**Exercise 6: Processing MovieLens data**

**Assignment Specification**

**Description**: This program will use Pandas to process MovieLens movie rating data.

**Input**: No user provided input. Data from files: movies.dat, users.dat, and ratings.dat

Please, refer to the movielens\_readme.txt file for details.

**Output**:

See details in the Procedure.

**Procedure**:

1. Starting on page 26 of the McKinney text, there is a description of a study of movie ratings data collected from the MovieLens website and available on the grouplens site [www.grouplens.org](http://www.grouplens.org)

The datasets for the million-ratings study have been uploaded to Canvas

1. Following the description in the text, load the three datasets movies.dat, users.dat, and ratings.dat into a pandas DataFrame
2. Print the FIRST 3 rows of each of the three DataFrames. Leave a blank line before each DataFrame listing, and print an appropriate heading
3. Merge the data into a single DataFrame named ***data,*** using pandas.merge (pd.merge) as described in the text
4. Print (with appropriate headings) the number of records in each of the 4 DataFrames (movies, users, ratings, data).
5. “Occupation” column has values from 0 to 20. Replace the numbers as:

0 other/not specified

1 academic/educator

2 artist

3 clerical/admin

4 college/grad student

5 customer service

6 doctor/health care

7 executive/managerial

8 farmer

9 homemaker

10 K-12 student

11 lawyer

12 programmer

13 retired

14 sales/marketing

15 scientist

16 self-employed

17 technician/engineer

18 tradesman/craftsman

19 unemployed

20 writer

1. Print the LAST 3 rows of the DataFrame ***data*** (with appropriate heading).
2. Find the 5 occupations that give higher ratings for movies on DataFrame ***data***.