DS-NYC-DAT-45 | Final Project 2: Project Design Writeup and Approval Template

"Can We Predict Hillary Clinton's Performance in the 2016 Election Based on ACA Exchange Plan Fluctuations?"

Project Problem and Hypothesis

- Political leanings aside, I believe the 2016 presidential election caught a lot of people by surprise, no less the loser of the election, Hillary Clinton.
- My thinking, which is not very different than those in the "industry", if you wanted label punditry an industry, is: she could have seen it coming. And why is that? There was a political groundswell which intuitively materialized between 2012, the year of Barack Obama's re-election over a wealthy Republican opponent and 2016, the year Hillary Clinton lost to a wealthy Republican opponent?
- What changed? Part of my hypothesis is, despite any flaws in Hillary's candidacy, there were macro issues at play, namely the amount of attention placed in the Republican platform to repeal the ACA, Barack Obama's crown domestic policy achievement during his term. And with that, specifically the steady rise in ACA (which we will define as "on_exchange" plans) health plans.
- What I want to tackle is, looking at Obama's performance on the county-level, taking into account Census-captured changes in demographics, could we have predicted Clinton's performance in counties by looking at a few features: any potential core changes in demographics; changes in ACA rates; changes in ACA plan-availability.

Datasets

- I am using a few datasets. One is from my employer and the other which I found on Kaggle.
- The employer-driven dataset, which I would like to not host on Git, since it is the core product we do sell.
- The Kaggle dataset contains Census data on county demographics and election results btween 2012 and 2016.
- I have sent to both Paul and Mason, at one point or another, the datasets over Slack.

Domain knowledge

- As a politico myself, having worked in both politics and health insurance in different points in my career, I think the union of the two is a rather interesting one, especially in how in the future we might want to try to model different macroeconomic changes to voter behavior.
- There is a lot of noise in the world of political/electoral analysis, but the concept of "all elections are local" isn't a new one, and what is more local than your out-of-pocket health costs?

Project Concerns

- There are some risks. For one, we might be finding some faulty correlations between health care price rises and Hillary's ultimate vote share. Though a price might have risen, say, five-fold, were there other causes that contributed to her loss on a county-level? It's possible, and we should try to account for that by looking at demographic data to ensure nothing too, too different occurred on a

county-level. Are we looking at similar populations? That's the purposes of the introduction of the Census-driven data.
- And what if it's inconclusive as a whole? Also possible.

Outcomes

- What I consider success is given the average price and vote fluctuation between the years we are measuring (health 2015-2017) and votes (2012 and 2016), can we predict whether Hillary will perform better or worse than Obama relative to the mean? Can we even predict her actual vote-total % in that county? The latter might be more difficult.
- I want to say success is being able to predict Hillary's performance in swing counties (ideally all counties); or at the very least highlight problematic counties Democrats could have anticipated would be problems, which would fit a certain criteria.
  - Obama barely won in 2012.
  - They experienced huge raises in health insurance costs over the last few years.
- This might be an interesting data visualization problem as well, and something I will explore.