CS 6350

ASSIGNMENT \_\_\_\_Project\_\_\_\_\_

Names of students in your group:

Qi Gao qxg150130

Yongxuan Tan ymt180001

Xiaoran Guo xxg180001

Number of free late days used: \_\_\_\_\_2\_\_\_\_\_   
Note: You are allowed a **total** of 4 free late days for the **entire semester**. You can use at most 2 for each assignment. After that, there will be a penalty of 10% for each late day.

This read-me file includes instructions to execute our program and sample input for final project.

Submitted files:

* build.sbt
* RunExample.scala
* SVD.scala
* projectx\_2.11-0.1.jar

Input files:

* Data can be downloaded from the below link by providing an email address:  
  <http://www.yelp.com/dataset_challenge>
* It should contain at minimal the following files:
  + business.json
  + review.json
  + user.json

Instruction:

1. To execute the SVD algorithm and generate output, submit Spark application with parameters:
   1. --class “SVD”
   2. File: projectx\_2.11-0.1-1.jar
   3. Parameters:
      1. <path-to-review.json>
      2. <path-to-output-directory>
      3. <sample-factor> - value between 0 and 1, used to set what fraction of data to run on; for all data, set to 1
2. To execute the GraphX algorithms and generate output, submit Spark application with parameters:
   1. --class “RunExample”
   2. File: projectx\_2.11-0.1-1.jar
   3. Parameters:
      1. <path-to-user.json>
      2. <path-to-output-directory>
      3. <sample-factor> - value between 0 and 1, used to set what fraction of data to run on; for all data, set to 1

For collaborative filtering:

1. Download dataset from Yelp, https://www.yelp.com/dataset

2. unzip the zip file, then you will get a bunch of datasets, which are json format file

3. open the project by inteillij

4. move the datasets to your project dictionary

5. run the program

6. you will be asked to input User ID, then input three user IDs

7. wait for 10 mins

8. you will get recommendations for those three users you just input