**CSE523 - Machine Learning**

**Movie Recommendation System using Machine Learning**

**Faculty - Prof. Mehul Raval**

**Weekly Report 2**

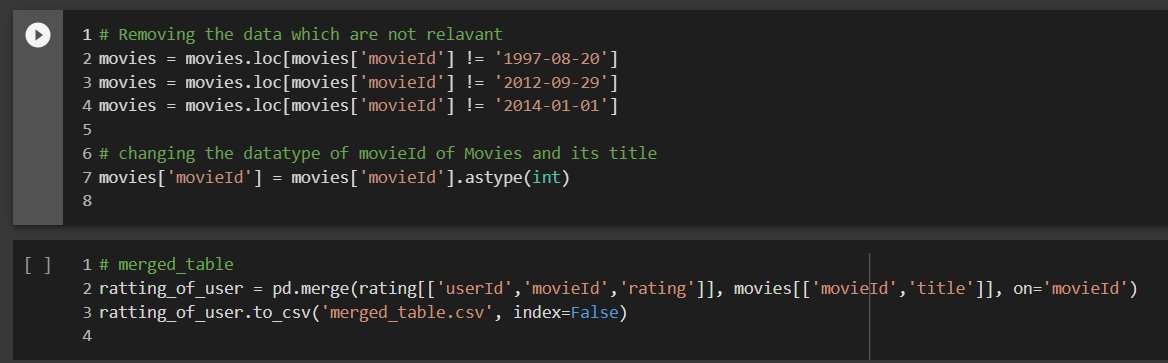
**Submission Date: 18-02-23**

**Group: Tech Titans**

| **Name** | **Roll Number** | **Email** |
| --- | --- | --- |
| Jainam Shah | AU2040186 | jainam.s4@ahduni.edu.in |
| Yash Chotaliya | AU2040193 | yashkumar.c@ahduni.edu.in |
| Akshay Parmar | AU2040199 | akshay.p@ahduni.edu.in |
| Shubham Bhatt | AU2040206 | shubham.b1@ahduni.edu.in |

Dataset Link: <https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset?select=movies_metadata.csv>

In this week, we have analyzed data



Data filtering is an essential process in data analysis and manipulation. Filtering data enables the user to extract and manipulate specific data subsets required for analysis. In Python, the Pandas library is commonly used for filtering data, as it offers a variety of functions that facilitate filtering data in various ways.

One common way of filtering data in Pandas is by selecting rows based on specific conditions. For example, you can select all rows where a particular column meets a specific criterion using the loc function. The syntax for this function is df.loc[df['column\_name'] condition], where df is the data frame’s name, column\_name is the column’s name, and the condition is the condition to be met. The condition can be an expression or a function that returns a boolean value.

Another way of filtering data in Pandas is by dropping rows or columns that are not required. The drop function can remove rows or columns from a data frame based on their index or label. For example, df.drop(columns=['column\_name']) would drop the column named column\_name from the data frame df. You can also drop rows based on a condition using df.drop(df[df['column\_name'] condition].index).

Additionally, Pandas offers to filter based on string manipulation. This allows you to filter rows based on the presence or absence of a specific string or substring. You can use the contains function to check for the presence of a string, which starts with and ends with functions to check if a string starts or ends with a specific substring.

In conclusion, data filtering is essential to data analysis, and Python provides various tools to facilitate this process. Pandas offer a variety of functions that enable users to filter data based on specific conditions, drop unnecessary rows or columns, and manipulate strings to extract relevant information. By mastering data filtering techniques in Python, users can effectively analyze large datasets and derive valuable insights.

