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 can be viewed through a very convenient accessibility feature provided by Tableau known as "Tableau Public"; a cloud based visualization platform 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. The first shows the relationship between income and monthly cost, and the second shows a correlation between tenure and census-reported 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 or population. This will filter the entire dataset, so a selection made on one graph will filter not only that graph but all of the graphs on the dashboard accordingly.
    -   The other 2 graphs are maps of the United States. The first represents regional median reported income as a ratio of the customers' income to the general population income. The second displays the customers per capita (count in every 10,000).

- These graphs can be filtered against by selecting one or multiple states on the map. 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 for a range of customer 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 most likely to come when only one or few filters are applied at a time. Too many strict filters on the data reported in the graphs will quickly become too narrow in scope to properly understand 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

Summary of video:
- Introduction of myself and my background
    - I am new to the company but I have 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 leadership at the company.

- Summary of both chosen data sets and the context in which they occur
    - Collected metrics of customer demographics alongside their churn status from internal company data storage (data collected at customer onboarding).
    - Collected census data for US Zip Code level populations and median incomes.

- Outline of key results from your analysis of the two data sets:
    - Monthly charges across the board were higher for customers who have churned.

- The customers who have not churned generally have below average incomes.
- The company has a pretty even spread of customers in almost all states. Which means that the states with larger populations are underrepresented proportionally.
- Trends seem to indicate customers living in highly populated areas tend to stay with the company's services for longer.

- Two different data representations to serve as supporting evidence for your results:
  - Each of the four graphs charted in the video and in the Tableau dashboard directly support those key results I outlined.

- Summary of actionable insights:
  - Insights in this data indicate that low income people living in high population areas would stay longer with the company and be the least likely demographic to churn. The data also indicates that the more populous regions of the United States are currently underutilized markets that could have huge potential. I believe targeting this demographic directly in sales and marketing efforts would increase overall revenue, despite those customers having lower monthly payments.

Part 3: Reflection Paper

The company 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 is a common business metric used to measure how quickly customers are leaving. The task of my data analysis was to try and understand why customers are leaving so fast and to come up with actionable insights to prevent future customers from leaving quickly, if at all. In other words, my task was to use data to lower the churn rate. My ideas for how I think this can be achieved is expressed in the dashboard I shared. It looks at the demographic attributes of our customers and how some of them indicate which customers are more likely to churn than others. Aligning our target demographic with the demographic least likely to churn will then help to greatly minimize our churn rate.

It also takes a look at these same customers but in a broader perspective, by comparing their demographics to the demographics of non-customers. This allowed me to find characteristics of our non-churning customers that can be used to more broadly target a specific demographic of people with a low likelihood of churning. This broader demographic data comes from census reports. The new data allows us to answer questions like, What economic class do our customers come from? and How populated is the city that they live in? Both of which insights are not available in the initial dataset alone and can be crucial to understanding churn prevention.

A couple of the visual representations in my dashboard indicate 2 key characteristics of our customer base. 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 is a count of customers per capita metric on each state (number of customers per 10,000 people). This gives a heatmap indication of what states most of our customers are located in proportional to the populations there. Executives can use both of these visuals to better understand their customers in terms that are relatable to a wider population.

In order to better apply that understanding of the customer population to actionable insights, other metrics were provided along with controls to interact with the visuals and allow the executives to fine tune the analysis of the data. These controls include a slider for the range of their customers' income. Adjusting this would greatly affect the income ratio graph as customers with higher incomes will compare differently to the populous average than customers with lower incomes. The other control is a filter for churn status which affects all graphs and is crucial for executives to see the effect different factors have had on the churn rate.

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. Using step color sequences instead of gradients helped as well.

In my presentation I was able to show these graphs to executives and stakeholders and assure them that a proper understanding of the data could help us lower the churn rate significantly. I shared the two maps previously mentioned and pointed out that high concentrations of our customers per capita are in the lower populated areas. I then provided further evidence that our customers in highly populated areas tend to last longer with the company than those from lower populated areas. So being less concentrated in those markets

was a big miss on the company's part. Next I pointed out that the economic class of the vast majority of our customers was the lower class. Some additional analysis showed that low class customers who were able to pay a low price for their services almost never left. The ones who did leave were the low class people paying too much. So instead of targeting a hard to reach mid to upper class market, I recommended actions around sales plans to offer more affordable options and open our opportunities to more customers than before.

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 or data terminology. 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. I included background context like where I got the data from and what most of the raw metrics meant, and I spoke about the difficulties of churn so that anyone could watch the presentation and have a good understanding of what I was talking about. This clarity of the presentation is useful as sometimes executives are out of touch with the company's problems, and proposing a solution to a problem they know nothing about could cause confusion.

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 rather than just another request in someone's inbox. All of this makes real change even 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_data_for_zip_codes/