DNSC 6315
Spring 2022
Assignment 5

A recent association rules study of consumers with infant children found the following raw data regarding the purchase of diapers and beer at grocery stores. Use this data to answer the questions below. Note there are 100,000 total transactions in the dataset, and, for example, the first 10,000 transactions are only {Beer}. **(If you are confused by this notation, reach out to jphall@gwu.edu. Don't turn in wrong answers because you don't understand the example.)**

Purchase Number	**Items**
1	{Beer}
⁝	⁝
10000	{Beer}
10001	{}
⁝	⁝
30000	{}
30001	{Beer, Diapers}
⁝	⁝
80000	{Beer, Diapers}
80001	{Diapers}
⁝	⁝
100000	{Diapers}

1. (1 pt.) What is the support of the rule **Beer ➔ Diapers** ?

2. (1 pt.) True or false: the confidence of the rule **Beer ➔ Diapers** and the rule **Diapers ➔ Beer** are always equal?

3. (1 pt.) What is the lift of the rule **Beer ➔ Diapers** ?

4. (1 pt.) Based on lift, do we expect consumers with infants who purchase beer to also purchase diapers?

5. (1 pt.) What is likely the most profitable application for this information in today's e-commerce environment?

6. (1 pt.) Name a key risk to be aware of when using recommendation engine technologies.

Use the Assignment 11 notebook to answer Questions 7—10. Before running the notebook, copy it to your own drive and follow these instructions carefully:

- Upload the u.data and u.item files, available from GitHub, in Cell 2.
- Set the number of features to extract to 15 in Cell 10.
- Set the random seed in Cell 10.
- Set the number of clusters to 50 in Cell 11.
- Set the random seed in Cell 11.
- Set the matrix to the matrix with all users in Cell 11.

7. (1 pt.) How many non-zero ratings exist in the ratings information?

8. (1 pt.) The *k* archetypes (columns) created in Cell 10 and displayed in Cell 12 can be considered the main types of movies or the main types of users in the ratings data?

9. (1 pt.) How many movies has user 5 already watched?

10. (1 pt.) What are the top 5 movies to recommend to user 5?