D211 - ADVANCED DATA ACQUISITION

Kimberly Hubacek

STUDENT ID 0012498236

Part A: Data Dashboards

A1. Datasets

Please see the CSV files submitted with my performance assessment for copies of the cleaned datasets used in this project. I used the preloaded WGU Churn dataset as well as the Telco Communications dataset created by IBM solutions downloaded from Kaggle.com. Additionally, there is a zip file that contains my Tableau dashboard, the TelcoChurn.csv file, a text file titled "CleaningChurnData" and a text file titled "CreatingTelcoTable". The text files contain the SQL code that is required to prepare the data in the Labs On Demand virtual machine.

A2: Dashboard Installation Instructions

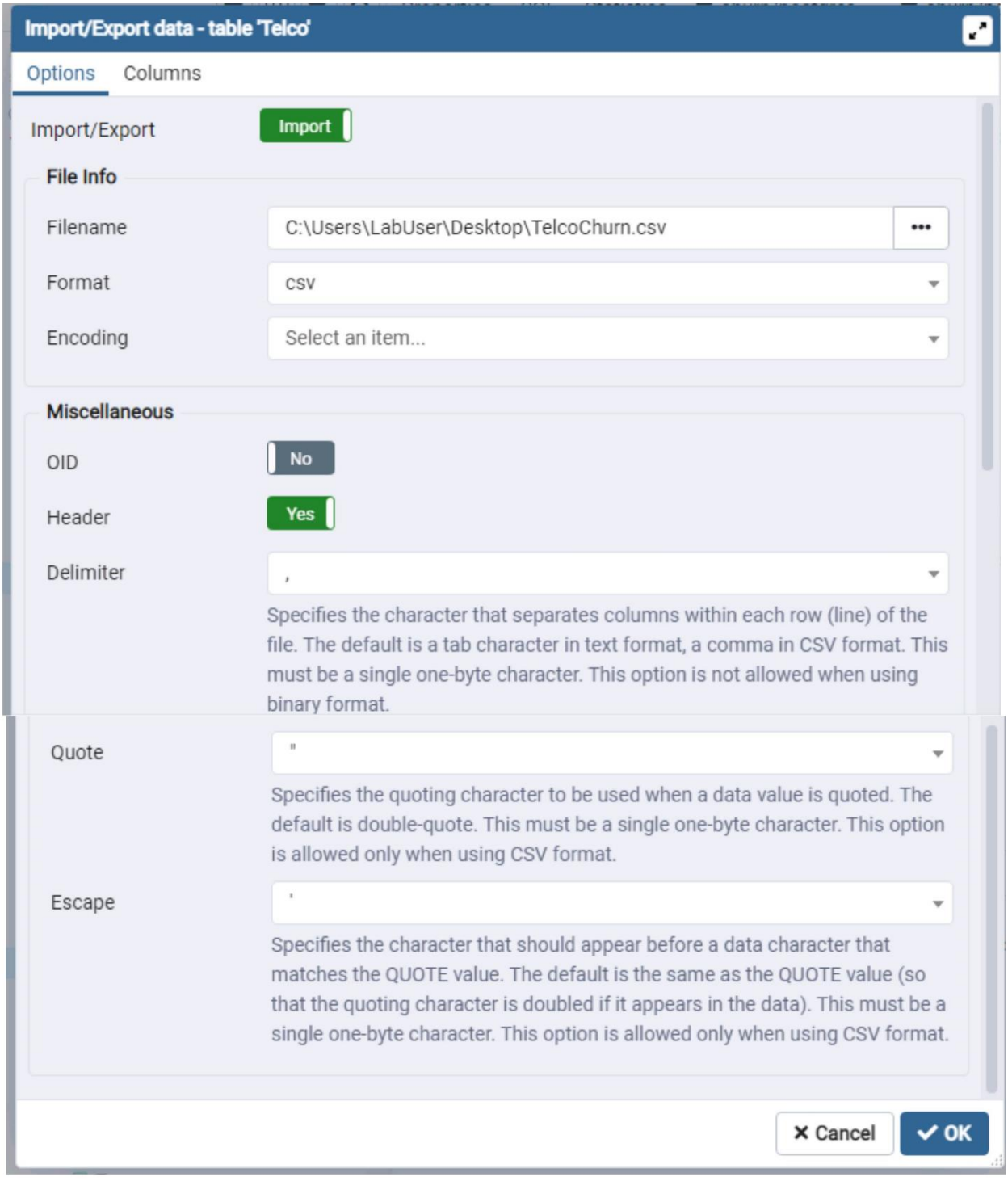
Downloading Files:

