**Building user-based recommendation model for Amazon.**

Project 3

DESCRIPTION

The dataset provided contains movie reviews given by Amazon customers. Reviews were given between May 1996 and July 2014.

**Data Dictionary**  
UserID – 4848 customers who provided a rating for each movie  
Movie 1 to Movie 206 – 206 movies for which ratings are provided by 4848 distinct users

**Data Considerations**  
- All the users have not watched all the movies and therefore, all movies are not rated. These missing values are represented by NA.  
- Ratings are on a scale of -1 to 10 where -1 is the least rating and 10 is the best.

**Analysis Task**  
- Exploratory Data Analysis:

* Which movies have maximum views/ratings?
* What is the average rating for each movie? Define the top 5 movies with the maximum ratings.
* Define the top 5 movies with the least audience.

- Recommendation Model: Some of the movies hadn’t been watched and therefore, are not rated by the users. Netflix would like to take this as an opportunity and build a machine learning recommendation algorithm which provides the ratings for each of the users.

* Divide the data into training and test data
* Build a recommendation model on training data
* Make predictions on the test data