Part b

Question 3

MSE, RMSE, MAE-

MSE- mean of all the squares of the induvial errors for every case. It is sensitive to values with great error as it squares them. It produces errors between 0 and infinity.

RMSE- is just the square root of MSE to bring the error to the same unit order as the data.

MAE- it is the mean of the absolute value of all errors. It also lies between 0 and infinity. It treats every error the same as there is no squaring involved.

Comparing both models with RMSE, MSE, MAE-

Model 1 with 80/20 split-

```
1  model = als.fit(training)
```

Show result

Cmd 12

```
1  predictions = model.transform(test)
```

Show result

Cmd 13

```
1  rmse = eval.evaluate(predictions)
2  mse = eval2.evaluate(predictions)
3  mae = eval3.evaluate(predictions)
```

Show result

Cmd 14

```
1  print("rmse-")
2  print(rmse)
3  print("mse-")
4  print(mse)
5  print("mae-")
6  print(mae)
```

```
rmse-
1.1501724845428571
mse-
1.322896744199489
mae-
0.7908617407083511
```

Model 2 with 70/30 split-

```
1  model2 = als.fit(training2)
```

Show result

Cmd 16

```
1  predictions2 = model2.transform(test2)
```

Show result

Cmd 17

```
1  rmse2 = eval.evaluate(predictions2)
2  mse2 = eval2.evaluate(predictions2)
3  mae2 = eval3.evaluate(predictions2)
```

Show result

Cmd 18

```
1  print("rmse-")
2  print(rmse2)
3  print("mse-")
4  print(mse2)
5  print("mae-")
6  print(mae2)
```

```
rmse-
1.259408289740908
mse-
1.5861092402681192
mae-
0.8795168348219209
```

We can see that model 1 has a lower RMSE with 1.150 while model 2 has RMSE of 1.259. This can be attributed to the fact that model 1 had more training data (80%) to train on there by creating a more competent model. Having more training data ensures the model has seen more possibilities while it trains thereby increasing the accuracy of the model.