A1.

Two datasets were used this performance assessment. The first is a churn dataset which was is comprised of the tables "contract", "customer", "job", "location", and "payment". The data from these tables has been combined and can be accessed in the attached "churn_clean.csv" file.

The second dataset is from the ACS community survey on types of computers and internet subscriptions used by households in each US state. The cleaned data can be accessed in the "computer_clean.csv".

A2.

To install the dashboard, please view the following steps. The steps used to create the dashboard can be found in "D211_Report.pdf" file.

1. Download the attached Tableau workbook titled "D211_Dashboard.twbx"
2. Open Tableau Desktop
3. Click "File" on the top ribbon
4. Click "Open"
5. Select "D211_Dashboard.twbx"
6. Click "Open"

A3.

The dashboard includes four visualizations. On the top left, there is a scatterplot which compares the percent of the households with cable, fiber optic, or DSL internet with the average tenure in each state. Each point on the scatterplot represents a US state. By hovering over a point, the viewer can compare the percent of households with cable, fiber optic, or DSL internet to the average tenure in that state.

On the top right, there is a map of the United States. This map indicates the percentage of the telecommunications market that the company owns. States with a blue coloring have a relatively high market share, while state with orange coloring have a relatively low market share. Additionally, there is a control on the right which allows the user to filter the states on the map by the number of households. For example, an input of 10 would display the top 10 states with the largest number of households.

On the bottom left, there is a horizontal barchart that compares the average broadband subscription rate between income brackets. The user can see subscription rates for households earning less than $20,000, households earning between $20,000 and $74,999, and households earning $75,000 or more. The user can hover over a bar to see the specific subscription rate for a given income bracket.

On the bottom right, there is a scatterplot which compares the percent of the households with tablets to the average amount of internet bandwidth used in each state. Each point on the scatterplot represents a US state. By hovering over a point, the viewer can compare the percent of households with at least one tablet to the amount of bandwidth used (in gigabytes).

Lastly, there are two key performance indicators on the right pane. The first indicator, is the overall percentage of internet market share in the US. This was calculated as the number of telecom customers divided by the total number of households in the country, expressed as a percentage.   The second indicator is the customer churn rate. This was calculates as the number of customers with a churn value of "Yes" divided by the total number of customers, expressed as a percentage. Both Indicators are shown in the color blue.

A4.

The code below was used to create the "computers" table. The data from "computers.csv" was then added using the import tool in pgAdmin.

```
CREATE TABLE public.computers
(
    state text COLLATE pg_catalog."default" NOT NULL,
    "total households" integer,
    "cellular data plan (total)" integer,
    "broadband such as cable, fiber optic, or dsl (total)" integer,
    "has one or more tablet (percent)" numeric,
    "broadband such as cable, fiber optic, or dsl (percent)" numeric,
    "less than $20,000 with broadband internet (percent)" numeric,
    "$20,000 to $74,999 with broadband internet (percent)" numeric,
    "$75,000 or more with broadband internet (percent)" numeric,
    CONSTRAINT computers_pkey PRIMARY KEY (state)
)
```

The code below was used to add a foreign key called "state_fkey" to the "location" table. It references the "state" column in the "computers" table. This was done to enforce referential integrity in the dataset.

```
ALTER TABLE public.location
ADD CONSTRAINT state_fkey FOREIGN KEY (state)
       REFERENCES public.computers (state) MATCH SIMPLE
       ON UPDATE NO ACTION
ON DELETE NO ACTION
       NOT VALID
```

