**Advanced Data Acquisition (SLM1 TASK 1)– D211**

**Performance Assessment**

**Western Governors University**

**By: Christian LeBlanc**

***Part 1:  Data Dashboards***

**A-1**

Please see submitted zip file Christian LeBlanc D211 files containing two csv files of data and the Tableau file.

**A-2**

The following instructions to the dashboard instructions are done with the assumption that whoever is installing the dashboard is doing so on the Labs On Demand environment because executives would have access to company data.

1. Open pgAdmin4.
2. Click the arrow next to Servers, then PostgreSQL, then Database, then right click on churn and click Query Tool.
3. Run the codes from A-4 that are called out for pgAdmin4 in the order they come in.
4. Click the arrow next to Schemas and then next to Tables. Right click on India and click on Import/Export. Switch to Import. Click the three dots next to the filename section and select the file location that you have the d211clean2 csv file then click select. Turn Header to Yes. Go to the Delimiter selection and make it a comma. Click OK.
5. Go to the folder where you have Christian LeBlanc D211 Dashboard v1 saved to and double click to open. Click Yes for the Custom SQL Warning. Type Passw0rd! in the password section and click sign in.

**A-3**

The dashboard is set up to be very customizable and easy to read and navigate. It is broken up into five different sections and all but the Churn by source can be used as a filter. This is done so if an executive wants to see the churn numbers of Males that have two children from the India data in the age range of 30-39 can simply click on each thing. As an example, doing this will show in the Source section that there are 29,123 customers that fall into the previously mention range. While the filters are still active you can hover your mouse over churn by source bar chart to see that 23,357 of them are still active and 5,766 fall under churn. Clicking on the same places will undo the filters. I would suggest viewing the data with the source filter to one or the other because the number of customers in the Kaggle data for India customers is so much larger that the WGU data customer that it makes the charts harder to read next to each other.

**A-4**

The following SQL code is ran in the pgAdmin4:

1. Adding a source column to the customer table:

ALTER TABLE public.customer

ADD COLUMN source text;

1. Filling in the source column with WGU:

UPDATE customer

SET source = 'WGU';

1. Create a new table with just the columns I need:

CREATE TABLE public."cft" AS

SELECT gender, age, children, income, churn, source

FROM customer;

ALTER TABLE public."cft"

OWNER to postgres;

1. Create India table for my external data:

CREATE TABLE public."India"

(

gender text,

age integer,

children integer,

income integer,

churn integer,

source text

);

ALTER TABLE public."India"

OWNER to postgres;

The following SQL code is ran in Tableau with picture after to show the join:

1. To create a custom SQL query with a Join:

SELECT "India"."age" AS "age (India)",

"cft"."age" AS "age",

"India"."children" AS "children (India)",

"cft"."children" AS "children",

"India"."churn" AS "churn (India)",

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

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

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

"India"."income" AS "income (India)",

"cft"."income" AS "income",

CAST("India"."source" AS TEXT) AS "source (India)",

CAST("cft"."source" AS TEXT) AS "source"

FROM "public"."cft" "cft"

FULL JOIN "public"."India" "India" ON (CAST("cft"."source" AS TEXT) = CAST("India"."source" AS TEXT))

A screenshot of a computer

Description automatically generated

***Part 2:  Demonstration***

**B**

Please see my video Christian LeBlanc D211 Panopto in Advanced Data Acquisition SLMx | D211 (Student Creators) [assignments]. The link is <https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=5db2412a-f9f3-41f6-80e3-b13b0172fd26>

***Part 3:  Report***

**C-1**

This section seems to have some confusion as in the requirements section it say “1. Explain how the purpose and function of your dashboard aligns with the needs outlined in the data dictionary associated with your chosen data set.” Then in the D211 PowerPoint it states “It needs to be rewritten to say, inspect the Churn database and the five tables. How are they related among themselves and then to the external dataset that you are using?” I will respond to both. This is to make sure I am providing all the necessary information.

I am using the D210 data dictionary since the D211 data is the same I will be assuming the purpose for the dashboard is similar. The purpose of my dashboard is to compare similar customer demographics between our customers and the customers of telecom companies in India. The EVP is tasked with strategic recruitment and is interested in broad categorization of customers. This dashboard does both for him. India is a big market that we are currently not in. The EVP can compare the churn rates between us and them with it being broken down by different attributes of the customer bases.

In the churn database, the customer table is the table that connects to the other four. The customer table has a primary key of customer\_id to ensure that each customer has a unique id. The customer table has four foreign keys. The contract\_id is the foreign key that connects to the contract table and is the primary key in the contract table. The job\_id is the foreign key that connects to the job table and is the primary key in the job table. The location\_id is the foreign key that connects to the location table and is the primary key in the location table. The payment\_id is the foreign key that connects to the payment table and is the primary key in the job table. Below is the ERD showing these connections.

A screenshot of a computer

Description automatically generated

The connection that is made between the external data and existing churn data is a full join done on a column I added to both that I named source. I added source to the churn customer table with all responses being WGU and to the external data I made the responses Kaggle. The response is the place where I received each dataset. I also created a new table containing only the columns I needed named cft and created a table called India to add my external data to. In Tableau I created a custom SQL query between cft and India with a full join on source.

