Course Syllabus for DS 700: Foundations of Data Science

NOTE: This syllabus document contains the basic information of this course. The most current syllabus is available in the full course.

Course Description

This course provides an introduction to data science and highlights its importance in business decision making. It provides an overview of commonly used data science tools along with spreadsheets, relational databases, statistics, and programming assignments to lay the foundation for data science applications.

Course Objectives

By the end of this course, you will be able to:

- Comprehend fundamental concepts in Data Science and Analytics.
- Understand fundamentals of SQL and apply them to Relational Databases.
- Use R Studio for programming, using packages, debugging and knitting .Rmd files.
- Effectively utilize packages from Tidyverse for R.
- Read/Write data of various formats using R.
- Create effective visualizations using ggformula.
- Perform linear and logistic regression, interpret and graph the model.
- Perform exploratory data analysis with dplyr package in R. / Perform data cleaning, management and manipulation with R.
- Use conditional statements, loops and control flow logic in R.
- Interpret p-values

Course Components

Homework 1-4: Include a quiz of 20–30 questions covering that lesson's course content, readings, and videos. You are allowed only 1 attempt, so please do not attempt the quiz until you have thoroughly worked through the lesson's material. The quizzes are graded automatically, so you should receive your score immediately.

Mini Project: The SQL mini-project assignment is designed to test your skills acquired in Lessons 2 to 4. You will be required to create and submit SQL queries for various scenarios using the sample database. SQL mini project will be out of 100 pts. Students will be submitting a .sql file and are allowed to submit it multiple times before the submission deadline. I will be grading their latest submission. The exact instructions for the submission will be included in the word document for the project.

Homework 5-11: First, solve the problems in the provided RMD file. When you are done, submit your answers to selected questions in the Homework Quiz for auto-grading. You are allowed 3 attempts. After the assignment is closed, everyone who submitted the assignment will gain access to a partial key, which you can use to check

your work on the open-ended questions. The questions which do not appear on the Homework Quiz will not be graded, but you are expected to understand them to help prepare you for the final project and other courses in the Data Science program.

Quiz on control flow: This quiz will assess your understanding of for loops, while loops, and if/else loops. You will have 60 minutes to complete 4 questions. You are allowed 2 attempts. (Your second attempt will involve a new random selection of questions.) The higher of your two scores will be counted. The quizzes are graded automatically, so you will receive your score immediately. The problems in this quiz follow a similar style to the last 3 problems on the Practice on Control Flow, so you are encouraged to complete that activity before taking this quiz.

Discussions: Discussions are your opportunity to contribute to and learn from your peers. We will engage in several discussions throughout the semester, as outlined in the Course Calendar. Note that every assessed discussion requires two posts: one for your original contribution to the discussion's central issues and at least two responses to your peers. Your initial post counts as 75% of each discussion grade, while your replies/peer responses count for the remaining 25%. In order to create a professional, open communication environment, you are expected to follow these online discussion guidelines.

Final Project: The Final project for DS700 simulates a real world scenario. Students will be utilizing the SQL skills they have acquired in the course to generate the dataset required for the analysis portion. Once the dataset is obtained, they will be using their R programming skills honed during the course to perform analytical operations and derive inferences.

Grading

Your mastery of course content is assessed using a variety of methods:

Homework 1-11 (5% each)	55%
Self-introductions	2%
Discussion 1 and 2 (5% each)	10%
SQL mini project	10%
Quiz on control flow	3%
Final project	20%
Total	**100%**

Final grades are assigned using the following scale:

90–100%	A
80–89%	B
60–79%	C
0–59%	F