Project Details

Students will work in a small group (max 3) or individually on a Data Analytics project. The students will use some open data set from sites such as www.kaggle.com or www.open.canada.ca.

The project should be implemented using Python and one or more of required data science libraries of Python (see resource folder on your blackboard). The project could look into one of the problem areas mentioned below (or you can come up with your own)

Problem area:

- Text analysis (understanding the meaning of text but **not** simple sentiment analysis such as twitter sentiment analysis)
- Analysis of structured / semi structured data (a lot of data set from kaggle will fit into this, but remember this has to be more challenging than 3360 project)
- Analysis of multimedia data (voice, images, video). Go for something beyond simple image recognition.
- AI/ML from a variety of data sources (think of combining different data to make a better system)

Steps to proceed:

1. Identify a problem area you wish to work in
2. Identify data set for that domain
3. Identify what analysis can you do
4. Specify data analytics model being used
5. Identify tools and libraries you need
6. Identify potential roadblocks.

Based on this, bring a single page initial draft to the class after reading break.

Initial Draft should include:

- A paragraph of the broad goals for your project,
- Data set you have chosen (with URL to accessing the dataset)
- Tools (libraries) you are planning to use
- Possible DA goals or insights you hope to draw (some examples include what you plan to predict, what you plan to classify, what relationship/s you are trying to model etc.)
- Keep it short, bullet points are fine your idea should be clear
- No points for this project deliverable; this is just a means for you to crystallize your idea