

# **MIE 1624 Introduction to Data Science and Analytics – Winter 2019**

## **Course Project**

**Deadline: Thursday, March 28, 11:59am**

### **Background**

Data science, analytics, AI, big data are becoming widely used in many fields, that leads to the ever-increasing demand of data analysts, data scientists and other data professionals. Due to that, data science education is now a hot topic for educators and entrepreneurs.

In this project, your team will first re-design the course curriculum for “MIE1624: Introduction to Data Science and Analytics” course at University of Toronto, such that students acquire an introduction to the most relevant topics and skills in data science. Second, your team will design a curriculum for a new professional “master of data science and analytics” program with technical emphasis, see, e.g., <http://www.ryerson.ca/graduate/datascience/> or <http://csds.uvu.edu.ua>, that contains optimal courses (and internships, projects, extra-curricular activities, etc.) for students to acquire required knowledge and skills to pursue a successful career as a data analyst/scientist. Third, your team will need to design a curriculum for a new professional “master of business and management of analytics and AI” program with both business, management, soft skills focus as well technical focus, see, e.g., [http://smith.queensu.ca/grad\\_studies/mmai/](http://smith.queensu.ca/grad_studies/mmai/), <http://mbai.kse.org.ua> or <http://www.rotman.utoronto.ca/Degrees/MastersPrograms/MMA>, that contains optimal courses (and internships, projects, extra-curricular activities, etc.) for students to obtain necessary skills to pursue a successful career as an analytics and data manager, business analyst, AI system designer, etc.

### **Learning Objectives**

- Develop the ability to work in a team on a consulting project. (You are required to work on the project in the same group as for your in-class presentation. Check the Quercus portal for the list of your group mates.)
- Improve on skills and competencies required for performing a full cycle of data science and analytics workflow, i.e., data collection and pre-processing, applying algorithms to analyze data, trend identification, storytelling based on analytics (writing a consulting report and delivering an oral presentation).

### **Tools Allowed**

- You can use the Python libraries mentioned in-class as well as any other libraries you find during your research. Note that you can only use Python 3.
- For visualizing the results in your report and presentation you may use Python or any other outside tools, e.g., IBM Watson Analytics, Tableau, etc.

## TO DO:

Finish the following four parts **based on your data analytics:**

### Part 1 - Course curriculum design:

Your team has been tasked with redesigning the course curriculum for MIE 1624: Intro to Data Science and Analytics. (**i.e., choose topics that should be covered in this course**). The topics selection can be based on the skills required for data analyst/scientist job postings from websites like <http://www.indeed.ca> (may need to do web-scraping), and Kaggle ML and Data Science Survey <http://www.kaggle.com/kaggle/kaggle-survey-2018> on curriculums of courses offered on online platforms, such as <http://CognitiveClass.ai>, Coursera, EdX, DataCamp, and/or any other data that your team can find or collect yourself.

### Part 2 – Technical Data Science program curriculum design:

Your task is to design a sequence of courses and curriculum for each of those that can be used by the Faculty of Applied Science and Engineering at UofT. They hope that your team will help them to **design a curriculum for their new professional technically-oriented program: Master of Data Science and Analytic (M.D.S.A.)**. Propose how many required and optional courses should be offered and design a curriculum for each course based on datasets from Part 1 and any other data that your team can find or collect. Propose what other activities should be part of the program, e.g., internships, projects, extra-curricular activities, etc.

### Part 3 – Managerial and Business Data Science program curriculum design:

Your task is to design a sequence of courses and curriculum for each of those that can be used by any university. The hope is that your team will help to **design a curriculum for a new professional business/technically-oriented program: Master of Business and Management in Analytics and AI (M.B.A.I.)**. Propose how many required and optional courses should be offered and design a curriculum for each course based on datasets from Parts 1-2 and any other data that your team can find or collect. Propose what other activities should be part of the program, e.g., internships, projects, extra-curricular activities, etc.

### Part 4 - Data Science education EdTech effort:

Now your team has become an experienced team in Data Science education and you are concerned about the development of Data Science education, namely your own EdTech startup. **Propose and answer your own question related to the design of educational efforts (something that helps you and others to find jobs) in data science**. It can be an online/MOOC specialization, a bootcamp or a series of bootcamps, meetups, which companies can be approached in Canada for internships in data science and why, etc.

**Note:** the scope of the question is quite wide, and it is advised that you narrow it down based on your interests and expertise. Make the work truly yours.

## Project Presentations

- Project presentations are open to the public and are scheduled for **Monday, April 1, 6:00-9:00pm, MP-203.**
- **Do not make your presentation overly technical.** Your audience is business-oriented and may know little about data science, people are interested in the insights that you got from your analysis and why your results can and should be used for decision-making.

## What to Submit via Quercus:

1. Your Jupyter notebook with appropriate documentation for every step as well as the relevant data files. Comment out the data retrieval procedure (from Kaggle, Coursera, Twitter API, New York Times API, etc.) in your code and replace it with code for reading the corresponding data from files. (**Submit all those data files together with your Jupyter notebook**). Make sure that your IPython notebook runs on Data Scientist Workbench portal with the Python 3 kernel and that all needed data files are included in your submission. If the size of the data files exceeds Quercus's capacity, those should be stored on a cloud drive (e.g., Dropbox, Google Drive), and the link to the directory should be included in the notebook.
2. A 4 to 5-page consulting report in PDF format that summarizes your findings and results (all graphs should have axes appropriately labelled, all visual materials should be understandable and the graphics of sufficient quality to be easily readable.) This report should be business oriented, and cover your problem more extensively than your presentation.
3. Your business-oriented presentation slides in PowerPoint and PDF formats. (Each group will present their findings and results during a 7 minute presentation with 1 minute for questions. Presentations will be timed and stopped after 7 minutes.)

## Marking

- The project is worth 30 points (15 points for your analysis and report and 15 points for your business-oriented presentation).
- The presentation will be graded as follows (15 total marks):
  - 5 marks for organization and delivery (e.g. clarity, enthusiasm, poise)
  - 5 marks for content (e.g. proper visuals, high-level ideas, answering questions)
  - 5 marks for the business pitch (e.g. recommendations, solution to the problem)
- The analysis and the report will be graded as follows (15 total marks):
  - 4 marks for identifying the problem and searching the relevant data
  - 7 marks for the analysis (e.g. cleaning the data, visualizations, applying algorithms)
  - 4 marks for discussion and insight (e.g. how your analysis contributes to the problem, making a decision, storytelling)
- Every group member gets the same mark for the project. It is your responsibility to determine how you split the work inside your group. At least half of your group needs to be present during the project presentations to present the results.

## Notes

- For the deliverables, consider the Jupyter notebook as what you would report to senior data scientists and machine learning engineers, and the consulting report and the presentation as what you would report to CEOs, VPs, PR managers, university officials, government officials, and journalists.
- The presentation would be a visual representation of the executive summary of your report.
- The audience for your presentation and report in particular is business-oriented and includes people who are interested in the insights you gathered from your analysis and how your results should be used for decision-making.