**Assignment 3**

The Netflix Prize data has over 100 million records of ratings. There are about 17k movies. Determine the correlations of each pair among all the movies. Report all the pairs with correlation >0.90.

To calculate the correlation, you need to define a vector for each movie. The vector can be all the ratings for that movie. Note some users have not rated that movie. You may treat the absence of ratings as a special number such as 0 or -1.

1. Print all the columns:

**Graphical user interface, table

Description automatically generated**

1. Create Two subset with two movies and ratings and fill one column with Zero.

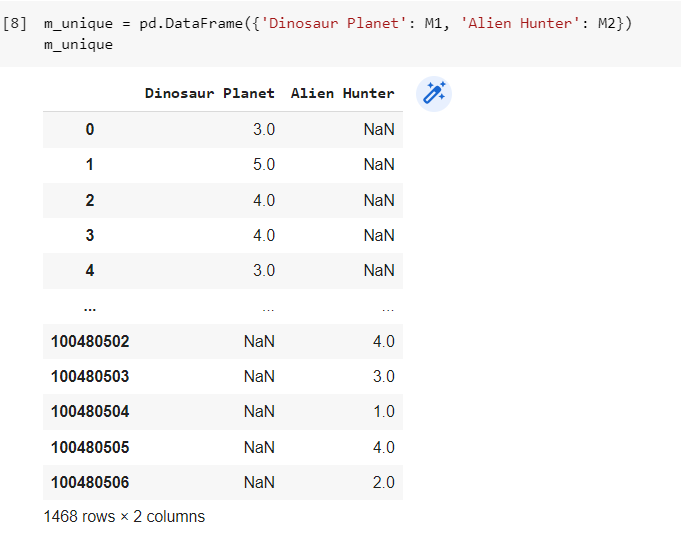
Table

Description automatically generated

1. Check the unique number of the movies

Graphical user interface

Description automatically generated with medium confidence

****

1. Check the missing values and fill it with zeros.

Table

Description automatically generated

1. Check the correlation between the selected two movies.

Graphical user interface, text, application

Description automatically generated

Graphical user interface, application

Description automatically generated

1. Check the correlation between all the movies.

Graphical user interface, text, application

Description automatically generated

A screenshot of a computer

Description automatically generated with medium confidence