# Homework1 Description

## Readme:

## Contents

## Description:

The Source Code of both task1 and task2 is in the folder Ridhi_Shika\Solution. My source code for task1 and task2 is called Ridhi_Shikha_task1.py and Ridhi_Shikha_task2.py.
All the 4 Output Files are inside the folder OutputFiles.

IMPORTANT:
**I have used jupyter notebook in Windows to solve this Homework, so please test in a windows environment.**

## How to run the task 1:

Open File Ridhi_Shikha_task1: Inside Ridhi_Shikha\Solutions

```
21
22    ratings = pd.read_csv("C:\\Users\\ridhi\\OneDrive\\Documents\\inf 553\\small\\ratings.csv")
23    df2= pd.DataFrame(ratings)
24    df2.columns = df2.columns.str.replace('\s+', '')  #remove whitespace
25    cols= df2[["movieId","rating"]]                    #choose columns
26    avg= cols.groupby("movieId")["rating"].mean()      #group by movieId and mean of rating
27
28    #write to file
29    File_Write = open('C:\\Ridhi_Shikha_HW1\\Ridhi_Shikha\\OutputFiles\\Ridhi_Shikha_result_task1_small.csv', 'w')
```

For small data set:

- Change the path of ratings to the input source of small data in ratings
- Change the path of File_Write to the path where you want the output file to be stored.

For big data set:

- Change the path of ratings to the input source of big data in ratings
- Change the path of File_Write to the path where the output file has to be stored.

## How to run the task 2:

Open File [Ridhi_Shikha_task2](#): Inside Ridhi_Shikha\Solutions

```
22   ratings1 = pd.read_csv("C:\\Users\\ridhi\\OneDrive\\Documents\\inf 553\\small\\ratings.csv")
23   tags1 = pd.read_csv("C:\\Users\\ridhi\\OneDrive\\Documents\\inf 553\\small\\tags.csv")
24   rating = pd.DataFrame(ratings1)
25   tag = pd.DataFrame(tags1)
26
27   #remove whitespace
28   rating.columns = rating.columns.str.replace('\s+', '')
5(
52   #write to file
53   File_Write = open('C:\\Ridhi_Shikha_HW1\\Ridhi_Shikha\\OutputFiles\\Ridhi_Shikha_result_task2_small.csv', 'w')
54   #add header
55   header = ['tag','ratings']
56   csvEdit = csv.writer(File_Write)
57   csvEdit.writerow(header)
```

### For small data set:

- Change the path of ratings1 and tags1 to the input source of small data for ratings.csv and tags.csv
- Change the path of File_Write to the path where the output file has to be stored.

### For big data set:

- Change the path of ratings1 and tags1 to the input source of big data in ratings
- Change the path of File_Write to the path where the output file has to be stored.