Import necessary pyspark and additional libraries

Intialize spark context and spark session

Read in the ratings csv to an RDD

Remove the header from the rdd using the filter

Split on comma and convert the userId movieId and rating to a pyspark Rating object

Create a new user 0 and read in their movie predictions

Create an RDD with the user 0 movie predictions and ratings converted to a pyspark Rating object

Union the Ratings rdd with the newUser RDD

Set the rank and numIterations and train the Alternating Least Squared model on the ratings + new user RDD training data and parameters

Read in a sample prediction csv for the new user

Predict the new model on the new sample test data and output the sample predictions for user 0