- I attached a D211.zip file with my submission. Please email this file to yourself.
- Connect to the Labs On Demand Virtual machine.
- Open Internet Explorer from the menu bar. Navigate to you email provider, sign in, and download the zip file you sent to yourself.
- Click "show in folder" when the file is downloading or navigate to C:\Users\LabUser\Downloads after the file has downloaded to locate the file.
- Copy the file and paste it onto the desktop.
- Right click the file and extract all files onto the desktop.
- Move the TelcoChurn.csv file from the folder to the desktop.

Creating Telco Table and Loading TelcoChurn.csv File:

- Navigate to the pgAdmin4 icon on the desktop and double click it to open it.
- When pgAdmin4 has loaded, navigate to Servers, and click the down arrow.
- Navigate to the Churn database in the left panel and right click it. Select "Query Tool" from the menu that pops up.
- Copy and paste the text from the "**CreatingTelcoTable**" text file that was previously downloaded and extracted into the Query Tool and run the query by clicking on the play button in the query menu.
- Refresh the database by right clicking on the churn database in the lefthand column and selecting "Refresh…" The Telco table will appear under "Schemas" > "Tables".
- Time to import the TelcoChurn.csv file into the Telco table. Right click the new Telco table in the lefthand menu and select "Import/Export…" from the menu that pops up.
- Toggle the slider to "Import."
- Click on the ellipses by filename and select the "TelcoChurn.csv" file. You may need to navigate to the desktop in the Select File menu if PGAdmin does not automatically direct you there to locate the TelcoChurn file.
- Double click on the TelcoChurn.csv file to select it.
- Toggle the header to "Yes".
- Select the comma "," in the Delimiter menu.

- **The final layout should look like this:**



- Click “OK”.

Cleaning the Churn Data:

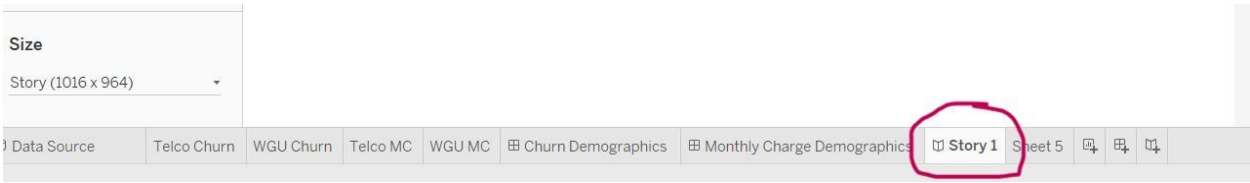
- Navigate to the churn database in the lefthand panel and right-click. Select “Query Tool” from the menu that pops up. Copy and paste the contents from the “**CleaningChurnData**” text file into the Query Tool. Click the pay button to run the query. The Churn data has now been cleaned.

Installing the Dashboard:

- Minimize PGAdmin and double click the white Tableau icon to open the program.
- If you receive an update prompt click “Skip This Update”
- In the upper left corner, click on “File” then “Open”.
- Navigate to the Desktop Folder and then the Unzipped folder featuring the downloaded files you emailed to yourself.
- Click on the “D211Final” Tableau packaged notebook and click “Open”.
- A username and password prompt will pop up. The username should be postgres. In the Password field, type in “Passw0rd!” without the quotation marks and click Sign In.
- The Tableau Story should appear. You are ready to navigate the dashboard.

