3460:489/589 Platforms for Big Data Programming **Spring 2020**
Project #1: **(50 pts)** **Due: 2/16/20**

Problems:

Use the MovieLens data sets to generate the following analytics results:

1. (10 pts) Show total number of rows with missing values (after join all files), and remove those rows from the DataFrame.
2. (5 pts) Show
   total number of movies,
   total number of genres,
   total number of users,
   total number of ratings.

3. (20 pts) Plot the historgrams for
   number of ratings per user,
   number of movies per year,
   number of movies per genra,
   average ratings per movie,
   average ratings per genra.
4. (15 pts)
   (1) Show the number of users whose number of ratings are greater or equal to the median of the number of ratings,
   (2) Show the top ten movies with title and genres rated by each user from (1),
   (3) Show the average rating per genra for each user from (1)

Submission:
Name your source file as p1_xxxx.py, where xxxx is the last four digits of your UA ID.
Submit your source file to the Brightspace by midnight of due day.