**MSCI: 6110 Fall 2019 Big Data Management and Analytics HW0**

**Submit on ICON Dropbox.**

**This is a background survey. Will Not Be Graded.**

1. Why do you take this course? What are your expectations?
2. Anything in the syllabus you are particularly interested in learning?
3. Anything you are interested in learning but NOT in the syllabus?
4. Do you have the prerequisites? Explain if not.
5. Check the database and SQL topics you know.

\_\_\_\_ CREATE/MODIFY/DROP Table \_\_\_\_\_\_ INSERT/DELETE/UDPATE data

\_\_\_\_ SELECT/FROM/WHERE \_\_\_\_\_\_ GROUP BY/HAVING/ORDER BY

\_\_\_\_ JOIN, OUTER JOIN, CROSS-JOIN \_\_\_\_\_\_ SELF-JOIN

\_\_\_\_ Sub-queries in WHERE/FROM \_\_\_\_\_\_ Physical models, Index

1. Describe your experiences with writing R code (e.g., project, work). What have you learned?
2. Do you have any experiences with the Linux operating system and its commands?
3. Never heard of it.
4. Have heard of but never used it.
5. Have some experiences using it.
6. Expert or professional users.
7. Are you familiar with Hadoop/Hive/Spark? Choose from below \_\_\_\_\_\_\_\_\_\_\_
8. Never heard of them.
9. Have heard of but never used them.
10. Have some experiences using them.
11. Expert or professional users.
12. Have you taken any data science or data mining courses (e.g., in this program)? Describe.
13. Check all the data science techniques that you have learned from previous courses (no worries if you don’t know some or any).

\_\_\_\_ Decision Tree \_\_\_\_\_\_ Artificial Neural Network \_\_\_\_\_\_\_ Naïve Bayes Classifier

\_\_\_\_ Random Forest \_\_\_\_\_\_ Linear Regression \_\_\_\_\_\_\_ Logistic Regression

\_\_\_\_ K-Means clustering \_\_\_\_\_\_ Association Rule Mining

\_\_\_\_ Others please name:

1. Have you learned how to access web service APIs and process XML or JSON data in other courses? Check the topics you already know at this point (no worries if you don’t):

\_\_\_\_\_\_\_\_\_\_\_\_\_ Generating word-cloud in R

\_\_\_\_\_\_\_\_\_\_\_\_\_ Using web service API to get twitter data (or other web data)

\_\_\_\_\_\_\_\_\_\_\_\_\_ JSON data format and R packages

\_\_\_\_\_\_\_\_\_\_\_\_\_ XML data format and R packages.

1. Other questions or concerns: