The implementation of this assignment was done using Python 3.6
You would need to import csv, pandas, numpy and pprint before you execute the main file.

This folder contains the following items-

1**. Notes_Join&Agrr**
This is a pdf file that contains notes for joins and aggregate functions along with the SQL query for the same.

2. Dataset files

There are two kinds of dataset files-:

Table1.csv and Table2.csv contains 5 rows and 5 columns and is used for testing purpose while building this code.
ratings_small.csv and tags.csv are the two main csv files used for testing this code with. ratings.csv contains 1200 records while tags.csv also contains 1200 records approximately. These csv files are taken from movielens.com.

3. main.py

This is the main file that implements two kinds of joins along with one aggregate function. This code is written as user specific so the user needs to enter based on what he/she wants to do. Note-There user entries are strictly option based so entering wrong input might prompt the code to ask for input again or it might end the execution based on where you are at that point of time.

First part of the code implements joins, Natural and Left Inner join along with Inner join for reference. My code uses Inner join results to formulate left inner join.

Second part of the code implements the aggregate function count() which counts the frequency of each entity on a column and returns the value.

4. Instruction to run the program

A. To run this code using terminal, download the zip folder, unzip it, go the

folder where you have saved it and then run **python main.py** to execute the code.

B. You can also run this code using Pycharm. Just load the folder and run main.py

5. Outputs

The outputs are logged separately to a .csv file as well apartment from displaying it on the screen. The names of the csv file are self explanatory to know which file contains what.

NaturalJoin.csv
InnerJoin.csv
LeftInnerJoin.csv
aggCount_T1.csv
aggCount_T2.csv

The output folder contains two sub folders containing the above 5 files for Table1,Table2.csv or ratings, tags.csv