**ISE 535 Summer 2022 Project**

**Due Aug 1st, 2022**

NC State University is conducting a funding record analysis of projects that have been awarded by the National Institutes of Health (NIH) and its various sub-institutes. Organization personnel have downloaded records post 2006 for 6 of the Institutes (NCATS, NIBIB, NHLBI, NIAMS, NIDCR and NIOSH). Results have been downloaded from the NIH REPORTER award record site in a .csv format. Your job is to analyze across the available dataset and report back to the Vice President of Research on the following questions:

1. Find the total amount of funding awarded by NIH across these 6 institutes based on ‘Total Cost IC’. Plot your results in a pie chart. **(10 Points)**
2. Prepare a list of the top 10 funded Principal Investigators based on awarded Total Amount (Column: ‘Total Cost IC’). **(10 Points)**
3. Prepare the top 20 popular ‘PROJECT TERMS’ that frequently appear across all funded projects. **BONUS:** Prepare a ‘Word Cloud’ visualization with key terms listed, perhaps broken down by each institute for more visual clarity. **(10 Points, BONUS: 5)**
4. For awards that have been made after Jan 1st 2010 (based on Project Start Date Column), how many awards and the total funding amount awarded (Total Cost IC) have been made to Research Triangle Area Universities (UNC-Chapel Hill, NC State University, Duke University, NC A&T). You may report results graphically to help communicate your results. **(20 Points)**
5. How many Principal Investigators have been awarded by two or more Institutes? Rank them based on number of awards received by multiple institutes. (NOTE: Duplicate names may exist across organizations. Since PI’s may not have unique IDs, identification can be done by a combination of their PI Name and Organization he/she belongs to.) **(20 Points)**
6. List the Engineering departments and their associated organizations that have been awarded by the NIH. (based on DEPARTMENT column). Group the results based on Organizations and show the total amount awarded for each organization). **(20 Points)**
7. Report on the top 20 funded Universities (Total Cost IC) by funds received from NIBIB and NIAMS since 2010. **(10 Points)**

**SUBMISSION**

1. Code file used. Make sure that code files are appropriately commented to enable me to easily identify your code block for each question.
2. Results displayed on a Powerpoint slide deck for each question. Any other sub-analysis performed beyond what was asked can also be shown.
3. The video that accompany your submission can be used to quickly go through certain portions of your results, particularly highlighting interesting insights and/or visualizations that you may have used.

**CONSIDERATIONS:**

1. You can use any Python library to analyze your results.
2. Either Spyder or Jupyter Notebooks can be used to demonstrate your results.
3. All analysis and visualizations performed must be using Python.
4. Reports may contain duplicated records.
5. Some rows contain missing values.
6. Data types may not always be in the form that you want them to be.

**LIMITATIONS:**You are not allowed to modify the CSV file in Excel sheets using Excel Functions. You may have however use its tools to check your answers.