D211 Task 1: Advanced Data Acquisition

Jeremy Dorrough

Western Governors University

WGU Student ID# 000994113

**A1.**

Please find both data sets attached.

**A2.**

Please use these step-by-step instructions to view the Tableau Presentation. These instructions have been provided to guide you through the installation and setup process.

1. Download and install Tableau Desktop from<https://www.tableau.com/products/desktop>. Alternatively, you can use Tableau Reader from<https://www.tableau.com/products/reader> if you only want to view the dashboards and not create or edit them. Make sure to download the software that corresponds to your operating system and ensure that your computer meets the technical requirements.
2. Open Tableau Desktop and log in with your account credentials, which may need to be created.
3. Download the Tableau workbook file to your computer. Click "Open" and select the file from the download location.
4. After opening, Tableau will attempt to connect to the data source. Assuming you have PGAdmin installed and setup with the correct database, you may need to update the data connection, which can be accomplished using the following steps.
   1. Navigate to “Data Source” and click "Edit Connection".
   2. In the "Edit Connection" window, select "PostgreSQL" from the "To a Server" options and click "Connect".
   3. In the "Connect" window, enter the following information:
      1. Server: localhost
      2. Port: 5432
      3. Database: churn
      4. Authentication: Username and Password
      5. Username: postgres
      6. Password: Passw0rd!
   4. Click "Sign In" to connect to the PostgreSQL database and pull the data into the workbook. You should be able to see the data in the dashboards and presentation.
   5. If you need to make any other changes to the data source, navigate back to the "Data Source" tab and select "Edit Connection". From there, you can change the server, database, and authentication information as needed.
      1. *Note: If the external dataset requires further evaluation within PGAdmin in the lab environment, the provided external data’s CSV file must be placed in the C:/Users/Public/Downloads/ folder.*
5. The Tableau presentation uses the "Presentation" story tab, where you can interact with the data visualizations and representations.

With these instructions, you should be able to view the Tableau presentation and connect to the PostgreSQL database using Tableau Desktop.

**A3.**

The presentation is divided into three dashboards within a Tableau story. The dashboards are separated by tab and labeled as KPI (Key Performance Index), Geographical, and Service. These contain information about company performance, customer geography, and service data analysis, respectively. These can be navigated by simply clicking on the desired tab and allowing the respective data to load. It is important to note that most presentation data points can be highlighted simply by clicking them and unhighlighted by simply clicking again. All visualizations also include text and numerical descriptions for specific information. Additionally, several interactive controls are provided in the Geographical tab. The below detailed instructions explain how to view and navigate within the tabs.

* KPI:
  + ‘Tenure` bar charts: Bars represent tenure length, and a longer bar indicates higher average tenure.
  + ‘Churn Rate’ pie charts: Slices show the share of customers who have or have not churned in that category.
* Service:
  + Tenure/Bandwidth area charts: Observe the colored areas which represent the change in values over time. Trend lines indicate overall trends.
  + Churn Rate by Contract Type Pie Charts: Slices represent churn rate by contract type, with larger slices indicating higher churn rates as detailed in the included text and numerical information.
  + Text Tables: The values in each cell represent summary information for the categories.
* Geographical:
  + Maps: Visualizes the location of customers using their provided geographical information.
  + Churn Status Filter: To use this filter, select an option (all, yes, or no) from the dropdown menu to filter based on churn status.
  + Customer Tenure Filter: Use the slider for selecting a desired range of values, which will filter cities and zip codes based on customer tenure in months.

**A4.**

The following PostgreSQL code was used within PGAdmin to transform and prepare the data for use in Tableau:

-- Create the table used for Tableau using only California data

CREATE TABLE wgu\_data\_for\_tableau AS

SELECT

customer.Customer\_id,

customer.Tenure,

customer.Churn,

customer.Lng,

customer.Lat,

location.ZIP,

location.City,

contract.duration AS Contract,

customer.Bandwidth\_GP\_Year,

customer.Monthly\_Charge,

location.State

FROM customer

JOIN contract ON customer.contract\_id = contract.contract\_id

JOIN payment ON customer.payment\_id = payment.payment\_id

JOIN location ON customer.location\_id = location.location\_id

WHERE location.State = 'CA';

-- Add the foreign key constraint to the existing table

ALTER TABLE wgu\_data\_for\_tableau

ADD CONSTRAINT fk\_data\_for\_tableau\_customer

FOREIGN KEY (Customer\_id) REFERENCES customer (Customer\_id);

-- Data transformations for provided data

ALTER TABLE wgu\_data\_for\_tableau

DROP COLUMN State;

ALTER TABLE wgu\_data\_for\_tableau

RENAME COLUMN Bandwidth\_GP\_Year TO Bandwidth\_GB\_Month;

UPDATE wgu\_data\_for\_tableau

SET Bandwidth\_GB\_Month = Bandwidth\_GB\_Month / 12;

