D211: Data Analysis

Advanced Data Acquisition

Jonathan R. Schmaltz

The purpose and function of Tableau Storyboard is to provide information to decision makers at the telecommunications company in a way that is easier to access visually. The data sets that were chosen were the US Census Employment data and the telecommunications churn data. The first dashboard compares the Average Employee pay by Gender that can be filtered by state and interaction with customer service resulting in either a churned or retained customer. This allows for a regional analysis of interacting customers. A pay comparison always for decision-makers to use this information to decide if pricing point of services may be resulting in additional churned customers. The second dashboard compares total US Census Employee counts with the telecommunications churned data set population counts. This allows decision makers at the telecommunications company the ability to see what sort of overall market penetration they have and whether they wish to further spend to increase this metric.

The purpose of using Tableau is the ease with which to make a visually compelling story. With a few clicks the data from both data sets can be added to a table and then that table can be converted to a bar graph or pie chart. By adding a filter and connecting them to the two worksheets, the filter can be used to modify both worksheets together. Creating a dashboard, the two worksheets can be dragged and dropped on to the dashboard and adding the filter means that both data sets will change in sync. SQL is used to organize and manage data. By adding Primary and Foreign keys to tables, referential integrity is maintained between the different tables and data points. SQL allows the telecommunications company to collect data and then use split relevant data for analysis.

To start the cleaning process, the first step is to import the churn data into a table called “census-raw.” From there, all rows are cleaned so that only the totals for each state are left. This is done by eliminating the individual NAICS codes, the ethnic, veteran, and race breakdowns, and removing any null values from the states. The final data table is then created named “census” with totals for employees, employee pay, and then the breakdown between male and female pay and employee counts. Each state’s relevant information is added to the final table (census) in the relevant columns. Finally, a primary key is added to the “census” table as the state column. Additional information from the customer table is added into the location table so that when creating the visuals, Tableau can reference that information directly. This information includes each customer’ churn status, sex, income, and tenure.

To create the dashboards, the first process was to connect the churn database to Tableau, and then drag the three tables into a cascading figure with Location as the primary point and then census and customer as the drop legs. Then, 2 work sheets are create first with tables on them. The first table is created with states by measure names and measure values. The measure values are Sum of AVG- femalePay from census, and the SUM of AVG MalePay. AVG is a calculated field made by doing (each individually) [SEX] employees divided by [SEX] payroll and multiplying by 1000 to get the average for both sexes, and because the census table has 1 value per state this results in the average pay for each sex. Then a filter is added so that each value can be broken down by state and then using the show me tab, the table is converted to a vertical bar graph. The process for creating the churn payroll is a little simpler. The rows are added by gender from the location table, and the value is changed from count to average. Then a filter is added and by right clicking on the filter the selected worksheets are added so that Census Payroll and Churn Payroll have the same state filter. Both worksheets are copied and the data for the Census Population is change from average pay to sum of [SEX] employees, and the same is done for Churn population with average being changed to count. The filter is kept linked so that each worksheet still changes together. A dashboard is created for payroll and population comparisons with the worksheets for both churn and census data added. The filter for state and a filter is added to each of the dashboard for churn status, as well. The two visuals are then modified by shape to be easier to read and fit in the storyboard and added numerical values for the color blind as well as legends. A text box is added to help users understand the filters and how the data can be changed. Finally, s storyboard is create with the two dashboards to make it easier for decision makers to navigate the data analysis.

The results for the data analysis show two distinctions. First, the income dashboard shows that the average income for customers is below the average for the state. Second, the population of interacting customers is a fraction of the total employees within a state. The first means that there is a possibility that a customer interaction could have some bearing on personal income so further study into the telecommunications price points could lead to increased positive interactions with customer service. The second means that there is a large population within each state and so further expansion, and customer recruitment drives could improve the overall volume of customers within each state.

One of the biggest limitations of the data presented is that the data only analysis customers that have interacted with customer service proactively. By taking a “backseat” to customer interactions, the data shows when a customer either a, looking to change their service or b, has an issue with their service. The data also doesn’t show what the net result of the interaction with the customer is either than churn or retained. Both of these limitations make it difficult to see an overall picture of the company’s customer satisfaction or the initial reasoning for interacting with customer service.