

Western Governors University

# Advanced Data Acquisition

D211

Allison Casey  
11-6-2024

---

## Part 1: Data Dashboards

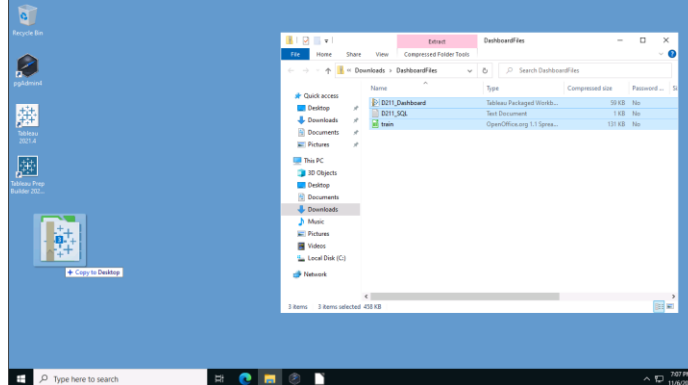
---

### A1: DATASETS AND DASHBOARD FILE

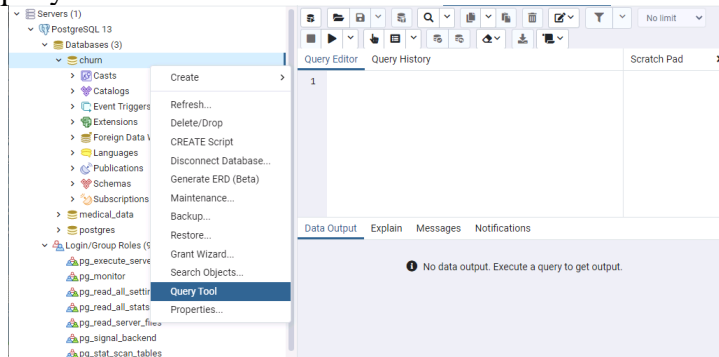
1. Data set 1: churn\_clean.csv
2. Data set 2: train.csv
3. Dashboard: D211\_Dashboard.twbx

### A2: DASHBOARD INSTALLATION

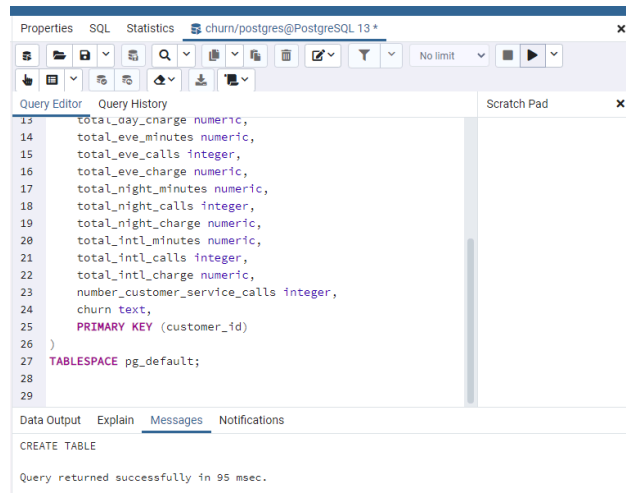
1. In the default version of Labs on Demand unzip the DashboardFiles.zip to the desktop



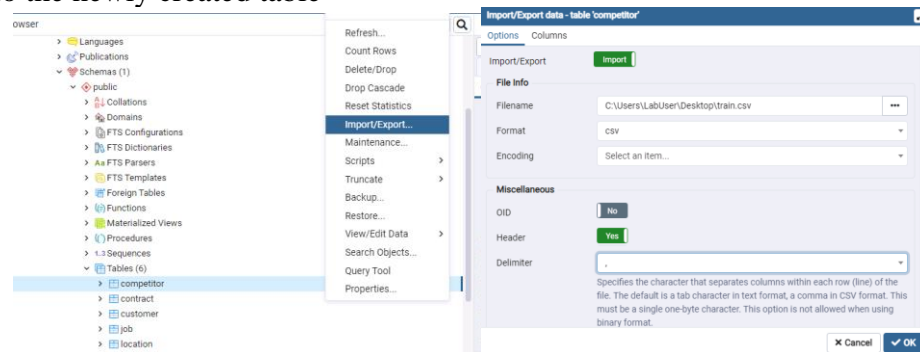
- a.
  - b. You should see three files on the desktop
    - i. D211\_Dashboard.twbx
    - ii. D211\_SQL.txt
    - iii. train.csv
2. Open pgAdmin4, navigate to the churn database, right click on the churn database and select query tool