UPDATE wgu\_data\_for\_tableau

SET bandwidth\_gb\_month = ROUND(bandwidth\_gb\_month);

ALTER TABLE wgu\_data\_for\_tableau

ALTER COLUMN Tenure TYPE numeric USING ROUND(Tenure);

ALTER TABLE wgu\_data\_for\_tableau

ALTER COLUMN Monthly\_Charge TYPE numeric USING ROUND(Monthly\_Charge);

UPDATE wgu\_data\_for\_tableau

SET Contract = 'One Year'

WHERE Contract = 'One year';

-- Add column to differentiate internal data

ALTER TABLE wgu\_data\_for\_tableau

ADD COLUMN source TEXT DEFAULT 'wgu';

-- Create table for external data

CREATE TABLE kaggle\_data\_for\_tableau (

"Customer ID" TEXT,

"Gender" TEXT,

"Age" INTEGER,

"Married" TEXT,

"Number of Dependents" INTEGER,

"City" TEXT,

"Zip Code" INTEGER,

"Latitude" NUMERIC,

"Longitude" NUMERIC,

"Number of Referrals" INTEGER,

"Tenure in Months" NUMERIC,

"Offer" TEXT,

"Phone Service" TEXT,

"Avg Monthly Long Distance Charges" NUMERIC,

"Multiple Lines" TEXT,

"Internet Service" TEXT,

"Internet Type" TEXT,

"Avg Monthly GB Download" NUMERIC,

"Online Security" TEXT,

"Online Backup" TEXT,

"Device Protection Plan" TEXT,

"Premium Tech Support" TEXT,

"Streaming TV" TEXT,

"Streaming Movies" TEXT,

"Streaming Music" TEXT,

"Unlimited Data" TEXT,

"Contract" TEXT,

"Paperless Billing" TEXT,

"Payment Method" TEXT,

"Monthly Charge" NUMERIC,

"Total Charges" NUMERIC,

"Total Refunds" NUMERIC,

"Total Extra Data Charges" NUMERIC,

"Total Long Distance Charges" NUMERIC,

"Total Revenue" NUMERIC,

"Churn Category" TEXT,

"Churn Reason" TEXT,

"Source" TEXT

);

-- Import the Kaggle Data (The Kaggle CSV must be placed in C:/Users/Public/Downloads/)

COPY kaggle\_data\_for\_tableau ("Customer ID", "Gender", "Age", "Married", "Number of Dependents", "City", "Zip Code", "Latitude", "Longitude", "Number of Referrals", "Tenure in Months", "Offer", "Phone Service", "Avg Monthly Long Distance Charges", "Multiple Lines", "Internet Service", "Internet Type", "Avg Monthly GB Download", "Online Security", "Online Backup", "Device Protection Plan", "Premium Tech Support", "Streaming TV", "Streaming Movies", "Streaming Music", "Unlimited Data", "Contract", "Paperless Billing", "Payment Method", "Monthly Charge", "Total Charges", "Total Refunds", "Total Extra Data Charges", "Total Long Distance Charges", "Total Revenue", "Churn Category", "Churn Reason","Source")

FROM 'C:/Users/Public/Downloads/TELECO~1.CSV'

DELIMITER ',' CSV HEADER QUOTE '"' ESCAPE '''';

-- Rename columns

ALTER TABLE kaggle\_data\_for\_tableau RENAME COLUMN "Customer ID" TO Customer\_id;

ALTER TABLE kaggle\_data\_for\_tableau RENAME COLUMN "Monthly Charge" TO Monthly\_Charge;

ALTER TABLE kaggle\_data\_for\_tableau RENAME COLUMN "Avg Monthly GB Download" TO Bandwidth\_GB\_Month;

ALTER TABLE kaggle\_data\_for\_tableau RENAME COLUMN "Tenure in Months" TO Tenure;

ALTER TABLE kaggle\_data\_for\_tableau RENAME COLUMN "Latitude" TO Lat;

ALTER TABLE kaggle\_data\_for\_tableau RENAME COLUMN "Longitude" TO Lng;

ALTER TABLE kaggle\_data\_for\_tableau RENAME COLUMN "Zip Code" TO ZIP;

-- Update Churned to Yes and other values to No

UPDATE kaggle\_data\_for\_tableau SET "Churn Category" = 'Yes' WHERE "Churn Category" = 'Churned';

UPDATE kaggle\_data\_for\_tableau SET "Churn Category" = 'No' WHERE "Churn Category" != 'Yes';

-- Remove customers using unlimited data service not used in analysis

DELETE FROM kaggle\_data\_for\_tableau WHERE "Unlimited Data" = 'Yes';

-- Only keep data used in the analysis and add source column

CREATE TABLE kaggle\_data\_for\_tableau\_temp AS

SELECT

Customer\_id,

Tenure,

"Churn Category" AS Churn,

Lng,

Lat,

ZIP,

"City",

"Contract",

Bandwidth\_GB\_Month,

Monthly\_Charge,

'kaggle' AS Source

FROM kaggle\_data\_for\_tableau;

