

D211 – Advanced Data Acquisition

Bernard James Connelly III

Master of Science, Data Analytics, Western Governors University

Dr. Sewell

November 16, 2024

A. How does the dashboard align and assist with business needs?

The primary concern of the dataset presented by the telecommunications company from WGU was to analyze and ultimately reduce customer churn, both to prevent the loss of revenue in raw customers and mitigate the cost of acquiring new customers vs. maintaining previous customers. As an external dataset, education levels in various states across the United States would help identify if those who were higher educated were less or more likely to remain with the company or leave. The dashboard would allow executives to target marketing in specific states or offer sales and other incentives to individual states to reduce overall customer churn. If any overall trends were identified in the individual states, additional resources could be spent to expand the analysis to regional effects or more granular details by narrowing down the education level to the individual county level, which the Kaggle dataset also had available.

B. Why was the business intelligence tool utilized?

Tableau was the tool that was utilized to complete this presentation for a variety of reasons. The simplest of these was the suggestion within the rubric that “it is recommended, but not required, that the business intelligence tool be Tableau.” (WGU, SLM2). More broadly, Tableau’s functionality allowed for effective cleaning and connection of the add-on database to the WGU Telecommunications Customer Churn database and built-in integration with the PGAdmin SQL databases.

C. How was the data cleaned for analysis?

After the dataset was selected from Kaggle and inserted into the virtual lab environment, the first steps taken were to narrow the dataset down to the relevant columns utilized. Tableau prep was initially used to accomplish this task. The columns for State, Area Name, City/Suburb/Town/Rural 2013, and the most recent education information from 2019 were selected for import. The cleaning steps taken were essentially to rename fields to titles that more closely aligned to what was utilized in the WGU dataset, for example, changing the name of the Area name field to county and changing the name of the *City/Suburb/Town/Rural 2013* field to area_type. Additionally, the education columns were renamed to utilize underscores in place of spaces to align more closely with the standard for an SQL database. The final step in Tableau prep was to export the data as a .csv file into the lab's environment.

The remainder of the data cleaning was completed in SQL. The exact annotated code is below:

```
-- Creation of an education table
```

```
CREATE TABLE education
```

```
(
```

```
    state varchar(7),
```

```
    area_type varchar(30),
```

```
    county varchar(100) NOT NULL,
```

```
    less_than_a_high_school_diploma int,
```

```
    high_school_degree_only int,
```

```
    some_college_or_associates_degree int,
```

```
        bachelors_degree_or_higher int,  
        percent_less_than_a_high_school_diploma float,  
        percent_high_school_degree_only float,  
        percent_some_college_or_associates_degree float,  
        percent_bachelors_degree_or_higher float  
    );
```

```
-- Importing data from the Tableau Prep cleaned dataset
```

```
COPY education
```

```
FROM 'C:\Users\Public\cleaned_education.csv'
```

```
DELIMITER ','
```

```
CSV HEADER;
```

```
-- Confirming import worked
```

```
SELECT * FROM education LIMIT(10);
```

```
-- Checking the table for null values
```

```
SELECT *
```

```
FROM education
```

```
WHERE state IS NULL
```

```
OR area_type IS NULL
```

```
OR county IS NULL
```

```
OR less_than_a_high_school_diploma IS NULL
```

OR high_school_degree_only IS NULL
OR some_college_or_associates_degree IS NULL
OR bachelors_degree_or_higher IS NULL
OR percent_less_than_a_high_school_diploma IS NULL
OR percent_high_school_degree_only IS NULL
OR percent_some_college_or_associates_degree IS NULL
OR percent_bachelors_degree_or_higher IS NULL;

-- Removing rows with null values for education

DELETE FROM education

WHERE less_than_a_high_school_diploma IS NULL;

-- Adding values to area_type that are null

UPDATE education

SET area_type = 'State' -- Setting area_type to state for US States + Territories

WHERE area_Type IS NULL;

UPDATE education

SET area_type = 'Country' -- Setting area_type to Country for whole country

WHERE state = 'US';

