D211: Advanced Data Acquisition

Part I: Data Dashboard

A. Data Dashboard

The dashboard can be viewed at the following link:
https://public.tableau.com/app/profile/ashley.munguia/viz/D211_dashboard_17335244879280/D211Dashboard?publish=yes

I have also attached it as "D211_dashboard.twbx".

A1. Data Sets & Dashboard File

The dashboard integrates data from two sources, one of which has been provided for the students that consists of tables in the Churn database located on Labs On Demand virtual machine.

The other source was an external, public data set that was found from Kaggle. The link to the external source can be found following the link:
https://www.kaggle.com/datasets/ebruiserisobay/telcods

A2. Dashboard Installation

The dashboard can be viewed at the link from part A through Tableau Public.

Alternatively, since Tableau Public is not permissible to use for this assignment, the following steps are how to access the dashboard through the virtual machine. This dashboard was created on the Labs On Demand virtual machine.
1. Download the d211.zip file and save it to C:\Users\Public\Downloads\d211.
2. Right click on the file and extract all to create a new file unzipped.
3. Open pgAdmin on the desktop.
4. Navigate to Servers > Databases > Churn.
5. Right click on Churn and select the Query Tool.
6. Click on the folder icon in the tool bar.
7. Go to C:\Users\Public\Downloads\d211. Go to the bottom right to change the format to "All Files" and select the "d211_sql.txt" file.
8. Press the play button to execute the query.
9. Once this is completed, close out the pgAdmin application.
10. Next, open the "Tableau 2021.4" on the desktop.
11. Click File > Open > C:\Users\Public\Downloads\d211 > "D211_dashboard.twbx".
12. Sign in to connect to the server. The username is "postgres" and the password is "Passw0rd!" Now it should be ready to use.

A3. Dashboard Navigation

To navigate through the dashboard, there are four visualizations represented by graphs to compare our company (WGU company) to the competitor as well as three tables to show the

WGU KPIs, external KPIs. And count of records. It is also very important take color blindness into consideration when choosing the color combinations for these visualizations. It is recommended to use blue/orange, blue/red, or blue/brown. Hence why I used blue/orange in all of my visualizations.

- KPIs and Count of Records tables can be found in the top left corner below the title. These tables summarize the averages of factors that could be affecting customer churn and the count of customer records of each company.
- Monthly Charge by Payment Method visualization can be found in the middle of the dashboard. You are able to hover over each individual circle to see the average monthly charge. The bigger the circle, the higher the monthly charge. There is also a legend to the left of the visualization that shows the competitor is represented by the color blue and WGU is represented by the color orange.
- Tenure by Customer Churn can be found in the top right corner. If you hover over the slice of the circle, it has what the color represents and the average tenure depending on the company. The orange represents customer churn, and the blue represents the customer staying with the company. I also have the averages labeled beside each slide to make it easier to read.
- Monthly Charge by Customer Churn can be found in the bottom left of the dashboard. It is self-explanatory since everything is already clearly labeled. It shows the average monthly charge of each company and what that monthly charge would be if there is customer churn or not.
- Monthly Charge by Contract can be found in the bottom right. It is similar to the Monthly Charge by Customer Churn histogram but instead, it is a stacked histogram instead of side-by-side. I created aliases for the labels since they were originally labeled as 1, 2, and 3 to make it easier to follow along with. The color blue represents competitor, and the color orange represents WGU.

A4. SQL Code

- SQL code for internal dataset (already in pgadmin):

```
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,
```

```sql
        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;
```

