**D211 Advanced Data Acquisition**

Nicole Reiswig

College of Information Technology, Western Governors University

Master of Science in Data Analytics

William Sewell

February 04, 2024

**Table of Contents**

**Part 1: Data Dashboard**

A. Dashboard 3

**Part 2: Demonstration**

B. Panopto recording 8

**Part 3: Report**

C. Report 8

D. Sources 13

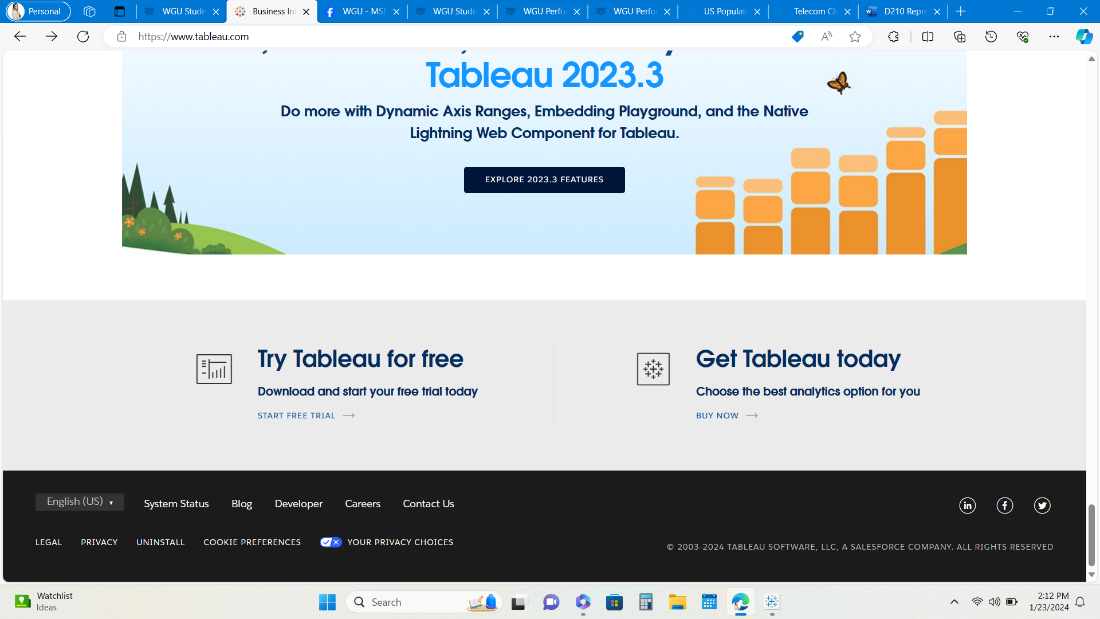
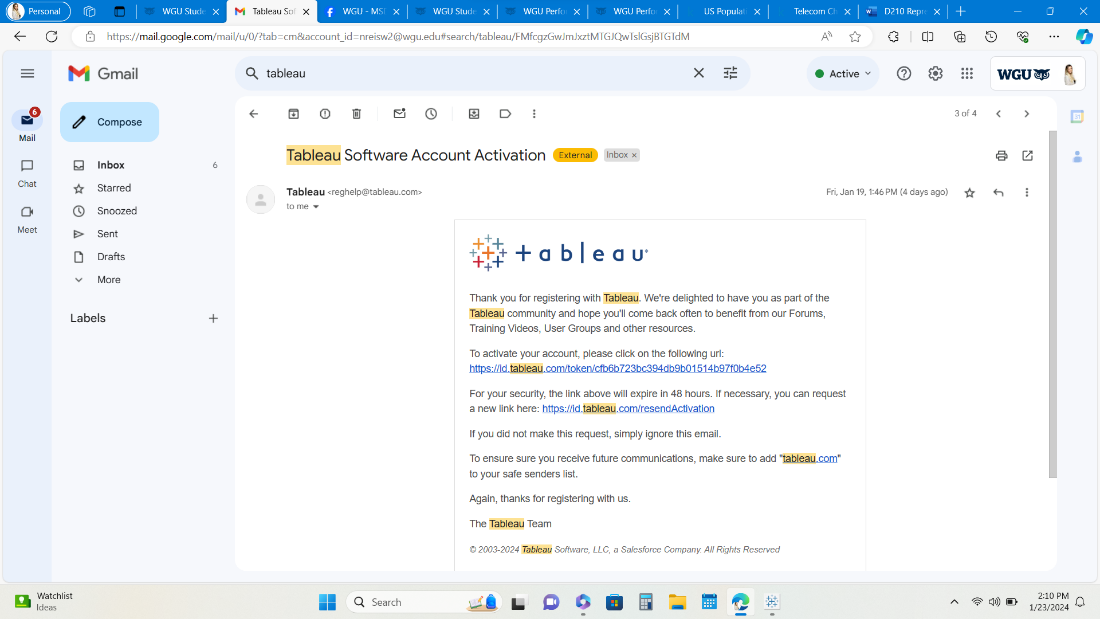
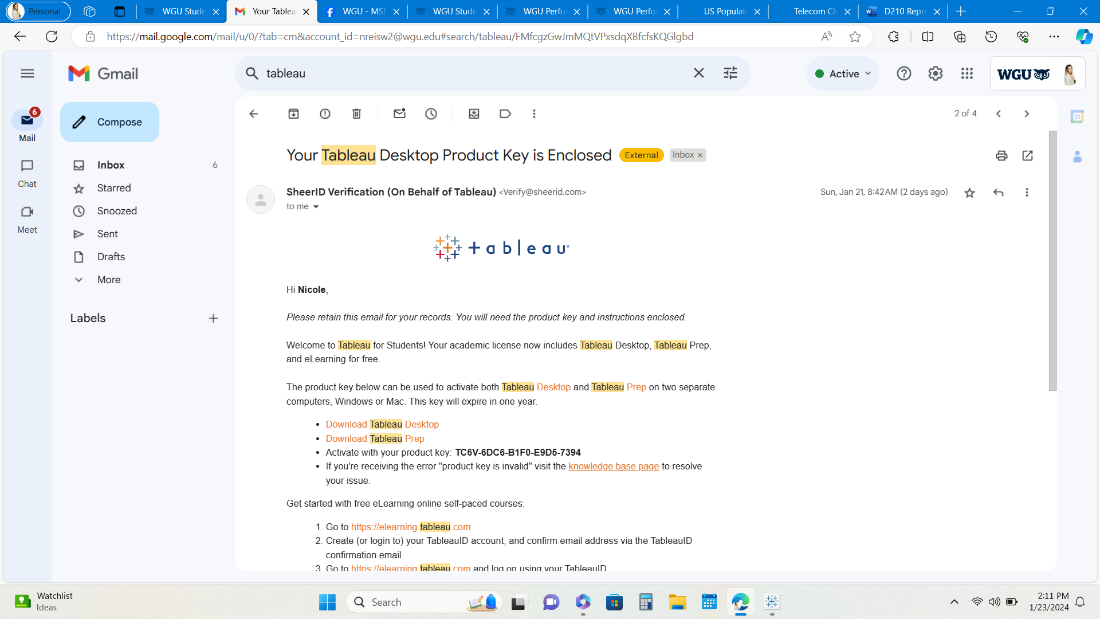
# Advanced Data Acquisition

