

IT370 Introduction to Analytics Fall 2022 Final Project Guidelines

Overview

In lieu of a final exam, students need to complete and submit a final group project, with **up to three students in one team**. However, it is also allowed if a student chooses to complete it as an individual project.

Overall, there are two options for students to complete this project. With either option, students need to complete a data-driven project by using Tableau to analyze a large data set and then write a project report. With **Option #1**, students can bring their own data. With **Option #2**, students can use a dataset provided by the instructor. Please refer to the project instruction below for details.

The instructor recognizes that some types or topics of projects take longer time to complete or are harder to complete, so the challenges imposed by your chosen project and your specific research questions will be taken into consideration when the project is graded. If you choose a particularly difficult project and research questions, it is possible that you get rewarded with extra points. (You can also somehow highlight the challenges you were facing when writing your final project report or its appendix as long as such highlights will not interfere with the primary logic flow of your report.)

Grading

The project is worth 30 points (out of 100 total points for the course). If it is completed as a group project, all the group members share the same grade unless there are complaints about free-riding (which will be investigated case by case). The same grading criteria will be applied no matter whether it is a group or individual project, or whether it is a two- or three-person project team.

Final Deliverables

Your final project deliverable need to include **two** files:

- Project report, in Word or PDF format
- The original Tableau project file, in ***.twbx** format

General Format Guidelines for the Project Report

Despite the option you choose, you need to submit a complete project report. The report needs to be logically coherent and concentrated on one theme. So, spend some time organizing your project report in logical ways.

Please follow the format requirements specified below

- A **maximum of 15 double-spaced pages** (**NO** minimum limit, as long as your project report is fully developed), inclusive of all supporting material (e.g., tables, figures, appendices, etc.)
- 12-point fonts
- 1" margins on all sides

- Professionally formatted (e.g., cover page, page numbers, section titles, etc.)
- Free of grammar, spelling, punctuation, and other writing problems.

Submission Guidelines

Due Time: The final project is due on by **11:59pm December 23 (Friday)**. Submit your work through Blackboard.

If working as a team, only one team member needs to make the submission, but make sure all the team members' names are included in the report's cover page.

Due to the time of pressure of grading and reporting your final grade to the university, **NO late submission will be accepted.**

Project Instructions

Despite which option students choose, the project submission requires both a written project report and the original Tableau project file. The Tableau project file needs to be submitted in **a *.twbx format**. While not absolutely required, it is also highly recommended that your Tableau solution includes an **interactive dashboard**.

Option #1 Bring your own project

You can propose your own project, with a substantive part of it related to data analyses. If you have access to some real-life data, and you have a general interest in understanding them better, it would be highly encouraged that you complete your course project based on them. Although optional, you are also encouraged to discuss your potential project topic with the instructor before you start the substantive work. The instructor will work with you to define the scope of the project and provide some general guidance on how to proceed with it.

There are a lot of example Tableau projects available online. You can of course learn from them. But please do **NOT** complete your project by simply modifying an existing Tableau project you could find from the Internet. Do **NOT** even just reverse-engineer (replicate) an existing Tableau project. By reverse engineering, you may complete the project step by step on your own, but you will miss the opportunity of practicing designing a project from scratch.

For the above considerations, you are recommended to bring your own data. You can still use resources from the Internet, but you can only use the Internet for getting raw data.

If you choose this option, you need to report in detail how and where you obtain the data used in your project report. You can refer to the project report guideline for Option #2 below with regards to other parts of your project report.

Option #2 Analyzing the Hospital Compare Dataset

• Background and Dataset Introduction

Hospital Compare (<http://www.medicare.gov/hospitalcompare/search.html>) is a consumer-oriented website, managed by Centers for Medicare and Medicaid Services (CMS), which provides information on how well hospitals provide recommended care to their patients. Hospital Compare provides periodically reported data on the care quality of U.S. Hospitals based on various quality measures. For this final project, we will use the publicly disclosed quality data for U.S. hospitals that reflects the actual quality status of U.S. hospitals in 2021. For simplicity, the instructor simplified the dataset to focus only on hospitals' cardiac care services.

The data set provided is mainly concerned with hospitals' quality in treating three types of cardiac care patients, including heart attack, heart failure, and artery bypass surgery. Particularly, we focus on the 30-day mortality rate of these patients. (However, it is worth noting that mortality alone may not be the best quality measure, because the outcomes of care are influenced by many factors not controllable by the hospital, such as patients' family medical history or the severity of their cases.)

