**Team Project TP**

[Project Description - Data Acquisition and Analytics](https://mycourses.cityu.edu/content/enforced/12884-12241131/TP_Data%20Acquisition%20and%20Analytics.pdf?_&d2lSessionVal=W04O20MVDpvf0tbBLT21SmXIQ&ou=12884)

Each student can select his or her team. Each team consists of three students.  A group of fewer than three students requires the instructor's approval. Each team will use an instructor approved topic relevant to the course.

Students are expected to use the assigned readings, videos, and other materials throughout the quarter. Students will need to utilize additional sources that were not assigned by the professor. While stylized after an industry report, nonetheless, students are expected to employ APA formatting of citations, footnotes, and bibliography. Students must cite the sources of all ideas, facts, and information used that are not their own, even if they have put the information into their own words. Failure to do so is plagiarism, although the oversight is unintentional.

**Four submissions are required according to the following schedule:**

* **Proposal** (1 page; 30 points) - Starting (Module 1) & Ending (Module 3)
* **Progress** Report (3-4 pages; 70 points; graded after the proposal has been submitted) - Starting (Module 4) & Ending (Module 7)
* **Final Report** (6-7 pages; 70 points; graded after the progress has been submitted) - Starting (Module 8) & Ending (Module 10)
* **Final PPT** (15+4 slides, 30 points; graded after the final report has been submitted) - Starting (Module 8) & Ending (Module 10)
* 1  
  DS522 Team Project (TP)  
  Project Description - Data Acquisition and Analytics  
  In this project, each team will select a topic of interest in a domain area (e.g., business, healthcare,  
  public sector, etc.), obtain interesting publicly available datasets, and apply core data acquisition  
  and pre-processing concepts. Each group will perform the initial phases of the project life cycle of  
  data science, including data acquisition, pre-processing, wrangling, exploration, and visualization.  
  Data analyses involving sophisticated machine learning models are not expected in this project.  
  More than one data source is required for this project, and Python and Power BI should be used  
  for each step of data acquisition and analytics.  
  Project Requirements  
  1) Each team member needs to contribute professionally and fairly to the project.  
  2) Each team needs to have a name that is short and fun.  
  3) A team captain must be selected. As the team leader, collaborator, and coordinator, the  
  team captain is the point of contact with the instructor and is responsible for assignment  
  submissions.  
  4) Each team must select a project topic and collect relevant data.  
  5) The dataset should be either precompiled from sites like Kaggle or crawled through its  
  own script. Some interesting datasets (in .csv, .xlsx, or .txt format) include, but are not  
  limited to:  
  a. FiveThirtyEight (https://github.com/fivethirtyeight/data)  
  b. UCI Machine Learning Repository (https://archive.ics.uci.edu/ml/index.php)  
  c. Data.gov (https://catalog.data.gov/dataset)  
  More than one data source is required for this project, and the dataset should be real and  
  large enough to practice data cleaning, wrangling, etc.  
  6) Each team must demonstrate applying at least two techniques from module 5 to 10 core  
  concepts to the dataset.  
  7) Each team must analyze the dataset using a simple statistical methodology. The use of  
  more complicated methodology is allowed in this project.  
  8) Data visualizations must be demonstrated for important findings.  
  9) A minimum of two papers or book chapters (outside textbook) is required per student.  
  10) All of these steps must be demonstrated by using Python and PowerBI.  
  11) Software code (in Python and PowerBI) should be submitted with documentation on  
  GitHub.  
  12) Workload assignment among members must be described.  
  Proposal Contents  
  1) Project Title  
  2) Team Information - Team member names, email addresses, and the team captain's name
* 2  
  3) Project Summary – A short description of the project  
  4) Project Description - Describe what you are trying to achieve in this project. What  
  problem do you want to solve? What kind of information do you intend to extract?  
  5) Usefulness - Describe how your project can be helpful. Who will benefit from your  
  project findings? Has this topic been studied before? How is your project differentiated  
  from the previous ones?  
  6) Dataset - You need a dataset either precompiled from sites like Kaggle or crawling  
  through your script. Please describe what your dataset is and how you intend to collect it.  
  7) Workload assignment - Describe the contribution of each member.  
  Progress Report Contents  
  It serves as a checkpoint before your final report, and the progress report should have the same  
  layout as the final report. There could be some sections in progress, but an introduction, literature  
  review, and data should be close to the final version.  
  1) Introduction  
  2) Literature Review  
  3) Data  
  4) Workload assignment - Describe what each member has contributed to the progress  
  report.  
  Final Report Contents  
  1) Abstract  
  2) Introduction – Describe the problem your team wants to solve. Why is it important?  
  Provide some context.  
  3) Literature Review – Describe related problems and where your problem is situated. What  
  is significant about your problem compared to existing ones? Two papers or book  
  chapters are required per student.  
  4) Data – Describe the data. What is the source of the dataset? What are the characteristics  
  of data? More than one data source is required.  
  5) Methodology - Describe your approaches in data pre-processing and data wrangling.  
  6) Result – Describe the finding of patterns or trends using a statistical methodology.  
  Demonstrate the visualization of data.  
  7) Conclusion – Describe what your paper has achieved, lessons learned, and the limitations  
  of the paper. Please also include how your team can improve the paper if given another  
  opportunity.  
  8) Reference  
  9) Workload assignment - Describe what each member contributed to the final report and  
  presentation slides.  
  10) GitHub address – Show the codes of both Python and PowerBI.