



Northeastern University, Khoury College of Computer Science

CS 6220 Data Mining | Assignment 0

Due: February 15, 2023(100 points)

Qingxuan Ma

<https://github.com/qingxuanmaa/CS6220DataMining>

Code Review

Question 1

The cardinality in "basket_data.csv" is 21,

Question 2

```
def all_itemsets(items, k):  
    res = []  
    helper(items, k, res, [], 0)  
    print("The list of all possible unique item sets: ", res)
```

```
def helper(items, k, res, temp, index):  
    if k == 0:  
        res.append(temp[:])  
    else:  
        for i in range(index, len(items)):  
            temp.append(items[i])  
            helper(items, k-1, res, temp, i+1)  
            temp.pop()
```

```
all_itemsets(["ham", "cheese", "bread", "milk"], 3) You, 17
```

Examining Our First Dataset

Question 3

- a) The total records of movie ratings in the entire dataset is 100480507
- b) The total unique users in the entire dataset is 480189
- c) The range of years that this data is valid over is 1999 to 2005

Question 4

- a) The number of movies with unique names is 17297
- b) The number of movie names refer to four different movies is: 7

Question 5

- a) The number of users who rated exactly 200 movies is: 605
- b) The lowest user ID is 508, and his favorite movies are {'The Accused', 'Maria Full of Grace', 'The Royal Tenenbaums', 'To Die For', 'American Beauty', 'To Be and To Have', 'Election', 'The Manchurian Candidate', 'Shakespeare in Love', 'Minority Report', 'Lord of the Rings: The Return of the King', 'Monty Python's The Meaning of Life: Special Edition', 'Monty Python's Life of Brian', 'Memento', 'Apocalypse Now', 'Cabaret', 'The Shawshank Redemption: Special Edition', 'Lost in Translation', 'Good Will Hunting', 'Eternal Sunshine of the Spotless Mind', 'Whale Rider', 'Amelie', 'Super Size Me', 'Raising Arizona', 'Band of Brothers', 'Downfall', 'Being John Malkovich', 'Taxi Driver', 'Lord of the Rings: The Two Towers', 'The Lord of the Rings: The Fellowship of the Ring: Extended Edition', 'Boys Don't Cry', 'Touching the Void', 'Raging Bull', 'Monster', 'Three Kings', 'Unforgiven', 'Garden State', 'Monty Python and the Holy Grail', 'Adaptation', 'Bowling for Columbine', 'Apocalypse Now Redux', 'High Fidelity', 'Gandhi', 'L.A. Confidential', 'Schindler's List', 'Harold and Maude', 'The Silence of the Lambs', 'Lord of the Rings: The Return of the King: Extended Edition', 'Lord of the Rings: The Fellowship of the Ring', 'The Pianist', 'The Usual Suspects', 'Days of Wine and Roses', 'Lord of the Rings: The Two Towers: Extended Edition', 'Vietnam: A Television History', 'This Is Spinal Tap', 'Sideways', 'Roger & Me'}