

D211: Advanced Data Acquisition

Performance Assessment – Task 1

Leng Yang

Student ID: 012298452

WGU MSDA

A. Data Dashboards

The dashboard is attached as a packaged Tableau file named “D211_PA_LY.twbx.”

A1. Datasets and Dashboard File

The internal dataset chosen was the Churn dataset, provided by WGU, and was already available in pgAdmin.

The external data set is submitted alongside the report as “telco_churn.csv.” This data set was sourced from Kaggle by user BlastChar (2018).

A2. Dashboard Installation

The dashboard from the Tableau file “D211_PA_LY.twbx” can be accessed in Lab on Demand with the steps listed below:

- Move the “D211_PA_LY.twbx” and “telco_churn.csv” files into the Labs on Demand environment using whichever preferred technique, such as using OneDrive or as email attachments.
- Move the “telco_churn.csv” file into the directory 'C:\Users\Public\Downloads\' location.
- Open pgAdmin4 and navigate to Servers > Postgres SQL 13 > Databases > churn > Schemas > public > Tables
- Use the Query Tool to open the Query Editor.
- Copy the first SQL code listed in Section A4 to create the table for the external dataset.
- Copy the second SQL code in Section A4 to import the data from the “telco_churn.csv” file into the database.
- Open the “D211_PA_LY.twbx” file.
- Enter the username and password.
 - Username: postgres
 - Password: Passw0rd!
- The dashboard should now appear on the screen.

A3. Dashboard Navigation

The dashboard overviews customer subscription data between WGU and a competitor telecom company, IBM. The dashboard contains four different data representations and one filter

control. The filter at the top can generate various customer base views. The “Churn” filter can filter views for customers who have churned and those who haven’t.

The data representation on the top left shows customer KPIs. The KPIs include the total number of customers, the average monthly cost, and customer tenure, with the value referring to the number of months. Clicking on either company will also filter for views specific to that company.

The following data representation, on the bottom left, is a heatmap displaying the contract types and the number of customers under those contracts. The size of the proportions pertains to the number of customers under the specified contracts, with more enormous proportions meaning more customers. Clicking on any proportion will filter for views of customers under the corresponding contract type.

The following data representation, on the top right, displays a bar chart of the number of customers by tenure in months. For example, 773 customers have a tenure of 1 month. Additionally, clicking any of the bars will filter for views specific to the tenure length.

At the bottom right, the last data representation is another bar chart displaying the average monthly charges by tenure in months. Clicking on any of the bars will filter for views specific to the tenure length.

A4. SQL Code

- SQL code used to create the table for external dataset

```
CREATE TABLE public.telco
```

```
(
```

```
  customer_id text,
```

```
  gender text,
```

```
  senior_citizen text,
```

```
  partner text,
```

```
  dependents text,
```

```
  tenure numeric,
```

```
  phone_service text,
```

```
  multiple_lines text,
```

```
  internet_service text,
```

```

online_security text,
online_backup text,
device_protection text,
tech_support text,
streaming_tv text,
streaming_movies text,
contract text,
paperless_billing text,
payment_method text,
monthly_charge numeric,
total_charge numeric,
churn text
);
ALTER TABLE public.telco
OWNER to postgres;

```

- SQL code used to import the external dataset into pgAdmin (Note: the csv must be placed in 'C:\Users\Public\Downloads' file directory for it to be processed in pgAdmin)

```

COPY telco
FROM 'C:\Users\Public\Downloads\telco_churn.csv'
DELIMITER ','
CSV HEADER;

```

- Underlying SQL code in Tableau

```

SELECT "t0"."Table Name" AS "Table Name",
       "t0"."age" AS "age",
       "t0"."bandwidth_gp_year" AS "bandwidth_gp_year",

