

Project 04

Data Science with Python — Real World Project

After learning about Data Science in depth, it is now time to implement the knowledge gained through this course in real-life scenarios. We will provide you with four scenarios where you need to implement data science solutions. To perform these tasks, you can use the different Python libraries such as NumPy, SciPy, Pandas, scikit-learn, matplotlib, BeautifulSoup, and so on.

Movielens Dataset Analysis

The GroupLens Research Project is a research group in the Department of Computer Science and Engineering in the University of Minnesota. The researchers of this group are involved in many research projects related to the fields of information filtering, collaborative filtering, and recommender systems. Here, we ask you to perform the analysis using the Exploratory Data Analysis technique. In particular, we want you to apply the tools of machine learning to predict the survivors of the tragedy.

The details of these projects and the scope of each project are listed in the sections below.

- Data acquisition of the movielens dataset
 - users dataset
 - rating dataset
 - movies dataset
- Perform the Exploratory Data Analysis (EDA) for the users dataset
 - Visualize user age distribution
 - Visualize overall rating by users
 - Find and visualize the user rating of the movie “Toy Story”
 - Find and visualize the viewership of the movie “Toy Story” by age group
 - Find and visualize the top 25 movies by viewership rating
 - Find the rating for a particular user of user id = 2696
 - Visualize the rating data by user of user id = 2696
- Perform machine learning on first 500 extracted records
 - Use the following features:
 - movie id

- age
 - occupation
- Use rating as label
- Create train and test data set and perform the following:
 - Create a histogram for movie, age, and occupation

The required resources for this project will be available in the Downloads section (Learning Tools > Downloads > Projects). Good Luck!