The data set is provided as one Excel file with three worksheets (in *.xlsx format), which contains descriptive information about 5,307 hospitals, the mortality data about their cardiac care patients, and the results of patient satisfactory survey (the survey was administrated to all patients of a hospital, not just cardiac care patients). The detailed column information of each worksheet is as follows:

[Sheet 1] "General_Infor"

Column Name	Description
Hospital_ID	A unique ID for each hospital. It is the provider number assigned by CMS to each hospital.
Hospital_Name	The name of a hospital
Address	The street address of a hospital
City	The name of the city where a hospital locates
State	The two-letter abbreviation of the state where a hospital locates
ZIPCode	The zip code of a hospital
County	The name of the county where a hospital locates
Hospital_Type	The type of a hospital, including four types: <ul style="list-style-type: none"> • A (Acute care hospital) • C (Children's hospital) • CA (Critical Access hospital) • M (Military General Hospital) • P (Psychiatric hospital)
Ownership	The ownership type of a hospital, including <ul style="list-style-type: none"> • DOD (Department of Defense) • GF (Government –Federal) • GH (Government – Hospital District or Authority) • GS (Government – State) • GL (Government – Local) • FP (Proprietary, for-profit)

	<ul style="list-style-type: none"> • NPC (Non-profit – Church) • NPP (Non-profit – Private) • NPO (Non-profit – Other) • P (Physician-owned) • T (Tribal)
EMS	Whether a hospital provides emergence services or not, including <ul style="list-style-type: none"> • Yes • No

[Sheet 2] “Cardiac_Care”

Column Name	Description
Hospital_ID	A unique ID for each hospital. It is the provider number assigned by CMS to each hospital.
AMI_count	The number of heart attack (acute myocardial infarction, AMI) patients treated annually in a hospital
AMI_mortality	the 30-day mortality rate for heart attack patients treated in a hospital, measured as the percentage of heart attack patients discharged from a hospital who die in 30 days
HF_count	The number of heart failure patients treated annually in a hospital
HF_mortality	the 30-day mortality rate for heart failure patients treated in a hospital, measured as the percentage of heart failure patients discharged from a hospital who die in 30 days
CABG_count	The number of patients who receive Coronary Artery Bypass Graft (CABG) surgery annually in a hospital
CABG_mortality	the 30-day mortality rate for patients who received CABG surgery in a hospital, measured as the percentage of CABG patients discharged from a hospital who die in 30 days

[Sheet 3] “Hospital_Rating”

Column Name	Description
Hospital_ID	A unique ID for each hospital. It is the provider number assigned by CMS to each hospital.
Star_rating	The rating of a hospital based on patient surveys, ranging from 1 (the poorest) to 5 (the best)

The data set provided is not fully cleaned. There might be many data quality problems. You are suggested to carefully review and clean the data as you see fit for your analyses. Unlike in our previous assignment, for the final project, you can clean the data in any way you want. You do not have to use Tableau for data cleaning. For example, you can prepare the data in Excel by using Excel functionalities. Discuss in your project report any data cleaning steps.

• Project Objectives

The major goals of this project are twofold: (1) understand the current status of hospitals offering cardiac care; (2) understand the variance of hospitals in their effectiveness of patient treatment. Students are encouraged to come up with their own research questions based on these datasets. Some exemplary questions may include:

- Geographically, where (state/county etc) do more cardiac care patients locate? Which states/counties have more cardiac care resources (e.g., # of hospitals)?
- Based on either mortality rates or patient surveys, who are the best hospitals in cardiac care (in your local area or a state of your choice)? Which state is better in cardiac care in general?
- Do different types of hospitals (e.g., governmental, non-profit, Proprietary) differ in their effectiveness in cardiac care? Do hospitals in different geographical regions (such as states, or groups of states as northeast, south, west, or mid-west) differ? Does having emergency services or not change a hospital's effectiveness in cardiac care?
- Is the quality of CABG surgery correlated with the effectiveness of heart attack/heart failure treatment in a hospital?
- Do hospital ratings, as measured based on patient survey, reflect their effectiveness in cardiac care? What could some other factors (such as location or ownership) that affect patients' ratings of a hospital?

You can be creative in proposing your own research questions. Your analyses need to be done in Tableau. However, you do not have to perfectly answer your research questions. If Tableau analyses are not sufficient in answering your questions, you can still report your analysis results and then discuss their implications.

- **Requirements on the final Deliverable**

Students will write a project report as the final delivery. While students can choose the specific format and the structure of their technical reports, the report needs to clearly cover the following contents (in addition to other contents the students choose to include), although not necessarily in this particular order.

- Who are the targeted audience of this report?
- What are the questions the report tries to answer? Why are they important questions?
- The process or procedures of data cleaning and preparation before analyses
- Detailed steps of data analyses.
- Detailed results, including text tables and graphics
- Result interpretation – what do the results mean?
- Discussion and conclusion – what are the implications of the studies? What are the conclusions or recommendations the author tries to make after the study?
- Submit your Tableau file (**in *.twbx format**) together with your project report.

- **Software tool**

As mentioned, students can clean and prepare the data by using any software tool he or she is familiar with. **The majority of the data analyses work, however, needs to be conducted by using Tableau.**