DROP TABLE kaggle\_data\_for\_tableau;

ALTER TABLE kaggle\_data\_for\_tableau\_temp RENAME TO kaggle\_data\_for\_tableau;

-- Delete rows where bandwidth is null

DELETE FROM kaggle\_data\_for\_tableau

WHERE Bandwidth\_GB\_Month IS NULL;

-- Round monthly charge to whole number

ALTER TABLE kaggle\_data\_for\_tableau

ALTER COLUMN Monthly\_Charge TYPE numeric USING ROUND(Monthly\_Charge);

-- Create a new data\_for\_tableau table as a UNION of the wgu and kaggle tables

CREATE TABLE data\_for\_tableau AS

SELECT \* FROM wgu\_data\_for\_tableau

UNION ALL

SELECT \* FROM kaggle\_data\_for\_tableau;

The following is SQL was generated by Tableau for visualizations:

-- Connecting Data

SELECT "data\_for\_tableau"."bandwidth\_gb\_month" AS "bandwidth\_gb\_month",

CAST("data\_for\_tableau"."churn" AS TEXT) AS "churn",

CAST("data\_for\_tableau"."city" AS TEXT) AS "city",

CAST("data\_for\_tableau"."contract" AS TEXT) AS "contract",

CAST("data\_for\_tableau"."customer\_id" AS TEXT) AS "customer\_id",

"data\_for\_tableau"."lat" AS "lat",

"data\_for\_tableau"."lng" AS "lng",

"data\_for\_tableau"."monthly\_charge" AS "monthly\_charge",

CAST("data\_for\_tableau"."source" AS TEXT) AS "source",

"data\_for\_tableau"."tenure" AS "tenure",

"data\_for\_tableau"."zip" AS "zip"

FROM "public"."data\_for\_tableau" "data\_for\_tableau"

-- Avg Tenure and Bandwidth

SELECT AVG("data\_for\_tableau"."tenure") AS "avg:tenure:ok",

"data\_for\_tableau"."bandwidth\_gb\_month" AS "bandwidth\_gb\_month"

FROM "public"."data\_for\_tableau" "data\_for\_tableau"

WHERE (CAST("data\_for\_tableau"."source" AS TEXT) = 'kaggle')

GROUP BY 2

-- WGU Customer Calculations

SELECT COUNT(DISTINCT (CASE WHEN (CAST("data\_for\_tableau"."source" AS TEXT) = 'wgu') THEN CAST("data\_for\_tableau"."customer\_id" AS TEXT) ELSE NULL END)) AS "TEMP(Calculation\_436849219392933897)(1913223596)(0)",

COUNT((CASE WHEN ((CAST("data\_for\_tableau"."source" AS TEXT) = 'wgu') AND (CAST("data\_for\_tableau"."churn" AS TEXT) = 'Yes')) THEN CAST("data\_for\_tableau"."churn" AS TEXT) ELSE NULL END)) AS "TEMP(Calculation\_436849219392933897)(1914332625)(0)",

AVG("data\_for\_tableau"."bandwidth\_gb\_month") AS "avg:bandwidth\_gb\_month:ok",

AVG("data\_for\_tableau"."monthly\_charge") AS "avg:monthly\_charge:ok",

COUNT(CAST("data\_for\_tableau"."customer\_id" AS TEXT)) AS "cnt:customer\_id:ok",

CAST("data\_for\_tableau"."contract" AS TEXT) AS "contract",

MAX("data\_for\_tableau"."bandwidth\_gb\_month") AS "max:bandwidth\_gb\_month:ok",

COUNT((CASE WHEN (CAST("data\_for\_tableau"."churn" AS TEXT) = 'Yes') THEN CAST("data\_for\_tableau"."churn" AS TEXT) ELSE NULL END)) AS "usr:Calculation\_436849219397222416:ok"

FROM "public"."data\_for\_tableau" "data\_for\_tableau"

WHERE (CAST("data\_for\_tableau"."source" AS TEXT) = 'wgu')

GROUP BY 6

-- WGU Churn by Contract

SELECT CAST("data\_for\_tableau"."churn" AS TEXT) AS "churn",

COUNT(CAST("data\_for\_tableau"."churn" AS TEXT)) AS "cnt:churn:ok",

COUNT(CAST("data\_for\_tableau"."contract" AS TEXT)) AS "cnt:contract:ok",

CAST("data\_for\_tableau"."contract" AS TEXT) AS "contract"

FROM "public"."data\_for\_tableau" "data\_for\_tableau"

WHERE ((CAST("data\_for\_tableau"."churn" AS TEXT) = 'Yes') AND (CAST("data\_for\_tableau"."source" AS TEXT) = 'wgu'))

