**Task 1: Data Analysis**

Alex Rivera

Western Governors University

D211 - Advanced Data Acquisition

October 31, 2023

**A: Data Dashboards**

A copy of the packaged dashboard (D211.twbx) can be found in the ZIP file attached along with the submission.

**A1: Both Data Sets**

Churn\_clean and ACSST1Y2022.S2802 CSVs can be found in the ZIP file attached along with the submission.

**A2: Dashboard Installation**

A copy of all files needed for installation is included in D211 Final Zip.zip is attached with the submission of the project. The steps to install the dashboard are below:

* In the virtual lab environment download the D211 Final Zip file to extract.
  + Extract to: C:\Users\Public\Downloads
  + A screenshot of a computer

    Description automatically generated
* Right-click and Open the D211 SQL text file in the folder.
* Go to the start menu and type “pgAdmin4” and left-click to start the PostgreSQL tool.
* Within pgAdmin4 go to the left pane to select under databases “churn” and right-click Query tool.
  + A screenshot of a computer

    Description automatically generated
* On the right-hand side, the query editor tool will appear where you can copy the contents of the D211 SQL text file opened earlier.
* Highlight all lines of code up to 134 and left-click the execute button to run the query to create the tables and import the data.
* A screenshot of a computer

  Description automatically generated
* Highlight lines 136 to 140 and execute to update the national\_percentage table.
* A screenshot of a computer

  Description automatically generated
* Highlight lines 142 to 146 to update the churn\_clean table.
* A screenshot of a computer

  Description automatically generated
* Highlight lines 148 to 150 and execute to update the national\_percentage table and drop the S2802 intermediary table.
* A screenshot of a computer

  Description automatically generated
* All SQL code needed for installation has been executed.
* Go to the below folder to find the Tableau dashboard file:
  + C:\Users\Public\Downloads
* Right-click and open the D211 Tableau Packaged file to run the dashboard.
* If it prompts you to update, skip the update.
* When it prompts you to edit the connection left click and under password type:
  + Passw0rd!
  + A screenshot of a computer

    Description automatically generated
* Update the data source by clicking update now.
  + A screenshot of a computer

    Description automatically generated
* The dashboard is now installed, got to the National/State AVG dashboard to use the dashboard.
* A screenshot of a computer screen

  Description automatically generated

**A3: Dashboard Navigation**

The interactive dashboard is comprised of a single dashboard called “National/State AVG”. The National/State AVG is comprised of four data representations and two filters on the right for Churn and State. The data representation “Customer National Internet AVG” is found on the top center and displays the following metrics: current customer national internet average, the census national internet average, and the variance from the national average. Second, the data representation “Customer State Internet AVG” is found right below the previous data representation and displays the following metrics: current customer state(selection) internet average, the census national internet average, and the variance from the national average of the state selection. Third, the bottom left is States Ranked by Internet AVG and this ranks all states by customer internet average in descending order. Finally, the bottom right is the Customer Counts by State this is a map representation that highlights customer counts by the size of the highlighted graphic. Churn and State filters on the right-hand impact all three-state data representation, the filters can be combined for deeper insights into the dataset.

**A4: SQL Code**

DROP TABLE IF EXISTS S2802;

DROP TABLE IF EXISTS national\_percentage;

DROP TABLE IF EXISTS churn\_clean;

CREATE TABLE S2802 (

col1 VARCHAR(255),

col2 VARCHAR(255),

col3 VARCHAR(255),

col4 VARCHAR(255),

col5 VARCHAR(255),

col6 VARCHAR(255),

col7 VARCHAR(255),

col8 VARCHAR(255)

);