- SQL code for external dataset:

```sql
CREATE TABLE public.competitor
(
        customerID text PRIMARY KEY,
        gender text,
        SeniorCitizen int,
        Partner text,
        Dependents text,
        tenure int,
        PhoneService text,
```

```
        MultipleLines text,
        InternetService text,
        OnlineSecurity text,
        OnlineBackup text,
        DeviceProtection text,
        TechSupport text,
        StreamingTV text,
        StreamingMovies text,
        Contract text,
        PaperlessBilling text,
        PaymentMethod text,
        MonthlyCharges numeric,
        TotalCharges text,
        Churn text
);

copy competitor
FROM 'C:\Users\Public\Downloads\d211\Telco-Customer-Churn.csv'
DELIMITER ','
CSV HEADER;

UPDATE competitor
SET "contract" =
  CASE
    WHEN "contract" IN ('1 year', 'One year') THEN '1'
    WHEN "contract" = 'Two year' THEN '2'
    WHEN "contract" = 'Month-to-month' THEN '3'
    ELSE "contract"
  END,
"paymentmethod" =
  CASE
    WHEN "paymentmethod" = 'Bank transfer (automatic)' THEN '1'
    WHEN "paymentmethod" = 'Credit card (automatic)' THEN '2'
    WHEN "paymentmethod" = 'Electronic check' THEN '3'
    WHEN "paymentmethod" = 'Mailed check' THEN '4'
    ELSE "paymentmethod"
  END;

ALTER TABLE competitor
RENAME COLUMN SeniorCitizen TO senior_citizen;

ALTER TABLE competitor
RENAME COLUMN partner TO marital;

ALTER TABLE competitor
RENAME COLUMN PhoneService TO phone;
```

```sql
ALTER TABLE competitor
RENAME COLUMN MultipleLines TO multiple_lines;

ALTER TABLE competitor
RENAME COLUMN InternetService TO internet_service;

ALTER TABLE competitor
RENAME COLUMN OnlineSecurity TO online_security;

ALTER TABLE competitor
RENAME COLUMN OnlineBackup TO online_backup;

ALTER TABLE competitor
RENAME COLUMN DeviceProtection TO device_protection;

ALTER TABLE competitor
RENAME COLUMN TechSupport TO tech_support;

ALTER TABLE competitor
RENAME COLUMN StreamingTV TO streaming_tv;

ALTER TABLE competitor
RENAME COLUMN StreamingMovies TO streaming_movies;

ALTER TABLE competitor
RENAME COLUMN Contract TO contract_id;

ALTER TABLE competitor
RENAME COLUMN Customerid TO customer_id;

ALTER TABLE competitor
RENAME COLUMN PaperlessBilling TO paperless_billing;

ALTER TABLE competitor
RENAME COLUMN PaymentMethod TO payment_id;

ALTER TABLE competitor
RENAME COLUMN MonthlyCharges TO monthly_charge;

ALTER TABLE competitor
RENAME COLUMN TotalCharges TO total_charge;

ALTER TABLE competitor
ADD COLUMN company text DEFAULT 'competitor';
```

```sql
ALTER TABLE customer
ADD COLUMN company text DEFAULT 'wgu';

ALTER TABLE competitor
ALTER COLUMN contract_id TYPE Numeric
USING contract_id::numeric;

ALTER TABLE competitor
ALTER COLUMN payment_id TYPE Numeric
USING payment_id::numeric;


CREATE TABLE public.combined AS
SELECT
    subquery.customer_id,
    subquery.gender,
    subquery.churn,
    subquery.tenure,
    subquery.monthly_charge,
    subquery.payment_id,
    subquery.contract_id,
    subquery.company
FROM (
    SELECT
        customer_id,
        gender,
        churn,
        round(tenure) as tenure,
        monthly_charge,
        payment_id,
        contract_id,
        company
    FROM customer

    UNION ALL

    SELECT
        customer_id,
        gender,
        churn,
        round(tenure) as tenure,
        monthly_charge,
        payment_id,
        contract_id,
        company
    FROM competitor
```

) subquery;

UPDATE customer SET company = 'wgu';

