.

CS 6220 Data Mining | Assignment 1
Due: September 19, 2023(100 points)

.

So Man Amanda Au-Yeung
https://github.com/amandaay/CS6220DataMining/tree/main

Question 1

Cardinality in 'basket_data.csv' is 21.

Question 2

[['ham', 'cheese'], ['ham', 'bread'], ['cheese', 'bread']]

Question 3

(a) Total ratings: 100480507
(b) Total Unique users = 479723
(c) Range of years: 1999 – 2005

Question 4

(a) Total ratings: 17359
(b) Number of movie names refer to four different movies = 5

Question 5

(a) Exactly 200 movies = 605
(b) Most liked movies among the 200 users =
    {'Adaptation',
     'Amelie',
     'American Beauty',
     'Apocalypse Now',
     'Apocalypse Now Redux',
     'Band of Brothers',
     'Being John Malkovich',
     'Bowling for Columbine',
     "Boys Don't Cry",
     'Cabaret',
     'Days of Wine and Roses',
     'Downfall',
     'Election',
     'Eternal Sunshine of the Spotless Mind',
     'Gandhi',

'Garden State',
'Good Will Hunting',
'Harold and Maude',
'High Fidelity',
'L.A. Confidential',
'Lord of the Rings: The Fellowship of the Ring',
'Lord of the Rings: The Return of the King',
'Lord of the Rings: The Return of the King: Extended Edition',
'Lord of the Rings: The Two Towers',
'Lord of the Rings: The Two Towers: Extended Edition',
'Lost in Translation',
'Maria Full of Grace',
'Memento',
'Minority Report',
'Monster',
'Monty Python and the Holy Grail',
"Monty Python's Life of Brian",
"Monty Python's The Meaning of Life: Special Edition",
'Raging Bull',
'Raising Arizona',
'Roger & Me',
"Schindler's List",
'Shakespeare in Love',
'Sideways',
'Super Size Me',
'Taxi Driver',
'The Accused',
'The Lord of the Rings: The Fellowship of the Ring: Extended Edition',
'The Manchurian Candidate',
'The Pianist',
'The Royal Tenenbaums',
'The Shawshank Redemption: Special Edition',
'The Silence of the Lambs',
'The Usual Suspects',
'This Is Spinal Tap',
'Three Kings',
'To Be and To Have',
'To Die For',
'Touching the Void',
'Unforgiven',
'Vietnam: A Television History',
'Whale Rider'}