



Northeastern University, Khoury College of Computer Science

CS 6220 Data Mining | Assignment 1

Due: January 18, 2023 (100 points)

Yichen Sun

<https://github.com/LAnselet/cs6220-datamining>

Topic: Coding Review

Question 1 [20 pts]

The cardinality in "basket_data.csv" is 21.


Question 2 [20 pts]

```
def all_itemsets(items, k):
    result = [[]]

    for item in items:
        result += [current + [item] for current in result]

    return [comb for comb in result if len(comb) == k]

all_itemsets(["ham", "cheese", "bread"], 2)
```

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

This approach uses a concept which is similar to the power set, but instead of returning all possible subsets, it filters them to include only those of a specific length.

Topic: Examining Our First Dataset

Question 3 [20 pts]

- There are 100480507 records of movie ratings in entire dataset.
- There are 480189 unique users in entire dataset.
- Range of years: 1999 to 2005.

Question 4 [20 pts]

- There are 17359 with unique names.
- There are 5 movies refers to four different movies.

Question 5 [20 pts]

a. There are 605 users rated exactly 200 movies.

b. The lowest user id is 508. The name of the movies that this person likes the most:

Lord of the Rings: The Two Towers

Cabaret

Harold and Maude

Amelie

Election

Lord of the Rings: The Two Towers: Extended Edition

Lord of the Rings: The Fellowship of the Ring

Taxi Driver

Monty Python and the Holy Grail

Monster

Gandhi

Maria Full of Grace

Raging Bull

This Is Spinal Tap

Monty Python's Life of Brian

The Accused

Lord of the Rings: The Return of the King

Super Size Me

Whale Rider

The Silence of the Lambs

Raising Arizona

Adaptation

To Be and To Have

The Pianist

High Fidelity

American Beauty

The Usual Suspects

The Lord of the Rings: The Fellowship of the Ring: Extended Edition

Band of Brothers

Minority Report

Schindler's List

Good Will Hunting

Eternal Sunshine of the Spotless Mind

Being John Malkovich

Touching the Void

Lost in Translation

Garden State

The Royal Tenenbaums

Downfall

Roger & Me

Sideways

Apocalypse Now

Boys Don't Cry

The Shawshank Redemption: Special Edition

L.A. Confidential

Unforgiven

Days of Wine and Roses

The Manchurian Candidate

Shakespeare in Love

Memento

Three Kings

Monty Python's The Meaning of Life: Special Edition

Vietnam: A Television History

Lord of the Rings: The Return of the King: Extended Edition

To Die For

Apocalypse Now Redux

Bowling for Columbine