GROUP BY 1,

4

-- WGU Avg Tenure and Bandwidth

SELECT AVG("data\_for\_tableau"."tenure") AS "avg:tenure:ok",

"data\_for\_tableau"."bandwidth\_gb\_month" AS "bandwidth\_gb\_month"

FROM "public"."data\_for\_tableau" "data\_for\_tableau"

WHERE (CAST("data\_for\_tableau"."source" AS TEXT) = 'wgu')

GROUP BY 2

-- Kaggle Customer Calculations

SELECT COUNT((CASE WHEN ((CAST("data\_for\_tableau"."source" AS TEXT) = 'kaggle') AND (CAST("data\_for\_tableau"."churn" AS TEXT) = 'Yes')) THEN CAST("data\_for\_tableau"."churn" AS TEXT) ELSE NULL END)) AS "TEMP(Churn Rate (Int) (copy)\_436849219398893586)(311974858)(0)",

COUNT(DISTINCT (CASE WHEN (CAST("data\_for\_tableau"."source" AS TEXT) = 'kaggle') THEN CAST("data\_for\_tableau"."customer\_id" AS TEXT) ELSE NULL END)) AS "TEMP(Churn Rate (Int) (copy)\_436849219398893586)(3483384996)(0)",

AVG("data\_for\_tableau"."bandwidth\_gb\_month") AS "avg:bandwidth\_gb\_month:ok",

AVG("data\_for\_tableau"."monthly\_charge") AS "avg:monthly\_charge:ok",

COUNT(CAST("data\_for\_tableau"."churn" AS TEXT)) AS "cnt:churn:ok",

CAST("data\_for\_tableau"."contract" AS TEXT) AS "contract",

MAX("data\_for\_tableau"."bandwidth\_gb\_month") AS "max:bandwidth\_gb\_month:ok",

COUNT((CASE WHEN (CAST("data\_for\_tableau"."churn" AS TEXT) = 'Yes') THEN CAST("data\_for\_tableau"."churn" AS TEXT) ELSE NULL END)) AS "usr:Churned (Int) (copy)\_436849219399159827:ok"

FROM "public"."data\_for\_tableau" "data\_for\_tableau"

WHERE (CAST("data\_for\_tableau"."source" AS TEXT) = 'kaggle')

GROUP BY 6

-- WGU Distinct Customers

SELECT COUNT(DISTINCT (CASE WHEN (CAST("data\_for\_tableau"."source" AS TEXT) = 'wgu') THEN CAST("data\_for\_tableau"."customer\_id" AS TEXT) ELSE NULL END)) AS "TEMP(Calculation\_436849219392933897)(1913223596)(0)",

COUNT((CASE WHEN ((CAST("data\_for\_tableau"."source" AS TEXT) = 'wgu') AND (CAST("data\_for\_tableau"."churn" AS TEXT) = 'Yes')) THEN CAST("data\_for\_tableau"."churn" AS TEXT) ELSE NULL END)) AS "TEMP(Calculation\_436849219392933897)(1914332625)(2)",

AVG("data\_for\_tableau"."bandwidth\_gb\_month") AS "avg:bandwidth\_gb\_month:ok",

AVG("data\_for\_tableau"."monthly\_charge") AS "avg:monthly\_charge:ok",

COUNT(CAST("data\_for\_tableau"."customer\_id" AS TEXT)) AS "cnt:customer\_id:ok",

CAST("data\_for\_tableau"."contract" AS TEXT) AS "contract",

MAX("data\_for\_tableau"."bandwidth\_gb\_month") AS "max:bandwidth\_gb\_month:ok",

COUNT((CASE WHEN (CAST("data\_for\_tableau"."churn" AS TEXT) = 'Yes') THEN CAST("data\_for\_tableau"."churn" AS TEXT) ELSE NULL END)) AS "usr:Calculation\_436849219397222416:ok"

FROM "public"."data\_for\_tableau" "data\_for\_tableau"

WHERE (CAST("data\_for\_tableau"."source" AS TEXT) = 'wgu')

GROUP BY 6

-- WGU Churned Customers

SELECT COUNT((CASE WHEN ((CAST("data\_for\_tableau"."source" AS TEXT) = 'kaggle') AND (CAST("data\_for\_tableau"."churn" AS TEXT) = 'Yes')) THEN CAST("data\_for\_tableau"."churn" AS TEXT) ELSE NULL END)) AS "TEMP(Churn Rate (Int) (copy)\_436849219398893586)(311974858)(2)",

COUNT(DISTINCT (CASE WHEN (CAST("data\_for\_tableau"."source" AS TEXT) = 'kaggle') THEN CAST("data\_for\_tableau"."customer\_id" AS TEXT) ELSE NULL END)) AS "TEMP(Churn Rate (Int) (copy)\_436849219398893586)(3483384996)(0)",

AVG("data\_for\_tableau"."bandwidth\_gb\_month") AS "avg:bandwidth\_gb\_month:ok",

