

# Big Data Analytics

## Assignment # 1

Due Data: 10 March 2022

Assignments are to be done individually. No late assignments will be accepted.

### HONOR POLICY

This assignment is a learning opportunity that will be evaluated based on your ability to think, work through a problem in a logical manner. You may however discuss verbally or via email the assignment with your classmates or the course instructor, and use the Internet to do your research, but the written work should be your own. Plagiarised reports or code will get a zero. If in doubt, ask the course instructor.

## Task

- Load the Data Source for the assignment into MongoDB LocalHost using Python.
  1. The dblp computer science bibliography provides open bibliographic information on major computer science journals and proceedings. Originally created at the University of Trier in 1993, dblp is now operated and further developed by Schloss Dagstuhl.
  2. <https://dblp.org/xml/>  
You have to download the above mentioned xml or any-other necessary file. Load this XML in an efficient manner such that it doesn't take too much of system memory and cpu.
  3. Your MongoDB Database name should be your Roll-no i20\_XXXX and Collection Name should be DBLP\_DATA.
- After Loading the Data then perform EDA on the above data using MongoDB queries to fetch the data from the database.
  1. Make Bar-chart for Publications per year and Explain your Findings.
  2. Make Pie showing Number of Bibliography based on Element Types i.e  
PHD THESIS  
INPROCESSDINGS  
WWW, ARTICLE  
etc.
  3. Make StackBarPlot for Number of Publication per year and Element Types.
- You have to make 2 Questions similar to above 3 of your own from the data and Answer them using plots and visualizations .

## Submission

You are required to submit a report with response to each question with supporting graphs, tables and stats. Py File for XML to MongoDB Storage. IPYNB File Used for EDA.

**NOTE : No need to submit XML or any other dtd Files. If submitted will result in marks reduction. For marking Assignment I will be using default XML File i.e [dblp.xml](#)**

GOD BLESS YOU 🙏