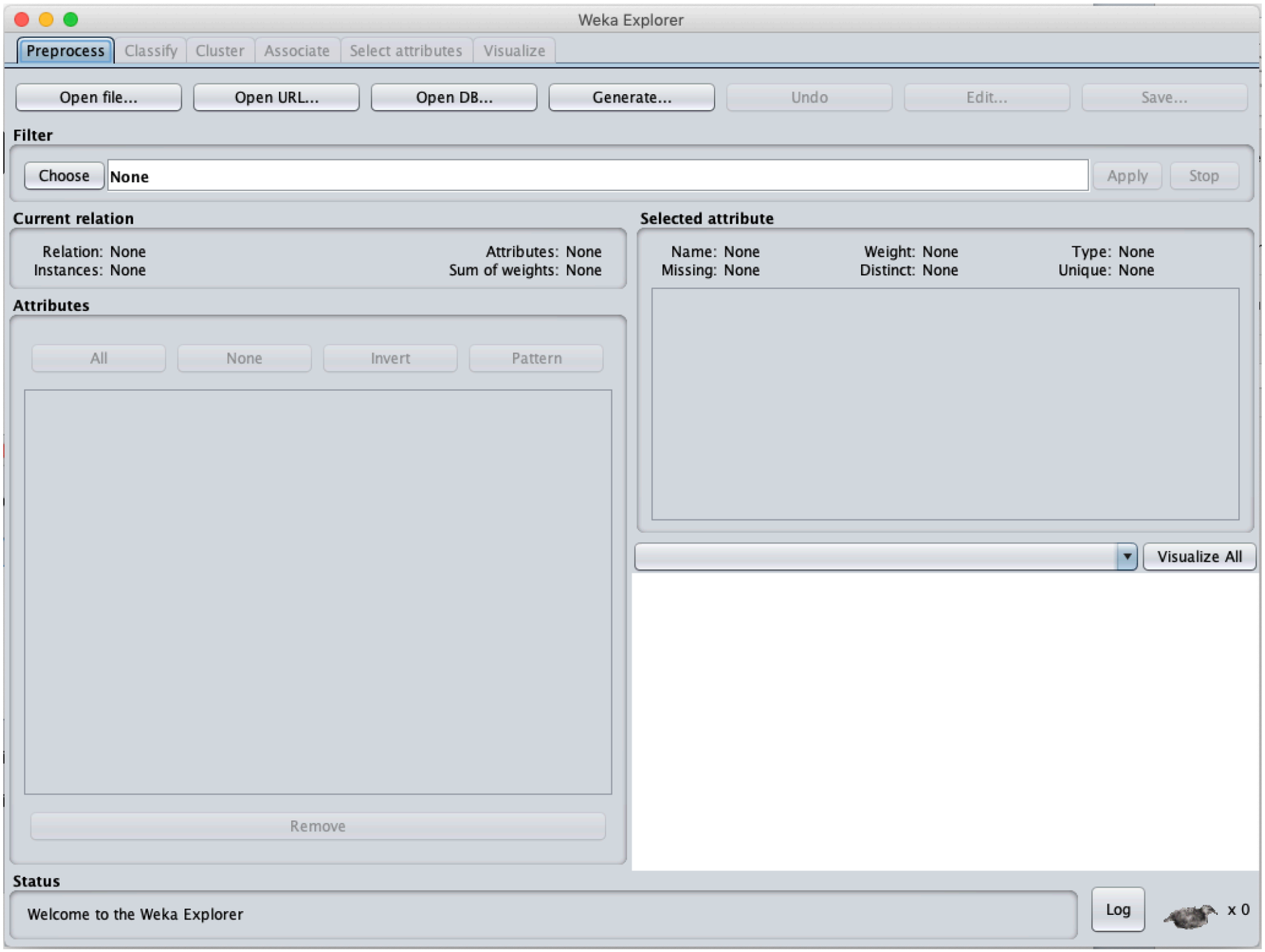
Additional Information

- Install the Weka tool and use it to check your results.
  - https://waikato.github.io/weka-site/index.html