```

"t0"."children" AS "children",
"t0"."churn" AS "churn",
"t0"."contacts" AS "contacts",
"t0"."contract" AS "contract",
"contract"."contract_id" AS "contract_id (contract)",
"t0"."contract_id" AS "contract_id",
"t0"."customer_id" AS "customer_id",
"t0"."dependents" AS "dependents",
"t0"."device_protection" AS "device_protection",
CAST("contract"."duration" AS TEXT) AS "duration",
"t0"."email" AS "email",
"t0"."gender" AS "gender",
"t0"."income" AS "income",
"t0"."internet_service" AS "internet_service",
"t0"."job_id" AS "job_id",
"t0"."lat" AS "lat",
"t0"."lng" AS "lng",
"t0"."location_id" AS "location_id",
"t0"."marital" AS "marital",
"t0"."monthly_charge" AS "monthly_charge",
"t0"."multiple_lines" AS "multiple_lines",
"t0"."online_backup" AS "online_backup",
"t0"."online_security" AS "online_security",
"t0"."outage_sec_week" AS "outage_sec_week",
"t0"."paperless_billing" AS "paperless_billing",
"t0"."partner" AS "partner",
"t0"."payment_id" AS "payment_id",

"t0"."payment_method" AS "payment_method",
 "t0"."phone_service" AS "phone_service",
 "t0"."population" AS "population",
 "t0"."port_modem" AS "port_modem",
 "t0"."senior_citizen" AS "senior_citizen",
 "t0"."streaming_movies" AS "streaming_movies",
 "t0"."streaming_tv" AS "streaming_tv",
 "t0"."tablet" AS "tablet",
 "t0"."tech_support" AS "tech_support",
 "t0"."techie" AS "techie",
 "t0"."tenure" AS "tenure",
 "t0"."total_charge" AS "total_charge",
 "t0"."yearly equip_faiure" AS "yearly equip_faiure"

FROM (

SELECT "t1"."Table Name" AS "Table Name", "t1"."age" AS "age",
 "t1"."bandwidth_gp_year" AS "bandwidth_gp_year", "t1"."children" AS "children", "t1"."churn"
 AS "churn", "t1"."contacts" AS "contacts", "t1"."contract" AS "contract", "t1"."contract_id" AS
 "contract_id", "t1"."customer_id" AS "customer_id", "t1"."dependents" AS "dependents",
 "t1"."device_protection" AS "device_protection", "t1"."email" AS "email", "t1"."gender" AS
 "gender", "t1"."income" AS "income", "t1"."internet_service" AS "internet_service",
 "t1"."job_id" AS "job_id", "t1"."lat" AS "lat", "t1"."lng" AS "lng", "t1"."location_id" AS
 "location_id", "t1"."marital" AS "marital", "t1"."monthly_charge" AS "monthly_charge",
 "t1"."multiple_lines" AS "multiple_lines", "t1"."online_backup" AS "online_backup",
 "t1"."online_security" AS "online_security", "t1"."outage_sec_week" AS "outage_sec_week",
 "t1"."paperless_billing" AS "paperless_billing", "t1"."partner" AS "partner", "t1"."payment_id"
 AS "payment_id", "t1"."payment_method" AS "payment_method", "t1"."phone_service" AS
 "phone_service", "t1"."population" AS "population", "t1"."port_modem" AS "port_modem",
 "t1"."senior_citizen" AS "senior_citizen", "t1"."streaming_movies" AS "streaming_movies",
 "t1"."streaming_tv" AS "streaming_tv", "t1"."tablet" AS "tablet", "t1"."tech_support" AS
 "tech_support", "t1"."techie" AS "techie", "t1"."tenure" AS "tenure", "t1"."total_charge" AS
 "total_charge", "t1"."yearly equip_faiure" AS "yearly equip_faiure"