A3: Dashboard Navigation Instructions

With the dashboard installed, navigate to the bottom menu bar, and select “Story 1” if it has not automatically navigated there:

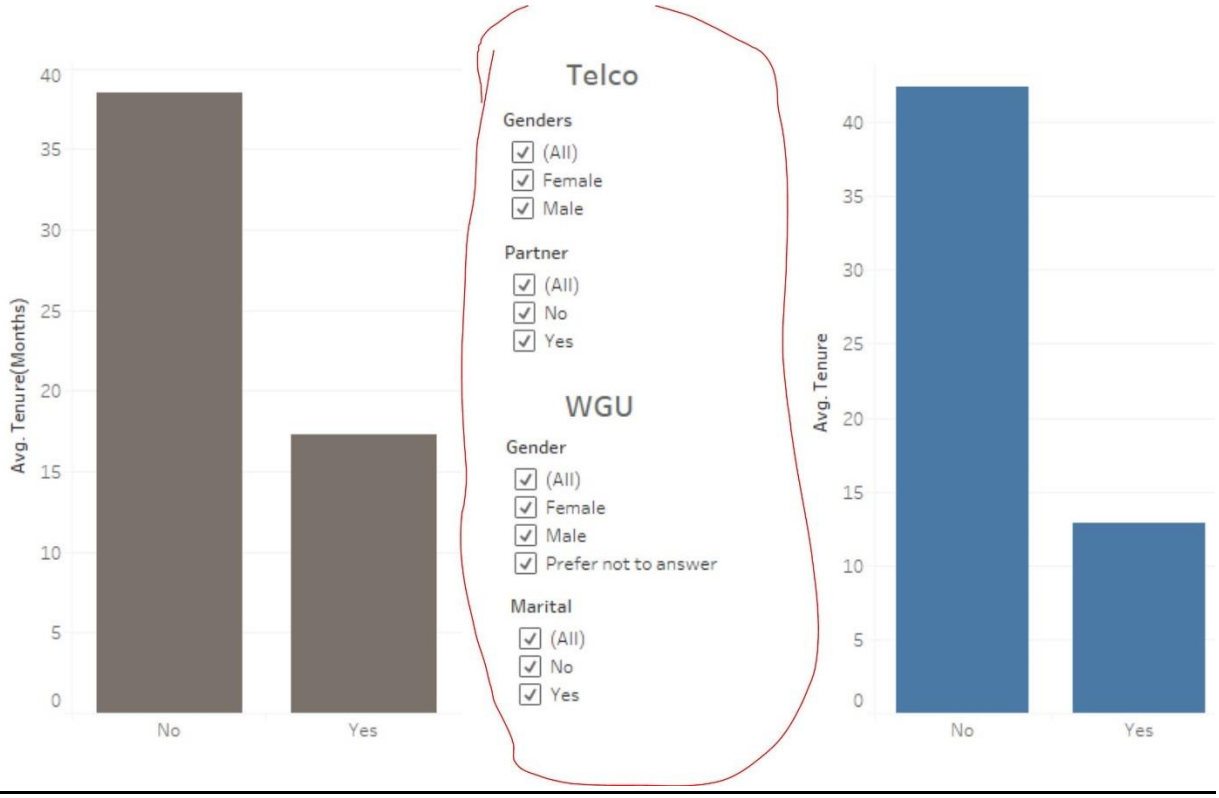


Scroll up and navigate to the “Introduction” panel for brief description of the business scenario that required the dashboard. Following the Introduction, navigate to the “Churn Demographics” page by clicking the right arrow after “Key Takeaways”.