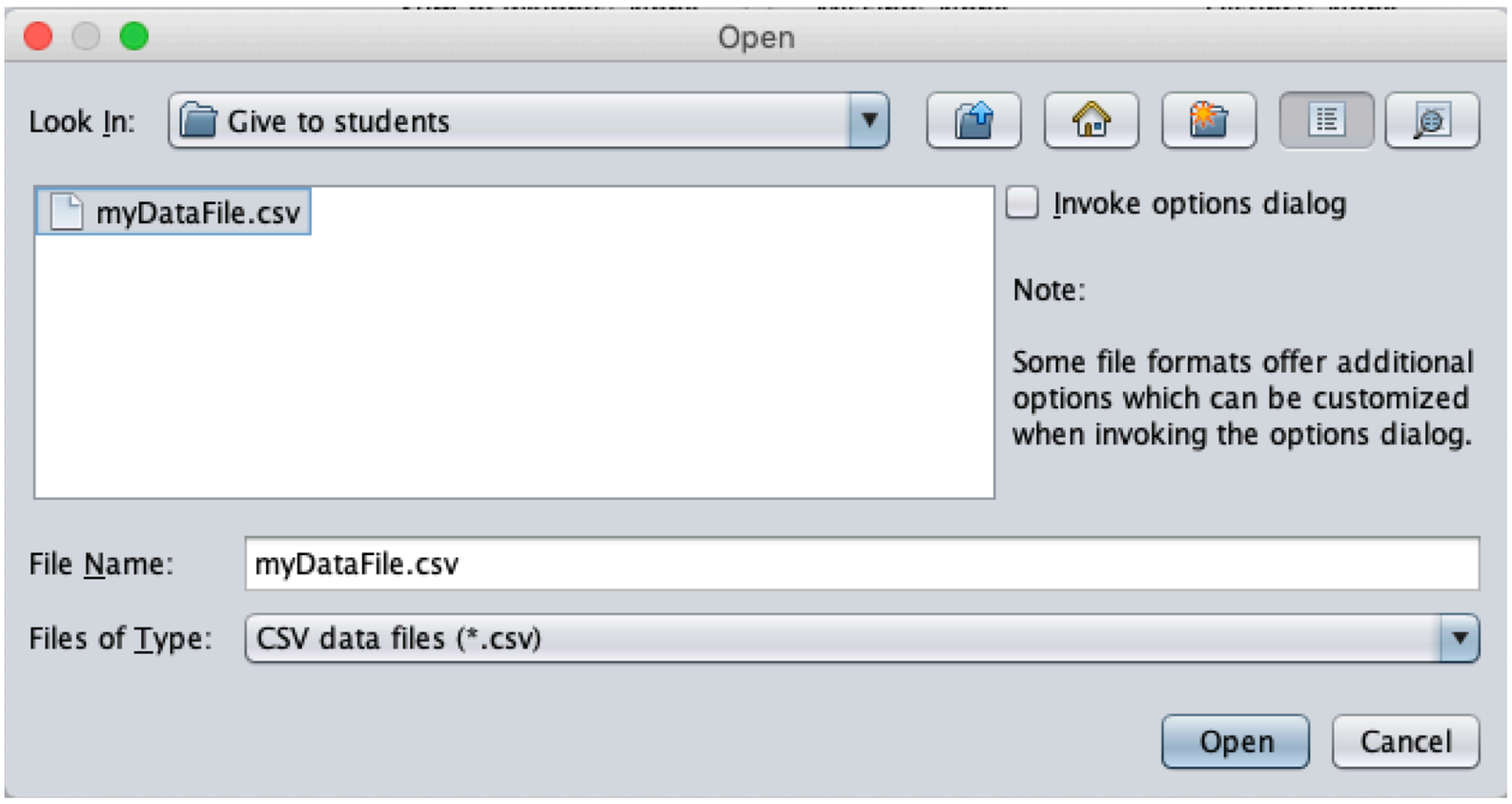
Click on Explorer



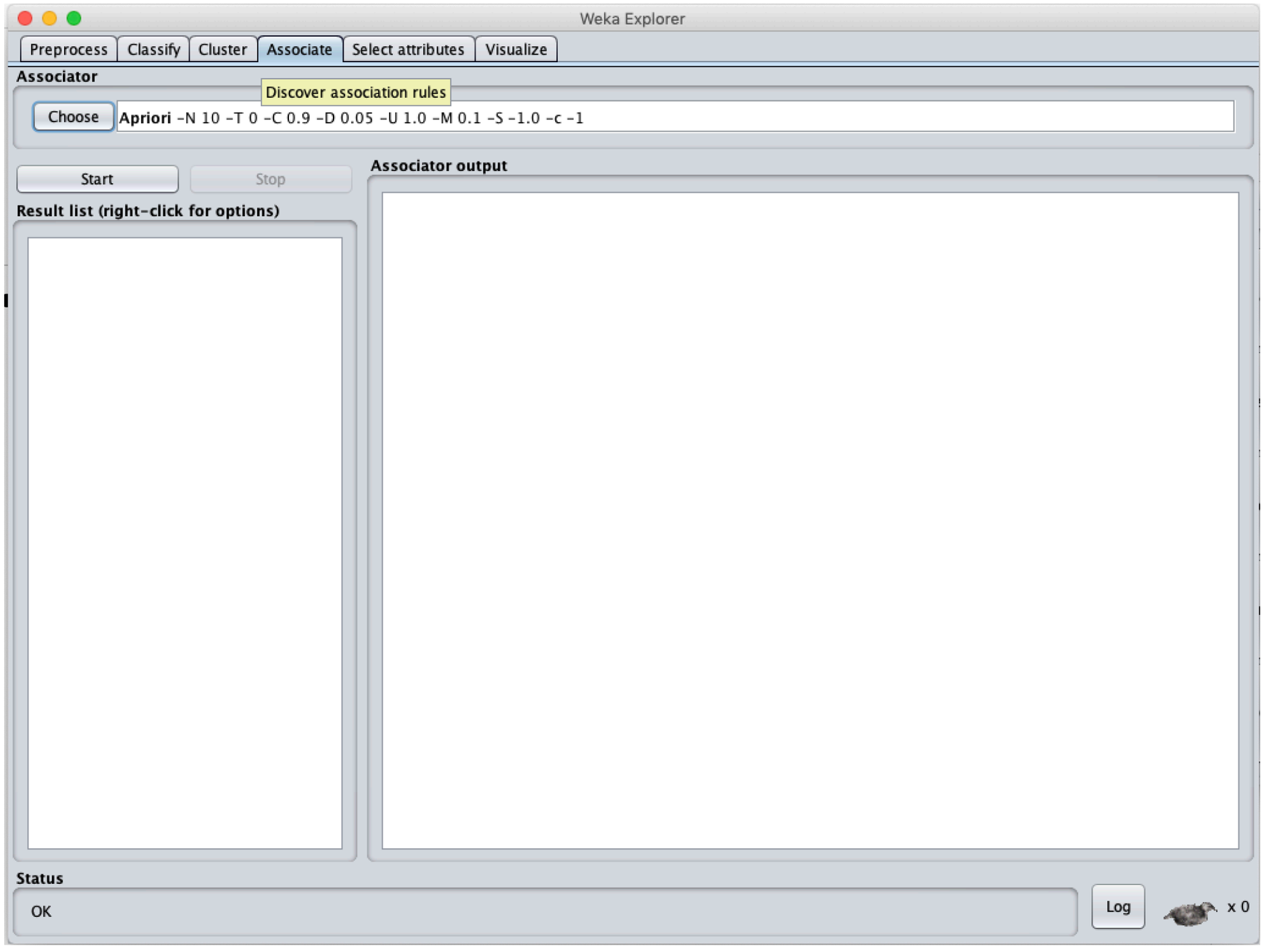
Click on Open File