FROM (

```
SELECT ('customer'::text) AS "Table Name",
"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",
CAST(NULL AS TEXT) AS "contract",
"customer"."contract_id" AS "contract_id",
CAST("customer"."customer_id" AS TEXT) AS "customer_id",
CAST(NULL AS TEXT) AS "dependents",
CAST(NULL AS TEXT) AS "device_protection",
"customer"."email" AS "email",
CAST("customer"."gender" AS TEXT) AS "gender",
"customer"."income" AS "income",
CAST(NULL AS TEXT) AS "internet_service",
"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",
CAST(NULL AS TEXT) AS "multiple_lines",
CAST(NULL AS TEXT) AS "online_backup",
CAST(NULL AS TEXT) AS "online_security",
"customer"."outage_sec_week" AS "outage_sec_week",
CAST(NULL AS TEXT) AS "paperless_billing",
CAST(NULL AS TEXT) AS "partner",
```

```

"customer"."payment_id" AS "payment_id",
CAST(NULL AS TEXT) AS "payment_method",
CAST(NULL AS TEXT) AS "phone_service",
"customer"."population" AS "population",
CAST("customer"."port_modem" AS TEXT) AS "port_modem",
CAST(NULL AS TEXT) AS "senior_citizen",
CAST(NULL AS TEXT) AS "streaming_movies",
CAST(NULL AS TEXT) AS "streaming_tv",
CAST("customer"."tablet" AS TEXT) AS "tablet",
CAST(NULL AS TEXT) AS "tech_support",
CAST("customer"."techie" AS TEXT) AS "techie",
"customer"."tenure" AS "tenure",
CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION) AS "total_charge",
"customer"."yearly equip_faiure" AS "yearly equip_faiure"
FROM "public"."customer" "customer"
) "t1"

```

UNION ALL

```

SELECT "t2"."Table Name" AS "Table Name", "t2"."age" AS "age",
"t2"."bandwidth_gp_year" AS "bandwidth_gp_year", "t2"."children" AS "children", "t2"."churn"
AS "churn", "t2"."contacts" AS "contacts", "t2"."contract" AS "contract", "t2"."contract_id" AS
"contract_id", "t2"."customer_id" AS "customer_id", "t2"."dependents" AS "dependents",
"t2"."device_protection" AS "device_protection", "t2"."email" AS "email", "t2"."gender" AS
"gender", "t2"."income" AS "income", "t2"."internet_service" AS "internet_service",
"t2"."job_id" AS "job_id", "t2"."lat" AS "lat", "t2"."lng" AS "lng", "t2"."location_id" AS
"location_id", "t2"."marital" AS "marital", "t2"."monthly_charge" AS "monthly_charge",
"t2"."multiple_lines" AS "multiple_lines", "t2"."online_backup" AS "online_backup",
"t2"."online_security" AS "online_security", "t2"."outage_sec_week" AS "outage_sec_week",
"t2"."paperless_billing" AS "paperless_billing", "t2"."partner" AS "partner", "t2"."payment_id"
AS "payment_id", "t2"."payment_method" AS "payment_method", "t2"."phone_service" AS
"phone_service", "t2"."population" AS "population", "t2"."port_modem" AS "port_modem",
"t2"."senior_citizen" AS "senior_citizen", "t2"."streaming_movies" AS "streaming_movies",
"t2"."streaming_tv" AS "streaming_tv", "t2"."tablet" AS "tablet", "t2"."tech_support" AS

```


"tech_support", "t2"."techie" AS "techie", "t2"."tenure" AS "tenure", "t2"."total_charge" AS
"total_charge", "t2"."yearly_equip_faiure" AS "yearly_equip_faiure"

FROM (

SELECT ('telco'::text) AS "Table Name",

CAST(TRUNC(CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION)) AS BIGINT)
AS "age",

CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION) AS "bandwidth_gp_year",

CAST(TRUNC(CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION)) AS BIGINT)
AS "children",

CAST("telco"."churn" AS TEXT) AS "churn",

CAST(TRUNC(CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION)) AS BIGINT)
AS "contacts",

CAST("telco"."contract" AS TEXT) AS "contract",

CAST(TRUNC(CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION)) AS BIGINT)
AS "contract_id",

CAST("telco"."customer_id" AS TEXT) AS "customer_id",

CAST("telco"."dependents" AS TEXT) AS "dependents",

CAST("telco"."device_protection" AS TEXT) AS "device_protection",

CAST(TRUNC(CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION)) AS BIGINT)
AS "email",

CAST("telco"."gender" AS TEXT) AS "gender",

CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION) AS "income",

CAST("telco"."internet_service" AS TEXT) AS "internet_service",

CAST(TRUNC(CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION)) AS BIGINT)
AS "job_id",

CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION) AS "lat",

CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION) AS "lng",

CAST(TRUNC(CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION)) AS BIGINT)
AS "location_id",

CAST(NULL AS TEXT) AS "marital",

```

"telco"."monthly_charge" AS "monthly_charge",
CAST("telco"."multiple_lines" AS TEXT) AS "multiple_lines",
CAST("telco"."online_backup" AS TEXT) AS "online_backup",
CAST("telco"."online_security" AS TEXT) AS "online_security",
CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION) AS "outage_sec_week",
CAST("telco"."paperless_billing" AS TEXT) AS "paperless_billing",
CAST("telco"."partner" AS TEXT) AS "partner",
CAST(TRUNC(CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION)) AS BIGINT)
AS "payment_id",
CAST("telco"."payment_method" AS TEXT) AS "payment_method",
CAST("telco"."phone_service" AS TEXT) AS "phone_service",
CAST(TRUNC(CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION)) AS BIGINT)
AS "population",
CAST(NULL AS TEXT) AS "port_modem",
CAST("telco"."senior_citizen" AS TEXT) AS "senior_citizen",
CAST("telco"."streaming_movies" AS TEXT) AS "streaming_movies",
CAST("telco"."streaming_tv" AS TEXT) AS "streaming_tv",
CAST(NULL AS TEXT) AS "tablet",
CAST("telco"."tech_support" AS TEXT) AS "tech_support",
CAST(NULL AS TEXT) AS "techie",
"telco"."tenure" AS "tenure",
"telco"."total_charge" AS "total_charge",
CAST(TRUNC(CAST(CAST(NULL AS TEXT) AS DOUBLE PRECISION)) AS BIGINT)
AS "yearly_equip_faiure"
FROM "public"."telco" "telco"
) "t2"
) "t0"
LEFT JOIN "public"."contract" "contract" ON ("t0"."contract_id" = "contract"."contract_id")

```

B – B7. Demonstration

The Panopto demonstration is submitted alongside the report and can also be found at:

<https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=5755b95d-f468-4243-8332-b1f8010198b9>

C. Written Report

The written report is submitted as “D211_PA_LY.pdf.”

C1. Dashboard Alignment

The dashboard allows stakeholders to look at high-level statistics of WGU’s telecom customer data compared to those of a competitor company. As customer retention is essential for business, the dashboard allows for deeper insights into the characteristics of customers who have churned. Additionally, including a competitor company provides comparisons of company performance. In summary, the data on customer background and the ability to compare between companies allows for actionable insight to reduce churn and increase tenure in the process.

C2. Business Intelligence Tool

The business intelligence tool chosen to create the dashboard was Tableau. Tableau was selected because it provides an easy-to-use interface for simple drag-and-drop actions. Additionally, visualizations can be quickly built as Tableau contains a tool for pre-built visualization types such as bar charts, line charts, pie charts, etc., using its “Show Me” option. This saves a lot of time when building out commonly used visualizations. Beyond building visualizations, the biggest reason is that Tableau’s visualizations allow interactivity. This interactivity with the visualizations allows for better analysis as it allows users to quickly and easily explore the same data in different ways, such as with filters that can drastically change how the same visualization is displayed.

C3. Data Cleaning

Only a little data needed to be cleaned, but preparation of the data was necessary before creating any visualizations in Tableau. After the external dataset was imported into pgAdmin, a connection was made in Tableau to the database.

In Tableau, the “customer” and “telco” tables were joined by a union operation since both tables contain customer records with similar variables. Afterward, a left join operation between the newly combined table and the “contract” table was performed as it contained the contract type each customer was under in WGU’s customer base. Joining the tables this way caused some data desync as column headers had different naming conventions for the same data type. For example, the “contract” table had the contract types under the “duration” column, whereas for the “telco” table, the contract types were under the “contract” column. Both variables were combined into one column within Tableau to correct this discrepancy using its capabilities. This preparation and its use case are further detailed in the next section, C4.

