Northeastern University, Khoury College of Computer Science

CS 6220 Data Mining | Assignment 1

Due: September 19, 2023 (100 points)

Spencer Solomon
https://github.com/spencersolomon6/DataMining_HW1

Coding Review

Question 1

The Cardinality of the “basket_data.csv” dataset is **21**.

Question 2

NA: Question only relates to code

Question 3

a) There are **100,480,506** total movie ratings across all of the combined_records files.
b) There are **480,189** total unique users in the dataset.
c) The years this dataset covers is **1999-2005**.

Question 4

a) The number of unique titles in the dataset is **17,355.**
b) The number of titles with 4+ movies is **8.**

Question 5

1. Of all the users, **605** users rated exactly 200 movies.
2. The list of movies this user liked most is:
   - **2.1. high fidelity**
   - **2.2. monty python's the meaning of life: special edition**

2.3. **american beauty**
2.4. **roger & me**
2.5. **eternal sunshine of the spotless mind**
2.6. **being john malkovich**
2.7. **vietnam: a television history**
2.8. **super size me**
2.9. **lord of the rings: the fellowship of the ring**
2.10. **this is spinal tap**
2.11. **the pianist**
2.12. **the silence of the lambs**
2.13. **sideways**
2.14. **whale rider**
2.15. **garden state**
2.16. **bowling for columbine**
2.17. **gandhi**
2.18. **apocalypse now redux**
2.19. **to die for**
2.20. **monty python's life of brian**
2.21. **the manchurian candidate**
2.22. **memento**
2.23. **amelie**
2.24. **apocalypse now**
2.25. **the usual suspects**
2.26. **lord of the rings: the two towers: extended edition**
2.27. **the lord of the rings: the fellowship of the ring: extended edition**
2.28. **touching the void**
2.29. **minority report**
2.30. **the royal tenenbaums**
2.31. **election**
2.32. **good will hunting**
2.33. **l.a. confidential**
2.34. **taxi driver**
2.35. **lord of the rings: the two towers**
2.36. **cabaret**
2.37. **adaptation**
2.38. **the accused**
2.39. **lost in translation**
2.40. **boys don't cry**
2.41. **to be and to have**
2.42. **schindler's list**
2.43. **raging bull**
2.44. **lord of the rings: the return of the king**
2.45. **monty python and the holy grail**
2.46. **raising arizona**
2.47. **the shawshank redemption: special edition**
2.48. **harold and maude**
2.49. **downfall**
2.50. **lord of the rings: the return of the king: extended edition**
2.51. **monster**
2.52. **band of brothers**
2.53. **three kings**
2.54. **unforgiven**
2.55. **maria full of grace**
2.56. **days of wine and roses**
2.57. **shakespeare in love**