**Dashboard for Monitoring Broadband Availability, Internet Reliability, and Customer Distribution**

Kelseyann Wright

Student ID: 00565814

College of Information Technology, Western Governors University

Advanced Data Acquisition - D211

Dr. William Sewell

September 15, 2023

**Dashboard for Monitoring Broadband Availability, Internet Reliability, and Customer Distribution**

**PART 1: DATA DASHBOARDS**

**A. DATA DASHBOARDS**

The tableau dashboard was created using Tableau Desktop on my personal computer. Then, the dashboard was uploaded to Tableau Public for sharing.

**A1. BOTH DATA SETS**

The two dataset sources for the dashboard are attached. The first data source, is the churn database in PostgreSQL provided internally. The second data source, “broadband\_data\_2020October\_ORIGINAL.csv” includes county level broadband internet availability and usage data released by Microsoft and the FCC (Thomas, 2022). This broadband dataset was pre-processed in PostgreSQL (see section A4). The processed data table was exported from PostgreSQL submitted (broadband\_processed.csv).

**A2. DASHBOARD INSTALLATION**

The dashboard can easily be accessed via Tableau Public by clicking this link: <https://public.tableau.com/shared/WG2YXG9S8?:display_count=n&:origin=viz_share_link>

**A3. DASHBOARD NAVIGATION**

The top of the dashboard contains the title on the left, and a comparison of average broadband use and broadband availability for the broadband dataset population. Below that, there are three bar graphs comparing the number of customers, outage seconds, equipment failure, and churn proportions for the different internet service types (DSL, Fiber Optic, Non). These bar graphs will allow analysts to determine if the type of internet service is related to outages, equipment failure, and customer churn.

Below the bar graph is a map of broadband availability and customer distribution. The legends can be found above the map. The map tools allow the user to zoom in/out on certain areas of the map and select specific map areas. Hovering the mouse pointer over data points allows the user to view a tool tip displaying the data.

**A4. SQL CODE**

The external broadband dataset (Thomas, 2022) was imported into the churn database and pre-processed in PostgreSQL. First, the dataset was imported into the database as a table titled “broadband”. Then, the query tool in PostgreSQL was used to pre-process the data. For pre-processing, the values in the “County Name” field of the broadband dataset were adjusted to match the “County” field in the customer churn dataset so that “county” serve as a key for joining the table to the other tables in the database for use in Tableau. In the broadband dataset, the county names included the word “county”, “municipality”, or “borough” after the name, which is inconsistent with the customer churn dataset “county” format. Therefore, those ending strings were removed from the broadband dataset “country name” column. The SQL code used was:

UPDATE broadband

SET county\_name = REPLACE (county\_name, ' County', '');

UPDATE broadband

SET county\_name = REPLACE (county\_name, ' City and Borough', '');

UPDATE broadband

SET county\_name = REPLACE (county\_name, ' Borough', '');

UPDATE broadband

SET county\_name = REPLACE (county\_name, ' Census Area', '');

UPDATE broadband

SET county\_name = REPLACE (county\_name, ' Municipality', '');

UPDATE broadband

SET county\_name = REPLACE (county\_name, ' Parish', '');

UPDATE broadband

SET county\_name = REPLACE (county\_name, ' city', '')

Custom SQL Queries were also used within Tableau after linking the PostgreSQL database server to Tableau. Four tables from the database were used: broadband (the external dataset), location, customer, and services. A join was used to establish the relationship between the tables.

