4. Evaluation Methodology

- Explain the dataset and its source that you employed to study this task. Any specific challenges to using this data for your task?
- List the metrics you employed to evaluate the output of the data mining task and/or questions investigated. Justify their choice from a real-world applications perspective.

Our retail dataset used in this was huge specifically, 20507 entries, over 9000 transactions, and 4 columns. The four columns were DataTime for a date and time stamp of the transactions made, Daypart representing if a transaction is made in the morning, afternoon, evening, or night, DayType classifying whether a transaction was made on a weekend or weekday and Transaction Numbers as the unique identifiers for every transaction. Before we performed the analysis, we were able to preprocess the dataset by removing irrelevant features and normalizing the transaction data. This helped get rid of data like transactions with improbable item counts or dates (noisy data). Normalizing the transaction data helped ensure consistency across items. To do all of this, we parsed the dataset using pandas python library.

In terms of the performance of the Association Rule Algorithm, we evaluated it using a couple metrics such as accuracy, precision, recall and F-1 score. These were calculated based on the predicted top10 itemsets and the actual top 10 itemsets for Confidence, Lift, and Support. In the next few paragraphs, we'll take a look at the results and provide an interpretation of the predictions as well. We willprovide more information on the findings, interesting trends from the data and analyze the algorithm performance. Some recommendations on future work and ways to improve results from similar work will also be briefly discussed.

///////////////////////////////////////	'//////////////////////////////////////	,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	///////////////////////////////////////	'//////////////////////////////////////
,,,,,,,,,	,,,,,,,,,,,,,,,	,,,,,,,,,,,,,,,,,,,,,,,,,		,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,

Results discussion goes here:

For extra information:

#Use graphs and charts to discuss:

Top 10 Itemsets by Month: For each month, we will discuss the top 10 itemsets that were predicted. We will explain any patterns or trends in the predictions.

Top 10 Itemsets Overall (You can get this from top 10 items every month average): We will discuss the top 10 itemsets across all months. We will explain any patterns or trends in the predictions.

Recommendations for future work:

When working towards similar goals in the future with such datasets, we recommend incorporating additional data resources such as customer demographics, weather patterns, and social media, among other categories to help improve the accuracy and performance of the predictions. Having more categories to classify the data caters for random error or sampling variability where the results would be subject to random fluctuations due to limited classification or sampling variability and may not be representative of the data. With this too, the FP-Growth algorithm can better identify frequent association rules and make more accurate predictions.

Feel free to add on recommendations for future work and reproducibility as well.