

Project 0:

Deadline: November 16, 2025, 11:59pm Eastern Time.

Description:

An organization has approached you for collaboration seeking to have you solve a problem that they have. The solution is likely to include data analytics approaches so that you can help them. As a data analytics expert, you need to follow a process to gather necessary information from the organization and other various sources (e.g., literature), and document this information as the project progresses. Initially, there is essential information that you need to collect. This project is about collecting such required information and planning for your probable solution. You are expected to write up to three pages for an initial report that includes your plan and the required information.

1. Define and explain the problem. Why is this problem important to solve and what is the significance of this problem to the organization/society/people? What are the goals that the organization wants to achieve with your solution?
2. Explain your solution and why your proposed solution should solve the problem. What kind of value your solution will bring to the organization, society, or lives of the people?
3. How does the organization currently work to solve the problem (if any)? How will your solution fit into their current process and help them improve the outcomes?
4. How will you acquire data? Will the organization provide, or will you find public data or will you need both? What will be the likely size of the data?
5. What data analytics components will your solution include? E.g., data acquisition/collection, data pre-processing, exploratory analysis, etc.
6. How will you process your data, in real-time or in batches? Give an overview of the data, e.g., data types, variables, labels, etc. What kind of preprocessing techniques will you apply and why?
7. What predictive analytics solutions will you propose to address the problem and why this approach is a good solution?
8. What will be your evaluation strategy? What metrics will you use to evaluate your predictive model and why? How will you evaluate the improvement that your solution provides to the outcomes of the organization? E.g., how KPI will change after your solution is integrated into a company's ecosystem?
9. How will the organization be expected to utilize your solution in their workflow? What is the capacity requirement for the company to utilize your solution?

Guidelines:

This is project 0 that will help each team solidify necessary information on their project and get a better understanding of the problem, the proposed solution, evaluation, and its likely integration with the real world mechanisms.

Project 0 will be graded according to the following guidelines:

- A score between 0 and 100 will be assigned.
- If it is not submitted before the specified deadline, then a grade of 0 will be assigned.
- Each question is worth 10 points.
- The team leader will submit a separate report that briefly describes the contribution of each team member to the joint effort. Also, the leader will provide a score between 0 and 5, where 0 indicates no effort and 5 indicates equal contribution.
- At the beginning of the project 0 report, please include the names of the team members specifying the team leader and GSU email addresses. Also, include the following text:
 - “Statement of Academic Honesty:
The following code represents our own work. We have neither received nor given inappropriate assistance. We have not copied or modified code from any source other than the course webpage or the course textbook. We recognize that any unauthorized assistance or plagiarism will be handled in accordance with Georgia State University's Academic Honesty Policy and the policies of this course. We recognize that our work is based on an assignment created by the Institute for Insight at Georgia State University. Any publishing or posting of source code for this project is strictly prohibited unless you have written consent from the Institute for Insight at Georgia State University.”

Important Note:

As you will need data for your project, you can explore public data repositories for your problem:

<https://archive.ics.uci.edu/ml/index.php>

<https://data.europa.eu/euodp/data/dataset>

<https://www.data.gov/>

<https://catalogue.data.govt.nz/dataset>

<https://data.gov.in/>

<https://www.opendatani.gov.uk/>

<https://msropendata.com/>

<https://registry.opendata.aws/>

<https://datasetsearch.research.google.com/>

<https://www.kdnuggets.com/datasets/index.html>

<https://dataverse.org/>

<https://zenodo.org/>

<https://figshare.com/>

<https://guides.library.cmu.edu/machine-learning/datasets>
<https://github.com/awesomedata/awesome-public-datasets>
<https://blog.bigml.com/list-of-public-data-sources-fit-for-machine-learning/>
<https://openml.github.io/OpenML/Data-Repositories/>
<https://www.kaggle.com/datasets>
<https://www.analyticsvidhya.com/blog/2018/03/comprehensive-collection-deep-learning-datasets/>
<https://pub.towardsai.net/best-datasets-for-machine-learning-data-science-computer-vision-nlp-ai-c9541058cf4f>
<https://index.quantumstat.com/>
<https://rubikscore.net/2021/07/19/top-23-best-public-datasets-for-practicing-machine-learning/>
<https://www.kdnuggets.com/2020/03/20-machine-learning-datasets-project-ideas.html>