Kiet Nguyen
UID: 001601720
Email: kngu179@wgu.edu

Analysis Report

Contents

Purpose

My objective for creating the dashboard is to show stakeholders the connection between cost of living, income, and churn rate. Important determinants of whether customers are prone to churn include their income and cost of living. The SVP would be interested in this information since they are curious about the traits and motivations of their customers.

The EVP, on the other side, is interested in consumer classification and the impact of demography on various regions. To fulfil the requirements of this stakeholder, I created a dashboard that offers a comprehensive perspective of the entire U.S. market. The stakeholder can dig down into a certain state and view the information at the same time.

The Regional VP is subject to the same requirements. The Regional VP would welcome being able to look at individual states as well as the regional turnover rate. This information is helpful to them because they oversee setting policy and managing operations. They can then use this information to determine which regions require special offers and new features to keep people from leaving.

Tool Selection

Most of the data involved in this analysis came in the form of relational databases. Since the analyst had to interact with this data by cleaning and transforming it, knowledge of SQL and how to use a relational database management system (RDBMS) is required. PostgreSQL is a great free and open-source RDBMS that supports the SQL standard. It is fast, stable, and have many features for both new and advanced users.

Moving beyond the database, Tableau is an excellent data visualization tool that takes cleaned data and tell a compelling story. Tableau is easy to use, even for people who are not technical. It allows analysts to present their findings to stakeholders with ease. The community behind Tableau is a great resource for anyone who wants to learn more about the tool. It is also possible to write custom SQL statements in Tableau for more advanced users.

Data Cleaning Process

The *churn* dataset provided in the database was already clean. There were no null or duplicate values in any of the columns. The external dataset containing the cost-of-living data was also clean.

To prepare the data for analysis, I performed the following process:

1. In pgAdmin 4, use SQL query to create a new table named *cost* in churn dataset for the external dataset.

```
CREATE TABLE cost (
    rank SMALLINT PRIMARY KEY,
    state TEXT,
    index NUMERIC,
    grocery NUMERIC,
    housing NUMERIC,
    utilities NUMERIC,
    transportation NUMERIC,
    health NUMERIC,
    misc NUMERIC
);
```

2. Import data from the CSV file into the new table.

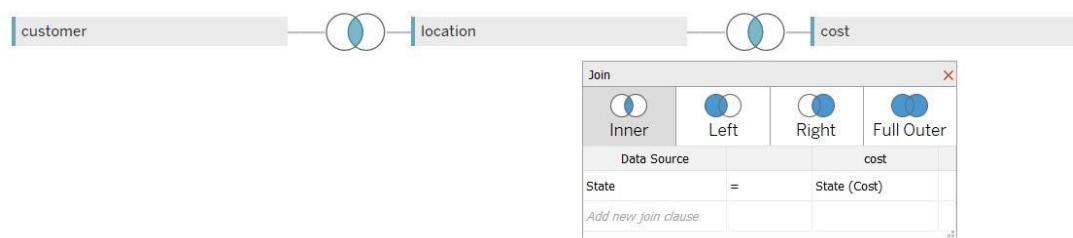3. In Tableau, join three tables together: *customer*, *location*, and *cost*
   a. For *customer* and *location* join, use *location_id* as the join clause.
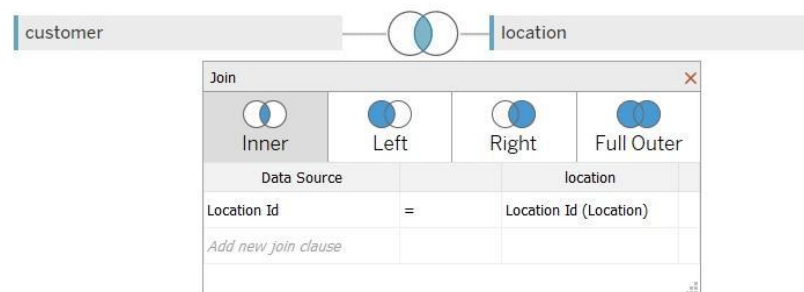
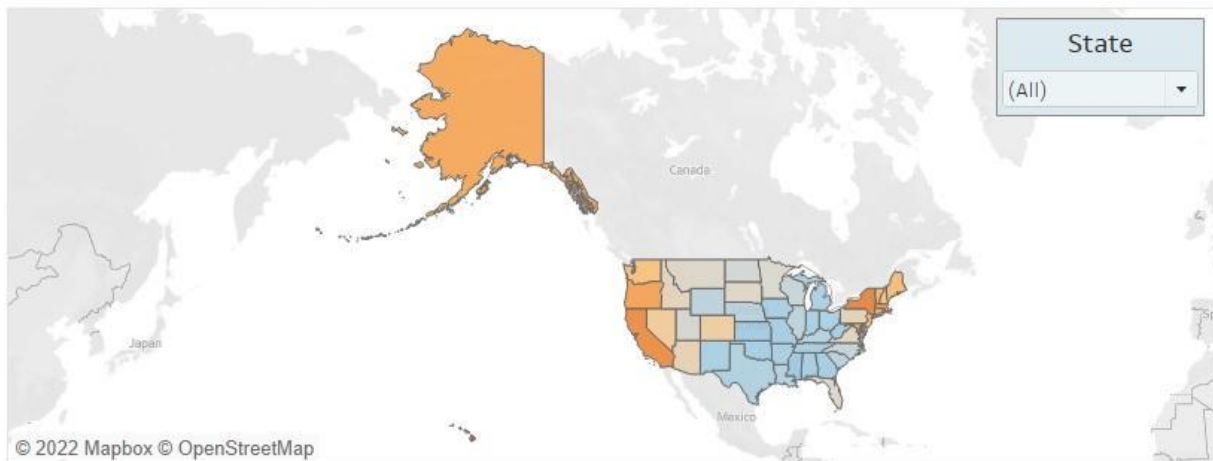   b. For *location* and *cost* join, use *state* as the join clause.