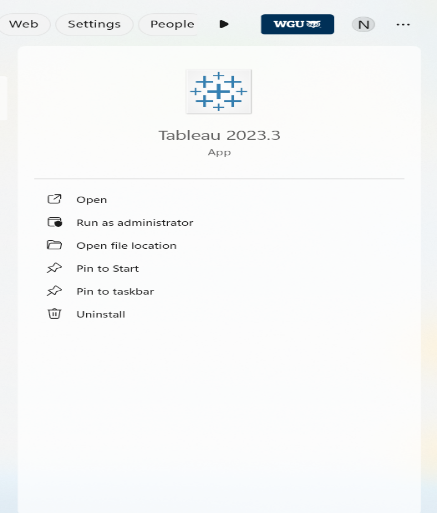
**Part 1: Data Dashboard**

A. See the attached Tableau dashboard that supports executive decision-making.

1. See the attached CSV files churn\_clean and population by zip that serve as the data source for the dashboard.

2. Instructions for Tableau installation:

* 1. Visit website tableau.com
  2. Download Tableau for desktop
  3. Follow download prompts
  4. Request student access and create an account 
  5. Utilize the passcode e-mailed to you to access Tableau Student for free
  6. Sign in to the Tableau desktop



1. For this course, we will utilize the labs' on-demand environment. You will navigate to your course performance assessment task overview page and scroll to the bottom. You will select the link labeled labs on demand.
2. Once you are in the labs on demand environment you will open pgAdmin and Tableau desktop apps. You will use the credentials server: localhost, port: 5432, database: churn, username: Postgres, password: Passw0rd! to log into Tableau.
3. In Postgres pgAdmin you will utilize tables in the churn database already provided. The tables utilized are customer and location. The SQL used to create these provided tables is as follows:

CREATE TABLE public.customer

(

    customer\_id text COLLATE pg\_catalog."default" NOT NULL,

    lat numeric,

    lng numeric,

    population integer,

    children integer,

    age integer,

    income numeric,

    marital text COLLATE pg\_catalog."default",

    churn text COLLATE pg\_catalog."default",

    gender text COLLATE pg\_catalog."default",

    tenure numeric,

    monthly\_charge numeric,

    bandwidth\_gp\_year numeric,

    outage\_sec\_week numeric,

    email integer,

    contacts integer,

    yearly\_equip\_faiure integer,

    techie text COLLATE pg\_catalog."default",

    port\_modem text COLLATE pg\_catalog."default",

    tablet text COLLATE pg\_catalog."default",

    job\_id integer,

    payment\_id integer,

    contract\_id integer,

    location\_id integer,

    CONSTRAINT customer\_pkey PRIMARY KEY (customer\_id),

    CONSTRAINT customer\_contract\_id\_fkey FOREIGN KEY (contract\_id)

        REFERENCES public.contract (contract\_id) MATCH SIMPLE

        ON UPDATE NO ACTION

        ON DELETE NO ACTION

        NOT VALID,

    CONSTRAINT customer\_job\_id\_fkey FOREIGN KEY (job\_id)

        REFERENCES public.job (job\_id) MATCH SIMPLE

        ON UPDATE NO ACTION

        ON DELETE NO ACTION

        NOT VALID,

    CONSTRAINT customer\_location\_id\_fkey FOREIGN KEY (location\_id)

        REFERENCES public.location (location\_id) MATCH SIMPLE

        ON UPDATE NO ACTION

        ON DELETE NO ACTION

        NOT VALID,

    CONSTRAINT customer\_payment\_id\_fkey FOREIGN KEY (payment\_id)

        REFERENCES public.payment (payment\_id) MATCH SIMPLE

        ON UPDATE NO ACTION

        ON DELETE NO ACTION

        NOT VALID

)

TABLESPACE pg\_default;

ALTER TABLE public.customer

    OWNER to postgres;

-- Table: public.location

-- DROP TABLE public.location;

CREATE TABLE public.location

(

    location\_id integer NOT NULL,

    zip integer,

    city text COLLATE pg\_catalog."default",

    state text COLLATE pg\_catalog."default",

    county text COLLATE pg\_catalog."default",

    CONSTRAINT location\_pkey PRIMARY KEY (location\_id)

)

TABLESPACE pg\_default;

ALTER TABLE public.location

    OWNER to postgres;

