VI Data Science Home Assignment — Instructions

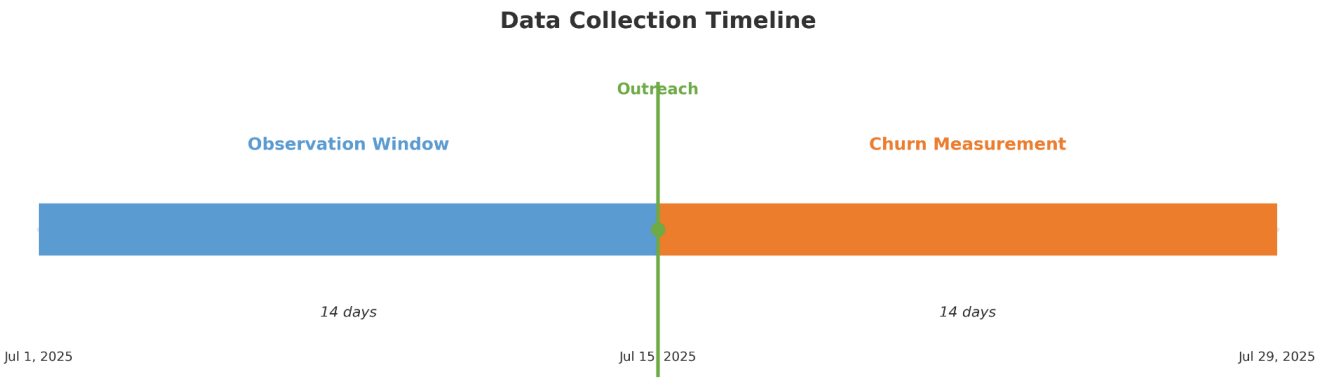
Overview

Our client, WellCo, is experiencing increased member churn and seeks our assistance in reducing it. Your primary objective is to provide a ranked list of 'n' members for prioritized outreach. Since outreach incurs a constant (unknown, and marginal) cost, you must also determine the optimal value for 'n'.

Data Provided

- Training Files:
  - web_visits.csv
  - app_usage.csv
  - claims.csv
  - churn_labels.csv
- Test files:
  - test_web_visits.csv
  - test_app_usage.csv
  - test_claims.csv
  - test_members.csv
- Reference:
  - wellco_client_brief.txt
- Schema files:
  - schema_web_visits.md
  - schema_app_usage.md
  - schema_claims.md
  - Schema_churn_labels.md

Note: Make your analyses on the training data and apply them to the test data.



Required Deliverables

- A public Git repository containing a reproducible end-to-end solution.
- A README file detailing setup and run instructions, along with a concise description of your approach.
- An executive presentation (3-5 slides) tailored for non-technical stakeholders.
- A CSV file containing a sorted list of the top 'n' members for outreach. This file must include, at minimum, member_id, a prioritization score, and the member's rank.

Note: Use the provided **test files** (`test_*.csv`) to evaluate your final model.

Evaluation Criteria
Your submission will be evaluated based on the following aspects:

- **Code Clarity and Readability**
- **Solution Robustness**
- **Visualization Quality**
- **Presenting Results**
- **Storytelling**

Additional Guidance
To help you focus your efforts, please address the following in your approach and documentation:

- **Feature Selection:** Explain which features you chose to use and why. Consider domain relevance, data quality, and predictive power**.**
- **Model Evaluation:** Describe how you evaluate model performance and justify your chosen metrics.
- **Using Outreach Data in Modelling:** The dataset includes an outreach event that occurred between the observation period and the churn measurement window. You are expected to incorporate this information into your modelling and explain how it influences your approach and results.
- **Selecting n (Outreach Size):** Describe how you determine the optimal outreach size. Is it driven only by cost, or are there other factors you considered?