Friend Recommendation System:

Code File Description:

1) GenerateRecommendation.java file generates upto top 10 recommendations. The concept used here for suggesting the recommendations is as below:

   For each user U, the algorithm recommends N = 10 users who are not already friends with U, but have the largest number of mutual friends in common with U. Only the second level friendship to find mutual friends is used here.

   Input: soc-LiveJournal1Adj.txt which has the input in the form <USER ID’S>TAB<FRIEND ID’S(comma separated)>
   Output: Below is a sample output (only for few users):

   Note: The recommendations here are less than 10 for most users because they have less than 10 second degree friends

```
924     11860,15416,2409,43748,439,45881,6995,
8941    8943,8944,8940,
8942    8939,8940,8943,8944,
9019    9022,317,9023,
9020    9021,9016,9017,9022,317,9023,
9021    9020,9016,9017,9022,317,9023,
9022    9019,9020,9021,317,9016,9017,9023,
9990    13134,13478,13877,34299,34485,34642,37941,
9992    9987,9989,35667,9991,
9993    9991,13134,13478,13877,34299,34485,34642,37941,
```

2) GetAddressDetailsOfMutualFriends.java file retrieves the mutual friends given two userID’s

   Input: soc-LiveJournal1Adj.txt which has the input in the form <USER ID’S>TAB<FRIEND ID’S(comma separated)> and userID’s (ex: 9,10)

   Output: Below is a sample output : 9 10 0

   (Indicating 0 is the mutual friend of 9 and 10)

3) GetAddressDetailsOfMutualFriends.java file also retrieves the mutual friends given two users but along with their name and zipcode . The name and zipcode information are stored in another input file which is loaded in-memory.
   Input:
   a) soc-LiveJournal1Adj.txt which has the input in the form <USER ID’S>TAB<FRIEND ID’S(comma separated)>
   b)userdata.txt (column1 : userid column2 : firstname column3 : lastname column4 : address column5: city column6 :state column7 : zipcode column8 :country column9 :username column10 : date of birth)
   c)userID’s (ex: 1234 ,4312)

Output: 1234 4312 [John:75075, Jane : 75252, Ted:45045]

Below are the instructions to run the code :

The Hadoop single node setup was used to run the programs and the input files were placed in the hdfs system and the below commands can be run (the path of the jar would need to be changed)

```
1)hadoop jar /home/supriya/WayFair/FriendRecommendationSystem/Recommendation.jar
Assignment1_part1 soc-LiveJournal1Adj.txt outputforpart1
```

```
hdfs dfs -cat outputforpart1/*
```

```
2)hadoop jar /home/supriya /WayFair/FriendRecommendationSystem/Recommendation.jar
Assignment1_part2 soc-LiveJournal1Adj.txt outputforpart2 1 6
```

```
hdfs dfs -cat outputforpart2/*
```

```
3)hadoop jar /home/supriya /WayFair/FriendRecommendationSystem/Recommendation.jar
Assignment1_part3 soc-LiveJournal1Adj.txt outputforpart1 1 6
```

```
hdfs dfs -cat outputforpart3/*
```

The dataset is placed in : WayFair\FriendRecommendationSystem\dataset