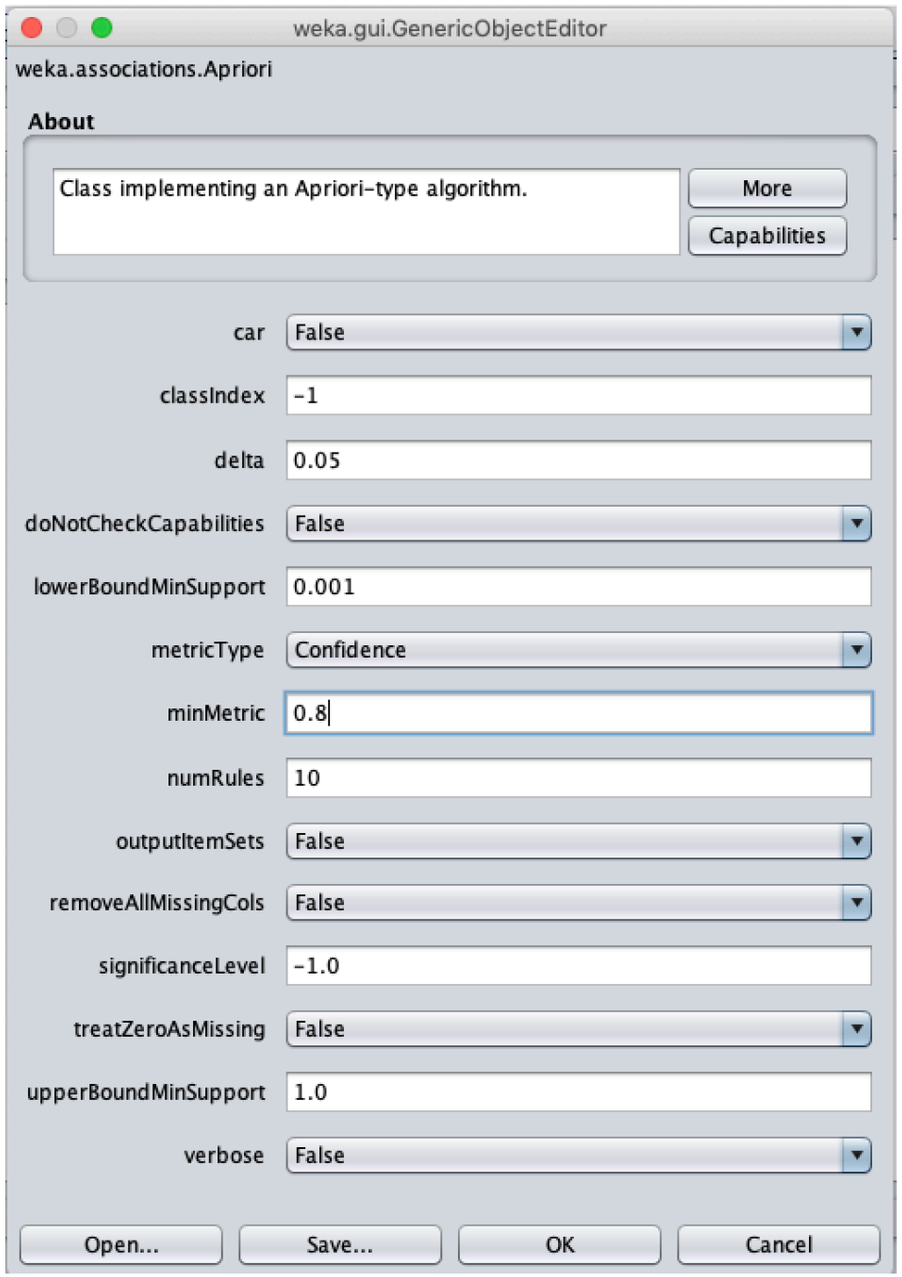
Open the provided myDataFile.csv



Click on the Associate tab, and click in the field next to the Choose button



Set the minimum support and minimum confidence



Finally, press start