C4. Dashboard Creation

KPIs Table:

- Create a new “KPIs” worksheet and change the view from standard to fit the entire view.
- Drag “Table Name” to Columns.
- Double click on “customer (Count),” “Monthly Charge,” and “Tenure.” These should all populate into the Measure Values shelf that pops up.
- In the Measure Values shelf, right-click on “SUM(Monthly Charge)” and change the measure to “Average.”
- Repeat the step above for “SUM(Tenure)” to change the measure to “Average.”
- Hide the “Table Name” title at the top of the table.
- Rename the “customer” column to “WGU” and the “telco” column to “IBM.”
- Format the worksheet so that the text of the columns is aligned to the center.
- Format the column headers to size 16 text and use Tableau Bold font.
- Format the Pane to have size 12 font.

Contracts Heatmap:

- Create a new worksheet called “Contracts Heatmap” and change the view from standard to fit the entire view.
- In the data pane, control-click “Duration” and “contract.” After having both highlighted, right-click and create a “Combined Field.”
- Right-click the new combined field “contract & Duration (Combined)” and create a new group.

- Group the Month-to-Month, One-Year, and Two-Year values together and name each group, respectively. Name the new Group field “contracts.”
- Control-click on the new “contracts” group and “customer (Count)” and then select “treemaps” from the “Show Me” tab at the top right to create the heatmap.
- Select “Churn” from the Data pane and show it as a filter.

Tenure Bar:

- Create a new “Tenure Bar” worksheet and change the view from standard to fit the entire view.
- Create a bin from “Tenure” called “Tenure (in Months)” with a bin size of 1.
- Drag “Tenure (in Months)” to the Columns field and “Tenure” to the Rows field.
- Change the measure of Tenure from Sum to Count.
- Rename the y-axis to “Number of Customers.”

Monthly Charge by Tenure:

- Create a new “Monthly Charge by Tenure” worksheet and change the view from standard to fit the entire view.
- Create a bin from “Tenure” called “Tenure (in Months)” with a bin size of 1.
- Drag “Tenure (in Months)” to the Columns field and “Monthly Charge” to the Rows field.
- Change the measure of Monthly Charge from Sum to Average.

Dashboard:

- Create a new dashboard sheet and set the size to automatic.
- Add the worksheets to the dashboard with the KPIs and Contracts Heatmap on the left and the bar charts on the right.
- Hide all worksheet titles.
- Allow each visualization in the dashboard to be used as a filter.
- Remove the “Count of customer” legend from the dashboard.
- Customize the “Churn” filter to “Single value (dropdown).” Also, change the filter to floating and place where appropriate.
- Remove the extra white space from the dashboard where the filters used to be.

C5. Data Analysis Results

The additional data set contains variables that are the same or very similar. For example, the additional data set includes variables also seen in the WGU churn data set, such as unique customer ID, gender, tenure, monthly charges, and churn. This similarity allows for direct

comparisons between the two data sets, enhancing the insights that can be drawn versus drawing insights from one data set alone.

For example, one data representation that supports decision-making is the KPIs table. The KPIs show calculated values of average monthly charge and average tenure. Both values provide insight into customer information. For example, WGU customers, on average, have longer tenure than IBM customers, but also seen is that WGU customers pay, on average, twice as much per month. The analysis from this representation can assist stakeholders in delving deeper into why customers stay longer with their services despite the higher charges.

Another data representation that supports decision-making is the heatmap containing the contract types and the number of customers under those contracts. Immediately, stakeholders can identify which contract type is the most popular and the number of customers under those contracts. Also, when filtering for customers who churned, an overwhelming majority were under a month-to-month contract. This analysis can provide direction into different marketing strategies that could move customers into other contracts and assist in decreasing churn.

C6. Analysis Limitations

One of the more prominent limitations is the legitimacy of the external dataset. It is essential to ensure that the data is accurate and complete. As the data hails from Kaggle, there is a chance that it could contain errors. Additionally, WGU's churn data comprises a nationwide customer base. In contrast, the external dataset did not include the demographic location of the customer base, which limits how well the data can be compared. This limitation is essential to note as customer characteristics may differ by region.

D. Web Sources

BlastChar. (2018, February 23). *Telco Customer Churn*. Kaggle.

<https://www.kaggle.com/datasets/blastchar/telco-customer-churn>

E. Sources

No other sources were used.