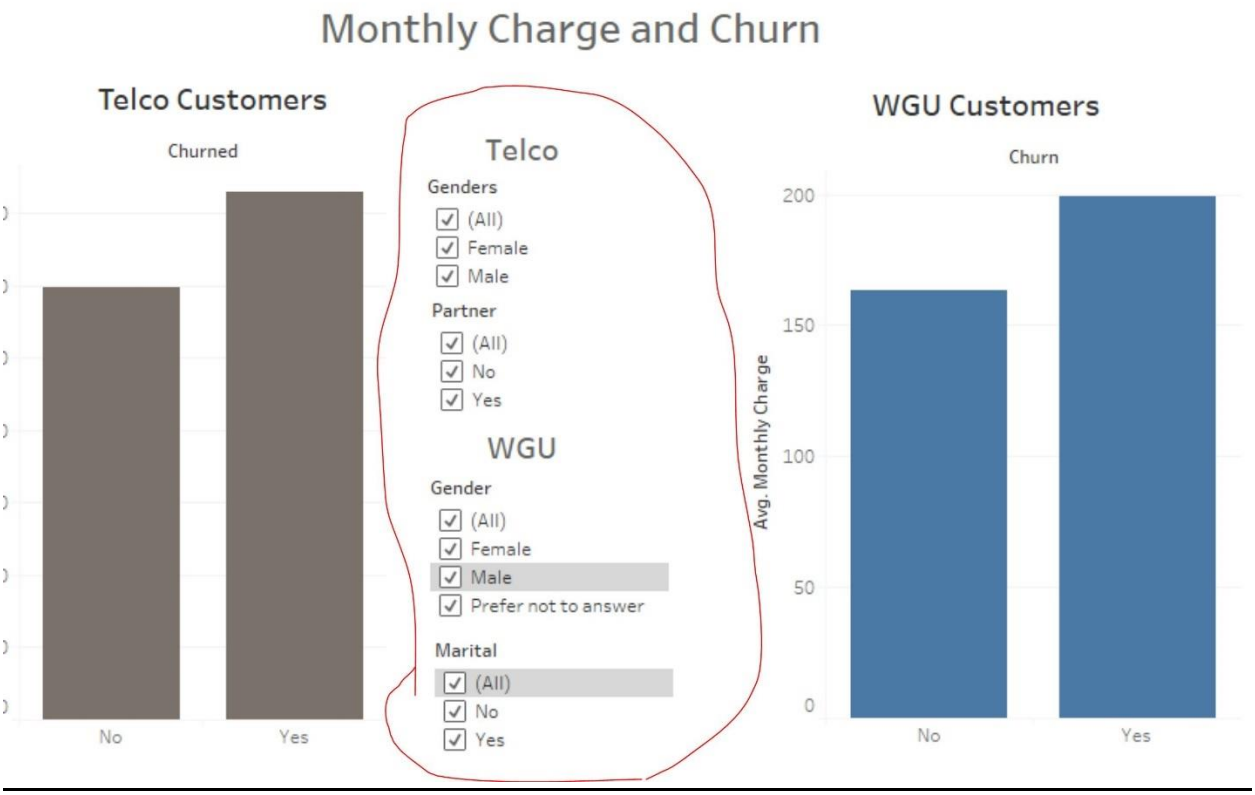
Story 1

Scenario:
WGU Telecomunications has aquired Telco communications in a merger. Management wishes to decrease churn while increasing tenure for both companies. The marketing department is trying to identify demographics for the next marketing campaign offering service discounts for signing long term contracts.

You will see two separate bar charts representing the Telco customer data and WGU customer data with filters in the middle. The audience can interact with the filters to see how the gender and marital status affects the tenure length and churn rates of Telco and WGU customers:



Click the right arrow next to “Key Takeaways” to navigate to the next page. The next page in the story measures the average monthly charge along with churn rates of Telco and WGU customers. The layout is identical to the previous page, with one bar chart representing the Telco customer data, one bar chart representing the WGU customer data and filters in the middle. The audience can interact with the filters in the same way to see how the customers’ monthly charges and churn rates differ between genders and marital statuses:



Click the right arrow next to “Key Takeaways” to go to the final page of the story. The final page of the story communicates key takeaways from the data and actionable insights for the business.

