PROJECT 2: Support for Ranked Choice Voting (Q20)

Overview

We want to understand which factors (such as political affiliation or confidence in elections) predict whether someone supports or opposes **ranked choice voting** (Q20).

Variables

- **Target**: Q20 (Support for ranked choice voting)
- **Predictors**:
    - **Demographics** (Q1–Q8: Ward, Age, Gender, Marital Status, Ethnicity, Religion, Education, Income)
    - **Political Affiliation** (Q9)
    - **Confidence/Direction** (Q10, Q11, Q14–Q16)
    - **Presidential vote/excitement** (Q12, Q13)
    - **Information Sources** (Q17)
    - **Reasons for In-Person** (Q18, optional)

Steps

1. **Clean and Prepare Data**
    - Handle missing values (e.g., label "No Response" as a separate category if Q20 is blank).
    - Encode data based on the mappings
2. **Set Up Model**
    - Classification of support (binary or multi-class).
    - Model options: Logistic Regression or Random Forest
3. **Train/Test Model**
    - Split the dataset into training (e.g., 80%) and test (e.g., 20%).
    - Perform k-fold cross-validation (k = 5) on the training set to fine-tune model parameters.
    - Apply the final model to the test set and compare predicted vs. actual Q20 responses (accuracy, F1-score, etc.).
4. **Interpret and Use the Results**
    - Examine feature importance or coefficients to see which questions (Q10, Q11, Q14, etc.) drive support for RCV.
    - Identify patterns (e.g., certain age groups, confidence levels, or political affiliations heavily favor or oppose)