greatlearning
Power Ahead

Project Debrief

Foundation of AIML

MovieLens Data Exploration

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.

Domain:

Internet and Entertainment

Objective:

To implement the techniques learnt as a part of the course.

Attribute Information

- We have three datasets: data.csv, item.csv, user.csv
- **Data.csv:** It contains information of ratings given by the users to a particular movie.
  - Columns: user id, movie id, rating, timestamp
- **item.csv:** File contains information related to the movies and its genre.
  - Columns: movie id, movie title, release date, unknown, Action, Adventure, Animation, Children's, Comedy, Crime, Documentary, Drama, Fantasy, Film-Noir, Horror, Musical, Mystery, Romance, Sci-Fi, Thriller, War, Western
- **user.csv:** It contains information of the users who have rated the movies.
  - Columns: user id, age, gender, occupation, zip code

Steps and Tasks

- Importing necessary Libraries
- Explore all the three datasets
- Number of movies in each genre
- Number of genres for each movie
- Dropping the movies where genre is unknown
- Univariate plots of columns: 'rating', 'Age', 'release year', 'Gender' and 'Occupation'
- Visualizing how popularity of different genres has changed over the years.
- Finding the top 25 movies according to average ratings such that each movie has number of ratings more than 100
- See gender distribution across different genres check for the validity of the below statements:

  Men watch more drama than women

  Women watch more Sci-Fi than men

  Men watch more Romance than women

Functions and Libraries used

- Numpy
- Pandas: Value Counts, Sort Values, Info, Describe
- Matplotlib
- Seaborn

Learning Outcome

- Exploratory Data Analysis
- Visualization using Python
- Pandas – groupby, merging

greatlearning
Power Ahead

Happy Learning !