Creating Dashboard

The dashboard consists of four sheets: *Cost of Living*, *Churn Rate*, *Price Difference*, and *Churn Income*.

Cost of Living

Cost of Living Index
*National average is 100

State
(All)

© 2022 Mapbox © OpenStreetMap

The *Cost of Living* sheet contains a map of the United States. This was created by dragging the *state* column of the *cost* table to the sheet. The color was formed by placing the *Index* column also from the *cost* table to the *Color* marks. The lower cost states are blue while the higher cost states are orange.

Churn Rate

State Churn Rate

MS	
NH	
LA	
UT	
DC	
VT	

0%   5%   10%   15%   20%   25%   30%   35%

Average
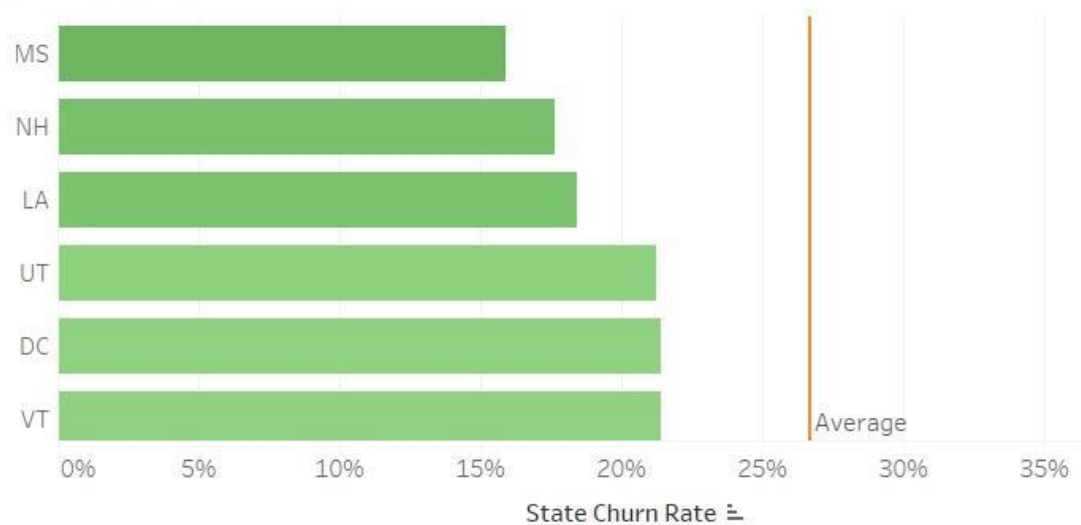
State Churn Rate ≞

The *Churn Rate* sheet contains the churn rate for each state. Two calculated fields were created for this sheet. The first calculated field is *State Churn*:

// Calculate churn customers

```
{ FIXED [State]: SUM(
    IF [Churn] == 'Yes'
    THEN 1
    ELSE 0
    END
) }
```

This calculated field assigns a value of 1 to each churn customer in each state, then sums up the total of those 1's to find the total number of churn customers.

The second calculated field is *State Churn Rate*:

```
{ FIXED [State]:
    SUM([State Churn]) / COUNT([customer])
}
```

This calculated field use the value of *State Churn* and divide it by the total number of customers. It effectively calculates the percentage of churn customer in each state. To create the horizontal bars, *State* variable was placed in the row while the *State Churn Rate* was placed in the columns. The bars were sorted, and an average line was added.