**C-2**

Tableau Desktop is the business intelligence tool that was used in this analysis. Tableau allows for the creation of in-depth visuals that can be combined with different visual on a dashboard. For this analysis it allows for an easy visual comparison of how our customer demographics compare to the customers from India and how much churn is in each demographic.

**C-3**

I did most of my cleaning and preparation for my outside data in D210 but will include the steps I did here to be thorough and complete. The following are the things I did to clean and prepare the data for analysis:

1. Using a Jupyter notebook I dropped off unnecessary columns, checked for null values and outliers, renamed the remaining columns to match the column names from the provided churn data, changed the gender responses from M or F to Male or Female, added source column and made the responses Kaggle, then exported it as a clean data set ready to use.
2. In pgAdmin in the Labs on demand workspace I added a source column to the customer table, made the values in the source column become WGU, made a new table cft with only the columns I needed from the customer table, I created a table called India and added my external data to it.
3. In Tableau I created a custom SQL query with a full join to make sure I have all the data from both my external data and the cft table, then combined the common measures to be able compare the two datasets in the same visual.

**C-4**

The steps used to create the dashboard are as follows:

1. Follow the first two steps described in C-3 to clean and prepare the data.
2. Connect to the PostgreSQL server to gain access to the data. Under the Connect section they have a “To a server” click More. From there click on PostgreSQL. Fill in localhost under Server, 5432 under Port, churn under Database. Username and Password are postgres and Passw0rd!.
3. Create custom SQL query that was provided in A-4.
4. Go to Sheet 1. While holding ctrl click on gender and gender (India) then right click go to Transform and the click on Merge Mismatched Fields. Repeat this on the remaining measures to create a combined source, age, children, income, and churn.
5. Drag the source measure into rows. Then drag the count measure into Text in the Marks box. Rename Sheet 1 to Source.
6. Add new sheet. Drag the gender and churn measure to the columns and the count to rows. Then drag churn measure into Color in the Marks box. Then click Color in the Marks box and edit colors. Under Select Color Palette click the down arrow then color blind and then apply. Rename Sheet 2 to Churn by gender.

Add new sheet. Right click on the age measure and under the create section click on Calculated field. Name the Calculated field “Age buckets”, and put this formula:

IF [age (India) & age] < 18 THEN 'Under 18'

ELSEIF [age (India) & age] >= 18 AND [age (India) & age] < 30 THEN '18-29'

ELSEIF [age (India) & age] >= 30 AND [age (India) & age] < 40 THEN '30-39'

ELSEIF [age (India) & age] >= 40 AND [age (India) & age] < 50 THEN '40-49'

ELSEIF [age (India) & age] >= 50 AND [age (India) & age] < 60 THEN '50-59'

ELSE '60+'

END

Then click Apply. Now drag count measure into Columns and Age buckets and source measures into Rows. Now drag churn measure into Color in the Marks box. Then click Color in the Marks box and edit colors. Under Select Color Palette click the down arrow then color blind and then apply. Rename Sheet 3 to Churn by age range.

1. Add new sheet. Drag the source measure into columns and count measure in Rows. Then drag churn measure into Color in the Marks box. Then click Color in the Marks box and edit colors. Under Select Color Palette click the down arrow then color blind and then apply. Rename Sheet 4 to Churn by source.
2. Add new sheet. Drag the children measure to both Columns and Rows. Drag count measure to Rows. In the Columns section click on the down arrow on the children measure and click on Dimension. Now go into the show me section and click on treemaps. Then click Color in the Marks box and edit colors. Under Select Color Palette click the down arrow then Orange-Blue Diverging and then apply. Rename Sheet 5 to Children.
3. Click New Dashboard. Add the sheets into the dashboard. Remove the keys that come with them as they are not needed to understand the data and just add clutter. Since the Source sheet is not a graph, right click on it and click on Floating and move it into an empty spot on the Dashboard. Now make each sheet other than Churn by source into a filter by clicking into the tile section of each and clicking the Use as Filter that appears when clicking in the tile section.
4. Save the Tableau file.

**C-5**

The dashboard that I created shows that in common attributes between our customers and the customers in India that the customer base in India generally have slightly lower churn. This would show the executives that expanding into the India market would grow our customer base with customers that are not showing signs of high churn. The number of customers that use telecommunications in India would double our current customer base with gaining as little as five percent of the market.

**C-6**

This data analysis is a good overview of comparing the churn of customers with similar attributes such as gender, age, and family size, but it does come with its limitations. Not having the billing information for the India customers is a big piece of the puzzle if it would be profitable to try to expand into India. This could lead for the whole expansion to fail if we come in at too high of a price for our services. I have discussed this in previous limitation sections, but I do not think that any business should be making decisions off a single snapshot of time when it comes to looking at customer’s patterns. I believe having a monthly snapshot for over a year could lend much better insight to churn patterns.

**D**

Suraj520. (n.d.). Telecom Churn Dataset. Kaggle. Retrieved from <https://www.kaggle.com/datasets/suraj520/telecom-churn-dataset>  
**E**

No sources were used for using in-text citations and references.  
**F**

This assignment was done with professional communications throughout.