PROJECT: Interactive Election Polling Analysis

You will work by groups of 3 and choose a name for your group.

1. **Data collection and web scraping**
   - Scrape your assigned Wikipedia polling pages
   - For each election, extract all the information available:
     - Poll dates
     - Results for each party/candidate
     - Prediction for each party/candidate at the poll date
     - Polling organizations / firms (if available)
     - Sample sizes (if available)
     - Confidence intervals (if available)
     - Other relevant information (if available)
- Important: the most important polls are national-wide for the first round of the election. You do not need to collect data on regional polls, primaries, etc.

2. **Data cleaning and processing using pandas**
   - Create a data cleaning pipeline that handles:
     - Date standardization + missing values + inconsistent formatting + outlier detection
     - Use two variables: identity_candidate (name of the candidate/party) + create a variable political_leaning_candidate indicating the political leaning (far-left,left, center, right, far-right, green, etc.) of each candidate/party based on your research.
       - For instance: Smer–SD in Slovakia is a left-wing party. Your final dataset should contain this information.
     - At the end of this part, you should create an excel file for each election. These excel files will have standardized names such as {country}_{year}_{electiontype}.xlsx
       - These files will contain the following variables: poll_date, sample_size (if available), polling_organization (if available) final_result_candidate1, prediction_result_candidate1,identity_candidate1, political_leaning_candidate1, final_result_candidate2, prediction_result_candidate2,identity_candidate2, political_leaning_candidate2, etc etc.

3. **Create an interactive interface combining all the data sets**.
   - Users should be able to navigate the dataset and obtain relevant information on each election they ask about. This information should contain:
     - A plot for each election showing the poll predictions over time using **matplotlib.** It should show polling averages over time with trend lines.
     - Mean/median/max/min (summary statistics) prediction for selected periods before the election date using **numpy/pandas**
   - Users should be able to enter data manually
     - Think about edge cases and error handling (this also applies to the rest but particularly here)

- You can build the interface directly on python. A better way would be to use designated libraries to build a more "user friendly interface" such as Streamlit, Dash or Tkinter.

4. **Demo during the last session (10 min per group)**. The demo should contain two moments:
   - Presentation of the interface and how it works
   - Presentation of the code and how it works
   - The presentation should be:
     - 10 minutes maximum
     - Without written notes

5. **After the last session, you should upload your code and datasets on AMETICE**

Additional information:
It is possible to complete these tasks in many different ways with python. Therefore:

- Creativity will be rewarded. You are free to use additional functionalities of python we have not covered in class as long as it allows you to complete the tasks.
- You should consider that you are producing a product that might be used by consumers. The quality of the visualization and the interface will also matter for the final grade.