AVG("data\_for\_tableau"."monthly\_charge") AS "avg:monthly\_charge:ok",

COUNT(CAST("data\_for\_tableau"."churn" AS TEXT)) AS "cnt:churn:ok",

CAST("data\_for\_tableau"."contract" AS TEXT) AS "contract",

MAX("data\_for\_tableau"."bandwidth\_gb\_month") AS "max:bandwidth\_gb\_month:ok",

COUNT((CASE WHEN (CAST("data\_for\_tableau"."churn" AS TEXT) = 'Yes') THEN CAST("data\_for\_tableau"."churn" AS TEXT) ELSE NULL END)) AS "usr:Churned (Int) (copy)\_436849219399159827:ok"

FROM "public"."data\_for\_tableau" "data\_for\_tableau"

WHERE (CAST("data\_for\_tableau"."source" AS TEXT) = 'kaggle')

GROUP BY 6

-- WGU Bandwidth and Charge

SELECT COUNT(DISTINCT (CASE WHEN (CAST("data\_for\_tableau"."source" AS TEXT) = 'wgu') THEN CAST("data\_for\_tableau"."customer\_id" AS TEXT) ELSE NULL END)) AS "TEMP(Calculation\_436849219392933897)(1913223596)(0)",

COUNT((CASE WHEN ((CAST("data\_for\_tableau"."source" AS TEXT) = 'wgu') AND (CAST("data\_for\_tableau"."churn" AS TEXT) = 'Yes')) THEN CAST("data\_for\_tableau"."churn" AS TEXT) ELSE NULL END)) AS "TEMP(Calculation\_436849219392933897)(1914332625)(3)",

AVG("data\_for\_tableau"."bandwidth\_gb\_month") AS "avg:bandwidth\_gb\_month:ok",

AVG("data\_for\_tableau"."monthly\_charge") AS "avg:monthly\_charge:ok",

COUNT(CAST("data\_for\_tableau"."customer\_id" AS TEXT)) AS "cnt:customer\_id:ok",

CAST("data\_for\_tableau"."contract" AS TEXT) AS "contract",

MAX("data\_for\_tableau"."bandwidth\_gb\_month") AS "max:bandwidth\_gb\_month:ok",

COUNT((CASE WHEN (CAST("data\_for\_tableau"."churn" AS TEXT) = 'Yes') THEN CAST("data\_for\_tableau"."churn" AS TEXT) ELSE NULL END)) AS "usr:Calculation\_436849219397222416:ok"

FROM "public"."data\_for\_tableau" "data\_for\_tableau"

WHERE (CAST("data\_for\_tableau"."source" AS TEXT) = 'wgu')

GROUP BY 6

-- WGU Churned Customers

SELECT COUNT(DISTINCT (CASE WHEN (CAST("data\_for\_tableau"."source" AS TEXT) = 'wgu') THEN CAST("data\_for\_tableau"."customer\_id" AS TEXT) ELSE NULL END)) AS "TEMP(Calculation\_436849219392933897)(1913223596)(0)",

COUNT((CASE WHEN ((CAST("data\_for\_tableau"."source" AS TEXT) = 'wgu') AND (CAST("data\_for\_tableau"."churn" AS TEXT) = 'Yes')) THEN CAST("data\_for\_tableau"."churn" AS TEXT) ELSE NULL END)) AS "TEMP(Calculation\_436849219392933897)(1914332625)(4)",

AVG("data\_for\_tableau"."bandwidth\_gb\_month") AS "avg:bandwidth\_gb\_month:ok",

AVG("data\_for\_tableau"."monthly\_charge") AS "avg:monthly\_charge:ok",

COUNT(CAST("data\_for\_tableau"."customer\_id" AS TEXT)) AS "cnt:customer\_id:ok",

CAST("data\_for\_tableau"."contract" AS TEXT) AS "contract",

MAX("data\_for\_tableau"."bandwidth\_gb\_month") AS "max:bandwidth\_gb\_month:ok",

COUNT((CASE WHEN (CAST("data\_for\_tableau"."churn" AS TEXT) = 'Yes') THEN CAST("data\_for\_tableau"."churn" AS TEXT) ELSE NULL END)) AS "usr:Calculation\_436849219397222416:ok"

FROM "public"."data\_for\_tableau" "data\_for\_tableau"

WHERE (CAST("data\_for\_tableau"."source" AS TEXT) = 'wgu')

GROUP BY 6

-- Kaggle Churned Customers

SELECT COUNT((CASE WHEN ((CAST("data\_for\_tableau"."source" AS TEXT) = 'kaggle') AND (CAST("data\_for\_tableau"."churn" AS TEXT) = 'Yes')) THEN CAST("data\_for\_tableau"."churn" AS TEXT) ELSE NULL END)) AS "TEMP(Churn Rate (Int) (copy)\_436849219398893586)(311974858)(4)",