-- Removing all values which do not align to the State level

DELETE FROM education

```
WHERE area_type != 'State';
```

```
--Checking to confirm cleaning worked
```

```
SELECT * FROM education;
```

```
-- Making State the Primary Key in education
```

```
ALTER TABLE education
```

```
ADD CONSTRAINT pk_state PRIMARY KEY (state);
```

```
--Creating Unique Constraint on State Column in education
```

```
ALTER TABLE education
```

```
ADD CONSTRAINT unique_state unique(State);
```

```
--Creating Foreign Key to Maintain Referential Integrity
```

```
ALTER TABLE location
```

```
ADD CONSTRAINT fk_education_id
```

```
FOREIGN KEY (state)
```

```
REFERENCES education (state);
```

```
--Syntax to check table constraints. Just change table name in quotes
```

```
SELECT tc.constraint_name, tc.constraint_type, kcu.column_name
```

```
FROM information_schema.table_constraints as tc
```

```
JOIN information_schema.key_column_usage as kcu
```

ON tc.constraint_name = kcu.constraint_name

AND tc.table_name = kcu.table_name

WHERE tc.table_name = 'education';

--Joining Education to Location + checking for null values

SELECT *

FROM location as l

LEFT JOIN education AS e

on l.state = e.state

WHERE e.state IS NULL

OR e.area_type IS NULL

OR e.county IS NULL

OR e.less_than_a_high_school_diploma IS NULL

OR e.high_school_degree_only IS NULL

OR e.some_college_or_associates_degree IS NULL

OR e.bachelors_degree_or_higher IS NULL

OR e.percent_less_than_a_high_school_diploma IS NULL

OR e.percent_high_school_degree_only IS NULL

OR e.percent_some_college_or_associates_degree IS NULL

OR e.percent_bachelors_degree_or_higher IS NULL;

D. What steps were taken to create the dashboard?

The dashboard was created in Tableau, each worksheet and the final dashboard were created using the below steps:

“Education Level” - Pie Chart

- Drag State (Education) to Rows Shelf



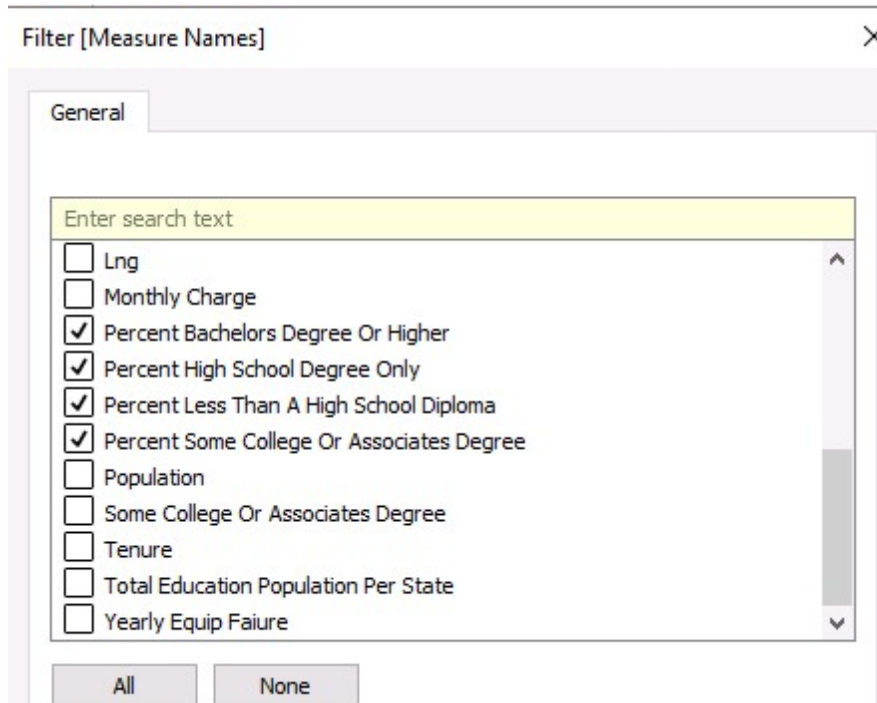
- Select the Pie chart in the marks section