The following SQL was generated using the "Convert to Custom SQL" option in Tableau. The code effectively joins all of the tables in the PostgreSQL database and allows the users to access all the data needed to create a dashboard.

```
SELECT "computers"."$20,000 to $74,999 with broadband internet (percent)" AS "$20,000
to $74,999 with broadband internet (percent)",
  "computers"."$75,000 or more with broadband internet (percent)" AS "$75,000 or more
with broadband internet (percent)",
  "customer"."age" AS "age",
  "customer"."bandwidth_gp_year" AS "bandwidth_gp_year",
```

```sql
  "computers"."broadband such as cable, fiber optic, or dsl (percent)" AS "broadband
such as cable, fiber optic, or dsl (percent)",
  "computers"."broadband such as cable, fiber optic, or dsl (total)" AS "broadband such
as cable, fiber optic, or dsl (total)",
  "computers"."cellular data plan (total)" AS "cellular data plan (total)",
  "customer"."children" AS "children",
  CAST("customer"."churn" AS TEXT) AS "churn",
  CAST("location"."city" AS TEXT) AS "city",
  "customer"."contacts" AS "contacts",
  "contract"."contract_id" AS "contract_id (contract)",
  "customer"."contract_id" AS "contract_id",
  CAST("location"."county" AS TEXT) AS "county",
  CAST("customer"."customer_id" AS TEXT) AS "customer_id",
  CAST("contract"."duration" AS TEXT) AS "duration",
  "customer"."email" AS "email",
  CAST("customer"."gender" AS TEXT) AS "gender",
  "computers"."has one or more tablet (percent)" AS "has one or more tablet
(percent)",
  "customer"."income" AS "income",
  "job"."job_id" AS "job_id (job)",
  "customer"."job_id" AS "job_id",
  CAST("job"."job_title" AS TEXT) AS "job_title",
  "customer"."lat" AS "lat",
  "computers"."less than $20,000 with broadband internet (percent)" AS "less than
$20,000 with broadband internet (percent)",
  "customer"."lng" AS "lng",
  "customer"."location_id" AS "location_id (customer)",
  "location"."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",
  "payment"."payment_id" AS "payment_id (payment)",
  "customer"."payment_id" AS "payment_id",
  CAST("payment"."payment_type" AS TEXT) AS "payment_type",
  "customer"."population" AS "population",
  CAST("customer"."port_modem" AS TEXT) AS "port_modem",
  CAST("location"."state" AS TEXT) AS "state (location)",
  CAST("computers"."state" AS TEXT) AS "state",
  CAST("customer"."tablet" AS TEXT) AS "tablet",
  CAST("customer"."techie" AS TEXT) AS "techie",
  "customer"."tenure" AS "tenure",
  "computers"."total households" AS "total households",
  "customer"."yearly_equip_faiure" AS "yearly_equip_faiure",
  "location"."zip" AS "zip"
FROM "public"."computers" "computers"
  RIGHT JOIN "public"."location" "location" ON (CAST("computers"."state" AS TEXT) =
CAST("location"."state" AS TEXT))
  INNER JOIN "public"."customer" "customer" ON ("location"."location_id" =
"customer"."location_id")
  INNER JOIN "public"."contract" "contract" ON ("customer"."contract_id" =
"contract"."contract_id")
  INNER JOIN "public"."job" "job" ON ("customer"."job_id" = "job"."job_id")
  INNER JOIN "public"."payment" "payment" ON ("customer"."payment_id" =
"payment"."payment_id")
```

B.

Please see the link below to access the Panopto presentation.

https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=ad3950d8-64bb-4c68-92e7-b159015abe1b

C1.

    For this assignment, I integrated data from the American Community Survey on the types of computers and internet subscriptions used by households in each US state. This data contains the totals and percentage of households that use various devices and services. The data is also segmented by income brackets. This aligns with the needs of the telecommunications company by allowing us to compare the company's performance relative to the larger telecom market. It also allows us to segment possible customers and compare them to the existing customer base.

C2.

    I selected Tableau to create the dashboard because of its versatility and ease of use. Tableau can connect to a wide variety of data sources with minimal effort. It was quite simple to connect to the PostgreSQL database to access the necessary data for the dashboard.   It also allowed me to access data across tables to perform calculations and gain deeper insights into the data.

    Additionally, Tableau makes it very easy to generate complex visualizations. The interface is very user-friendly and does not require a significant amount of technical knowledge to navigate.

C3.

    decided to clean the provided and external datasets using Python. I imported the pandas, seaborn, and numpy libraries to manipulate and explore the data before making any changes. The attached file, "D211_data_cleaning.pdf" contains all of python code used to clean the data.