COUNT(DISTINCT (CASE WHEN (CAST("data\_for\_tableau"."source" AS TEXT) = 'kaggle') THEN CAST("data\_for\_tableau"."customer\_id" AS TEXT) ELSE NULL END)) AS "TEMP(Churn Rate (Int) (copy)\_436849219398893586)(3483384996)(0)",

AVG("data\_for\_tableau"."bandwidth\_gb\_month") AS "avg:bandwidth\_gb\_month:ok",

AVG("data\_for\_tableau"."monthly\_charge") AS "avg:monthly\_charge:ok",

COUNT(CAST("data\_for\_tableau"."churn" AS TEXT)) AS "cnt:churn:ok",

CAST("data\_for\_tableau"."contract" AS TEXT) AS "contract",

MAX("data\_for\_tableau"."bandwidth\_gb\_month") AS "max:bandwidth\_gb\_month:ok",

COUNT((CASE WHEN (CAST("data\_for\_tableau"."churn" AS TEXT) = 'Yes') THEN CAST("data\_for\_tableau"."churn" AS TEXT) ELSE NULL END)) AS "usr:Churned (Int) (copy)\_436849219399159827:ok"

FROM "public"."data\_for\_tableau" "data\_for\_tableau"

WHERE (CAST("data\_for\_tableau"."source" AS TEXT) = 'kaggle')

GROUP BY 6

-- WGU Distinct Customers and Churned Customers

SELECT COUNT(DISTINCT (CASE WHEN (CAST("data\_for\_tableau"."source" AS TEXT) = 'wgu') THEN CAST("data\_for\_tableau"."customer\_id" AS TEXT) ELSE NULL END)) AS "TEMP(Calculation\_436849219392933897)(1913223596)(0)",

COUNT((CASE WHEN ((CAST("data\_for\_tableau"."source" AS TEXT) = 'wgu') AND (CAST("data\_for\_tableau"."churn" AS TEXT) = 'Yes')) THEN CAST("data\_for\_tableau"."churn" AS TEXT) ELSE NULL END)) AS "TEMP(Calculation\_436849219392933897)(1914332625)(5)",

AVG("data\_for\_tableau"."bandwidth\_gb\_month") AS "avg:bandwidth\_gb\_month:ok",

AVG("data\_for\_tableau"."monthly\_charge") AS "avg:monthly\_charge:ok",

COUNT(CAST("data\_for\_tableau"."customer\_id" AS TEXT)) AS "cnt:customer\_id:ok",

CAST("data\_for\_tableau"."contract" AS TEXT) AS "contract",

MAX("data\_for\_tableau"."bandwidth\_gb\_month") AS "max:bandwidth\_gb\_month:ok",

COUNT((CASE WHEN (CAST("data\_for\_tableau"."churn" AS TEXT) = 'Yes') THEN CAST("data\_for\_tableau"."churn" AS TEXT) ELSE NULL END)) AS "usr:Calculation\_436849219397222416:ok"

FROM "public"."data\_for\_tableau" "data\_for\_tableau"

WHERE (CAST("data\_for\_tableau"."source" AS TEXT) = 'wgu')

GROUP BY 6

-- Kaggle Churn and Customer Count

SELECT COUNT(CAST("data\_for\_tableau"."churn" AS TEXT)) AS "TEMP(TC\_)(3605431932)(0)",

CAST("data\_for\_tableau"."churn" AS TEXT) AS "churn",

COUNT(CAST("data\_for\_tableau"."customer\_id" AS TEXT)) AS "cnt:customer\_id:ok"

FROM "public"."data\_for\_tableau" "data\_for\_tableau"

WHERE (CAST("data\_for\_tableau"."source" AS TEXT) = 'kaggle')

GROUP BY 2

-- Kaggle Average Tenure

SELECT AVG("data\_for\_tableau"."tenure") AS "avg:tenure:ok"

FROM "public"."data\_for\_tableau" "data\_for\_tableau"

WHERE (CAST("data\_for\_tableau"."source" AS TEXT) = 'kaggle')

HAVING (COUNT(1) > 0)

-- WGU and Kaggle Distinct Customers and Churned Customers

SELECT COUNT(DISTINCT (CASE WHEN (CAST("data\_for\_tableau"."source" AS TEXT) = 'wgu') THEN CAST("data\_for\_tableau"."customer\_id" AS TEXT) ELSE NULL END)) AS "TEMP(Calculation\_436849219392933897)(1913223596)(0)",

COUNT((CASE WHEN ((CAST("data\_for\_tableau"."source" AS TEXT) = 'wgu') AND (CAST("data\_for\_tableau"."churn" AS TEXT) = 'Yes')) THEN CAST("data\_for\_tableau"."churn" AS TEXT) ELSE NULL END)) AS "TEMP(Calculation\_436849219392933897)(1914332625)(0)",

