

```
In [13]: ### import needed libraries
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd
import numpy as np
```

```
In [9]: ### load data in from csv to pandas dataframe
book_data = pd.read_csv('C:/Users/ODsLaptop/Documents/CUNY/DATA_612_rec_system
s/project1/project1/book_data.csv')
```

```
In [10]: ### create a user-item matrix from the dataframe
book_matrix = book_data.pivot_table(index='Reviewer', columns='Book',
                                     values='Score')
```

```
In [11]: ### break ratings into training and testing datasets
from sklearn.model_selection import train_test_split
train, test = train_test_split(book_matrix, test_size=0.2)
```

```
In [12]: ### calculate the raw average of training data
total_sum = (train['Hamlet'].sum(skipna=True) +
             train['Count of Monte Cristo'].sum(skipna=True) +
             train['The Aneid'].sum(skipna=True) +
             train['The Odyssey'].sum(skipna=True) +
             train['The Great Gatsby'].sum(skipna=True))

total_entries = (~np.isnan(train)).sum(1).sum()

train_mean = total_sum/total_entries

print "training raw average: ", train_mean

training raw average:  2.7058823529411766
```

```
In [14]: ### calculating rmse for training data
stan_err = []
for i in train.columns[1:5]:
    for j in train[i]:
        if np.isnan(j) == False:
            stan_err.append(j-train_mean)

train_rmse = sum(stan_err)/len(stan_err)

print "training RMSE: ", train_rmse

training RMSE:  -0.2443438914027151
```

```
In [15]: stan_err = []
        for i in test.columns[1:5]:
            for j in test[i]:
                if np.isnan(j) == False:
                    stan_err.append(j-train_mean)

        test_rmse = sum(stan_err)/len(stan_err)

        print "testing RMSE: ", test_rmse
```

testing RMSE: 0.04411764705882337

```
In [16]: # calculating the bias for each user

        for user in train.columns[1:5]:
            print user, train[user].mean(skipna=True)
```

Hamlet 2.66666666667
The Aneid 2.66666666667
The Great Gatsby 2.25
The Odyssey 2.33333333333

```
In [17]: print train.mean(axis=1, skipna=True)
```

Reviewer
Roger 1.75
Sandy 4.00
Diane 3.25
Molly 2.00
dtype: float64