D210
Landen Bailey

Part 1: Interactive Data Dashboard

1. The datasets used for the dashboard can be found in the following attached files:
   - Churn Dataset: churn_clean.csv
   - Census Data: populationPerZip.csv and MI_ZIP.csv (census dataset was split into two separate files to better accommodate Tableau uploads)

2. The Dashboard itself can be viewed through a very convenient accessibility feature provided by tableau known as "Tableau Public", a cloud based visualization system that allows any user with an internet connection to access my Dashboard along with its visualizations, metrics, and data at this link: https://public.tableau.com/app/profile/landen.bailey/viz/D210_17014111413950/Dashboard1

   - To see and interact with the dashboard one simply has to click on the link and explore the dashboard from there.

3. Upon arriving at the dashboard the user will see 4 separate graphs:
   - The first 2 are histograms with well labeled data defining the relation between income and monthly cost and between tenure and census defined population.
     - If further exploration is desired in these histograms the user can select one or more columns in the bar chart to filter the dataset to those select bins of income/population to filter not only that graph but all of the graphs to only include data from the selected bins.
   - The other 2 graphs are maps of the United States, the first representing region ratios of customers median reported income to general population median incomes, and the second displaying regional customer per capita counts.
     - These graphs can similarly be filtered against by selecting one or multiple states on the map. Likewise, filters applied to the maps will also filter down the relevant data for the other graphs.
   - A final set of features can be viewed in the very top right corner where additional filters can be set to control specifically for a range of customer reported Incomes as well as for the current churn status of customers.
   - All of the interactive filter options can be used simultaneously, though the best understanding of the dashboard is more likely to come when only one or few filters are applied at a time. With too many strict filters the insights reported in the graphs will quickly become just single data points, which is bad for real understanding of generalized trends.

Part 2: Storytelling with Data

B. The Panopto video is attached and a summary of the points discussed is included below, but it can be viewed at this link:
https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=b49fdb64-f827-4b51-a906-b10500a84bfe

- an introduction of yourself and your background
    - New to the company, but has many years of experience with data analysis, so when I got here I was excited to be able to provide some quick actionable insights to leaders at the company

- a summary of both chosen data sets and the context in which they occur
    - Collected metrics of customer demographics alongside their churn status
    - Collected census data about us zip code level populations and median incomes.

- an outline of key results from your analysis of the two data sets
    - Monthly charges were across the board were higher for customers who have churned.
    - The customers who have not churned are generally the customers who have a below average income.
    - The company has a pretty even spread of customers in almost all states. Which means that the states with larger populations are not pulling their weight.
    - General trends seem to indicate customers living in highly populated areas tend to stay with the company's services longer.

- two different data representations to serve as supporting evidence for your results
    - Each of the four graphs charred in the video and in the Tableau dashboard directly supports those key results I outlined

- a summary of actionable insights based on your results
    - Insights in this data supports that lower class people living in high population areas would stay longer with the company and be the least likely demographic to churn which I believe would increase overall revenue, despite those customers having lower individual contributions to the companies success. The data also indicates that the more populous regions of the United States are currently underutilized markets that could have huge potential if we build a game plan around this specific target demographic.

Part 3: Reflection Paper

The company I am performing this data analysis for has had a big problem with customer churn rate. It is very expensive to acquire new customers and it gets even more expensive if those customers don't stick around for a very long time. The churn rate, or the rate at which customers leave a service is a common business metric used to indicate the rate at which customers are leaving. The task of my data analysis was to try and understand why it is that customers are leaving and to come up with actionable insights as to how to prevent future customers from leaving so quickly, if at all. Or in other words, my task was to use data to lower the churn rate. My idea for how I think this can be achieved is expressed in the dashboard I shared. It looks at the kind of demographic attributes that our customers have, and how some of them indicate which customers are more likely to churn than others. Which can be used to discuss which kind of new customers we should be targeting or what kind of similar plan options and offerings we should put out in order to minimize our churn rate.

It also takes a look at these same customers, but in a broader perspective, now not only from the context of how do these customers compare to other customers, but how do these customers compare to other non-customers. I found demographic characteristics of our customers who have not churned that can be used to more broadly target a specific demographic of non-customers (yet) who are the most likely to last long with our services which with time will lower the churn rate. This broader demographic data comes from census data which I acquired to more effectively demonstrate that we have to know what kind of people will stay for the long haul before we just go out and blindly acquire more expensive customers who will leave after a month. This new data allows us to look at things like, What economic class do our customers come from? and How populated is the entire city that they live in? Both of which insights are not available in the initial dataset and can be crucial to understanding churn prevention on a broader perspective.