A4: SQL Code

Creating Telco Table:

```
CREATE TABLE public."Telco"
(
    "CustomerID" text,
    "Gender" text,
```

```
    "Partner" text,
    "Dependents" text,
    "Tenure(Months)" numeric,
    "PhoneService" text,
    "MultipleLines" text,
    "InternetService" text,
    "OnlineSecurity" text,
    "OnlineBackup" text,
    "DeviceProtection" text,
    "TechSupport" text,
    "StreamingTV" text,
    "StreamingMovies" text,
    "Contract" text,
    "PaperlessBilling" text,
    "PaymentMethod" text,
    "MonthlyCharges" numeric,
    "Churn" text,
    CONSTRAINT "Telco_pkey" PRIMARY KEY ("CustomerID")
);

ALTER TABLE public."Telco"
OWNER to postgres;
```

Import the Telco CSV into the Telco table:

```
--command " "\\copy public.\"Telco\" (\"CustomerID\", \"Gender\", \"Partner\",
\"Dependents\", \"Tenure(Months)\", \"PhoneService\", \"MultipleLines\",
\"InternetService\", \"OnlineSecurity\", \"OnlineBackup\", \"DeviceProtection\",
\"TechSupport\", \"StreamingTV\", \"StreamingMovies\", \"Contract\",
\"PaperlessBilling\", \"PaymentMethod\", \"MonthlyCharges\", \"Churn\") FROM
'C:/Users/LabUser/Desktop/TELCOC~1.CSV' DELIMITER ',' CSV HEADER QUOTE '\"'
ESCAPE '"';"
```

Cleaning WGU Churn Dataset:

Changing “marital” column in Customer table to "yes/no"

```
UPDATE public.customer
SET marital = 'Yes'
WHERE marital = 'Married';
UPDATE public.customer
SET marital = 'No'
WHERE marital = 'Separated';
```

```
UPDATE public.customer
SET marital = 'No'
WHERE marital = 'Never Married';
UPDATE public.customer
SET marital = 'No'
WHERE marital = 'Divorced';
UPDATE public.customer
SET marital = 'No'
WHERE marital = 'Widowed';
```

Changing “children” column in Customer table from integer to "yes/no"

```
ALTER TABLE public.customer
ALTER COLUMN children
TYPE text;
UPDATE public.customer
SET children = 'No'
WHERE children = '0';
UPDATE public.customer
SET children = 'Yes'
WHERE children = '1';
UPDATE public.customer
SET children = 'Yes'
WHERE children = '2';
UPDATE public.customer
SET children = 'Yes'
WHERE children = '3';
UPDATE public.customer
SET children = 'Yes'
WHERE children = '4';
UPDATE public.customer
SET children = 'Yes'
WHERE children = '5';
UPDATE public.customer
SET children = 'Yes'
WHERE children = '6';
UPDATE public.customer
SET children = 'Yes'
WHERE children = '7';
UPDATE public.customer
SET children = 'Yes'
WHERE children = '8';
UPDATE public.customer
SET children = 'Yes'
WHERE children = '9';
UPDATE public.customer
```

```
SET children = 'Yes'
WHERE children = '10';
```

Tableau Join:

