**Reflection paper: Advance data acquisition**

D211

By Josue Gonzalez

**A1: Data Sets**

The dashboard containing my data analysis is attached to this submission. I would’ve published it to Tableau Public, but that appears to not be an option for this class. This data analysis is identical to the one performed in D210. Thus it uses materials generated for that project, though it accomplishes this through a different means to satisfy the requirements of this particular project for D211. Again, two

datasets are used for this analysis:

* The WGU Churn dataset, provided by WGU for this program
* The 2023 census states median income

A2: Dashboard Installation

The dashboard is provided in a .twbx format, which can be opened with Tableau Reader. Because

it is not clear to me what evaluators are or are not able to do with the virtual machines, I have included

instructions using either the D211.zip file submitted alongside this report (if the evaluator can easily

transfer this file to the virtual machine) as well as a link to an identical version of D211.zip that can be

downloaded via one drive link provided to the virtual machine. Which process is used is up to the evaluator.

To do this on the Labs on Demand virtual machine, the evaluator must do the following:

* IF POSSIBLE: Take the D211.zip file included alongside this submission, and open File
* Explorer in the virtual machine, navigating to ‘C:\Users\Public\Downloads’ and pasting the
* provided D211.zip file to this folder.
* ALTERNATIVELY: click this link [D211 JG resources files.zip](https://westerngovernorsuniversity-my.sharepoint.com/:u:/g/personal/jgo1435_wgu_edu/EYY4fc3w7BxGtwJDr_bi_v4BkV3ZbDfQvXFx3yknShhDvQ?e=7pKPDd)
* Click the download icon in the top right of the screen to download the file. Open File Explorer
* and navigate to the Downloads folder where this file was saved. Right Click on D211.zip to cut
* the file, and then navigate to ‘C:\Users\Public\Downloads’, where you’ll paste the file.
* Right click on D211.zip and select “Extract All…”. Extract the files to
* ‘C:\Users\Public\Downloads’.
* On the desktop, double click on pgAdmin to open it. In the left side navigation pane, navigate
* down to the “churn” database (one level below “Databases”).
* Right click on the “churn” database and select Query Tool.
* When the Query Tool panel opens on the right side of the pgAdmin window, select the “Open
* File” button (second button from left).
* In the Select File dialog box, navigate to ‘C:\Users\Public\Downloads’ and select the SQL
* Commands(1).txt file.
* It may be necessary to change the bottom-right “Format” drop-down menu to reflect “All
* Files”, rather than just SQL files.
* Once the Query Tool has imported the contents of the SQL Commands.txt file, hit the Play button
* in the top right corner to execute the query.
* Once the query is complete (this may take several seconds), close pgAdmin and return to the
* desktop.
* Double click on Tableau 2021.4 to open Tableau Desktop.
* At the top left, click File > Open and navigate to ‘C:\Users\Public\Downloads’. Select the
* D211 finished workbook.twbx file to open it.
* When prompted to sign in to complete the connection to a server, use the username “postgres” and the password “Passw0rd!”

If all works you should be able to interact with the dashboard would be“installed”.

**3.Dashboard Usage**

After opening the provided dashboard, I recommend navigating to the **"Churn by State"** tab at the bottom of the screen. This can be made full-screen by selecting the projector icon in the upper-right corner. The dashboard is divided into multiple visualizations to help stakeholders understand the relationship between customer churn, service usage, and regional factors.

The **"Churn by State"** visualization on the left is a bar chart that displays the count of churned versus non-churned customers across different states. Each state is represented with two bars: one for customers who churned and one for those who didn’t. The blue bars represent customers who stayed, while the orange bars indicate those who churned. You can interact with this chart by clicking on a specific state to filter the data for that state across the dashboard.

On the right side, the **"Bandwidth Usage by State and Median Income"** tree map provides a visual representation of customer bandwidth usage, broken down by state and their corresponding median income bins. Each block represents a state, with the size and color indicating the total bandwidth usage and median income. By hovering over each block, you can see detailed information about that state’s bandwidth usage and median income. This helps stakeholders assess whether higher bandwidth usage correlates with specific income ranges or regions.

The **"KPI Breakdown by State and Median Income"** tab provides a more detailed analysis of churn rates and tenure across different states and income ranges. At the top, a scatter plot visualizes the relationship between churn rate and customer tenure, with points representing different states. The size of each point correlates with customer tenure, while the position shows churn rate. The filter controls on the right allow you to adjust the analysis based on **Gender**, **State**, **Median Income**, and **Customer Tenure**, providing a dynamic and interactive view of the data.