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 (customer)",

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",

broadband.broadband\_avail,

broadband.broadband\_use,

broadband.county\_name,

broadband.state,

services.Internetservice

FROM "public"."customer" "customer"

JOIN services

ON customer.customer\_id = services.customer\_id

JOIN location

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

JOIN broadband

ON location.county = broadband.county\_name AND location.state = broadband.state

**PART 2: DEMONSTRATION**

**B. PANOPTO PRESENTATION**

The Panopto multimedia presentation can be accessed here: <https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=91726bcd-f846-4501-bbe0-b0a201164ecc>

**PART 3: REPORT**

**C1. DASHBOARD ALIGNMENT**

A primary need indicated by the data dictionary is retaining profitable customers and reducing customer churn. To meet this need, it is important to identify customer characteristics that are associated with churn and longer tenure. During data exploration, a strong relationship between customer tenure and bandwidth use was discovered. In addition, longer contract lengths are associated with less churn. This led to the question, “How does internet use and availability impact customer retention?” Incorporating information about broadband internet availability into the dashboard will allow us to 1) identify patterns between internet availability and customer churn and 2) identify areas where there is internet available for expanding our current customer base. Finally, the dashboard also includes the distribution of customer contract types so that we can monitor customer churn for different contracts.

**C2. BUSINESS INTELLIGENCE TOOL**

Tableau was the chosen business intelligence tool for generating the dashboard. Tableau is a widely used business intelligence tool, with both public and professional/private servers. This allows dashboards generated in Tableau to be easily shared among collaborators. In addition, Tableau allows for direct connection to PostgreSQL databases, making it an ideal tool for the use with the provided Churn database. Connecting to the database allows for efficient and immediate dashboard updates with changes to the database if a “live” connection is enabled.

A screenshot of a computer

Description automatically generated**C3. DATA CLEANING**

PostgreSQL was used to check for duplicates in the customer table of the churn database. No duplicate records were found. See section A4 of this document and the attached SQL script (BroadbandData\_PreProcessing.sql) for a description of how the external dataset was prepared using PostgreSQL and the associated code.

**C4. DASHBOARD CREATION** The dashboard was created using the following steps.

1. In Tableau, a connection was added to the Churn database in PostgreSQL. This allows access to the data tables in the PostgreSQL database from within Tableau.
2. The tables to be used in the dashboard (broadband, customer, location, and services) were dragged into the data workspace in tableau.
3. Custom SQL code was used within Tableau to join the data tables (see section A4 for code).
4. Data visualizations were created on separate worksheets in Tableau. The visualizations included 3 bar graphs, a map, and a data table.
5. The visualizations were added to a dashboard, organized to tell a cohesive story with the data. The dashboard used the colorblind color palette for accessibility of all audience members.
6. The Tableau dashboard was exported as a packaged workbook and uploaded with this performance assessment.

**C5. DATA ANALYSIS RESULTS**

Key findings from the data analysis were:

1. The percent of the population using broadband is less than the amount of broadband available to customers, meaning there is enough internet access available to allow for expansion of the customer base.
2. Most customers have fiber optic internet service. The type of internet service (DSL, fiber optic, or none) is does not appear to be strongly associated with customer churn rates.
3. The average outage seconds per week and annual equipment failure is not dependent on the type of internet service, indicating that both DSL and fiber optic are equally as reliable.
4. The map shows that there are many counties with high broadband availability and low customer counts, indicating areas for ideal expansion of the customer base.

**C6. ANALYSIS LIMITATIONS**

One limitation of this analysis is that the broadband dataset is collected at the county level, meaning that all analyses involving that data in conjunction with the churn dataset must include grouping or averaging customers by county, or applying county-wide data to specific customers. This reduces the specificity of our dataset and limits our ability to provide insights on individual customers or smaller geographic areas with this dataset. In addition, the broadband dataset does not include data for Puerto Rico, so we are unable to use those customers in this analysis. The broadband dataset is also from 2020, which is fairly recent, but it would be better to have current data to compare with our customer database.

Another issue with this analysis is that county name had to be used as the key for linking the churn dataset with the broadband dataset. Using text fields as keys can be problematic and increase the probability of error when associating the datasets. Therefore, careful pre-processing was required to relate the two datasets.

**D. WEB SOURCES**

Thomas, A. 2022. Broadband usage in US. Dataworld. https://data.world/amberthomas/broadband-usage-in-us

**E. SOURCES**

No other sources were used for this work.