Performing a Left Join with the data from the Churn customer's Children table with Telco's Dependents column:

```
SELECT CAST("Telco"."Churn" AS TEXT) AS "Churn",
 CAST("Telco"."Contract" AS TEXT) AS "Contract",
 CAST("Telco"."CustomerID" AS TEXT) AS "CustomerID",
 CAST("Telco"."Dependents" AS TEXT) AS "Dependents",
 CAST("Telco"."DeviceProtection" AS TEXT) AS "DeviceProtection",
 CAST("Telco"."Gender" AS TEXT) AS "Gender",
 CAST("Telco"."InternetService" AS TEXT) AS "InternetService",
 "Telco"."MonthlyCharges" AS "MonthlyCharges",
 CAST("Telco"."MultipleLines" AS TEXT) AS "MultipleLines",
 CAST("Telco"."OnlineBackup" AS TEXT) AS "OnlineBackup",
 CAST("Telco"."OnlineSecurity" AS TEXT) AS "OnlineSecurity",
 CAST("Telco"."PaperlessBilling" AS TEXT) AS "PaperlessBilling",
 CAST("Telco"."Partner" AS TEXT) AS "Partner",
 CAST("Telco"."PaymentMethod" AS TEXT) AS "PaymentMethod",
 CAST("Telco"."PhoneService" AS TEXT) AS "PhoneService",
 CAST("Telco"."StreamingMovies" AS TEXT) AS "StreamingMovies",
 CAST("Telco"."StreamingTV" AS TEXT) AS "StreamingTV",
 CAST("Telco"."TechSupport" AS TEXT) AS "TechSupport",
 "Telco"."Tenure(Months)" AS "Tenure(Months)",
 "customer"."age" AS "age",
 "customer"."bandwidth_gp_year" AS "bandwidth_gp_year",
 CAST("customer"."children" AS TEXT) AS "children",
 CAST("customer"."churn" AS TEXT) AS "churn",
 "customer"."contacts" AS "contacts",
 "customer"."contract_id" AS "contract_id",
 CAST("customer"."customer_id" AS TEXT) AS "customer_id",
 "customer"."email" AS "email",
 CAST("customer"."gender" AS TEXT) AS "gender",
 "customer"."income" AS "income",
 "customer"."job_id" AS "job_id",
 "customer"."lat" AS "lat",
 "customer"."lng" AS "lng",
 "customer"."location_id" AS "location_id",
 CAST("customer"."marital" AS TEXT) AS "marital",
 "customer"."monthly_charge" AS "monthly_charge",
 "customer"."outage_sec_week" AS "outage_sec_week",
 "customer"."payment_id" AS "payment_id",
 "customer"."population" AS "population",
```

```
  CAST("customer"."port_modem" AS TEXT) AS "port_modem",
  CAST("customer"."tablet" AS TEXT) AS "tablet",
  CAST("customer"."techie" AS TEXT) AS "techie",
  "customer"."tenure" AS "tenure",
  "customer"."yearly_equip_faiure" AS "yearly_equip_faiure"
FROM "public"."customer" "customer"
  LEFT JOIN "public"."Telco" "Telco" ON (CAST("customer"."children" AS TEXT) =
CAST("Telco"."Dependents" AS TEXT))
```

Part B: Panopto Presentation

Please follow the link below to view my Panopto presentation:

https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=04e9138c-26fa-44ef-a3ce-b1100045180a

Part C: Written Report

C1: Dashboard Alignment

My dashboard aligns with a fictional scenario that would be normal in the business world. The WGU churn data comes from a fictional telecommunications company like Comcast and AT&T. The dataset contains information about customer demographics and what service offerings customers have purchased. The Telco data set contains similar customer information and information about services offered. Telecommunications companies seek to minimize churn, or customers discontinuing service. I put myself in the position of a marketing analyst at WGU Churn, a company who just purchased Telco Telecommunications. I was looking at both datasets to create an ad campaign with promotions targeted at preventing current customers from churning and increasing customer tenure.

I wanted to get an understanding of the customer demographics from both companies and compare them. How do churn rates, monthly charges, and tenure change among different genders? Is there different rate of churn, length of tenure, or monthly charges among married customers versus non-married customers? Are there any trends the data can reveal about customers' genders or marital status?

My first dashboard contained customer demographics. My second dashboard displayed the monthly charges of customers at both companies. My last dashboard was focused on tenure, or length of time customers have had services with each company.