Below, the **"Customer Tenure Value"** chart uses bubbles to represent customer tenure for each state, further breaking down how long customers in specific regions stay with the company. By selecting specific filters or interacting with these visuals, you can tailor the analysis to answer specific questions about regional or demographic churn behavior.

Each tab offers a comprehensive and interactive way to explore the factors driving customer churn, giving stakeholders powerful tools to identify and act on key insights.

**4. SQL Code**

-- Step 1: Replace NULL values in the 'Children' column with 0

UPDATE customer

SET Children = 0

WHERE Children IS NULL;

-- Step 2: Replace NULL values in the 'Age' column with the median age

WITH MedianAge AS (

SELECT PERCENTILE\_CONT(0.5) WITHIN GROUP (ORDER BY Age) AS median\_age

FROM customer

WHERE Age IS NOT NULL

)

UPDATE customer

SET Age = (SELECT median\_age FROM MedianAge)

WHERE Age IS NULL;

WITH MedianIncome AS (

SELECT PERCENTILE\_CONT(0.5) WITHIN GROUP (ORDER BY Income) AS median\_income

FROM customer

WHERE Income IS NOT NULL

)

UPDATE customer

SET Income = (SELECT median\_income FROM MedianIncome)

WHERE Income IS NULL;

-- Step 4: Verify the update

SELECT state, city, county, zip FROM location LIMIT 10;

-- Step 4: Replace NULL values in 'Marital' and 'Gender' with 'Unknown'

UPDATE customer

SET Marital = 'Unknown'

WHERE Marital IS NULL;

UPDATE customer

SET Gender = 'Unknown'

WHERE Gender IS NULL;

-- Step 5: Replace NULL values in 'Churn' with 'Unknown'

UPDATE customer

SET Churn = 'Unknown'

WHERE Churn IS NULL;

-- Step 6: Replace NULL values in numeric fields like 'Outage\_sec\_perweek' and 'Yearly\_equip\_failure' with 0

UPDATE customer

SET outage\_sec\_week = 0

WHERE outage\_sec\_week IS NULL;

UPDATE customer

SET yearly\_equip\_faiure = 0

WHERE yearly\_equip\_faiure IS NULL;

-- Step 7: Replace NULL in 'Techie' with 'Unknown'

UPDATE customer

SET techie = 'Unknown'

WHERE techie IS NULL;

CREATE TABLE state\_abbreviation\_mapping (

state\_abbr VARCHAR(2) PRIMARY KEY,

state\_full VARCHAR(255)

);

-- Insert state abbreviations and their corresponding full names

INSERT INTO state\_abbreviation\_mapping (state\_abbr, state\_full)

VALUES

('AL', 'Alabama'), ('AK', 'Alaska'), ('AZ', 'Arizona'), ('AR', 'Arkansas'),

('CA', 'California'), ('CO', 'Colorado'), ('CT', 'Connecticut'), ('DE', 'Delaware'),

('FL', 'Florida'), ('GA', 'Georgia'), ('HI', 'Hawaii'), ('ID', 'Idaho'),

('IL', 'Illinois'), ('IN', 'Indiana'), ('IA', 'Iowa'), ('KS', 'Kansas'),

('KY', 'Kentucky'), ('LA', 'Louisiana'), ('ME', 'Maine'), ('MD', 'Maryland'),

('MA', 'Massachusetts'), ('MI', 'Michigan'), ('MN', 'Minnesota'), ('MS', 'Mississippi'),

('MO', 'Missouri'), ('MT', 'Montana'), ('NE', 'Nebraska'), ('NV', 'Nevada'),

('NH', 'New Hampshire'), ('NJ', 'New Jersey'), ('NM', 'New Mexico'), ('NY', 'New York'),

('NC', 'North Carolina'), ('ND', 'North Dakota'), ('OH', 'Ohio'), ('OK', 'Oklahoma'),

('OR', 'Oregon'), ('PA', 'Pennsylvania'), ('RI', 'Rhode Island'), ('SC', 'South Carolina'),

('SD', 'South Dakota'), ('TN', 'Tennessee'), ('TX', 'Texas'), ('UT', 'Utah'),

('VT', 'Vermont'), ('VA', 'Virginia'), ('WA', 'Washington'), ('WV', 'West Virginia'),

('WI', 'Wisconsin'), ('WY', 'Wyoming'), (‘DC’, ‘District of Columbia’), (‘PR’, ‘Puerto Rico’);

