

## Milestone 2

1. Group #51
2. We have met and communicated with each other. Our group has scheduled regular meetings 2/3 times a week.
3. We have met with our TF (Henlin Jin) and we have scheduled regular meetings with him every other week.
4. Our team has not yet decided on a single question to answer in the project, based on the project description and references. So far, we have suggested/posed numerous possible questions. We are presenting the narrowed down list of possible questions in addition:
  - a. How to collect and model data so that our prediction is closer to the results than 2016's predictions? What can we learn from the data collection and modeling process from the 2016 models to improve the accuracy of our model?
  - b. Regarding the complexity of the prediction, and that little differences in our models can lead to consequent different outcomes, how can we narrow down the confidence intervals and better communicate the results?
  - c. How can we combine the fundamentals based approach (leveraging some metadata directly related to macro indicators), with a weighted blend of state polls and per-state modelling to account for the difference between national votes and electoral college votes?
  - d. How can we incorporate the per-state modeling approach, in order to take into account the voting trends amongst states with similar demographics, correlation between states and the importance of swing states?
5. Tasks to do:

In order to answer the set of questions we have listed upwards, we might need to get some specific data, on two scales: the macro scale with some socio-economics indicators, and the micro one directly related to voter's opinions. Here is presented how we decided to split the work among the group. We should keep in mind that getting good-quality data is a pillar in order to perform modelling the right way, explaining why this task is crucial.

  - a. Get macro data and perform EDA on it: David, deadline by November 10
  - b. Get human data (polls) and perform EDA: Tale and Raphaëlle, deadline by November 15
  - c. Write data description (summary of data selected, attributes and why we selected it): David, deadline by November 15
  - d. Write EDA summary: Tale, deadline by November 16
  - e. Write final version of Milestone 3: Raphaëlle, deadline by November 18