**CS 6350.001 Big Data Management and Analytics**

**Instructor: Prof. Latifur Khan**

**Homework 2**

Submitted By:

**Shyam Patharla (sxp178231)**

First, upload the given datasets (soc-LiveJournal1Adj.txt, userdata.txt, business.csv, review.csv) to the DataBricks file system

For each question, copy the code from the given scala files and to a Databricks notebook and run it.

**Output for Question 1:**

userA\_id userB\_id mutual friend count

24423 31131 6

168 20874 2

913 7831 0

2276 24032 0

2650 4494 0

12226 31249 14

…

…

**Output for Question 2:**

The output here is displayed in csv format. Actual output in Databricks was represented in a dataframe format

mutual friend count,userA First Name,userA Last Name,userA Address,userB First Name,userB Last Name,userB Address

100,Norma,Newton,2343 Zimmerman Lane,Jodi,Fuller,2073 Poling Farm Road

100,Teresa,Sylvester,1038 Peck Street,Thomas,Carini,178 Bartlett Avenue

100,Sabrina,Jones,1246 Don Jackson Lane,Terri,Zielke,3952 Whitman Court

100,Richard,Bishop,2559 Lincoln Street,Jennie,Price,2311 Meadow View Drive

100,Kenneth,Burger,1398 Park Avenue,John,Pritchard,2557 Diamond Cove

100,Becky,Austin,2405 Elkview Drive,Lesa,King,2543 Bombardier Way

100,Brenda,Edgar,4757 Henry Ford Avenue,Paul,Griffin,721 Heron Way

100,Margaret,Jones,3119 Wayside Lane,Jose,Shelton,3673 Elkview Drive

100,James,Riley,3403 Hickory Ridge Drive,Brian,Joachim,4057 Sugarfoot Lane

100,Jeff,Torres,1370 Riverside Drive,David,Segura,164 Reppert Coal Road

**Output for Question 3:**

The output here is displayed in csv format. Actual output in Databricks was represented in a dataframe format

user\_id,rating

nrifXqTX6a5mJlv2SzYiDA,5.0

xEzPJcx5Ny6NLm0p974dJA,5.0

HKjRVQiv95HXXeVaInOB5A,5.0

fV63MN6UMMTCtJstLBlyxg,5.0

IEggTMAmURzWvrVlP6DliA,5.0

ateRggfsSgc7A-rAoq4B7g,5.0

SZm0DtAR79nGTLFdgKH6aw,4.0

kZCZKDKdSHvamRT2rjbnRg,5.0

qYAbwD50JJM8HMw60isQTA,5.0

…

…

**Output for Question 4:**

The output here is displayed in csv format. Actual output in Databricks was represented in a dataframe format

business\_id,address,categories,avg rating

WXx3gZnrIysGFnpGMCskwg,"79 Vandenburgh AveTroy, NY 12180","List(Shopping, Drugstores)",5

1KWaczlXfDi9Z3urqlh\_Yw,"1080 Danby RdIthaca, NY 14850","List(Food, Bakeries, Coffee & Tea, Cafes, Restaurants)",5

JfZY98Jopy8OvYrXqKrF0Q,"408 E Upland RdIthaca, NY 14850","List(Optometrists, Health and Medical)",5

dkGTj5FVVqN4DbR7uWi8rw,"128 E State StIthaca, NY 14850","List(Shopping, Jewelry, Antiques)",5

nL6fL-kFpNSCrLdYM7qNZg,"3066 BroadwayMorningside HeightsManhattan, NY 10027","List(Barbers, Beauty and Spas)",5

\_ckbnVhwrD3vEowWVN8SnQ,"119 S Cayuga St Ste 301Ithaca, NY 14850","List(Active Life, Yoga, Fitness & Instruction)",5

JMQx87SPt0vt71MKniTz6Q,"179 Graham RdSte FIthaca, NY 14850","List(Massage, Beauty and Spas)",5

KNfrxyHr1oIGaRYAJK-W8Q,"414 Elmira Rd#6155Ithaca, NY 14850","List(Hardware Stores, Shopping, Home & Garden)",5

oBkgwRPRVa3Mnh9oP7CloQ,"3102 7th AveTroy, NY 12180","List(Automotive, Body Shops, Towing)",5

5BM0cmJvxaFlFuh0L-b4Ow,"217 N Greenbush RdTroy, NY 12180","List(Hardware Stores, Shopping, Home & Garden)",5