-- Update the 'location' table by replacing abbreviations with full state names

UPDATE location

SET state = state\_abbreviation\_mapping.state\_full

FROM state\_abbreviation\_mapping

WHERE location.state = state\_abbreviation\_mapping.state\_abbr;

-- Verify that the updates took place

SELECT state, city, county, zip FROM location LIMIT 10;

-- Step 1: Modify the table to accept state codes instead of abbreviations

CREATE TABLE state\_median\_income (

state\_code INT, -- Use state codes instead of abbreviations

state\_full VARCHAR(255), -- Full state name

median\_income FLOAT -- Median income

);

-- Step 2: Import CSV with state codes instead of abbreviations

COPY state\_median\_income(median\_income, state\_full, state\_code)

FROM 'C:/Users/Public/Downloads/state\_median\_income.csv'

DELIMITER ',' CSV HEADER;

SELECT

location.state AS state\_from\_location,

state\_median\_income.median\_income

FROM

location

JOIN

state\_median\_income

ON

location.state = state\_median\_income.state\_full;

**B. Panopto Presentation**

[D212 support files.zip](https://westerngovernorsuniversity-my.sharepoint.com/:u:/g/personal/jgo1435_wgu_edu/EWNAk_MfIWJPseqp_LPO6RQBK_FeuYgfTkChiM2ilhf3YA?e=dtco7E)

**Part 3: Data Exploration and Analysis Report**

**1. Purpose and Function of the Dashboard**

The purpose of the customer churn dashboard is to provide insights into customer retention and churn patterns across different demographic and service-based factors. The dashboard addresses the needs of stakeholders, such as the marketing and customer service teams, who require a clear and actionable understanding of which factors are driving customer churn. By visualizing churn data by state, income, and service usage, the dashboard helps these teams identify at-risk customer segments and develop strategies to reduce churn.

The functionality of the dashboard allows users to interact with the data dynamically, filtering by state, income, and customer demographics, which enables stakeholders to explore specific regions or segments that may require targeted interventions. The ability to visualize customer tenure and bandwidth usage also helps identify correlations between service consumption and churn, further assisting teams in prioritizing areas for action.

**2. Justification for Business Intelligence Tool Selection**

Tableau was chosen as the business intelligence tool for several reasons:

* **Ease of Use**: Tableau’s user-friendly interface allows for the creation of powerful visualizations without requiring extensive programming knowledge. This was critical in developing an interactive and accessible dashboard for stakeholders who may not have a technical background.
* **Data Integration**: Tableau integrates seamlessly with the PostgreSQL database used for this project, allowing real-time access to the churn data. This integration ensures that any updates to the database are immediately reflected in the dashboard.
* **Interactivity**: The ability to apply filters and interact with visualizations is one of Tableau’s strengths. Stakeholders can filter the data by state, income, or tenure, making it easy to explore different dimensions and identify patterns in churn behavior. This interactivity is vital for data-driven decision-making.
* **Advanced Visualizations**: Tableau offers a wide range of visualization options, which made it possible to present complex data in a visually engaging and easy-to-understand format, such as bar charts, tree maps, and scatter plots.

**3. Data Cleaning and Preparation**

Several steps were taken to clean and prepare the data for analysis:

* **Handling Missing Values**: Initial exploration revealed missing values in key columns such as 'Age,' 'Income,' and 'Customer Satisfaction.' These were addressed by replacing NULL values with median values for numerical fields (e.g., median income) and using default values for categorical fields (e.g., 'Unknown' for gender or marital status).
* **Standardizing State Data**: The customer dataset included a mix of state abbreviations and full state names. A SQL query was used to update the state column in the location table to ensure consistency by converting all state abbreviations to full names. This step was essential for linking customer data with external datasets, such as state median income data.
* **Joining Datasets**: The customer churn data was joined with external state-level data, such as median income. This was done using SQL joins, ensuring that the income data could be used in conjunction with the churn data for deeper analysis of demographic and geographic factors.
* **Preparing Service Usage Data**: Service usage data, such as bandwidth consumption and monthly charges, was normalized to ensure it could be effectively analyzed alongside customer demographics.

**4. Dashboard Creation Process**

Following are the directions for creating the dashboards used in my analysis, including the worksheets that were imported into the dashboards:

Churn by State:

1. Create a bar chart to visualize churn rates by state:
   * Drag the State field to Columns.
   * Drag the Churn field to Color and Rows, and set the measure to Count.
   * Adjust the colors so that "No" churn is displayed in blue and "Yes" churn is displayed in orange for clear contrast.
   * Change the title to "Churn by State", using Tableau Bold, and center align it.
   * Right-click on State and Churn and apply filters to allow users to interact with the data by filtering states and churn statuses.
   * Adjust the tooltips to display the Number of Churned Customers: <CNT(Churn)> customers churned in <State>.
   * Set the visualization to Fit Width for better screen adaptability.