    I started off by importing the initial census dataset, named "'ACSST1Y2022.S2801-Data.csv', into a pandas dataframe called "df_initial". After selecting the columns I wanted to use in the dashboard, I included them in a new dataframe called "df_computer". I chose not to deal with any potential outliers in the dataframe because I did not want to alter any of the census data.

    I then checked for missing values using the "isnull()" and "sum()" functions. The output indicated that there were no missing values in any of the columns.

    Next, I decided to rename the columns because to make them easier to interpret. To do this I used the "rename()" function on df_computer and entered new names for each of the columns.

    The last step I took to clean the external data was to change state names to abbreviations. I did this by creating a dictionary called "dict_state" and assigning abbreviations to each state name in the dataset. I then used the "replace()" function to swap the values in df_computer. This was done

so that the values in the external dataset would match those in the "location" table of the PostgreSQL database.

After cleaning the census data, I exported it to a file called "computer_clean.csv". I then moved on to cleaning the churn dataset.

I began by exporting each table from the churn database in pgAdmin as a csv file. I created five separate dataframes called "df_location", "df_job", "df_customer", "df_payment", and "df_contract". I again checked for missing values in each table using the "isnull()" and "sum()" functions from the pandas library. The output indicated that there were no missing values in any of the tables.

I then checked for duplicates using the "duplicated()" and "value_counts()" functions. The output showed that there were no duplicate rows in any of the tables.

Lastly, I checked for outliers in the "customer" table. To do this, I generated boxplots for each of the numerical features using the "boxplot()" function from the seaborn library. The output indicated that there were outliers in the "Population", "Children", "Income", "Outage_sec_week", "Email", "Contacts", and "Yearly_equip_faiure" variables. I decided not to remove or impute any of these values because they seemed reasonable, and I did not want to skew the dataset.

After reviewing the churn database, I found that I did not need to modify it in any way. There were no missing values or duplicate rows in any of the tables. Additionally, the outliers were not egregious enough to exclude or impute.

The final step I took was to upload the newly cleaned census data into the PostgreSQL database. Using the code snippet in section A4, I created a table called "public.computers". I uploaded the data from "computer_clean.csv" using the import tool in pgAdmin. I created a foreign key in the "location" table which referenced the "state" field in order to reinforce referential integrity. The data was now successfully cleaned and ready to be analyzed.

C4.

Please see the steps below to create each section of the Tableau Dashboard.

Churn Rate KPI
-Click "New Worksheet" on bottom ribbon
-Rename sheet to "Churn Rate KPI"
-Create a calculated field called "Churn Rate KPI"
-Enter COUNT(IF[Churn] = 'Yes' THEN 1 END) / COUNT([Customer Id]) and click OK
-Drag churn rate KPI to text marks
-Right click Churn Rate KPI > Default Properties > Number Format > Select Percentage > 2 Decimal Places > Click OK
-Right click AGG(Churn Rate) in the marks panel > Click Format > Change default font size to 14 > Change default color to top shade of blue

Internet Market Share KPI
-Click "New Worksheet" on bottom ribbon

-Rename sheet to "Internet Market Share KPI"
-Create a calculated field called "Internet Market Share KPI"
-Enter COUNT([Customer Id]) / SUM([Broadband Such As Cable, Fiber Optic, Or Dsl (Total)])
-Drag internet market share KPI to text marks
-Right click Internet Market Share KPI > Default Properties > Number Format > Select Percentage
> 6 Decimal Places > Click OK
-Right click AGG(Internet Market Share KPI) in the marks panel > Click Format > Change default
font size to 14 > Change default color to top shade of blue

Subscription Rate vs Tenure by State
-Click "New Worksheet" on bottom ribbon
-Rename sheet to "Subscription Rate vs Tenure by State"
-Drag "Broadband Such As Cable, Fiber Optic, Or DSL (Percent)" to columns > click arrow on the
right > change measure to average
-Drag "Tenure" to rows > click arrow on the right > change measure to average
-Drag state field from computers table to detail in marks panel

Market Share by State
-Click "New Worksheet" on bottom ribbon
-Rename sheet to "Market Share by State"
-Drag state field from computers table to detail in marks panel
-Drag "Internet Market Share KPI" to color in marks panel
-Click "Color" in the marks panel > click "Edit Colors" > Select "Orange-Blue Diverging" from
Palette dropdown > Click OK

