



Northeastern University, Khoury College of Computer Science

---

## CS 6220 Data Mining | Assignment 1

Due: January 18, 2023(100 points)

---

Yixuan Huang

Link:

Question 1: the cardinality in “basket data.csv” is 21

Question 2:

```
def all_itemsets(items, k):  
    """  
    Generate all possible unique itemsets with non-repeating k items  
    from a list of unique items.  
    """  
    def generate_itemsets(start, current_set, remaining_items):  
        if len(current_set) == k:  
            result.append(current_set[:])  
            return  
  
        for i in range(start, len(remaining_items)):  
            current_set.append(remaining_items[i])  
            generate_itemsets(i + 1, current_set, remaining_items)  
            current_set.pop()  
  
    result = []  
    generate_itemsets(0, [], items)  
    return result
```

Question 3:

- a) Total number of movie rating records: 100480507
- b) Total number of unique users: 480189
- c) The range of years that the data is valid over is: 1999 to 2005

Question 4:

- a) Total number of unique movie names: 17297
- b) Number of movie titles that refer to exactly four different movies: 7

Question 5:

- a) Number of users who rated exactly 200 movies: 605
- b) Favorite movies of user 508:

High Fidelity  
Monty Python's The Meaning of Life: Special Edition  
American Beauty  
Roger & Me  
Eternal Sunshine of the Spotless Mind  
Being John Malkovich  
Vietnam: A Television History  
Super Size Me  
Lord of the Rings: The Fellowship of the Ring  
This Is Spinal Tap  
The Pianist  
The Silence of the Lambs  
Sideways  
Whale Rider  
Garden State  
Bowling for Columbine  
Gandhi  
Apocalypse Now Redux  
To Die For  
Monty Python's Life of Brian  
The Manchurian Candidate  
Memento  
Amelie  
Apocalypse Now  
The Usual Suspects  
Lord of the Rings: The Two Towers: Extended Edition  
The Lord of the Rings: The Fellowship of the Ring: Extended Edition  
Touching the Void  
Minority Report  
The Royal Tenenbaums  
Election  
Good Will Hunting  
L.A. Confidential  
Taxi Driver  
Lord of the Rings: The Two Towers  
Cabaret  
Adaptation  
The Accused  
Lost in Translation  
Boys Don't Cry  
To Be and To Have  
Schindler's List

Raging Bull  
Lord of the Rings: The Return of the King  
Monty Python and the Holy Grail  
Raising Arizona  
The Shawshank Redemption: Special Edition  
Harold and Maude  
Downfall  
Lord of the Rings: The Return of the King: Extended Edition  
Monster  
Band of Brothers  
Three Kings  
Unforgiven  
Maria Full of Grace  
Days of Wine and Roses  
Shakespeare in Love