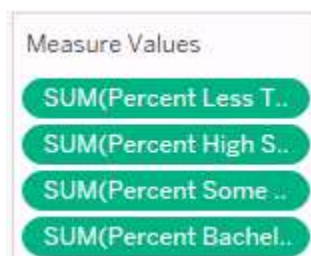
- Drag Measure Names > Measure Values From Tables to Angle on the Marks Card
- Drag Measure Names > Measure Values from Tables to Label on the Marks Card
- Drag Measure Names from the marks card to color



- Edit the filter to include only
 - Percent Less than a High School Diploma
 - Percent High School Degree Only
 - Percent Some College or Associates Degree
 - Percent Bachelor's Degree Or Higher

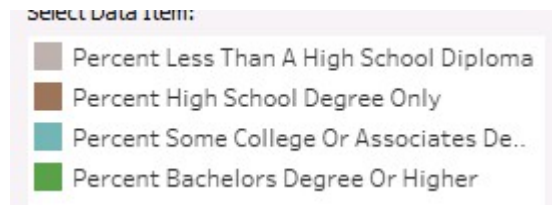


- Drag the measure values into an appropriate logical order under Measure Values:
 - Less than HS
 - HS Only
 - Some College or Associates
 - Bachelor's or higher



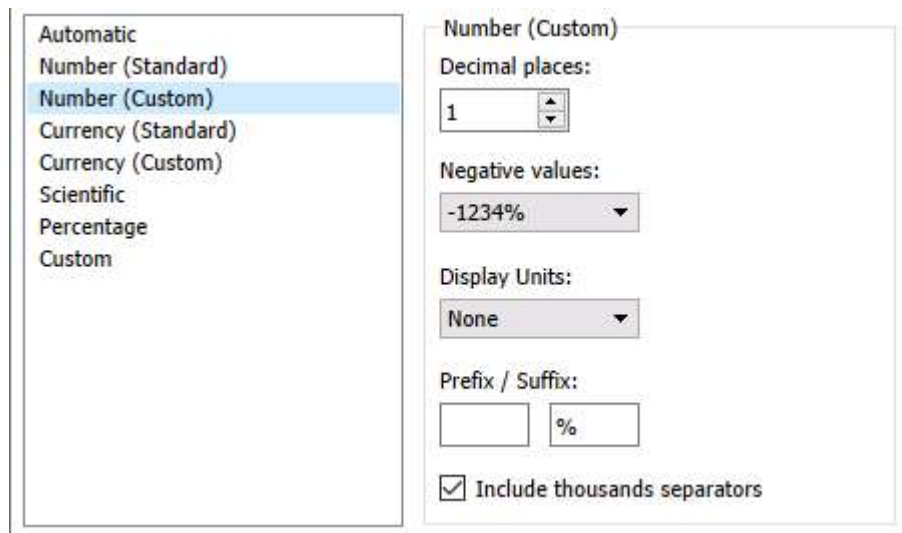
- Select Color on the Marks Card > Edit the colors of the labels:
 - Percent Less than a High School Diploma (Gray)
 - Percent High School Degree Only (Brown)
 - Percent Some College or Associates Degree (Aqua)

- Percent Bachelor's Degree Or Higher (Green)



- Select Measure Values for Tooltips > Format

- Click Numbers Dropdown
- Click Number (Custom)
- Change Decimal Places to 1
- Change Suffix to %

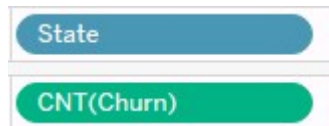


- Right Click State > Hide Field Labels for Rows
- Change the name of this Worksheet to "Education Level"

Education Level

“Churn by State” - Bar Chart

- Drag Location > State to Columns
- Drag Churn to Rows
 - Convert Churn to Measure > Count



