



Northeastern University, Khoury College of Computer Science

CS 6220 Data Mining | Assignment 1

Due: September 19, 2023(100 points)

Chun Wei Tseng

<https://github.com/Chun-Wei-Tseng/CS6220-Assignment-1>

Question X [20 pts]

1. Cardinality in basket_data.csv file: 21
- 3a. Total records of movie ratings in the entire dataset: 100498273
- 3b. Total number of unique users 497955
- 3c. Range of years that this data is valid over: 1999 ~ 2005
- 4a. Number of movies with unique name: 17297
- 4b. Number of movie names referring to four different movies: 7
- 5a. Number of users rated exactly 200 movies: 605
- 5b. Names of the movies that they (User ID: 508) liked the most: ["Dinosaur Planet", "Les Dames du Bois de Boulogne", "Blue Juice", "Sarfarosh"]