Ambrapali Electrotech Pvt Ltd.
Raunak Anand-IIT Roorkee

This document is regarding the procedure and detailing of assignment 1 files being submitted for the test.

The folder contains three files(CSV file not included due to large size):
1. IMDB.ipynb - an IPython file for the first 5 parts of the problem statement.
    a. It preprocesses the entire data.
    b. Extracts all the entities in reviews that have Entity Labels- "PERSON" or "GPE".
    c. Problem asks to extract user names which is highly unlikely to be available in reviews as no user would mention their own names in the revies generally. So, its all the names in the reviews that are extracted(mainly, Actors,characters,places).
    d. GPE - gives country, state and cities. So all are clubbed as cities.
    e. It generates a json file containing all the names and their entities in the first 5000 reviews using SpaCy NER(Named Entity Rectifier).

2. 'result.json' - the json file generated that stores the names and entities extracted from the model.

3. 'app.py' - this python file is a script that has to be run to access the simple webapp based on POST api.
    a. It takes the name as Input from the user.
    b. It returns if the name is a person or city(based on searching json file.)

P.S. - I had a weak problem understanding whether to make a new nlp based model to return the entity of the input name or use it from the extracted json itself. So, i went on with POSTing requests based on the json data due to the progress of problem as it was all linked to the previous problem in its generality.