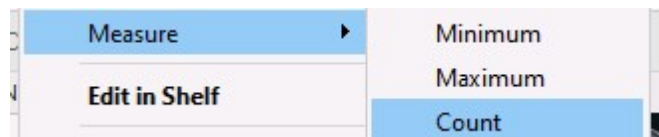
- Right click title
 - Center, bold, and underline the title
- Right-click Column title > Hide field labels for columns
- Change the view to Fit the entire view



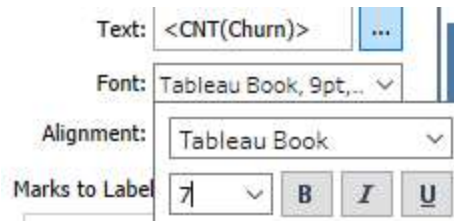
- Click the tooltip in Marks
 - Remove the Count of Churn
 - Insert > Sheets > Education Level

```
State: <State>  
<Sheet name="Education Level" maxwidth="300" maxheight="300"  
filter="<All Fields>">
```

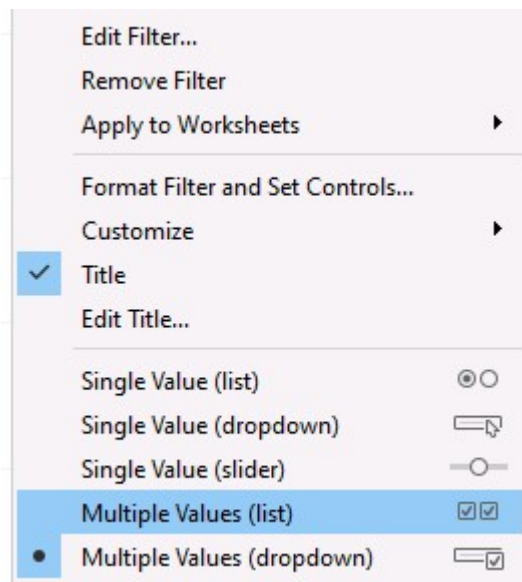
- Drag Churn to Label
 - Change to Measure > Count



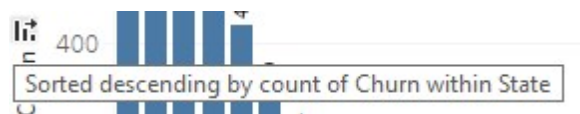
- Change the font size of the label to 7



- Drag Location > State to Filters
 - Select all values
 - Select Show filter
 - Change the format of the filter to Multiple Values (dropdown)



- Drag Churn to Filters, check "Yes," show filter
- Click the edge of "Count of Churn" on the axis to arrange in descending order



- Change the name of this worksheet to "Churn by State"

Churn by State

“Churn Data Table” – Data Table

- Create Calculated Fields:
 - Higher Education
 - Drag Education > Percent Bachelors Degree or Higher + Percent Some College or Associates Degree

Higher Education

[Percent Some College Or Associates Degree]+[Percent Bachelors Degree Or Higher]

- High School or Below
 - Drag Education > Percent High School Degree Only + Percent Less Than A High School Diploma

High School or Below

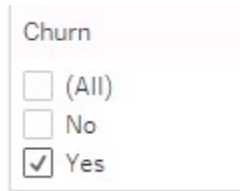
[Percent High School Degree Only]+[Percent Less Than A High School Diploma]

- Drag location > State to rows
- Drag Higher Education and High School or Below calculated fields to Columns
- Drag Churn to columns, then convert to Measure > Count

Columns	Measure Names
Rows	State

Measure Values
CNT(Churn)
SUM(High School or ..
SUM(Higher Educati..