CREATE TABLE churn\_clean (

col1 VARCHAR(255),

col2 VARCHAR(255),

col3 VARCHAR(255),

col4 VARCHAR(255),

col5 VARCHAR(255),

State VARCHAR(255),

col7 VARCHAR(255),

col8 VARCHAR(255),

col9 VARCHAR(255),

col10 VARCHAR(255),

col11 VARCHAR(255),

col12 VARCHAR(255),

col13 VARCHAR(255),

col14 VARCHAR(255),

col15 VARCHAR(255),

col16 VARCHAR(255),

col17 VARCHAR(255),

col18 VARCHAR(255),

col19 VARCHAR(255),

Churn VARCHAR(255),

col21 VARCHAR(255),

col22 VARCHAR(255),

col23 VARCHAR(255),

col24 VARCHAR(255),

col25 VARCHAR(255),

col26 VARCHAR(255),

col27 VARCHAR(255),

col28 VARCHAR(255),

Internet VARCHAR(255),

col30 VARCHAR(255),

col31 VARCHAR(255),

col32 VARCHAR(255),

col33 VARCHAR(255),

col34 VARCHAR(255),

col35 VARCHAR(255),

col36 VARCHAR(255),

col37 VARCHAR(255),

col38 VARCHAR(255),

col39 VARCHAR(255),

col40 VARCHAR(255),

col41 VARCHAR(255),

col42 VARCHAR(255),

col43 VARCHAR(255),

col44 VARCHAR(255),

col45 VARCHAR(255),

col46 VARCHAR(255),

col47 VARCHAR(255),

col48 VARCHAR(255),

col49 VARCHAR(255),

col50 VARCHAR(255)

);

COPY S2802(col1, col2, col3, col4, col5, col6, col7, col8) from 'C:\Users\Public\Downloads\ACSST1Y2022.S2802-2023-10-09T071926.csv' DELIMITER ',' CSV HEADER;

COPY churn\_clean(col1, col2, col3, col4, col5, State, col7, col8, col9, col10, col11, col12, col13, col14, col15, col16, col17, col18, col19, Churn, col21, col22, col23, col24, col25, col26, col27, col28, Internet, col30, col31, col32, col33, col34, col35, col36, col37, col38, col39, col40, col41, col42, col43, col44, col45, col46, col47, col48, col49, col50) from 'C:\Users\Public\Downloads\churn\_clean.csv' DELIMITER ',' CSV HEADER;

ALTER TABLE churn\_clean

DROP COLUMN col1,

DROP COLUMN col2,

DROP COLUMN col3,

DROP COLUMN col4,

DROP COLUMN col5,

DROP COLUMN col7,

DROP COLUMN col8,

DROP COLUMN col9,

DROP COLUMN col10,

DROP COLUMN col11,

DROP COLUMN col12,

DROP COLUMN col13,

DROP COLUMN col14,

DROP COLUMN col15,

DROP COLUMN col16,

DROP COLUMN col17,

DROP COLUMN col18,

DROP COLUMN col19,

DROP COLUMN col21,

DROP COLUMN col22,

DROP COLUMN col23,

DROP COLUMN col24,

DROP COLUMN col25,

DROP COLUMN col26,

DROP COLUMN col27,

DROP COLUMN col28,

DROP COLUMN col30,

DROP COLUMN col31,

DROP COLUMN col32,

DROP COLUMN col33,

DROP COLUMN col34,

DROP COLUMN col35,

DROP COLUMN col36,

DROP COLUMN col37,

DROP COLUMN col38,

DROP COLUMN col39,

DROP COLUMN col40,

DROP COLUMN col41,

DROP COLUMN col42,

DROP COLUMN col43,

DROP COLUMN col44,

DROP COLUMN col45,

DROP COLUMN col46,

DROP COLUMN col47,

DROP COLUMN col48,

DROP COLUMN col49,

DROP COLUMN col50;

UPDATE "churn\_clean" SET "internet" = 0 WHERE "internet" = 'None';

UPDATE "churn\_clean" SET "internet" = 1 WHERE "internet" = 'DSL';

UPDATE "churn\_clean" SET "internet" = 1 WHERE "internet" = 'Fiber Optic';