• SQL code generated for ERD (and a screenshot of the ERD):

```
BEGIN;

CREATE TABLE public.combined
(
    customer_id text,
    gender text,
    churn text,
    tenure double precision,
    monthly_charge numeric,
    payment_id numeric,
    contract_id numeric,
    company text
);

CREATE TABLE public.competitor
(
    customer_id text NOT NULL,
    gender text,
    senior_citizen integer,
    marital text,
    dependents text,
    tenure integer,
    phone 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_id numeric,
    paperless_billing text,
    payment_id numeric,
    monthly_charge numeric,
    total_charge text,
    churn text,
    company text,
    PRIMARY KEY (customer_id)
```

```sql
);

CREATE TABLE public.contract
(
    contract_id integer NOT NULL,
    duration text,
    PRIMARY KEY (contract_id)
);

CREATE TABLE public.customer
(
    customer_id text NOT NULL,
    lat numeric,
    lng numeric,
    population integer,
    children integer,
    age integer,
    income numeric,
    marital text,
    churn text,
    gender text,
    tenure numeric,
    monthly_charge numeric,
    bandwidth_gp_year numeric,
    outage_sec_week numeric,
    email integer,
    contacts integer,
    yearly_equip_faiure integer,
    techie text,
    port_modem text,
    tablet text,
    job_id integer,
    payment_id integer,
    contract_id integer,
    location_id integer,
    company text,
    PRIMARY KEY (customer_id)
);

CREATE TABLE public.job
(
    job_id integer NOT NULL,
    job_title text,
    PRIMARY KEY (job_id)
);
```

```sql
CREATE TABLE public.location
(
    location_id integer NOT NULL,
    zip integer,
    city text,
    state text,
    county text,
    PRIMARY KEY (location_id)
);

CREATE TABLE public.payment
(
    payment_id integer NOT NULL,
    payment_type text,
    PRIMARY KEY (payment_id)
);

ALTER TABLE public.customer
    ADD FOREIGN KEY (contract_id)
    REFERENCES public.contract (contract_id)
    NOT VALID;


ALTER TABLE public.customer
    ADD FOREIGN KEY (job_id)
    REFERENCES public.job (job_id)
    NOT VALID;


ALTER TABLE public.customer
    ADD FOREIGN KEY (location_id)
    REFERENCES public.location (location_id)
    NOT VALID;


ALTER TABLE public.customer
    ADD FOREIGN KEY (payment_id)
    REFERENCES public.payment (payment_id)
    NOT VALID;
```
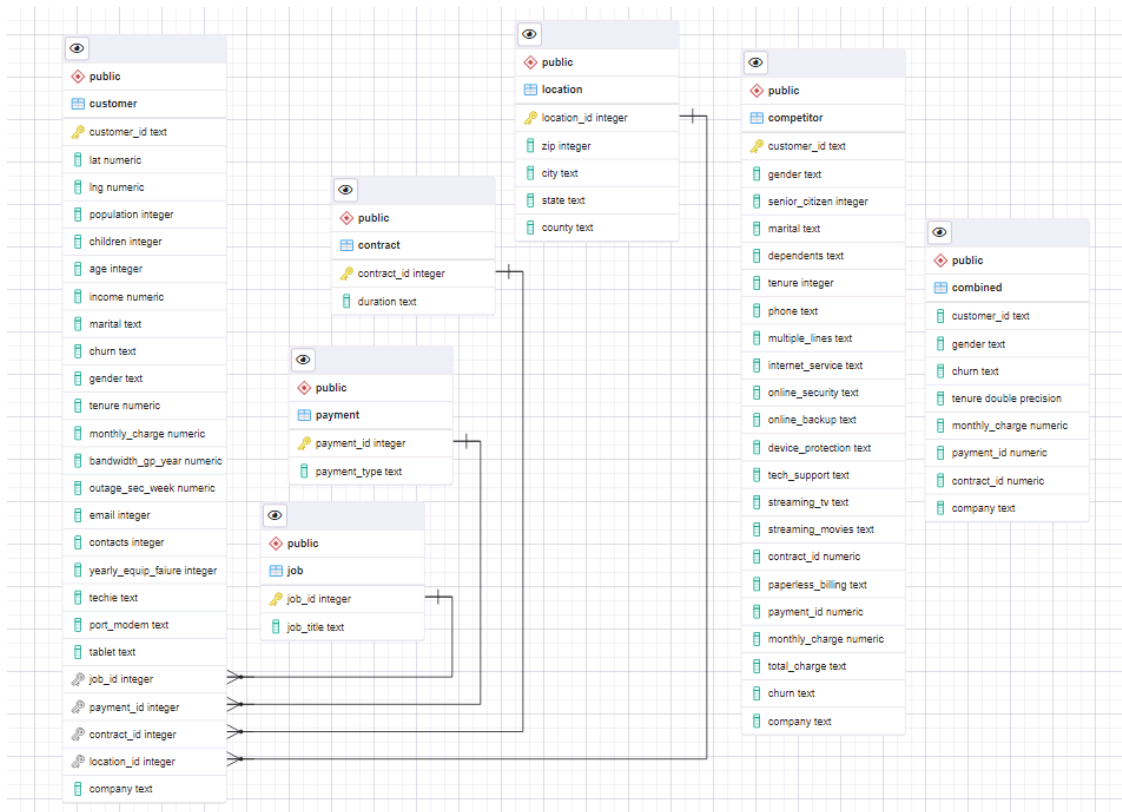
END;

Part II: Demonstration

B. Panopto Video Link

https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=41d4d641-e43b-4753-9d12-b23f000d5e8b

Part III: Report

C1. Dashboard Alignment

The purpose of the dashboard is to analyze the churn database along with the five tables included to determine the relationship between them and the external dataset. The external dataset will be considered the competitor. Therefore, I compared the customer demographic data between our company and the competitor to investigate the reasoning behind customer churn.

C2. Business Intelligence Tools

I will be using SQL and Tableau, the business intelligence tool, for this assignment. For SQL, PostgreSQL is a great tool for managing large amounts of data and running SQL queries whereas Tableau offers a variety of features to visualize this data in a more appealing, easy to

understand way to present to people and organizations. Tableau will help the executive leaders answer a specific research question regarding an organizational need for the company.

