

Mid-project update

Market Basket Analysis

Progress so far

- **What have you completed so far?**

As mentioned in the project proposal, we have completed the data processing and additional feature generation. We verified that the dataset that is being used does not have any invalid values by removing negative quantities and prices. The data read from the excel file also had invalid datatypes. We also made sure that proper types are used by manually setting them.

We generated various new attributes like time in hours, total price etc., and used that for exploratory analysis. We were able to come up with several questions that helps better understand the customer behaviour and we were able to successfully answer all of them with the cleaned data.

- **Provide a link to your data so far**

We got the data from <http://archive.ics.uci.edu/ml/datasets/online+retail>. We did some processing and created a csv formatted dataset.

Link to our Github Data

<https://github.com/The-devolper/Market-Basket-Analysis/tree/main/market-basket-analysis-r/data>

- **Provide a link to your code so far (e.g., github repository, etc)**

<https://github.com/The-devolper/Market-Basket-Analysis.git>

Challenges

- **Have you run into any challenges so far? Has anything taken longer than expected?**

We did not face any major challenges so far. Both the data clean-up as well the exploratory analysis were straight forward and we were able to complete them as proposed. Dealing with excel file was new and we had to explore to find a suitable solution.

- **How have you overcome these challenges?**

We found a easy to use library *readxl* that was widely used. We installed and started using the library.

- **If there are any current challenges you have not yet overcome, how do you plan to overcome them?**

As we are done with the exploration part, the next step is to finish the project's major goal, which is developing association rules. For which we need to again convert data into transaction format. We will be writing R logic to handle that conversion.

Other than that choosing the validation mechanism is a challenge. We will explore various choices that are available for validation of rule mining algorithms and pick the one that best suits the dataset.

Collaboration (if applicable)

- **How frequently does your group meet?**
 - We try to meet every 2 days on either zoom or offline and discuss the ideas, issues and progress of the project
- **Does everyone feel like each group member is contributing meaningfully to the project? If not, how does your group plan address this problem?**
 - Yes, each of our teammates are sharing the workload equally.

Next steps

- **What do you still need to complete for your project?**

We would need to convert the dataset into transaction format and feed it to rule mining algorithms. Based on the performance we would tune parameters and try to improve it.

We also need to come up with a proper model validation approach.

- **What is your plan to complete your project?**

We will start implementing the R code required for conversion of dataset into transactions. We will be using multiple rule association algorithms and compare performance based on the validation approach that is chosen. Once we have a stable output, we will further tune the parameters to see if we can improve its performance.

- **Do you foresee any potential challenges?**

The biggest challenge that we foresee is the model validation approach. We will explore more and pick the one that best suits the dataset.