COUNT((CASE WHEN ((CAST("data\_for\_tableau"."source" AS TEXT) = 'kaggle') AND (CAST("data\_for\_tableau"."churn" AS TEXT) = 'Yes')) THEN CAST("data\_for\_tableau"."churn" AS TEXT) ELSE NULL END)) AS "TEMP(Churn Rate (Int) (copy)\_436849219398893586)(311974858)(0)",

COUNT(DISTINCT (CASE WHEN (CAST("data\_for\_tableau"."source" AS TEXT) = 'kaggle') THEN CAST("data\_for\_tableau"."customer\_id" AS TEXT) ELSE NULL END)) AS "TEMP(Churn Rate (Int) (copy)\_436849219398893586)(3483384996)(0)"

FROM "public"."data\_for\_tableau" "data\_for\_tableau"

HAVING (COUNT(1) > 0)

-- Kaggle Zip Codes

SELECT "data\_for\_tableau"."zip" AS "zip"

FROM "public"."data\_for\_tableau" "data\_for\_tableau"

WHERE ((CAST("data\_for\_tableau"."source" AS TEXT) = 'kaggle') AND ((("data\_for\_tableau"."tenure" >= 1) AND ("data\_for\_tableau"."tenure" <= 72)) OR ("data\_for\_tableau"."tenure" IS NULL)))

GROUP BY 1

-- WGU Zip Codes

SELECT "data\_for\_tableau"."zip" AS "zip"

FROM "public"."data\_for\_tableau" "data\_for\_tableau"

WHERE ((CAST("data\_for\_tableau"."source" AS TEXT) = 'wgu') AND ((("data\_for\_tableau"."tenure" >= 1) AND ("data\_for\_tableau"."tenure" <= 72)) OR ("data\_for\_tableau"."tenure" IS NULL)))

GROUP BY 1

-- Kaggle City and Customer Count

SELECT CAST("data\_for\_tableau"."city" AS TEXT) AS "city",

COUNT(CAST("data\_for\_tableau"."customer\_id" AS TEXT)) AS "cnt:customer\_id:ok"

FROM "public"."data\_for\_tableau" "data\_for\_tableau"

WHERE ((CAST("data\_for\_tableau"."source" AS TEXT) = 'kaggle') AND ("data\_for\_tableau"."tenure" >= 1) AND ("data\_for\_tableau"."tenure" <= 72))

GROUP BY 1

-- WGU City, Churn, and Customer Count

SELECT CAST("data\_for\_tableau"."city" AS TEXT) AS "city",

COUNT(CAST("data\_for\_tableau"."churn" AS TEXT)) AS "cnt:churn:ok",

COUNT(CAST("data\_for\_tableau"."customer\_id" AS TEXT)) AS "cnt:customer\_id:ok"

FROM "public"."data\_for\_tableau" "data\_for\_tableau"

WHERE ((CAST("data\_for\_tableau"."source" AS TEXT) = 'wgu') AND ((("data\_for\_tableau"."tenure" >= 1) AND ("data\_for\_tableau"."tenure" <= 72)) OR ("data\_for\_tableau"."tenure" IS NULL)))

GROUP BY 1

**B1 -B7.**

Please see attached Panopto presentation: <https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=df009116-a990-4ef5-aa31-afca01778bed>

**C1.**

The purpose of my Tableau story and associated dashboards is to provide a comparison of the telecom company’s internet service with a competitor in the state of California, leading to customer growth, reduced churn, and increased tenure. My dashboards provided executive decision-makers at my telecom company with information related to KPI, customer geography, and highlighted service information such as bandwidth and contract type.

The function of my dashboard was to provide and compare the customer churn and tenure behavior of the provided telecom company data with the competitor. The provided telecom company is referred to as “Internal” and the competing telecom company which was selected is referred to as “External”. The External data was selected from the “Telecom Customer Churn Prediction” dataset, which was acquired on Kaggle, and transformed in PGAdmin using PostgreSQL along with the provided dataset to be used in Tableau (Zhuang, n.d.).

The dashboards align with the needs outlined in the chosen data dictionary by appropriately utilizing provided features including information related to churn, tenure, cost, service, and geography. My dashboard provides a visualized representation of key features within the dataset which can be used for executive decision-making.

**C2.**

Tableau was an appropriate business intelligence tool to use for the analysis for a variety of reasons, including its powerful data representation, visualization, and interactivity features, as well as its ability to use a live PostgreSQL database. Tableau allowed me to create professional interactive dashboards that can be used for executive decision-making at the company. Tableau’s ability to connect to a live database ensured that my data representations remained accurate and up to date if changes were made to the supporting PostgreSQL tables. Additionally, Tableau’s visualization, dashboard, and story tools helped me clearly represent patterns in the data in an engaging format for business decision-making. An example of how this was useful is how I was able to summarize key observations into three visual dashboards within the tabbed Tableau Story used for the presentation.

**C3.**

I used two datasets for the analysis: the provided telecom churn data stored in a series of PostgreSQL tables in PGAdmin and the Kaggle dataset containing matching churn data for a competing telecom company in the state of California (Zhuang, n.d.).