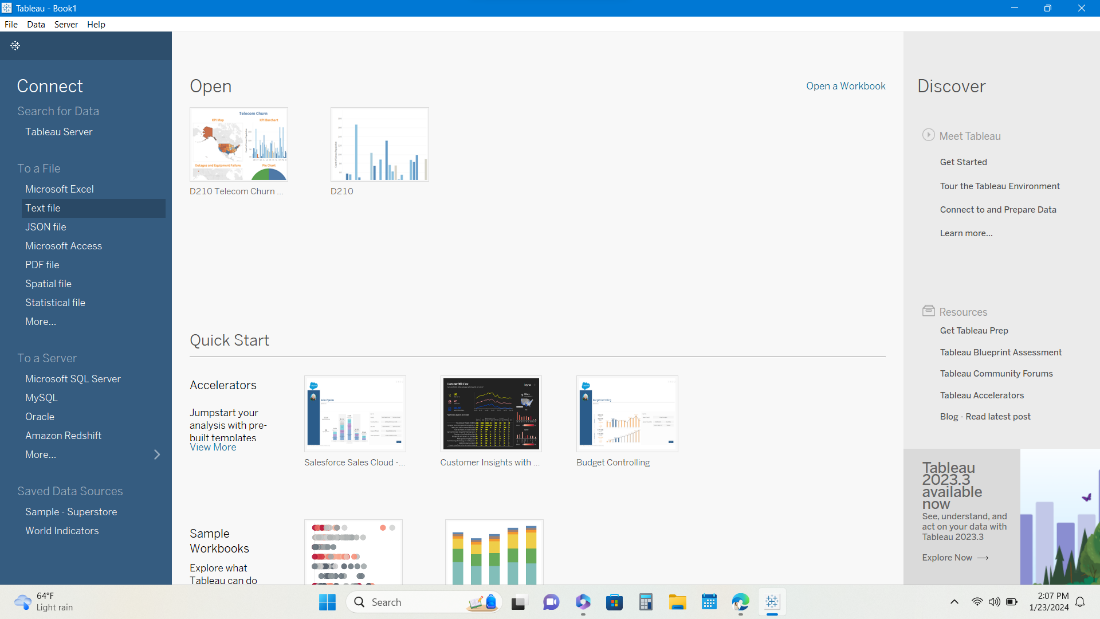
Once signed into Tableau to connect to Postgres pgAdmin you will scroll down to “servers” on the left side in the dark blue and select Postgres. If you do not see that option expand the menu A screenshot of a computer

Description automatically generated

1. Then, you will connect to the CSV file population\_by\_zip.csv by selecting the text option. Once you have successfully connected Postgres pgAdmin and the CSV file to your Tableau desktop you can then recreate the visualizations or open the .twbx file with the visualizations already created.
2. To open the .twbx file provided in my submission, save the file to your computer, navigate to Tableau, select File, and then select Open. Search your computer for the saved .twbx file and select it. This will open the Tableau dashboards.

3. Instructions to navigate the dashboard:

* 1. Select the type of file you’d like to utilize from the column on the left. I connected to pgadmin under the server option.



* 1. Select the second data source you’d like to utilize. I selected a text file.
  2. Open a sheet from the bottom left tabs and begin creating your visualization. I created 4 sheets with 4 visualizations, 1 on each sheet.
  3. Open a dashboard from the bottom left tabs and then select the sheets from the list in the left column that you’d like to display on your dashboard. Customize to create a visually appealing dashboard that meets the needs of your project.

4. See the attached SQL code.

**Part 2: Storytelling with Data**

B. See attached Panopto recording.

**Part 3: Reflection Paper**

C.

1. The goals of the telecom company are to both acquire new customers and retain highly profitable customers. It is more important for the telecom company to retain customers that bring in high profits. This is because it costs 10 times more to acquire a new customer than it does to retain an existing one. The purpose and function of my dashboard align with the needs of the telecom company. In my dashboard, we look at three charts and one map. In these visualizations, we can compare the count of customers who churn and who do not, we look at each state's count of equipment failures, percentage of equipment failures, where they rank in equipment failures compared to other states, and then we look at each state count of outage seconds per week and their rank compared to other states. In the map we can compare the total telecom customers by location and the total population for that state by incorporating the population\_by\_zip census data. This helps us to identify locations that are doing substantially well in providing services and areas that the business may need to focus on improving as well as locations we could expand.