Bandwidth Usage by State and Median Income (Tree Map):

1. Create a tree map to compare bandwidth usage across states, factoring in median income:
   * Drag the Bandwidth Usage field to Size and State to Color.
   * Drag the Median Income field into Label and use the bins option to categorize income into ranges (e.g., $40K, $60K, $80K).
   * Adjust the color palette to differentiate income ranges. Higher income categories are colored darker to signify higher bandwidth usage.
   * Edit the title to "Bandwidth Usage by State and Median Income", using Tableau Bold and center align it.
   * Adjust the tooltips to display Bandwidth Usage and Median Income: <Bandwidth Usage> in <State> with Median Income <Median Income>.
   * Set to Fit Width for best display.

Churn Rate KPI by State:

1. Create a scatter plot for churn rate and customer tenure:
   * Drag Churn Rate to Columns and Customer Tenure to Rows.
   * Drag State to Detail and Churn Rate to Size to size the dots according to churn rate.
   * Filter the scatter plot to show different States, allowing users to interact with each state’s churn metrics.
   * Adjust the title to "Churn Rate KPI Breakdown by State", using Tableau Bold and center align.
   * Modify the tooltip to show churn rate and tenure details for each state: <State> has a churn rate of <Churn Rate> and a tenure of <Tenure>.
   * Set to Fit Width for responsive design.

Customer Tenure Value:

1. Create a bubble chart to display customer tenure:
   * Drag Customer Tenure to Size and State to Color.
   * Use circles to represent tenure by state, where larger bubbles indicate longer tenure.
   * Adjust the color palette for states to clearly differentiate each region.
   * Set the title to "Customer Tenure by State", using Tableau Bold and center align it.
   * Edit tooltips to show the exact tenure value for each state: <Customer Tenure> tenure for <State> customers.
   * Set to Fit Width for proper visualization on any screen size.

Filters:

1. Add filters for interactivity:
   * Right-click the State field and add it as a filter, allowing users to select one or more states to focus on.
   * Right-click the Churn field and add it as a filter, allowing users to toggle between customers who churned and those who stayed.
   * Add a Median Income (bin) filter to allow users to explore bandwidth usage by income ranges.
   * Right-click the Customer Tenure field and add it as a filter to allow users to analyze churn based on how long customers have been with the service.
   * Set the filters to Apply to all using this data source for uniform functionality across all tabs.

**DASHBOARD 1: "Churn by State and Bandwidth Usage"**

1. **Import Worksheets**:
   * **Churn by State**: Drag the **Churn by State** worksheet to the left side of the dashboard.
   * **Bandwidth Usage by State and Median Income**: Drag this worksheet to the right side of the dashboard, adjusting the size of both charts for a clean, even look.
   * **Churn Rate KPI**: Import the **Churn Rate KPI** visualization into the lower portion of the dashboard if needed for additional KPI insights.
2. **Floating Filters**:
   * Right-click the **State** filter and select **Floating**. Move it to the empty space below the **Bandwidth Usage** chart for easy access.
   * Similarly, right-click the **Churn** filter and make it floating, placing it above the **Churn by State** chart for a more compact layout.
   * Make the **Income Range** filter floating as well, placing it near the **Bandwidth Usage** chart to allow users to adjust the income bins.
3. **Title and Instructions**:
   * Add a text box at the top of the dashboard with the title **"Churn by State and Bandwidth Usage by Median Income"** using **Tableau Bold** font and center-align it.
   * Insert another text box with directions for users, such as **"Use the filters on the right to explore churn and service usage trends by state and income levels."**
4. **Activate “Use As Filter”**:
   * Right-click the **Churn by State** visualization and select **Use As Filter**. This allows users to click on a state within the chart, automatically filtering the rest of the dashboard to that state's data.
   * Repeat this step for the **Bandwidth Usage by State** visualization to ensure that both charts interact with one another.
   * Ensure that the **Churn Rate KPI** visualization is also set to be used as a filter for cross-dashboard filtering.
5. **Format and Finalize**:
   * Right-click the column headers, format them to **Tableau Medium 12 pt**, and center align the text.
   * For all visualizations, right-click and format the tooltips, so they read: **"<State>: <Churn Rate> churned, <Bandwidth Usage> bandwidth used at <Median Income>."**
   * Set all elements to **Fit to Width** for a responsive dashboard layout that adapts to screen sizes.

