# דו"ח תרגיל 5 עדן כהן 318758778

## חלק 1

1. 1. מספר המשתמשים הייחודיים- 5000
   2. מספר המוצרים הייחודים – 500
   3. מספר הדירוגים- 399664
   4. מספר הדירוגים המינימלי שניתן למוצר – 0.0
   5. מספר הדירוגים המקסימלי שניתן למוצר – 0.5
   6. מספר הדירוגים המינימלי שמשתמש דירג- 53
   7. מספר הדירוגים המקסימלי שמשתמש דירג- 114
   8. במקרה בו יש מוצרים ללא דירוגים או משתמשים (חדשים לדוגמא) שלא דירגו מוצרים, המערכת תתקשה למצוא להם המלצה טובה כיוון שהמערכת לא תוכל למצוא משתמשים שדומים למשתמש זה (עם דפוס דירוג דומה). וכן עבור פריט ללא דירוג או עם דירוג מועט, המערכת לא תוכל לקבוע את הדירוג שלו לפי דירוגים של משתמשים אחרים (כי הם לא דירגו אותו) וכן לא תוכל גם למצוא פריטים שדומים לו.
   9. עבור מוצרים חדשים או מוצרים עליהם אנו מעוניינים להמליץ למרות מיעוט נתוני הדירוג נשתמש במערכת Content Based Filtering **שמתבססת על תכונות המוצרים ולפי זה מתאימה למשתמש פריטים שדומים לפריטים אחרים שדירג גבוה.  
      עבור משתמשים חדשים שאין להם דירוגים נשמש גם** במערכת Content Based Filtering **המתבססת על תכונות המשתמש ולא על הדירוגים שלו, כלומר מוצאת משתמשים דומים למשתמש החדש על פי תכונותיו (גיל, מגדר, מקום מגורים ועוד) ובהתאם לכך ממליצה למשמש על פריטים לפי ההמלצות של המשתמשים הדומים לו.**
2. ['B0001Y74VS', 'B0002B0R14', 'B0009EXONC', 'B0002JKQ8K', 'B0007CXX82']
3. ['B000052WYN', 'B000052ZSS', 'B0000534VO', 'B0000536M2', 'B0000632EN']

|  |  |
| --- | --- |
| RMSE |  |
| 0.54692 | User based |
| 0.55211 | Item based |
| 0.60155 | Mean based (bench-mark) |



|  |  |  |
| --- | --- | --- |
| Recall@20 | Percision@20 |  |
| 0.24706 | 0.04747 | User-based CF |
| 0.21285 | 0.04028 | Highest-rank (benchmark) |