- Drag Churn to Filters, select only yes as the default setting
 - Show filter



Churn

☐ (All)

☐ No

☒ Yes

- Sort by count of churn
- Resize and align to clean up, edit aliases to include %
- Hide title
- Change the name of this worksheet to "Churn Data Table"

"Churn / Education Dashboard" - Dashboard

- Select Size - Automatic
- Drag Churn by state onto the Dashboard
- Drag the Churn Data Table onto the Dashboard on the right-hand side
 - Hide title
- Make the Churn filter floating, apply to all worksheets using related data sources, modify to a list in a dropdown, move to the top left of the visualization
- Make the State filter floating, apply to all worksheets using related data sources, move to the left of the Churn filter
- Add a slider of Count of Churn to the worksheet as floating, set to apply to all worksheets using related data sources, and place it in the middle of the visualization.
- Make both the Churn Data Table and Churn by State charts filters for all worksheets
- Drag the Education Level Worksheet onto the dashboard as a floating item
 - Only need the legend here, resize the list of pie charts to be so small as to be unseen, and move the legend itself to the top right corner of the dashboard

- Move to the right of the filters
- Rename the dashboard to "Churn / Education Dashboard"

E. What were the results, and how did they support the purpose and function?

Reviewing the results of the dashboard, it wasn't easy to directly correlate education level and any specific levels of churn within the WGU dataset. Most of the groupings based upon higher education vs. high school education had a small range of variability in the high 50s to mid-60s of percentage points for higher education. There were several outliers, DC and Colorado on the high end of % Higher Education and Puerto Rico and West Virginia on the lower end, which were omitted from the analysis on trends. The states with the highest higher education values were MN, WA, MA, ND, OR and UT. When looking at the list organized by the number of customers churned, these states were dispersed relatively evenly throughout the entire dataset, not clustered in any way to denote a trend. Similarly, when looking at the states with the five highest churn counts vs. the states with the lowest five churn counts, the percentage of those with higher levels of education demonstrated a similar spread, ranging between 58% and 63% for the higher grouping, and 58 and 65% for the lower grouping.

There was a slight shift in relevance when viewing the data through the lens of customers who did not churn, in as much as customers with higher levels of education trended slightly more toward retention, but not so significantly as to make an executive decision based upon the data. Additionally, there was a similar dispersion of results through the dataset with regard to high vs. low levels of higher education when sorting specifically by the count of churn.

F. What were the limitations of the analysis?

A few key factors highlighted the overall limitations of the analysis. To make the best use of the data in the current dashboard, additional demographic information would be helpful to both the customers and the states they live in. For example, the WGU dataset houses 10,000 records, each belonging to an individual customer, whereas the education levels dataset utilizes the population of the entire United States. The differential between the overall counts of both datasets provides a somewhat skewed analysis. Even within the WGU dataset, when making a comparison between a state with high churn, such as California, vs. a state with low churn, such as Delaware, even though Delaware had only 13 customers churn compared to California's 382, Delaware also only had 21 total customers compared to California's 526. This makes raw counts of churn a less valuable statistic than an overall churn percentage. Further, from the perspective of an executive making decisions on this dataset, the overall population of California is much higher than that of Delaware. This means deciding where to spend more resources to reduce churn may be more effective overall if the population and, thus, potential customers reached are considered.

A more granular approach to the education levels is another factor to consider when looking at limitations. If details were mapped at a lower level, such as at the county level instead of the state level, additional insights could be gleaned from the data. This is particularly true when a company tries to target its resources for the most effective retention method. Grouping the data regionally could also provide additional understanding, as the education levels of the southern United States vs the Northern United States may produce more coherent results. A final point would be adding further

details to compare the education levels instead of just looking at churn levels. Overall education could result in higher or lower cases of customers producing additional services or being willing to pay more for the service, which could result in increased profits for the company if considered.

G. Web Sources & Citation

Western Governor's University. *SLM2 — SLM2 Task 1: Data Analysis*.

<https://tasks.wgu.edu/student/012300379/course/34320018/task/4303/overview>

Telecommunications Churn Dataset. Western Governor's University. Data Analytics Department.

Bauman, Val. (2021) *USA Unemployment & Education Level*. Retrieved 11/09/2024.

<https://www.kaggle.com/datasets/valbauman/student-engagement-online-learning-supplement/data>