**DASHBOARD 2: "KPI Breakdown by State and Median Income"**

1. **Import Worksheets**:
   * **Churn Rate KPI**: Drag the **Churn Rate KPI** worksheet to the upper portion of the dashboard, ensuring it's large enough for clear readability.
   * **Customer Tenure Value**: Place this worksheet below the **KPI** visualization for easy comparison between churn rate and tenure across states.
   * **Age vs Churn KPI** (Optional): If using an additional comparison, drag a related worksheet (like **Customer Tenure by Age**) to the bottom section of the dashboard.
2. **Insert Horizontal Containers**:
   * Add a **Horizontal Container** to the top of the dashboard.
   * Adjust the width and height of the container and drag the **Churn Rate KPI** and **Customer Tenure Value** worksheets into it.
   * Resize the container and worksheets as needed to maintain visual balance.
3. **Legends and Filters**:
   * Import the **State Legend** and **Median Income Legend** as separate worksheets and make them **floating**.
   * Position the **State Legend** next to the **Churn Rate KPI** visualization and the **Median Income Legend** next to the **Customer Tenure Value** visualization, both under the associated plots.
   * Activate **Use As Filter** for both charts so that clicking on a state in either chart updates the entire dashboard.
4. **Titles and Descriptions**:
   * Add a text box with the title **"KPI Breakdown by State and Median Income"** and center-align it in **Tableau Bold**.
   * Insert another text box for instructions: **"Filter by State, Gender, and Median Income using the controls on the right."** Position this near the top of the dashboard.
   * Drag the **State Filter**, **Gender Filter**, and **Income Range Filter** into a vertical container on the right side of the dashboard for user interaction.
5. **Vertical Container for KPIs**:
   * Create a **Vertical Container** on the left side of the dashboard.
   * Drag in the **KPI values** (for instance, the churn rate for each state or customer tenure) into this container, aligning each KPI value with its associated chart.
   * Adjust the spacing between the KPIs and charts using **"Distribute Evenly"** for a consistent look.
6. **Formatting and Finalization**:
   * Right-click the **Axis Titles** for each chart and format them to **Tableau Medium 10 pt**, aligning the text to center.
   * Format the tooltips for each KPI and chart, ensuring they display detailed churn rate and income insights: **"<State> has a churn rate of <Churn Rate> with median income of <Income Range>."**
   * Adjust column borders to resize appropriately, giving adequate space between the charts for readability.
   * Set all elements to **Fit to Width** for better scalability across devices and screen sizes.

**5. Results of the Data Analysis**

The analysis provided several important insights that supported the purpose and function of the dashboard:

* **Geographic Churn Patterns**: The bar charts showed clear patterns of higher churn in certain states, such as Georgia and Alabama, which may warrant targeted retention strategies in these regions.
* **Income and Churn**: The tree map revealed that states with lower median income tend to have higher churn rates, indicating that economic factors may play a role in customer decisions to leave.
* **Service Usage Correlation**: The scatter plots showed that customers with higher bandwidth usage were more likely to churn, possibly due to dissatisfaction with service or higher costs associated with heavier usage.
* **Customer Tenure**: The bubble charts illustrated that churn is more prevalent among newer customers, highlighting the importance of improving the onboarding experience to reduce churn early in the customer lifecycle.

These results directly aligned with the stakeholders' need to identify high-risk customer segments and regions and provided actionable insights for developing churn-reduction strategies.

**6. Limitations of the Data Analysis**

While the analysis provided valuable insights, there are some limitations to consider:

* **Limited Scope of External Data**: The analysis included external state-level income data, but other external factors—such as local competition, service outages, or customer feedback—were not incorporated. Including more external data could provide a fuller picture of the factors contributing to churn.
* **Data Completeness**: Some customer satisfaction metrics, such as detailed reasons for churn, were not available in the dataset. These metrics could provide deeper insights into customer dissatisfaction and better inform retention strategies.
* **Static Income Bins**: The income bins used in the dashboard were static and based on predefined ranges. A more dynamic analysis of income distribution within states could provide more granular insights into the relationship between income and churn.

Despite these limitations, the dashboard provides a strong foundation for stakeholders to begin exploring and addressing customer churn, with the potential for further refinement as more data becomes available.

***D:Sources***

[***https://www.census.gov/library/publications/2024/demo/p60-282.html***](https://www.census.gov/library/publications/2024/demo/p60-282.html) ***you have to make an api key to get a data set***

*my D210 task 1 was used throughout this project.*