- a.
3. Copy the contents of D211\_SQL.txt up to the section that fills in the table into the query editor a click the execute button. This will create the table.



- a.
- b. Note that the SQL to fill the table is also included, but I found that there would occasionally be permissions issues with the file when downloading it so this way those permissions issues are avoided by using the import/export tool
4. Open the import/export tool for the competitor table that was just created that will now be under the Schemas/public/Tables dropdown and use that tool to import the data from the csv into the newly created table



- a.
- b. The database should now be properly set up
5. Go back to the desktop and open D211\_Dashboard.twbx
  - a. A pop up should come up requesting the password for the database which is Passw0rd!
  - b. Once this is entered the dashboard should be ready to navigate and use

### A3: DASHBOARD NAVIGATION

The dashboard is comprised of five visualizations with two interactive controls. The first two visualizations display the rate of churn for both providers where in this case the data from the external public dataset is being considered the competitor. Hovering over any of the bars for the first two graphs will display the count and percentage of the total that either churned or did not churn. The next two visualizations display the average number of customer contacts compared to whether the customer churned for both data sets and the exact numbers can be shown by hovering over the bars. To the right of these the legend shows the colors indicating churn and no churn for the top four visualizations as well as the interactive state filter which can filter the results displayed by state or states selected for all five visualizations. There is also a churn filter to filter all the results for churn and no churn. The bottom visualization displays a map colored by state to easily see the different states and only shows data for the main data set and not the

competitor. Each state can be hovered over to view the average monthly charge, the average yearly equipment failure, and the average tenure for customers in that state.

#### **A4: SQL CODE**

SQL code provided in D211\_SQL.txt as well as added here:

```
-- SQL to create table from external dataset -
CREATE TABLE public.competitor
(
    customer_id integer NOT NULL,
    state text,
    account_length integer,
    area_code varchar,
    international_plan text,
    voice_mail_plan text,
    number_vmail_messages integer,
    total_day_minutes numeric,
    total_day_calls integer,
    total_day_charge numeric,
    total_eve_minutes numeric,
    total_eve_calls integer,
    total_eve_charge numeric,
    total_night_minutes numeric,
    total_night_calls integer,
    total_night_charge numeric,
    total_intl_minutes numeric,
    total_intl_calls integer,
    total_intl_charge numeric,
    number_customer_service_calls integer,
    churn text,
    PRIMARY KEY (customer_id)
)
TABLESPACE pg_default;
--fill the table-
COPY public.competitor (customer_id, state, account_length, area_code,
international_plan, voice_mail_plan, number_vmail_messages, total_day_minutes,
total_day_calls, total_day_charge, total_eve_minutes, total_eve_calls,
total_eve_charge, total_night_minutes, total_night_calls, total_night_charge,
total_intl_minutes, total_intl_calls, total_intl_charge,
number_customer_service_calls, churn) FROM 'C:/Users/LabUser/Desktop/train.csv'
DELIMITER ',' CSV HEADER QUOTE '\"' ESCAPE '\"';
```

---

### **Part 2: Demonstration**

---

#### **B: PANOPTO PRESENTATION**

<https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=67121677-e890-4cca-a9aa-b22100288865>

---

### **Part 3: Report**

---

#### **C1: DASHBOARD ALIGNMENT**

The scenario for the performance assessment explains that the task is to build a data dashboard to be able to explore the data for this company that serves customers in all regions of the United States. The data dictionary also explains churn and the importance of this metric since retaining customers is much cheaper than having to acquire new ones. The dashboard meets these needs through being an easy to navigate, interactive visual of this key information coming from the data set. The dashboard focuses on the churn rate and other factors that have been determined as some of the more statistically significant features regarding churn and tenure in previous performative assessments. Along with the company data the dashboard also incorporates the data of a theoretical competitor so that there can be a comparison between the two.

