Course Name: Quality in Information Systems   Course Code: IS411
Bioinformatics Program   Information Systems Department
Faculty of Computers and Information   Assiut University
Projects   Grade: 20 points
December 19, 2021   4th level

1. Describe your dataset in details (number of attributes, number of records)
2. Clean your data with OpenRefine (i.e. remove all missing values or incomplete data)
3. Apply t-test on your data
4. Visualize your data, explaining your graphs and why you used these graphs type.
5. Create your own ontology or use existing one using Protégé to annotate your data using Right Field
6. Prepare a Power point presentation, containing the following:
   - Team members
   - Link to your dataset
   - Results of visualization
   - Determine various types of queries that you need to answer using SPARQL
   - Determine what did you need exactly in the semantic data integration phase
   - Results of Right Field

Possible datasets you can use:

1. Datasets of your own choice that is related to scientific data (can be found at Kaggle, or UCI machine learning repository)

Note:

   - Your team 4-6 students NOT NECESSARILY FROM THE SAME SECTION

Deadline: January 3, 2022

Good luck

Prof. Taysir Hassan A. Soliman