C2: Business Intelligence Tools

I used PGAdmin and Tableau Desktop for this project. PGAdmin is a management tool for relational databases that utilizes PostgreSQL. The graphic user interface makes managing databases and writing SQL queries easy and intuitive, especially compared to using SQLshell command prompts (Page, 2023).

Tableau Desktop is a business intelligence tool used to create visualizations. Tableau allows users to create interactive dashboards from multiple sources. Tableau can mine data from live connections, meaning the visualizations are always current This provides up to date information in an easy-to-understand way without requiring employees to spend valuable time recreating dashboards to display the current data (Dambrosio, 2023).

C3: Data Cleaning

The WGU churn dataset preloaded into PGAdmin needed to be cleaned before I could use it in the Tableau dashboards. Specifically, I needed to change the Marital and Children columns to match the data types in the Telco table's Partner and Dependents columns.

The Marital column contained text descriptors of the customer marital status such as married, divorced, or widowed, while the Telco Partner column indicated if a customer was married with a yes or a no. To convert the Churn marital column to a yes or no answer, I used SQL code to update the customer table and set the marital column data to "Yes" under the condition (WHERE command) the marital column data indicated the customer was married. Then I updated the Churn marital column again and set the marital column data to "No" with the condition that the marital column data contained "Divorced," "Separated", "Widowed", or "Never Married".

Next, I needed to change the data in the Churn database's Children column from numerical data to "Yes" or "no" text responses to align with the Telco table Dependents column. First, I needed to change the data type from integer to text. I did this by using SQL code to alter the children column to the text data type. Following this, I updated the customer table by setting the Children column data to "No" under the condition of number of children was zero. I then altered the Children column by setting the Children column data to "Yes" under the conditions where the number of children was from one through ten. The data was now ready to be utilized in the Tableau dashboards.

C4: Dashboard Creation

I created two dashboards for this analysis. The first dashboard visualized if customers churned along with the tenure length of the customers. I created two sheets for this dashboard. The first sheet represented Telco customers' churn rate and average tenure. I created a bar graph with the Telco table's Churned column data in the columns section and the Telco table's Tenure(Months) column data in the rows sections. Tableau automatically displayed the sum of the tenure length. I changed the measurement of tenure length from sum to average. I then created filters for gender and marital status. The gender filter was created using the data from the Telco table's "Genders" column. I included all gender options (Male and Female) when creating the filter. The marital status filter was created using data from the Telco table's "Partner" column. I included all marital statuses (Yes and No) when creating the filter. Finally, I changed the color of the bar graph from the default blue to brown to differentiate the bar graph from the WGU bar graph.

Next, I created the second sheet for the first dashboard. This represented WGU customers' churn rate and tenure length. I created a bar graph with the WGU Churn column data in the columns section and the WGU customer's Tenure column data in the rows sections. I changed the measurement of tenure

length from sum to average. I then created filters for gender and marital status. The gender filter was created using the data from the WGU churn "Gender" column. I included all gender options (Male, Female, and Prefer not to answer) when creating the filter. The marital status filter was created using data from the WGU churn "Marital" column. I included all marital statuses (Yes and No) when creating the filter. I left the bar graph color as the default blue.

The worksheet for the first dashboard contained the bar graph from the first worksheet (Telco data) on the left side. The WGU bar graph from the second worksheet was placed on the right side. The filters from both worksheets were placed in the middle.

The second dashboard visualized if customers churned along with customers' monthly charges. I created two sheets for this dashboard. The first sheet represented Telco customers' churn rate and monthly charges. I created a bar graph with the Telco table's Churned column data in the columns section and the Telco table's Monthly Charges in the rows sections. Tableau automatically displayed the sum of the monthly charges. I changed the monthly charges' measurement from sum to average. I then created filters for gender and marital status. The gender filter was created using the data from the Telco table's "Genders" column. I included all gender options (Male and Female) when creating the filter. The marital status filter was created using data from the Telco table's "Partner" column. I included all marital statuses (Yes and No) when creating the filter. Finally, I changed the color of the bar graph to brown to differentiate it from the WGU bar graph.