## **C2: BUSINESS INTELLIGENCE TOOL**

The business intelligence tool used for this was Tableau. It was chosen because it is relatively easy to use and interfaces well with the PostgreSQL database. Overall, it is a very accessible tool with loads of built in features that make it a great choice especially to create interactive data visualizations.

## **C3: DATA CLEANING**

As both data sets were already clean and the data used from each set was consistent and accurate there was no need for additional cleaning.

## **C4: DASHBOARD CREATION**

To create the dashboard the first step once the data source was connected was to drag the necessary tables into the data area and set up the relationships between them so that they were connected based on the common data shared which was the state related to each record. The three tables used were the competitor table and the location table which were matched up based on the state column and the customer table was matched to the location table based on the location id which created the relationship between the three tables. Then the sheets could be created for each visualization using the connected data sources and dragged into the dashboard once completed. Since the data was set up connected then it was easy to have the filters adjust all the visualizations on the overall dashboard.

## **C5: DATA ANALYSIS RESULTS**

The purpose and function of the dashboard is to ideally help prevent churn. The dashboard helps accomplish this through being an easy to navigate, interactive visual of this key information coming from the data set. The dashboard focuses on the churn rate and other factors that have been determined as some of the more statistically significant features regarding churn and tenure in previous performative assessments. Along with the company data the dashboard also incorporates the data of a theoretical competitor so that there can be a comparison between the two. This secondary data set is helpful because it allows the company to understand its competitors, identify market gaps and opportunities for growth, and make strategic decisions to help the company better itself over its competitors (Revelio Labs, 2024). One of the clear points from the analysis using the dashboard is that the competitor has a lesser percentage of churn but also a greater average number of customer contacts. This shows that the company can work towards lowering its churn and potentially needing to get a better understanding of how the customer contacts at the competitor seem to have such a different impact on the churn outcome. The dashboard also allows this to be compared by state so we can see that some states are doing

better than others in terms of how the company is doing with churn compared to the competitor while others are doing worse.

## C6: ANALYSIS LIMITATIONS

There are a few key limitations to note for this data analysis. One of the major limitations is the competitor data is limited in quantity and has much less than the churn data set which is also relatively small. More data in both data sets would contribute to better analysis. Another limitation is not having a lot of background information about the competitor or other fields that match up with the churn data set. This prevents further analysis when comparing the two because only a few fields can be looked at for comparison.

---

## Sources

---

“How Does Your Company Compare to Your Competitors?” *Revelio Labs*,  
[www.reveliolabs.com/blog/how-does-your-company-compare-to-your-competitors/](http://www.reveliolabs.com/blog/how-does-your-company-compare-to-your-competitors/).  
Accessed 22 Sept. 2024.

Jeffrey Shaffer COO and VP of Information Technology and  
Analytics. “5 Tips on Designing Colorblind-Friendly Visualizations.” *Tableau*, 20 Apr.  
2016, [www.tableau.com/blog/examining-data-viz-rules-dont-use-red-green-together](http://www.tableau.com/blog/examining-data-viz-rules-dont-use-red-green-together).

M&ouml;lbius. “Telecom Churn Dataset.” *Kaggle*, 12 Apr. 2023,  
[www.kaggle.com/datasets/arashnic/telecom-churn-dataset](http://www.kaggle.com/datasets/arashnic/telecom-churn-dataset).

*Section508.Gov*, [www.section508.gov/develop/universal-design/](http://www.section508.gov/develop/universal-design/). Accessed 22 Sept. 2024.