-- Tutorial referenced for replace function below (PostgreSQL replace function ,2023)

SELECT CAST(REPLACE(REPLACE(REPLACE("col4", ' ', ''), '%', ''), '\n', '') AS FLOAT)

as national\_percentage

INTO national\_percentage

FROM S2802

WHERE "col1" = 'Total population in households'

LIMIT 1;

ALTER TABLE national\_percentage

ADD COLUMN "National" VARCHAR(255);

UPDATE national\_percentage

SET "National" = 'National'

ALTER TABLE churn\_clean

ADD COLUMN "National" VARCHAR(255);

UPDATE churn\_clean

SET "National" = 'National'

UPDATE "national\_percentage" SET "national\_percentage" = "national\_percentage"/100 WHERE "National" = 'National';

DROP TABLE IF EXISTS S2802;

**B: Panopto Presentation**

https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=6328d695-a1c9-4d11-9949-b0ad00485d78

**C: Written Report**

**C1: Dashboard Alignment**

Task 1 of D210 had a similar rubric task that was referenced for this section (Rivera, 2023). In the data dictionary, three takeaways guided me in the path of building the interactive dashboard in the project (D210 Churn Data Consideration and Dictionary, 2023):

* Given that it costs 10 times more to acquire a new customer than to retain an existing one, customer retention has now become even more important than customer acquisition.
* Build a data dashboard to enable executive leaders to explore the data, identify trends, and compare key metrics.
* Present broad and understandable insights from the data that are relevant to their perspective.

I focused on the internet subscription average as my focus for the dashboard as these are current customers that we can retain at a lower cost and offer additional services. The chosen supplemental dataset provides national internet subscription data for the creation of metrics. I built the interactive dashboard with the executive team's focuses in mind when incorporating the data representations. The interactive controls help focus on areas that their perspective teams might need to dive deep to analyze.

**C2: Business Intelligence Tool**

The Tableau website provides concise reasons to choose its software, “As the market-leading choice for modern business intelligence, the Tableau platform is known for taking any kind of data from almost any system and turning it into actionable insights with speed and ease” (Why choose Tableau? 2023). The Tableau business intelligence tool is incredibly intuitive to use and includes all the connectors to communicate with other data source systems. All data sources were connected within 10 minutes and creating visualizations is natural with the tool having progressed through the MSDA program.

**C3: Data Cleaning**

* Create a table called S2802 with 8 columns to load the census data from the CSV file
* Create a table called churn\_clean with 50 columns to load the churn\_clean data from the CSV file
* Load census data from ACSST1Y2022.S2802 CSV file
* Load churn\_clean data from churn\_clean CSV file
* Drop 47 columns that are not needed for analysis from the churn\_clean table, keep the following columns:
  + State
  + Churn
  + Internet
* In the churn\_clean table update the internet column to 0 where internet = ‘None’
* In the churn\_clean table update the internet column to 1 where internet = ‘DSL’
* In the churn\_clean table update the internet column to 1 where internet = ‘Fiber Optic’
* From the S2802 table select the fourth column where column one is ‘Total population in households’, replace spaces and % cast to float. Add the results into a new table called national\_percentage
* Add column “National” to national\_percentage
* Add column “National” to churn\_clean
* Update the national\_percentage column by dividing by 100 to create a percentage.
* Drop S2802 intermediary table.

**C4: Dashboard Creation**

Tableau coursework materials referenced (Billen & Sulmont, 2023)

**Connect to database and data sources**

1. Open Tableau Desktop
2. Connect to a PostgreSQL server an fill in the following
   1. A screenshot of a computer

      Description automatically generated
   2. Password is **Passw0rd!**
   3. Sign in
3. Drag churn\_clean and national\_percentage tables to create relationship.
   1. A screenshot of a computer

      Description automatically generated
   2. Click on line/relationship to adjust referential integrity and change to “all records match”
   3. A screenshot of a computer