A couple of the visual representations of this data in my dashboard indicate 2 key characteristics of our customer basis and specifically those who do not churn. Both of them are maps of the United States with metrics applied to each state. The first shows the ratio of our customers' median income compared to the median income of all people in the same zip codes. This gives an indication on a regional basis to executives of what economic class most of our customers come from? The second gives a customers per capita metric of each state (number of customers per 10,000 people). This gives a heatmap style indication of what parts of the states most of our customers are located in. Both of these visuals Executives can use to better understand their customer populations in terms that are more relatable to a wider population.

In order to better apply that understanding of the customer population to actionable insights other metrics were provided as well as some controls to interact with the visuals and allow the executives to fine tune the analysis of data they might need to perform in a specific context. These controls include a slider for the range of their customers' income. Adjusting this would affect the income ratio graph as obviously customers with a higher income will compare differently to the populous average than customers with lower incomes. Both of which comparisons can be useful in attempts to understand the market though. The other control is a

filter for Churn specifically. This control affects all graphs and metrics and it is crucial for executives to see the effect different factors have had and will continue to have on Churn unless something else is done.

It was important to me that these visualizations be accessible to all people, and especially that it not be confusing for colorblind people to see and understand. Being colorblind myself it was easy to pick color palettes for the maps that did not blend together, and then simply using step color sequences instead of gradients helped as well.

In my presentation I was able to show all of these graphs in the dashboard to possible executives in sharing the story of our company's battle with churn and how the start to the end of the war would be found only by first understanding the situation. Which is done through data. I shared the two graph maps previously mentioned and pointed out that the economic class of the vast majority of our customers was the lower class. And that the high concentrations of our customers per capita are in the lower populated areas. I provided further evidence as well that our customers in more highly populated areas always seemed to last longer with the company than those from low population areas. So being less concentrated in those markets was a big miss on the company's part. One might assume the same would apply for the low class demographic as well, that all our customers are low class and they are not staying so we have to try something new. But I would disagree there. The data story I was able to tell with some additional analysis was that low class customers who were able to pay a low price for their services almost never left. The ones who left were the low class people who were paying too much. So instead of targeting a hard to reach mid-upper class market, I recommended actions around sales tactics and controls to get a lot of new customers by giving people exactly what they need and nothing more. Losing out on a short term financial gain in order to get a lot more customers for a longer time.

I adapted my presentation to my audience of hypothetical executives by not speaking in too technical of terms as I shared with them my findings. I did still make sure that they knew the backstory and context of this project (ie who I was, why I was there, and where I got all this data from). But I did not attempt to over explain anything with complex math. I simply let the numbers do the talking, and even threw in some pictures to help the numbers make more sense.

This is how I adapted my presentation for the executives specifically. But really the presentation I gave could have been understood by anyone. Since I included background elements like where I got the data from and what most of the raw metrics meant. And I spoke about the difficulties of churn and suggested ways to mitigate it, really anyone could watch that presentation and have a good understanding of what I was talking about. This aspect of the presentation is often incredibly useful as more often than not executives are out of touch with some of the company's problems, and proposing a solution that they don't know anything about could often just make things worse.

This is why I presented my data like a story and I used elements of storytelling like setting up a background (the company struggles with churn), establishing a scene (I have these two sources of data that do these two different things), and mounting a conflict (we can defeat churn). Displaying it this way makes it seem like a mission to be accomplished which helps them listen and understand better, and it makes a call to action more exciting than just another request in someone's inbox too, making real change more likely.

Web resources used on this project

Apply filters to multiple worksheets. Tableau. (n.d.-a).
  https://help.tableau.com/current/pro/desktop/en-us/filtering_global.htm#

The best charts for color blind viewers: Blog. Datylon. (n.d.).
  https://www.datylon.com/blog/data-visualization-for-colorblind-readers#

Bureau, U. S. C. (n.d.-a). *Median Income By Zip code*. Explore census data.
  https://data.census.gov./table/ACSST5Y2022.S1901?q=median+income&g=010XX00US
  %248600000

Bureau, U. S. C. (n.d.-b). *Total Population*. Explore census data.
  https://data.census.gov/table/DECENNIALDHC2020.P1?g=010XX00US%248600000

Create bins from a continuous measure. Tableau. (n.d.-b).
  https://help.tableau.com/current/pro/desktop/en-us/calculations_bins.htm

*Display the total average as a reference line on a bar chart showing the top n average sales in
  each subcategory: Tableau Software*. Tableau. (n.d.-c).
  https://kb.tableau.com/articles/HowTo/display-the-total-average-as-a-reference-line-on-a-
  bar-chart-showing-the-top-n-average-sales-in-each-subcategory

R/datasets on reddit: How to obtain median income data for ZIP codes. Reddit.com. (n.d.).
  https://www.reddit.com/r/datasets/comments/hixfeo/how_to_obtain_median_income_dat
  a_for_zip_codes/