Person A: Data Gathering, Exploration & Preprocessing

- **Task 1: Data Collection & Preprocessing**

  – Download the "Presidential general election polls (current cycle)" data.
  – Explore the dataset to understand its structure, variables, and context.
  – Perform data cleaning (e.g., handle missing values, outliers).
  – Document any high-level cleaning steps and ensure a reproducible data preparation workflow.

- **Task 2: Exploratory Data Analysis (EDA)**

  – Generate summary statistics for all relevant variables.
  – Create visualizations (e.g., histograms, boxplots, scatter plots) to understand variable distributions and relationships.
  – Identify key predictors to include in the model, and justify their selection.

- **Task 3: Pollster Deep-Dive for Appendix A**

  – Pick a pollster from the data sample.
  – Research the methodology of this pollster, focusing on survey design (population, frame, sample recruitment, etc.).
  – Write a detailed overview of the survey's strengths, weaknesses, and key features.

- **Task 4: Data Section Write-Up**

  – Write a section on the dataset, including background, context, and detailed examination of variables.
  – Include all visualizations and summary statistics in a clear format.

Person B: Model Building, Analysis, & Validation

- **Task 1: Model Development**

  – Decide on the appropriate model to use (linear or generalized linear model).
  – Justify the model choice based on the dataset and research goals.
  – Implement the model in R, Python, or a combination, and ensure reproducibility.

- **Task 2: Model Explanation & Assumptions**

  – Clearly define all model components using mathematical notation and plain English explanations.
  – Discuss and document underlying assumptions, limitations, and potential issues.
  – Validate model performance using techniques like test/train splits, RMSE calculations, or sensitivity analyses.

- **Task 3: Results Section Write-Up**

  – Create tables, graphs, and visualizations to present the model's findings.
  – Write a clear and concise interpretation of the model results.

- **Task 4: Discussion Section Write-Up**

  – Discuss key takeaways from the model and results.
  – Highlight what the paper contributes to understanding the election forecast.
  – Address any weaknesses and propose future steps or improvements.

Person C: Appendices, Documentation & Reproducibility

- **Task 1: Idealized Survey Design for Appendix B**

  – Design an ideal methodology and survey for forecasting the U.S. presidential election on a $100K budget.
  – Detail all aspects of the methodology (sampling approach, recruitment, data validation).
  – Create the survey on a platform like Google Forms, include a link in the appendix, and export the survey content as a PDF.

- **Task 2: Reproducible Workflow & Documentation**

  – Create an organized folder structure in the GitHub repo (e.g., `data/`, `scripts/`, `figures/`, `docs/`).
  – Write a detailed README file explaining the project, data sources, model, and workflow.
  – Document any use of LLMs or other tools, and include a separate usage text file if necessary.

- **Task 3: Final Paper Integration & Formatting**

  – Combine and format sections from all team members into a cohesive paper.
  – Add cross-references to all figures, tables, and equations.
  – Ensure all required sections are present and adhere to the rubric guidelines.

- **Task 4: Quality Checks & Submissions**

  – Proofread the paper for typos, grammar, and formatting issues.
  – Verify that all elements are properly cited (R, data sources, references).
  – Check that the paper, repo, and all files do not show any evidence of being a class project.

Timeline

Week 1: October 7th - October 13th

- **Person A: Data Gathering & Preprocessing**

    - **By October 9th:** Download and inspect the "Presidential general election polls" dataset. Familiarize yourself with its structure and context.
    - **By October 11th:** Complete initial data cleaning (handle missing values and outliers).
    - **By October 13th:** Document all preprocessing steps and ensure the workflow is reproducible. Provide a summary of any challenges encountered and adjustments made to the data.

Week 2: October 14th - October 20th

- **Person A: Exploratory Data Analysis (EDA)**

    - **By October 15th:** Generate summary statistics for all key variables (e.g., mean, median, distribution).
    - **By October 17th:** Create visualizations (e.g., histograms, boxplots) to explore the distributions and relationships between predictors and outcomes.
    - **By October 19th:** Identify and justify key predictors to be used in the model based on EDA findings.
    - **By October 20th:** Draft a summary of the findings from the EDA and prepare the data section's visualizations and statistics.

- **Person C: Initial Workflow Setup**

    - **By October 18th:** Set up an organized GitHub repo structure (`data/`, `scripts/`, `figures/`, `docs/`), and add all current progress and documentation.

Week 3: October 21st - October 22nd

- **Person A: Pollster Deep-Dive for Appendix A**

    - **By October 21st:** Choose a pollster and research their methodology in-depth, focusing on their survey design and approach.

– **By October 22nd:** Draft the Appendix A content with a detailed overview of the pollster's methodology, strengths, weaknesses, and key features.

- **Majority of Data Tasks Due on October 22nd**

  – Data gathering, exploration, and pollster analysis should be completed and ready for initial review.

Week 4: October 23rd - October 29th

- **Person B: Model Building & Initial Analysis**

  – **By October 25th:** Decide on and justify the type of model (linear or generalized linear) based on research goals.
  – **By October 26th:** Implement the model in R or Python and ensure the reproducibility of the code.
  – **By October 28th:** Begin validating the model by conducting performance checks (e.g., test/train split, RMSE).
  – **By October 29th:** Write a draft of the results section, including visualizations and tables.

- **Person C: Idealized Survey Design for Appendix B**

  – **By October 27th:** Develop a survey methodology for forecasting the election, based on a $100K budget.
  – **By October 29th:** Create the survey on Google Forms, add the link to the appendix, and export a PDF version.

Week 5: October 30th - November 4th

- **Person B: Finalize Model & Discussion Section**

  – **By October 31st:** Finalize model performance analysis and diagnostics.
  – **By November 1st:** Write and refine the discussion section, including key takeaways, weaknesses, and future steps.

- **Person C: Final Integration, Quality Checks & Review**

  – **By November 2nd:** Integrate all sections (data, model, discussion) into a cohesive document. Ensure all cross-references, figures, and tables are correctly placed.

– **By November 3rd:** Conduct quality checks on grammar, formatting, citations, and ensure that there is no evidence of this being a class project.
– **By November 4th:** Final proofreading and submission of the polished paper.