

## STAT 515 – Final Project

This is a group project. This requires forming groups of 2 members.

You are required to find a dataset from the web or another source, form multiple research questions and answer them using the data modeling techniques that we used in class. Please see the details below.

The project includes, updating your website with the final project, accompanied by a 15-minute in-class presentation.

### **Purpose**

In a standard final project, the team selects, visualize, explores, and describes different aspects of a data set using statistical models (and Hypothesis tests) discussed in class.

### **Submission Details:**

- Group member allocations must be completed by **November 15<sup>th</sup>**.
- Before **November 24<sup>th</sup>**, every team must submit the **Progress Report** on the final project. All teams must meet your instructor in person, during office hours to discuss your progress on the project and need to submit a **hard copy** of your progress report. The meeting will last roughly 10min.

It is recommended that groups submit their progress report **As Soon As Possible**.

On the day of the presentation (**December 10<sup>th</sup>, 2025**), you need to submit following documents:

1. 15 min in-class presentation using the website created.
2. URL of your Website with the Final Project.
3. Codes of your Project, embedded to your website as a separate Webpage.
4. URL of your GitHub Repository, that contains your website.

*No changes should be made to your GitHub repository after the mid project deadline and until your projects are graded.*

### **Data Set:**

- You can use the same data as the redesign project or a new one. (If you want).
- **Do not** use data sets from 'Kaggle.com'.
- **Do not** use data sets from "UCI Machine Learning Repository" (<https://archive.ics.uci.edu>)
- **Time series, survey data** generally requires special methodology for analysis (not covered in class). If you already have time series/survey data, and you want to use it, **let me know**. It may be possible to formulate a regression or classification problem out of it.
- There are some links on Canvas where you can download data.

**Objective:**

- You should formulate **three or more** relevant research questions about your data set.
- Need to answer your research questions by using the statistical learning methods that taught in class.

**Data Analysis:**

- It is a good idea to have a short exploratory analysis (summary statistics, histogram/density plot) before fitting a statistical model. Your model may be guided by this. For example a skewed density plot may suggest you should transform a variable.
- Your analysis should answer the research questions that you formulate about the data. Do not just use **all** the methods taught in class on your dataset unless they answer the questions in a meaningful way. It is perfectly fine to have more than two question about the data.
- Visualization of the data can be used to identify behavior and patterns of the variables. Also, it can be used to put relevant graphs (e.g. diagnostic plots, plot of accuracy measure for model selection) with your analysis.
- Once you formulate two or more questions about the data, you can use multiple methods to answer the same question. If you do, make sure to make a comment comparing them.
- Draw a conclusion from your analysis. The project should reflect your understanding of the statistical methods, not just the ability to apply them.

**Writing:** This is the writing(report) proportion of your data analysis on the Website. This report should have the following components:

- A description of the data (context, source) and questions you are trying to address.
- A clear description of the statistical model or methodology you are going to use to answer your questions. For example, if you are going to use a linear regression model or a classification model, what is your response, what are the possible predictors, if you are using any transformation of your variables, etc.
- The relevant details of the analysis. Include relevant output, tables and figures. Do not copy paste output produced in R. The tables and figures should be well-formatted and clearly labeled.
- The conclusion(s), their implications, as well as future questions/ways to further analyze the problem.
- List of references (including the data source) used in the report.
- **Do not include R code, raw data or R outputs in the main report.**
- You can include additional outputs and any other relevant details in the appendix.

I expect a writing based on academic writing standards not a bulleted point report! This is a graduate level class so the quality of your writing should meet the requirements of academic writing of proceeding or peer-reviewed articles. Every resource used needs to be references properly (preferably APA format). Failing to reference books, documents, R, R packages and codes, will result in losing points. Copy pasted text (without quoting and fully referencing) is considered plagiarizing.

**Presentation:** Each group will have 15 minutes to present. Each group member should present some material during the presentation. The group will not be graded on their presentation skill, but points may be deducted if the presentation shows a lack of understanding of the material.

**Grading:** 95 out of 100 points will be determined by the project website and writing. The project will be evaluated in the following broad areas: (Please see the rubric for more information).

- 1) Adequate variety and correct use of methods taught in class
- 2) Appropriate interpretation of models and their output
- 3) Level of effort
- 4) Writing quality

5 out of 100 points will be determined by individual contribution from each group member. For the individual contribution component, each group will have 5 points per member (10 points for 2-person groups and 15 points for 3-person groups) to divide among group members as they see fit (similar to the redesign project). Each group member can be allocated between 0 and 10 points (half point increments are fine) and should be discussed and unanimously agreed upon. It is possible to exceed 100 points for the project. If a group cannot come to a consensus, Instructor will meet individually with each group member and then decide on an appropriate allocation.

**Group work expectations:**

- **ALL GROUP MEMBERS ARE EXPECTED TO CONTRIBUTE ROUGHLY EQUAL AMOUNTS OF TIME AND EFFORT TO THE PROJECT.** The form of the contributions may vary according to each group member's individual roles, responsibilities, and expertise. However, there **SHOULD NOT** be extreme disparities in the amount of time and effort contributed to the project among group members. On the other hand, those that do go above and beyond to make outsized contributions to their projects should be recognized and rewarded accordingly. Taken together, this is why some flexibility in the total individual grades is allowed, but not complete flexibility.
- Good communication is key to ensuring each person understands their individual roles and responsibilities, dependencies across each member's workflows, and expected deliverables and deadlines. **You are expected to make every effort possible** to attend project meetings, meet agreed upon deadlines, and coordinate with other group members depending upon your work to complete their work. **Unresponsiveness and failing to meet deadlines previously agreed upon are unacceptable.** Please report any such behavior directly to instructor.
- You may need to allocate time to explaining your individual work to others in the group. This is part of what makes group projects beneficial: 1) learning by teaching others, 2) learning by listening to others, and 3) realizing the limitations of your own understanding through 1) and 2) and pushing for a better understanding. This aspect is what makes the group project truly cohesive rather than a loose collection of individual components.
- If you encounter any difficulties within your group, please raise all concerns to your instructor in a timely manner, so that Instructor may address them before they become problematic.