2.  Tableau is the business intelligence tool utilized because it is a powerful data visualization platform that allows users to connect to other servers and data sources and then customize the visualizations' look and feel to meet their business needs.

3.  To clean and prepare the data sets the steps that must be taken are to check for missing or null values, duplicates, and outliers. If there are missing, null values, duplicates, or outliers the decision must be made to impute, retain, or exclude this from the analysis while also preserving the sample size and integrity of the data. The churn\_clean data set was provided already cleaned so no preparation was necessary. The population by zip data set was examined in RStudio for missing, nulls, and duplicates which did not contain any, therefore we moved forward with analyzing the data set.

4. In creating the dashboard, the first step was to log into the labs' on-demand environment. I then signed into Postgres pgAdmin. In this environment, the Churn data set was already provided. In Tableau I created a join for two of the tables customer and location from the Churn data set that I needed for the analysis utilizing the location\_id feature. I then created a relationship between the custom sql query with the customer and location tables to the population\_by\_zip.csv using the zipcode feature from both datasets. Once all of my data was successfully connected to Tableau, I created a relationship between the two data frames utilizing the column location\_id.

SELECT "customer"."age" AS "age",

  "customer"."bandwidth\_gp\_year" AS "bandwidth\_gp\_year",

  "customer"."children" 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",

   CAST("location"."city" AS TEXT) AS "city",

  CAST("location"."county" AS TEXT) AS "county",

  "location"."location\_id" AS "location\_id (location)",

  CAST("location"."state" AS TEXT) AS "state",

  "location"."zip" AS "zip"

FROM "public"."customer" "customer"