Next, I created the second sheet for the second dashboard. This represented WGU customers' churn rate and monthly charges. I created a bar graph with the WGU customer's Churn column data in the columns section and the WGU customer's Monthly Charge column data in the rows sections. I changed the measurement of Monthly Charge from sum to average. I then created filters for gender and marital status. The gender filter was created using the data from the WGU churn "Gender" column. I included all gender options (Male, Female, and Prefer not to answer) when creating the filter. The marital status filter was created using data from the WGU churn "Marital" column. I included all marital statuses (Yes and No) when creating the filter. I left the bar graph color as the default blue.

Like the first dashboard, the second dashboard contained the bar graph from the first worksheet (Telco data) on the left side. The WGU bar graph from the second worksheet was placed on the right side. The filters from both worksheets were placed in the middle.

C5: Analysis Results

My analysis revealed many things about WGU and Telco customers. For WGU customers, gender and marital status did not have an impact on the length of tenure. Marital status and gender did have an impact on WGU customer churn rate and monthly charges. Married WGU customers who did not disclose their gender had the highest monthly charges and the highest churn rates.

Marital status and gender impacted Telco customers' tenure and churn rates. Married Telco customers are more likely to churn than non-married customers. Married Telco customers also had longer tenure before churning compared to non-married customers. Additionally, male Telco customers had longer tenure before churning compared to female Telco customers.

If I were an executive in charge of marketing, I would use this information to create a marketing campaign targeted at married couples offering service discounts to customers who sign long term service contracts.

C6: Analysis Limitations

There are some limitations to this analysis that are worth addressing. First, not all WGU customers chose to disclose their gender. This was problematic when customers who did not disclose their gender had higher rates of churn and higher monthly charges. It would have been helpful for the marketing team to know which gender tended to have higher monthly charges.

Additionally, it would have been helpful if there was data that represented the total charges customers paid for the year and the charges customer paid for each individual service offering such as internet service and phone service. The business could have used this information to discern which services to offer at a discounted rate when customers signed a long-term contract.

Finally, there were no dates in the Churn data or Telco data other than the length of tenure. The data could be outdated and not reflect the current usage of customers. I suspect this is the case with the only gender options being "Male" or "Female." I would expect there to be a non-binary gender option for customers to choose from in a modern questionnaire. Perhaps this why many WGU customers preferred not to disclose their gender. It can be costly for companies to make executive decisions based on outdated information.

D: Data and Code Sources

Telco Communications Data Set by IBM Solutions Downloaded Through Kaggle:

https://www.kaggle.com/datasets/blastchar/telco-customer-churn/discussion

Dropping Columns in PGAdmin:

https://www.commandprompt.com/education/how-to-drop-columns-from-a-table-in-postgresql/#:~:text=The%20syntax%20given%20below%20is,a%20column%20to%20be%20dropped.

Updating Columns in PGAdmin:

https://www.w3schools.com/sql/sql_update.asp

Importing Database into PGAdmin:

https://www.pgadmin.org/docs/pgadmin4/development/import_export_data.html#:~:text=Move%20the%20Import%2FExport%20switch,directory%20and%20select%20a%20file.

https://medium.com/@cheshtadhamija/importing-and-exporting-database-in-postgres-using-pgadmin-ce83e0c9002f

Renaming Columns in PGAdmin:

https://www.techonthenet.com/postgresql/tables/alter_table.php#:~:text=The%20syntax%20to%20rename%20a,of%20the%20table%20to%20modify.

E. Academic Sources

Dambrosio, L. (2023). What is Tableau & What is it used for? An Analyst's Guide. Mode. https://mode.com/blog/what-is-tableau

Page, D. (2023). *pgAdmin FAQ*. pgAdmin Official Website. https://www.pgadmin.org/faq/#:~:text=pgAdmin%20is%20a%20management%20tool,the%20Features%20and%20Screenshots%20pages.