Price Difference
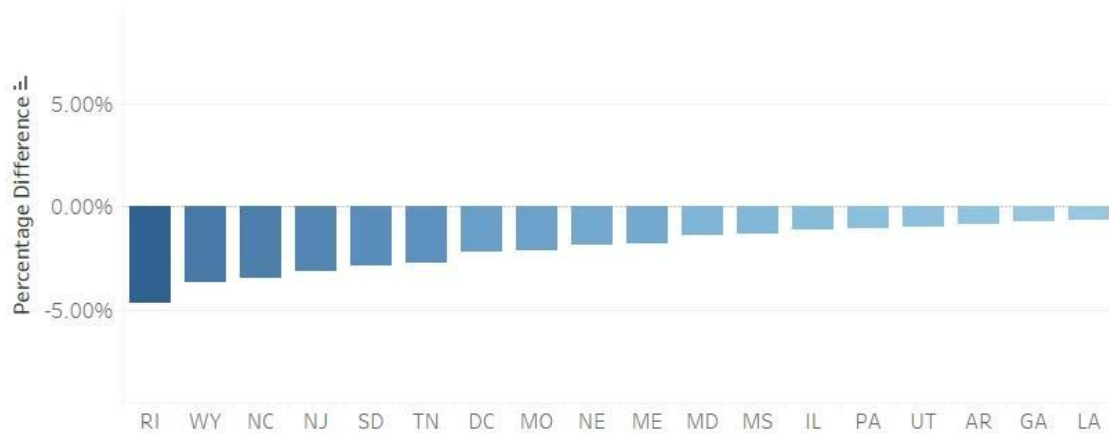
The *Price Difference* sheet contains the difference in monthly payment in each state from the mean. This sheet was created using a calculated field *Price Difference*:

```
(AVG([Monthly Charge]) - 172.6) / AVG([Monthly Charge])
```

The *State* variable was placed in the column, while the *Price Difference* field was placed in the row. To separate the color, the *Price Difference* was also used in the *Color* marks. With blue showing states that paid less than average while orange showing states that paid more.

Churn Income

The *Churn Income* sheet contains the average income of two groups in each state, churn customers and current customers. To create this sheet, the *State* and *Churn* variables were placed in the column, while the average *Income* variable was placed in the row. The *Churn* variable was also used in the *Color* marks to denote the colors of *Yes* and *No* groups for *Churn*.

Analysis Results

According to the analysis, the company was suffering from pricing difference and was losing both customers and profits. After conducting my investigation, I discovered that leaving consumers were paying substantially more than remaining customers. Customers in California, for example, pay 0.97 percent more than the national average. There is a significant difference between these two groups if we split the difference down by churn category. Current customers pay 3.37 percent less than average, whereas churning customers pay 10.9 percent more.

The SVP and EVP are both interested in client traits and how they may be classed across geographies. This meant that my dashboard had to include all the regions where the organization does business. As another matter the SVP requested, income and pricing were my choices for a characteristic that might influence customer behavior.

My dashboard might not provide the Regional VP with any new information because they oversee managing operations. However, my dashboard does let them see how other regions are doing in comparison to their own in terms of performance. Due to their familiarity with the industry, they could use this data to generate their own insights.

Limitations of Analysis

One limitation was that this analysis was performed based on the state level. There is still a large discrepancy in term of churn rate and income between different cities inside a state. A state with a high churn rate in urban cities might have much lower churn rate in more rural area, or vice versa. Income can also vary wildly since customers with extremely high or low income can influence the average.

Another limitation of this analysis was that the data is based on individual customers. Even customers within a specific region can have many different characteristics. A better way to perform churn analysis would be on households instead. This may provide a better picture regarding how and why some households are more likely to churn.

References

Acharya, G., & Whitney, H. (n.d.). *Creating Dashboards and Storytelling with Tableau*. Coursera. Retrieved June 14, 2022, from https://www.coursera.org/learn/dataviz-dashboards

Billen, S., & Sulmont, L. (n.d.). *Analyzing Data in Tableau*. DataCamp. Retrieved June 18, 2022, from https://app.datacamp.com/learn/courses/analyzing-data-in-tableau

Council for Community & Economic Research. (2021). *Cost of Living Data Series*. Missouri Economic Research and Information Center. Retrieved June 22, 2022, from https://meric.mo.gov/data/cost-living-data-series

Sulmont, L., & Fryar, C. (n.d.). Connecting Data in Tableau. DataCamp. Retrieved June 19, 2022, from https://app.datacamp.com/learn/courses/connecting-data-in-tableau

Rose, D. (2017, February 19). *Learning Data Science: Tell Stories with Data*. LinkedIn. Retrieved June 12, 2022, from https://www.linkedin.com/learning/learning-data-science-tell-stories-with-data