      Description automatically generated
4. Select churn\_table and change datatype of “internet” column to number(whole)
   1. A screenshot of a computer

      Description automatically generated

**Customer National Internet AVG Sheet**

1. Change sheet name to” Customer National Internet AVG”
2. Change sheet fit to “Fit Width”
3. Convert internet variable to measure
4. Right click data pane to create calculated field
   1. Name it “Variance from National AVG”
   2. Formula = (AVG([Internet])-AVG([National Percetnage]))
   3. A screenshot of a computer program

      Description automatically generated
5. Add Measure Names to sheet columns
6. Add Measure Values to Marks Text, selections and measure function below:
   1. A green box with white text

      Description automatically generated
7. Format measure values in the marks pane to change to percentage
   1. A screenshot of a computer

      Description automatically generated
8. Right click measure names under columns and select edit aliases
   1. Avg Internet = “Customer Internet AVG”
   2. Avg National Percentage = “National Internet AVG”
9. Under churn\_clean table grab the churn dimension and move it to the filer
   1. Left click on filter and select show filter
   2. Filter will show on the right and select options and customize deselect show “all” value
10. Right click on the dashboard title to edit/format the title
    1. Highlight <Sheet Name> center and bold it
11. The ribbon at the top select format and alignment to adjust
    1. Sheet area select center for pane
12. The ribbon at the top select format and format to adjust
    1. Worksheet font change to size 12
13. Final output of the dashboard should like the following:

A screenshot of a computer

Description automatically generated

**Customer State Internet AVG Sheet**

1. Duplicate the Customer National Internet AVG dashboard to use as a base
2. Rename sheet to “Customer State Internet AVG”
3. Grab State dimension from churn\_clean table and add it to the filter
   1. Select all values
   2. Show filter
   3. On the right hand side with the filter showing select the options and select multiple values dropdown
4. Final sheet should look like the below:

A screen shot of a computer

Description automatically generated

**States Ranked by Internet AVG Sheet**

1. Create new sheet called “States Ranked by Internet AVG”
2. Add internet measure to the columns
3. Add state dimension to the rows
4. Sort the values in bar chart by ascending order by clicking the following button
   1. A screenshot of a computer

      Description automatically generated
5. Grab State dimension from churn\_clean table and add it to the filter
   1. Select all values
   2. Show filter
   3. On the right hand side with the filter showing select the options and select multiple values dropdown
6. Under churn\_clean table grab the churn dimension and move it to the filer
   1. Left click on filter and select show filter
   2. Filter will show on the right and select options and customize deselect show “all” value
7. Right click on the dashboard title to edit/format the title
   1. Highlight <Sheet Name> and bold it
8. Right click on the axis to edit the axis
   1. Axis Title changed to “Customer Internet AVG”
9. Final sheet should look like the below:

A graph of a number of blue and white lines

Description automatically generated with medium confidence

**Customer Counts by State Sheet**

1. Create new sheet called “Customer Counts by State”
2. To create map:
   1. Add longitude dimension to columns
   2. Add latitude dimension to rows
   3. Add state dimension to detail in marks
3. Add state dimension to size in the marks pane
   1. A screenshot of a computer

      Description automatically generated
   2. Adjust the size by right clicking and adjusting to 50%
4. Adjust the color in the marks pane by right clicking
   1. Select orange and opacity to 76%
   2. A screenshot of a computer

      Description automatically generated
5. Grab State dimension from churn\_clean table and add it to the filter
   1. Select all values
   2. Show filter
   3. On the right hand side with the filter showing select the options and select multiple values dropdown
6. Under churn\_clean table grab the churn dimension and move it to the filer
   1. Left click on filter and select show filter
   2. Filter will show on the right and select options and customize deselect show “all” value
7. Right click on the dashboard title to edit/format the title
   1. Highlight <Sheet Name> and bold it
8. Final sheet should look like the below:

A map of the united states

Description automatically generated

**National/State AVG Dashboard**

1. Create dashboard named “National/State AVG”
2. Change size to:
   1. A black text on a white background

      Description automatically generated
3. Drag Customer National sheet to the top of the dashboard
4. Drag Customer State sheet right below the previous sheet on the dashboard
5. Drag States Ranked by Internet AVG into the bottom left hand corner
6. Drag Customer Counts by state into the bottom righ hand corner
7. On the right hand side remove duplicate churn and state filters
   1. Once removed go to Churn filter options
   2. Apply to worksheets
   3. Selected worksheets
   4. A screenshot of a computer

      Description automatically generated
   5. Go to the state filter options
   6. Apply to worksheets
   7. Selected worksheets
   8. A screenshot of a computer

      Description automatically generated
8. Adjust the vertical height of Customer National Internet AVG by hovering over and reducing the size of the sheet on the dashboard
9. Adjust the vertical height of Customer State Internet AVG by hovering over and reducing the size of the sheet on the dashboard
10. Final complete dashboard should look like the below:

A screenshot of a computer screen

Description automatically generated

**C5: Data Analysis Results**

Task 1 of D210 had a similar rubric task that was referenced for this section (Rivera, 2023).

* The customer national internet average is 78.7%, 14.2% below the national average.
  + All states below the national average

Customer National Internet AVG is made up of aggregate customers with an internet subscription and the variance to the national average. This data representation is made to refresh anytime the data is updated to act as a pulse check of how the company's internet subscription is currently doing. Executives can use this average to determine the company's health and if any strategies need to be put in place to impact the metric.

* Churn No internet average is 77.8%, 15.1% below national average.
  + All states below the national average
* Churn Yes internet average is 81.3%, 11.6% below national average.
  + All states below the national average besides Rhode Island at 100%
  + 100% internet average for churn customers should be reviewed.

States Ranked by Internet AVG is a data representation that breaks down and sorts the Internet average by state in descending order. Executives and Regional VPs can use this data representation to take a deep dive on states that not performing as well as the national internet average.

**C6: Analysis Limitations**

The additional dataset added the insight needed to compare the customer internet average to a national average for comparison, while this creates metrics it does not illustrate trends. A method for collecting results for comparison over time would need to be implemented. The executive team would benefit greatly from being able to analyze an upward or downward trend. Second, the limitation of the census data was only being able to import the national average. Currently, if states are selected in the filter it will still be compared to the national average and not the average of that specific state. An updated dataset with state internet averages would help achieve deep state insights.

**D: Web Sources**

*PostgreSQL replace function*. PostgreSQL Tutorial. (n.d.). <https://www.postgresqltutorial.com/postgresql-string-functions/postgresql-replace/>

*Why choose Tableau?*. Tableau. (n.d.). https://www.tableau.com/why-tableau

**E: Sources**

Billen, S., & Sulmont, L. (n.d.). *Connecting Data in Tableau*. DataCamp. https://app.datacamp.com/learn/courses/connecting-data-in-tableau

Rivera, Alex. (2023). Task 1: Data Dashboard And Storytelling. Unpublished manuscript, Western Governors University.

Sewell, W. (2023, October). *SQL Sunday Postgres D211*. *https://my.wgu.edu/*. Salt Lake City; UT. Retrieved October 21, 2023, from https://westerngovernorsuniversity-my.sharepoint.com/:p:/g/personal/william\_sewell\_wgu\_edu/EQtfWHAWXOxOtRGnIjWRVhMBS39Tsi1wzaJ9P0ohQhxzAA?e=A06I0O.

U.S. Census Bureau. (2022). Types of Internet Subscriptions by Selected Characteristics. American Community Survey, ACS 1-Year Estimates Subject Tables, Table S2802. Retrieved October 18, 2023, from https://data.census.gov/table/ACSST1Y2022.S2802?q=Telephone,+Computer,+and+Internet+Access&moe=false.

Western Governors University. (2023). D210 Churn Data Consideration and Dictionary. Millcreek.