

## BUS502: AI Business Cases

### Topic 2: Elections

**Question 1.** You are a data scientist at a shadow research center for the conservative party. The secret agency has long standing reputation of most accurate predictions on almost all political elections in your country, but for the last 5 years with zero winning in all nation-wide political events, party insiders have gradually become skeptical about your center's analysis and policy suggestions. Your job has been to prove that it is not so much the center's fault for the serial defeats as it is the outcome of ill-positioned or mis-delivered strategies. For the last 6 months during presidential campaign, nonetheless, you have confronted with party seniors paying no attention to your analysis at all.

Today is the presidential election day. There has been 35% mail-in ballots, which is exceptionally large number due to global pandemic, but the voter turnout does not seem as high as your expectation. Given the previous elections and newfound accessibility of mail-in ballots, the party expects over 80% voting in total, but at 1pm, 5 hours before the closer, it is only 60% nation-widely, which gives you a serious doubt if it will ever reach 75%.

What concerns you the most is the huge difference in vote counts between the party seniors' and yours. By the exit poll up until 1pm, your party enjoys a 7.5% margin over the democratic party's candidate that senior members of your party think it is similar in mail-in ballots, because the most recent poll right before the election was nearly identical.

For the past few days, your research center has conducted a secret poll by phone calls to measure the tendency in mail-in ballots, which is in fact 9% in favor of the counterpart's candidate. Adding up both numbers until 1pm,  $35\% \times (-9\%) + 25\% \times 7.5\%$ , even if the total voter turnout reaches 75%, you see that your party is likely to lose by 1.5% margin.

You are in a rush to circulate the back-of-the-envelope calculation to every member of the party emphasizing that everyone should push their friends and neighbors to raise the voter turnouts at least upto 77%. But, as was for the past few months, no one in your party listens to you. They think they already have won big time, and are talking about who should be given what secretary positions in what order for how long.

It has been your dream to work at the Presidential Office since your childhood. But, without pressing persuasion, the dream is likely a distant future. You badly want it tomorrow.

- 1 Assuming the election day turnouts will be homogeneous between the morning and the afternoon. How much voter turnouts in total does your party need? Let's assume that there can be some heterogeneity. Add some variability in your prediction and expand your analysis. Is it rational to add heterogeneity in your prediction?
- 2 Your data science team's boss asks you why the mail-in ballots are over 15% disfavor to your party from election day ballots. She firmly believes the earlier poll should be the right reference number, and the poll for mail-in ballots has noise in it, because of shy conservatives. You do not have full record of how the poll for mail-in was conducted, but let's assume that you believe statisticians did it right with minimal bias. If you are the statistician, how would you conduct the poll? You have variables called region, gender, and age groups by 10. Your country's minimum voting age is 20.
- 3 While listening to your quick presentation, your boss discovered from the National Election Commission (NES) webpage that the exit polls are conducted in systematic sampling, and the poll for mail-ins was done by cluster sampling, both of which should have no sampling bias in ideal cases. She wonders why then they use different sampling strategies and why there can be differences. Kindly educate your boss.
- 4 Since your team's boss is a trained data scientist, she understood all your arguments on possible heterogeneity in the afternoon votes and actual heterogeneity between mail-ins and election day turnouts. But, party elders expresses strong belief on significant similarities in ex-ante polls and actual counts. Your boss has been benchpressed by seniors' extensive election experience. It is now your turn to explain why polls

can be different from actual results.

- 5 One of the party elders claims that your argument is a contradiction in that on the one hand you argue polls could be wrong, while on the other hand, you trust the poll for mail-in. How can you reconcile the contradiction?
- 6 They are not still persuaded because they do not understand the term "heterogeneity". They firmly believe each region, gender, and age group has strong tendency, thus your argument on the differences in mail-ins and election day ballots are not accepted. Re-write your discussion in their language or in their "factor"s.
- 7 As poorly-educated as they are, still they are afraid of losing, and finally accepted your argument. They have decided to spend some extra budget to push voters. But in terms of strategy, you again see differences. The elders claim that they should only spend the budget for region/gender/age groups that are significantly in favor of them, while you believe it does not have to be that specific. In regards to factor analysis, defend your logic.
- 8 Since the elders are unlikely persuaded, your team's boss suggests you that deep-learning style model be used in your argument. They will believe if it is done by Artificial Intelligence, even if it is not really AI. Provide your strategy to use neural network models to strengthen or at least re-write your earlier arguments by factor analysis.
- 9 Assume that you have more variables such as ethnicity, annual income levels per household, housing prices, education levels, and etc. Will more of these variables help accurately measure your prediction? Why or why not? What's the advantage of sampling methods over multi-variate regressions in election analysis?
- 10 The election did end up with 77.5% voter turnouts, which is 0.5% above your threshold for winning, but winner predictions by exit-poll are varying by TV stations. One of them expects your party's wining with 0.6% margin, while the other station says 0.8% edge for the counterparty's candidate. You are told that their predictions are based on systematic sampling and cluster sampling, respectively. Explain the differences to your bosses and predict the winner.

**Bonus** Your party won by 0.7%, the smallest margin in your country's all historical presidential elections. You expected you would win every bit of credit for forecasting the most accurate outcome, but people are busy talking about government positions in the new administration. Given the upcoming battle to win the core position in the Presidential Office, you would like to remind them that your analysis and last minute voter push were the keys for winning. Provide one pager summary with convincing argument that can persuade even the non-experts in data scientific analysis.