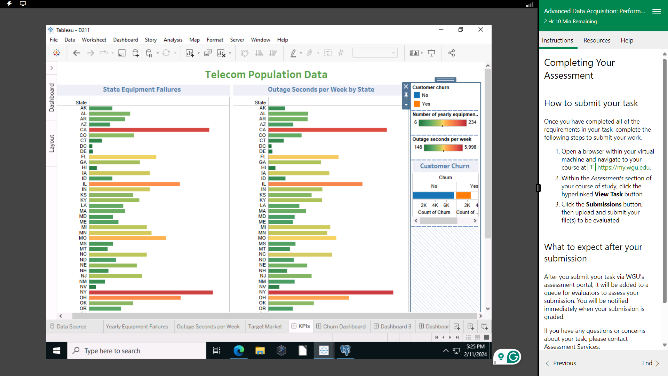
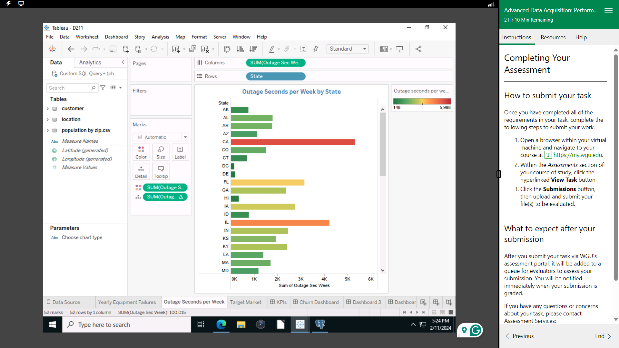
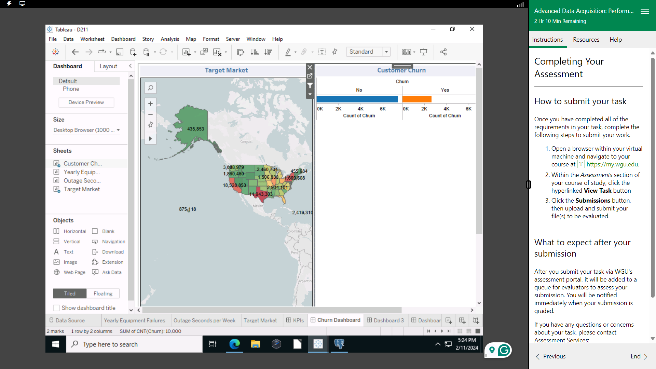
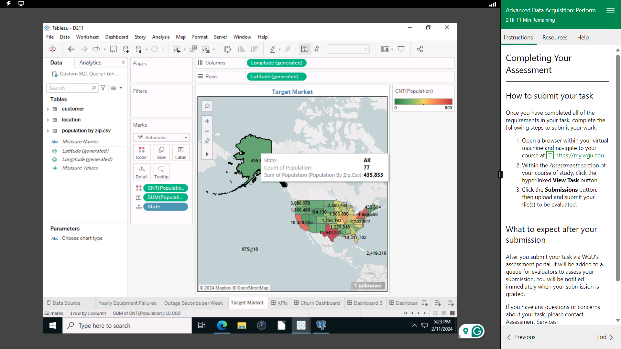
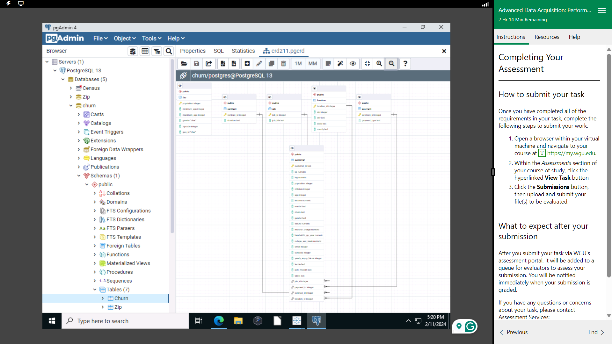
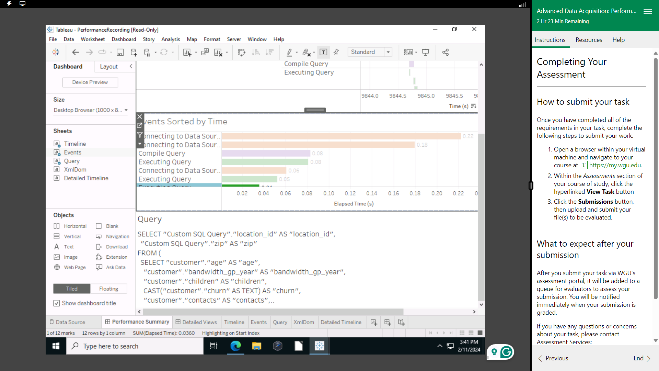
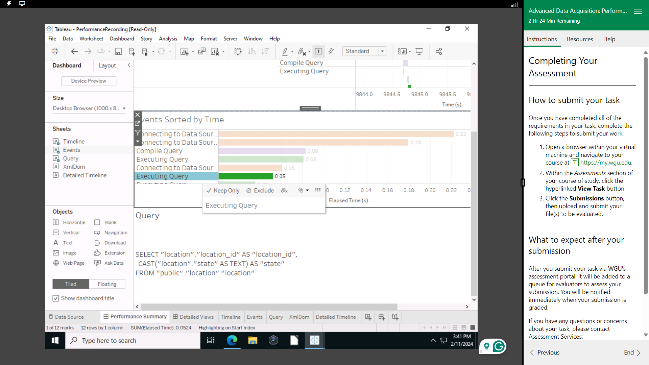
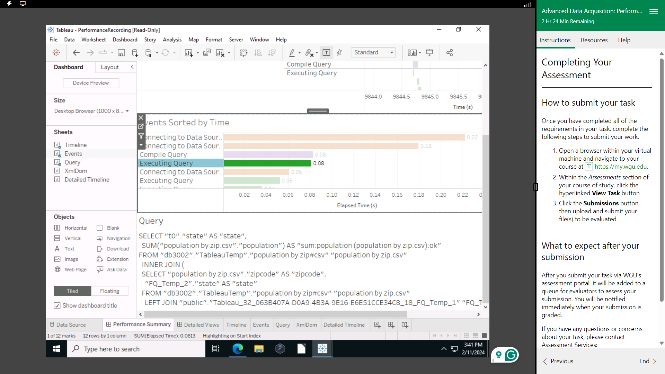
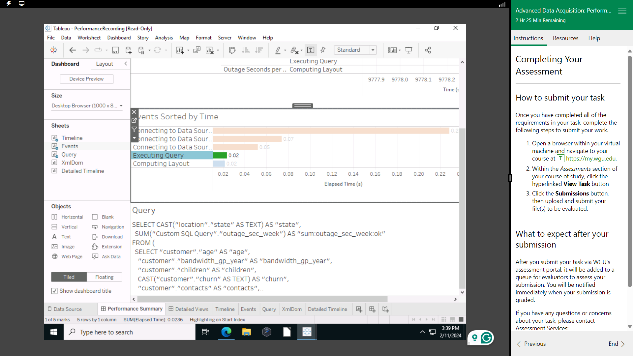
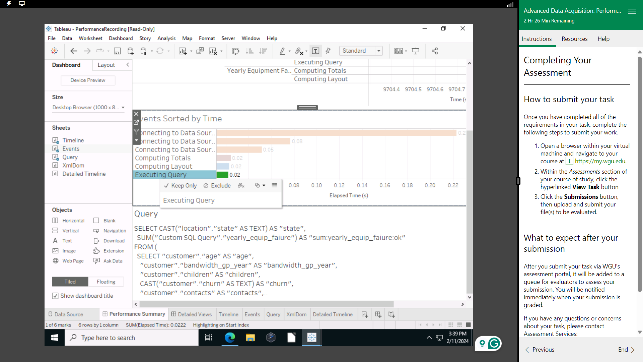
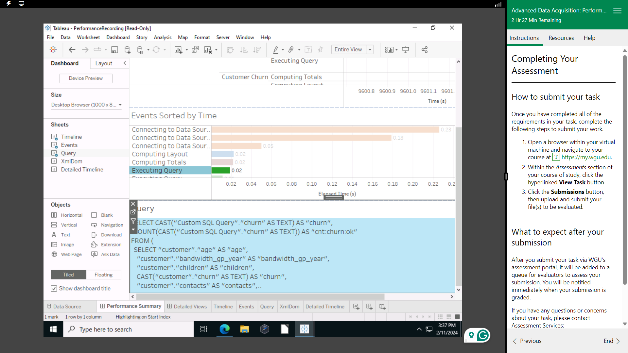
LEFT JOIN location

ON customer.location\_id = "location".location\_id

A screenshot of a computer

Description automatically generated

I opened a new sheet and began creating my visualizations. Once they were created to answer the business questions they were then added to a dashboard for presentation.



5.  The data representations in my presentation support the story I wanted to tell because it gives us the ability to see the areas in which we have room to improve and the locations where it is needed the most. There is room for improvement and increasing revenue. These visualizations give us the KPI data to make business decisions. In these charts, we visualize the number of customers we retain and the number of customers that churn. We then analyze two product performance areas yearly equipment failures and outage seconds per week and compare these figures by state. Making business decisions based on this data and improving performance will ultimately help us retain our current customers and know how much room we can expand in each state.

6.  The limitation of this analysis is that we have not yet looked at the percent of the population these figures represent.

D. No sources were cited in the creation of this paper.

[US Population By Zip Code (kaggle.com)](https://www.kaggle.com/datasets/census/us-population-by-zip-code)

<https://www.kaggle.com/datasets/census/us-population-by-zip-code>