It is now time to begin working on your class project. Your project will be in two parts.

1 Part 1

This part is to be done **individually by each student**. You are required to produce a dashboard to investigate customer churn in Tableau. Deadline for submission of this part is **Friday May 9th, 2025.**

2 Part 2

Part 2 is to be **done in your respective groups**. This is a comprehensive document on my expectations from you for your class project. Your project contributes to 30% of your final grade. Your submitted project should include a **well written** and **detailed** description of your analysis of your chosen data set. You are to consider and report on the following:

1. Problem definition
2. Data description and exploration
3. Data Analysis
4. Model specification and fitting
5. Model validation and diagnostics

Some guidelines for the project

- As you see fit, incorporate the R output and graphics into your report. Every graphic produced should be relevant to some part of your discussion, hence do not incorporate too many graphics, especially if they are unnecessary. You should include phrases like "based on Figure 2,..." if you are including a particular figure or table. You can summarize your findings from your data exploration without attaching too many graphs and or tables.
- Your document should be double spaced. Also, divide your project onto sections with titles.
- Aim for a report length of **at most** 12 pages, including figures, tables. There is no minimum number of pages you should attain.

- Use a professional and legible font, such as Times New Roman, size 12.
- Maintain consistent 1-inch margins on all sides.
- All figures and tables should have clear labels and captions. Once again, ensure they are referenced in the text.
- Your report should be error-free. Make sure to edit it as many times as necessary, at least 5 times before you submit your final report.
- Do not plagiarize anybody's work.
- Some modeling guidelines:
  - Check the nature of your target variable. That would inform which model to use
  - Use only 90% of the data for training and testing(Split this up for training and testing). Predict on the last 10% and comment on how well your model did. You are considering this 10% as future data, independent of the test set.
  - Be sure to deal with missing values - remove them if they are not too many. Note there are several other ways of dealing with missing values but you are not obliged to explore them for the purpose of this project.
  - Use as many of the models as we have discussed in the class that you see fit for your specific use case. If you end up fitting more than one model, compare them and choose your final model.
  - Employ everything you have learnt in this class, from Week 1 to Week 6. Only use models discussed in this class
  - Be creative.
  - Submit your written report and your R file you used.

Written project outline

Your work should be organized as follows with the relevant section titles, and add titled subsections if needed:

1. Have a title page.
2. **Introduction**: This is where you define your problem and explain why it is important to get a model for that purpose. Give the reader important background information on why it is important for you to investigate the problem and come up with a solution. Also give information on why the variables considered may be important for the specific use case.
3. Perform steps 2 to 5 listed above. Give a detailed exposition on any method you are using, any observations and results from your analysis.

4. Discuss your results and conclusions.
5. **Bibliography**: If you make use of any references, make sure to cite them.
6. Appendix- Put a well commented code for your entire analysis making sure they fit in the required margins. You can put supplementary figures here as well

Be sure to have fun doing this. You are combining all that you have learned in the class so far demonstrating your understanding.

Evaluation Criteria

Your report will be assessed based on:

- **Content Depth:** Comprehensive coverage of your topic and its application to the dataset.
- **Analysis Quality:** Rigor and correctness of the data analysis process.
- **Interpretation and Insight:** Ability to draw meaningful conclusions and insights from the results.
- **Presentation and Clarity:** Organization, structure, and clarity of writing.

DUE: Thursday May 8th 2025 at 11:59PM. Any other information will be duly communicated to you. Do not hesitate to forward all questions to me and schedule zoom meetings if yu need any help.

3 Use cases

The Iranian Telecom company churn data is linked on blackboard. Get the data description and variable names by following the link and scrolling down the page.

3.1 Evaluating Tariff Plan Preference

Target Variable: Tariff Plan (binary: 1 for Pay as you go, 2 for contractual)
Predictors: Call Failures, Complains, Subscription Length, Seconds of Use, Frequency of use, Frequency of SMS, Distinct Called Numbers, Age Group, Status, Customer Value.

3.2 Analyzing Subscription Length

Target Variable: Subscription Length (continuous data)
Predictors: Call Failures, Complains, Charge Amount, Seconds of Use, Frequency of use, Frequency of SMS, Distinct Called Numbers, Age Group, Tariff Plan, Status, Customer Value.

3.3 Predicting Customer Churn

Target Variable: Churn (binary: 1 for churn, 0 for non-churn)
Predictors: Call Failures, Complains, Subscription Length, Charge Amount, Seconds of Use, Frequency of use, Frequency of SMS, Distinct Called Numbers, Age Group, Tariff Plan, Status, Customer Value.
Predicting churn is a common application in the telecommunications industry, where it is critical to understand factors that may contribute to a customer's decision to leave the company.

3.4 Modeling Frequency of network use

Target Variable: Frequency of Use (count data, as it's the total number of calls made by a customer).
Predictors: Subscription Length, Charge Amount, Seconds of Use, Frequency of SMS, Distinct Called Numbers, Age Group, Tariff Plan, Status, Customer Value.
Here, we'd be examining how different factors affect how often customers use their phone for calls. For instance, customers with longer subscription lengths may call more frequently due to habit or necessity, while those with higher customer value may have different usage patterns.

3.5 Assessing Complains

Target Variable: Complains (binary: 0 for No complaint, 1 for complaint)
Predictors: Subscription Length, Charge Amount, Seconds of Use, Frequency of use, Frequency of SMS, Distinct Called Numbers, Age Group, Tariff Plan, Status, Customer Value. Here, we want to predict whether a customer would make complaints or not based on these predictors.

3.6 Modeling Call Failures

Target Variable: Call Failures (count data)
Predictors: Complains, Subscription Length, Charge Amount, Seconds of Use, Frequency of use, Frequency of SMS, Distinct Called Numbers, Age Group, Tariff Plan, Status, Customer Value.
Here, we want to understand the factors influencing network quality or customer experience issues.

4 Group allocations to use cases

Group 1 and 9: 1.4 and 1.5

Group 2: 1.1 and 1.4

Group 3: 1.5 and 1.6

Group 4: 1.4 and 1.5

Group 5: 1.1 and 1.2

Group 6: 1.2 and 1.3

Group 7: 1.3 and 1.6

Group 8: 1.1 and 1.6