Top States by Number of Households Control
-Drag state field from computers table to filter
-Click "Top" tab
-Click "By Field"
-Select "Top" and click "Create New Parameter"
    -Set name of parameter to "Top States by Number of Households"
    -Set "current value" to 52
    -Set "allowable values" to range
    -Set "minimum" to 1
    -Set "maximum" to 52
    -Set "step size" to 1
    -Select "fixed"
    -Click OK
-Back in the "By field" section change the second row to "Total Households" and "Sum"State
-Search for "Top States by Number of Households" in the "Parameters" section of the Data Pane
-Right click the "Top States by Number of Households" Parameter and select "Show Parameter"

Tablet Ownership vs Bandwidth Usage
-Click "New Worksheet" on bottom ribbon
-Rename sheet to "Tablet Ownership vs Bandwidth Usage"
-Drag "Bandwidth Gp Year" to rows > click arrow on the right > change measure to average
-Drag "Has One or More Tablet (Percent)" to columns > click arrow on the right > change
measure to average

-Drag state field from computers table to "Detail" in marks panel

Broadband Subscription Rates by Income Bracket
-Click "New Worksheet" on bottom ribbon
-Rename sheet to "Broadband Subscription Rates by Income Bracket"
-Drag "Less Than $20,000 With Broadband Internet (Percent)" to Rows > Right click on field > Change measure to Average
-Drag "$20,000 to $74,999 With Broadband Internet (Percent)" to Rows > Right click on field > Change measure to Average
-Drag "$75,000 Or More With Broadband Internet (Percent)" to Measure Values > Right click on field > Change measure to Average
-Click "Show Me" button on top right of Tableau
-Select Horizontal Chart
-Move "Measure Names" to columns
-Move "Measure Values" to rows

Dashboard
-Click "New Dashboard" on the bottom ribbon
-Drag "Subscription Rate vs Tenure by State" sheet to empty pane
-Drag "Market Share by State" to the right side of the pane
-Drag "Broadband Subscription Rates by Income Bracket" under "Subscription Rate vs Tenure by State"
-Drag "Tablet Ownership vs Bandwidth Usage by State" under "Market Share by State"
-Drag "Churn Rate KPI" Under "Top States by Number of Households" Parameter on the right side of the dashboard
-Drag "Internet Market Share KPI" under "Churn Rate KPI"

C5.

     Based on the results of the analysis, we can support executive decision-making in a few ways. When looking at "Subscription Rate vs. Tenure by State" scatterplot. We can see that there is a positive relationship between broadband subscription rates and average customer tenure. This indicates that the telecommunications company should target states with high broadband subscription rates. Customers in these states are likely to remain loyal for longer periods of time.

     From the map visualization on the top right, we can see where the telecommunication company has the strongest present. The company has a relatively large market share in Montana, North Dakota, South Dakota, Nebraska, and Iowa. There is also a strong presence in Alaska, West Virginia, and Vermont. The telecommunications company should continue to build a regional presence and expand outward.

     From the scatterplot on the bottom left, we can see how different income brackets differ in subscription rates. It appears that households with less than $20,000 in income are underutilized compares to the other two brackets. Because of this, the company should consider providing a more affordable package to meet the needs of these potential customers.

     Lastly, the "Tablet Ownership vs Bandwidth Usage" scatterplot on the bottom right shows another relationship between two variables. The visualization shows that states with a higher

percentage of tablet ownership tend to use more bandwidth. If the telecommunication company knows that a potential customer owns a tablet, they should try to upsell them to a higher bandwidth package.

C6.

The analysis was limited by the number of relevant fields in the external dataset. Out of the 34 columns in the initial dataset, I only used 9 to create the dashboard. This limited the insights that the visualizations and KPIs could provide. In the future, it might be worthwhile to look for a census dataset with information that can provide more insight.

Additionally, the data the datasets were joined on the "State" field, which limited the scope of the analysis. If the census data contained information by zip code or city, the analysis might have been able to provide even more detailed insights.