Data Description:

MovieLens data sets were collected by the GroupLens Research Project

at the University of Minnesota.

Domain:

Entertainment and Internet

Context:

The GroupLens Research Project is a research group in the Department of Computer Science and Engineering at the University of Minnesota. The data is widely used for collaborative filtering and other filtering solutions. However, we will be using this data to act as a means to demonstrate our skill in using Python to "play" with data.

Attribute Information:

- Download the zip file from data source
- Extract the zip file and you will find a folder named ml-100k
- Go through the README file that you will find in the folder from the above step where you will find the information about the attributes in the three datasets

Learning Outcomes:

- Exploratory Data Analysis
- Visualization using Python
- Pandas

Objective:

Demonstrate your skill in python for data analysis.

Steps and tasks:

- You will need to import 3 files from the folder as data frames into your Jupyter notebook
    - u.data
    - u.item
    - u.user

    (You might encounter some trouble importing the data, you are expected to figure out on your own)

- Display univariate plots of the attributes: 'rating', 'age', 'release date', 'gender' and 'occupation', from their respective data frames
- Visualize how popularity of Genres has changed over the years. From the graph one should be able to see for any given year, movies of which genre got released the most.
- Display the top 25 movies by average rating, as a list/series/dataframe.
  Note:- Consider only the movies which received atleast a 100 ratings
- Verify the following statements (no need of doing a statistical test. Compare absolute numbers):
    - Men watch more drama than women
    - Men watch more Romance than women
    - Women watch more Sci-Fi than men

References:

- https://movielens.org/