

AIT580 Data Analytics Project Guide

Objectives

- Select a suitable dataset for analysis
- Apply tools and methods of data analytics presented in this course
- Develop skills in applying different data analytics techniques to particular tasks
- Develop a preliminary data analysis system based on the requirements
- Demonstrate analysis using R, Python, SQL, NoSQL/MongoDB, Tableau, AWS, or other tools to produce statistical summaries and visualizations that support conclusions about the meaning and value of the dataset.
- Learn how to evaluate the data analysis system
- Learn how to write scientific/technical documentations
- Learn some project management techniques

General Instructions

The Big Data Analytics Project is a course-long effort consisting of Four (4) separate deliverables. You are strongly encouraged to work ahead on these deliverables. You may submit them early, if you would like, but they will not be assessed until after their respective due dates.

This project consists of **Four (4)** deliverables:

- Deliverable 1 – **Project Proposal** (3-5 pages)
- Deliverable 2 – **Project Presentation** (10-15 slides and 10-15 min presentation)
- Deliverable 3 – **Final project Report** (15+ pages)
- Deliverable 4 – **A Working System** (one zipped file)

Due dates: check the latest class schedule on blackboard.

Specific Requirements for Deliverables

Read the sections below for the specific requirements of each deliverable. **Use the following structures and the [Template of Data Analysis Project Report](#) as templates for your reports and submissions.**

1. **Proposal** (*submitted in **one Word/PDF file**, 3-5 pages*)
 - **Cover**
 - The Project Title
 - Team Number, Student(s)'s Name(s)
 - Course Professor's Name
 - University Name, Course Number and Name
 - Date
 - **Introduction**

- Briefly explain the project and its scope
- Briefly give some backgrounds and/or related work
- **Objectives**
 - Briefly describe/list the objectives of the overall project you proposed
- **Proposed selected dataset**
 - The # of instances/records **>= 3000** and the # of the features/attributes **>= 8**
 - Briefly describe the overall selected dataset
 - Briefly describe the features/columns you would like to select for your data analysis
- **Description of proposed system**
 - Briefly describe the conceptual system architecture you proposed and draw a diagram for it.
 - Briefly describe the data analytics methods you will use or propose (*each method should consider the features you will select*):
 - data visualization
 - data analysis
 - Descriptive analytics
 - Inferential analytics
 - Advanced analytics (ML, etc.) (*optional*)
- **Proposed development platforms**
 - Briefly describe the software and hardware development platforms
- **References**
 - Provide appropriate citations and references.
 - Be sure to include a citation and link(s) for the dataset(s)
 - See <http://infoguides.gmu.edu/citingdata>
- **Appendix**
 - show a piece of data
- *Other contents could be added if necessary*

2. Presentation (*submitted in one PPT file, 10-15 slides*)

- Power point slides (*submitted in PPT*)
- In-Class presentation (**10-15 minutes**)

3. Final Technical Report (*submitted in one Word/PDF file, 15+ pages*)

- **Cover**
- **Abstract** (*200-300 words*)
- **Introduction**
- **Objectives**
- **The Dataset(s)**
 - Selection
 - Description
 - Schema
 - Pre-processing
- **The System**

- Architecture
- Data Processing
- Data Analytics Algorithm(s)
- SW/HW Development platforms
- **Experimental results and analysis**
 - **Explore and present analysis of the dataset** using relevant tools discussed in the course (, Python, R, SQL, etc.)
 - **Prepare relevant analysis and visualizations** for selected data items
 - Statistical analysis (correlation, hypothesis testing, etc.
 - Machine learning
 - **Interpret the results**
 - What conclusions can be supported for each analysis?
 - This should reflect answers to the specific questions specified in the “Need” section.
 - Describe the value obtained from the study
- **Conclusions**
 - Draw conclusions for the overall project
 - Lessons Learned
- **References and/or Explain/Define Terms**
- **Appendix**
- *Other contents could be added if necessary.*

4. A Working System *(submitted in one zipped file)*

- The system will be checked if working or NOT via re-compiling and re-running on instructor’s computer for grading.
- Write a **README** txt file to concisely describe how to set up and run your system, dataset link(s), and any other info you need to tell others to re-compile and re-run your system.
- All the source code, datasets, README, and related files should be submitted in **one zipped file: AIT580_Team#_LastNames_sys.zip**.
- If the file is too big to submit, please contact instructor or GTA for help. If the original dataset is too big, please provide the link(s) for downloading.

Submission

Please submit all your files with the file name “**AIT580_Team#_LastNames_*.***” (e.g., ***.*** could be **proposal.docx/pdf, final.pptx, finalReport.docx/pdf, sys.zip**). Go to the Blackboard for submission.

Grading

The whole project will be graded based upon scope, complexity, quality and deliverable.