To begin data cleaning and preparation, I created a new table called "wgu\_data\_for\_tableau" by selecting relevant data from various tables using JOIN operations on their respective IDs. I added a foreign key constraint to ensure referential integrity and removed unnecessary columns, filtering only those relevant to the Tableau analysis. I converted the bandwidth column from annual to monthly by dividing values by 12 and deleted rows where the state column was not "CA" as the analysis focused only on a comparison with a competitor in California. I then dropped the state column and rounded the bandwidth and monthly charge numbers for analysis purposes. The "source" column was added to differentiate between the two datasets used in the analysis, allowing for easy filtering of the final table within Tableau and I labeled the internal data as “wgu”.

Next, I created a new table called "kaggle\_data\_for\_tableau" to hold the external dataset acquired from Kaggle and added a “source” column with the default value of “kaggle”. I added the necessary columns to match the external dataset, imported the data, and renamed column titles. I updated the rows in the "Churn Category" column to "Yes" or "No" depending on their previous values, and I filtered the table to keep only the necessary columns. I then created a new table called "kaggle\_data\_for\_tableau\_temp," with only columns being used in the analysis imported into the table, dropped the old table, and renamed the new table "kaggle\_data\_for\_tableau". I also deleted rows with null values in the bandwidth column as they were not relevant to the analysis.

Finally, I combined the two tables using UNION to create a new table called "data\_for\_tableau.”

**C4.**

The dashboards were designed to highlight key features of the dataset related to churn, tenure, cost, service, and geography.

I created a series of worksheets to develop a variety of data representations including pie charts, service summary tables, Key Performance Indicators (KPI), maps of ZIP code served, and bar charts of the most served cities. I divided my presentation into three tabbed dashboards within a Tableau story labeled KPI, Geographical, and Service. The worksheets were placed in the dashboards according to the category of the data to best emphasize the presentation’s overall priorities.

Each of the tabs had visualizations which provided information about company performance, customer geography, and service details for both the provided telecom company and the external competitor. In the Geographical tab, I included interactive controls such as the churn status filter and the customer tenure slider that allowed users to filter cities and zip codes based on customer tenure in months.

As noted, I selected an external dataset with the same features as the provided dataset, including demographic, service, churn, payment, and tenure data. All the data representations in the dashboards included side-by-side comparisons of the provided data as well as the external competitor data. This allowed for one-for-one comparisons of customer trends and preferences.

**C5.**

Through my data analysis, I was able to create informative and visually appealing dashboards that supported executive decision-making. As a result of the data analysis, executive decision-makers were provided with clear and actionable insights into customer behaviors and preferences such as the relationship between increased bandwidth and increased tenure. Additionally, the analysis highlighted opportunities for growth and improvement, such as the opportunity to expand to the many unserved ZIP codes or the opportunity to promote the much larger maximum bandwidth offerings, approaching 600/mbps compared to only 85/mpbs for the competitor.

Using visualizations and interactive controls made it easier to interpret the data and identify trends and patterns, enabling business decision-makers to make informed decisions to positively impact customer satisfaction and improve company performance. As an example, the pie charts labeled 'Churn Count by Contract Type' illustrated the comparatively high churn rate of customers on month-to-month contracts, which was a valuable observation that can be used for decision-making related to contract offerings. According to the Centers for Disease Control and Prevention (2018), a pie chart can be useful for displaying categorical data in an easily interpreted format. Another example is the 'ZIP Codes Served' maps which highlight the opportunity to service more ZIP codes and compete with the established competitor, which highlighted growth opportunities, a potential area for business improvement.

My analysis provides executives with a clear view of key performance indicators, customer geography, and service data analysis, empowering them to make more informed decisions which can positively impact customer satisfaction, reduce churn, increase tenure, and ultimately improve company performance.

**C6.**

My analysis had several limitations including a lack of detailed customer data including demographic data, customer interactions, or information about the executive decision maker’s priorities, past marketing, or retention efforts. Details about income and education level, for example, may have provided insights into customer behavior in different areas. Including more information regarding customer interactions could have provided useful information about customer churn or shown customer service activity that could be adjusted to improve tenure. More information about the company would have allowed for an analysis more reflective of executive priorities and included an evaluation of existing business strategies and their effectiveness. The limitations of my data analysis likely impacted the accuracy of its findings. Any future analysis would need to take these limitations into consideration.

**E.**

Zhuang, S. L. (n.d.). Telecom Customer Churn Prediction. Kaggle. <https://www.kaggle.com/datasets/shilongzhuang/telecom-customer-churn-by-maven-analytics>

Centers for Disease Control and Prevention. (2018, August). Evaluation Briefs. Using graphs and charts to illustrate. Retrieved from <https://www.cdc.gov/healthyyouth/evaluation/pdf/brief12.pdf>