C3. <u>Data Cleaning</u>

The churn dataset provided was already clean as well as the external dataset. However, there are a few steps that I had to perform in order to get the datasets prepared to be joined and transferred over to Tableau to create the dashboard. I created a "competitor" table with the external dataset to compare to our company's customer data. I renamed the external columns to match the internal columns, created a text column named "company" to differentiate between the competitor (external dataset) and WGU company (internal dataset), and then changed the "contract" and "paymentmethod" columns to numeric values. The numeric values and what they represent can be found below:

- Contract:
    - 1 year, One year = 1
    - Two year = 2
    - Month-to-month = 3
- Payment Method:
    - Bank transfer (automatic) = 1
    - Credit card (automatic) = 2
    - Electronic check = 3
    - Mailed check = 4

C4. <u>Dashboard Creation</u>

WGU KPIs
1. Create a new worksheet.
2. Drag Churn (customer) to Rows.
3. Drag Measure Names to Columns and Filters section.
4. Drag Measure Values to the Text box under the Marks section.
5. From there, you are able to pick and choose what columns you want included in the visualization. If you do not want something included, you can right click it, and press Remove.
6. Change all the measures to AVG.
7. Rename the title to WGU KPIs.

External KPIs
1. Create a new worksheet.
2. Drag Churn (competitor) to Rows.
3. Drag Measure Names to Columns and Filters section.
4. Drag Measure Values to the Text box under the Marks section.
5. From there, you are able to pick and choose what columns you want included in the visualization. If you do not want something included, you can right click it, and press Remove.
6. Change all the measures to AVG.
7. Rename the title to External KPIs.

Count of Records
1. Create a new worksheet.
2. Drag Company to Rows
3. Drag combined (Count) to the Text box under the Marks section.
4. Rename the title to Count of Records.

Monthly Charge by Payment Method
1. Create a new worksheet.
2. Drag Payment id to the Text box under the Marks section.
3. Drag Monthly Charge to the Size box under the Marks section.
4. Right click on Monthly Charge and change the Measure type to AVG.
5. Drag Company to the Color box under the Marks section.
6. On the top right of the application, click on Show Me. This will expand a bunch of different visualizations that you are able to choose from. I selected the packed bubbles visualization which is the last option on the bottom right. You can also go under the Marks section and click the drop-down menu to select which visualization you want that way. I used the visualization called Circle.
7. Changed Payment id from 1, 2, 3, and 4 to the specific payment methods by right clicking on each number > Edit Alias.
8. Rename the title to Monthly Charge by Payment Method.

Tenure by Customer Churn
1. Create a new worksheet.
2. Drag Company to Rows.
3. Drag Tenure to the Angle box under the Marks section.
4. Drag Churn to the Colors box under the Marks section.
5. Under the Marks section, click the drop-down menu to select the Pie visualization.
6. Under the Marks section, click Label box > check Mark Labels.
7. Rename the title to Tenure by Customer Churn.

Monthly Charge by Customer Churn
1. Create a new worksheet.
2. Drag Churn and Company to Columns.
3. Drag Monthly Charge to Rows.
4. Right click on Monthly Charge and change the Measure type to AVG.
5. Drag Company to the Colors box under the Marks section.
6. Rename the title to Monthly Charge by Customer Churn.

Monthly Charge by Contract
1. Create a new worksheet.
2. Drag Contract id to Columns.
3. Drag Monthly Charge to Rows.
4. Right click on Monthly Charge and change the Measure type to AVG.
5. Drag Company to the Colors box under the Marks section.
6. Rename the title to Monthly Charge by Contract.

C5. Data Analysis Results

My focus was on the average monthly charge because customers tend to leave a company when they are getting charged too much and swap to a different company that has the same services. The WGU customers' average monthly charge is nearly triple the competitors when it comes to contract, the different payment methods, and customer churn. The average tenure between the companies is not a huge difference. However, compared to the competitor that has a $6 difference between the month-to-month and the two-year contracts, I recommend creating some type of payment difference or discount if the customers sign a two-year contract. The company would not only lock down current customers longer, making the tenure increase, but would also increase the income because other customers would not want to miss out on a great deal. Perhaps even lowering the price slightly would retain customers as well.

C6. Analysis Limitations

There were a few limitations I noticed while completing this assignment. Perhaps it is because I am used to using Python, I feel as if using SQL was much more complex with these larger datasets. I also originally tried to link my personal Tableau to PostgreSQL to use futuristically and was having much trouble. There were some compatibility issues, and it was saying that I didn't download the server correctly. For the analysis, I would have preferred to compare the monthly charges to the online add-ons like streaming tv, movies, online security, etc. But I was able to since I chose the customer table to compare to my outside source. I was much more limited in my options.

D. Web Sources

N/A

E. Sources

N/A

F. Professional Communication