Big Data

Big Data is another topic in Data Science where it is a course by itself. Big Data is applied to disclose hidden patterns, market inclinations and consumer preferences, for the benefit of organizational decision making.

Chapter 17 of the textbook used in our class includes and demonstrates Big Data in various themes such as Database, Hadoop, Spark, and Internet of Things. There are multiple exercises available in the textbook that we can delve into to practice further. In this assignment, we will choose one of the given themes mentioned and recruit our best exertion to understand Big Data in the hope that we can comprehend the subject matter and be able to explore more on our own at a later time.

Let's start with watching the following two short videos on Big Data:

Big Data Explained (from YouTube 11:40 min)
https://www.youtube.com/watch?v=ps6HweFb5LI

Big Data's Volume, Velocity, and Variety (3 Vs) (from YouTube 3:29 min)
https://www.youtube.com/watch?v=wVAWAeOIIII

Your Assignment (there are three parts in this assignment)

Please read chapter 17 of the class textbook and practice some of the example programs given in the chapter on your own to get a better understanding of the subject matter.

I. Write a paper explaining what is Big Data? Why is Big Data important? What are the types of Big Data? An industry analyst Doug Laney articulated the now-mainstream definition of Big Data as the three V's. Explain what are the three V's.

II. Do the given exercise on section 17.2 of chapter 17 of your textbook. Before you start the exercise, create a directory called ch17 on your computer. Copy the two files; *books.db* and *books.sql* given on the blackboard into directory ch17 that you have just created. Follow the instruction on section 17.2. Store your program and the output into your GitHub account. Since you have to submit the program to your GitHub account, you need to create a file instead of using iPython.

III. Complete and submit exercise #17.1 (Books Database) on page 799. Store your program and the output into your GitHub account. The exercise stated that you use iPython to write the program. Since you have to submit the program to your GitHub account, you need to create a file instead of using iPython.

Write a Learning Report Summary (LRS)

Using Microsoft Word, write a summary report (not a bullet items) with a minimum of 100 words explaining how you completed your assignment. *Please describe your responses, not just yes/no answers.*

1. Did you successfully get your assignment done? Did it run? Any error? Did you get the correct result? Did you test your program thoroughly?
2. How much time did you spend to complete your assignment?
3. Did you find the assignment easy or challenging for you?
4. Did you write the program yourself? Did you get any help from anyone?
5. When you encountered obstacles to complete your program, how did you resolve the issues? Did you use Google to get help? Describe how Google was abled or not able to assist you?
6. What did you learn from doing this assignment?
7. Any other information you would like to share with your instructor? Make sure you provide program output on each option.

What to submit on blackboard

Your written document on Big Data.
Don't forget to include your GitHub URL in your submitted document.
Your Learning Report Summary.
DO NOT submit your programs and the program output results on blackboard. They should be stored on your GitHub account.