A1: Datasets

For the interactive Data Dashboard which contains my data analysis that is published at Tableau for general viewing. The two datasets I am using for my analysis are below:

- Churn_clean.csv
- acs2017_census_tract_data_unix.csv

I will provide these in my submission as well in case they are needed. The below are the steps to create my two tables and copy the data from my CSV files provided in my D211_File.zip. You will need to copy the D211.zip file to the "C:\Users\Public\Downloads" folder. Then right-click on the D211_Files.zip and selecting "Extract All..." to the C:\Users\Public\Downloads folder.

- Open pgAdmin4 on the VM Desktop.
- Expand the Servers
- Expand the Databases
- Expand the Churn Databases
- Expand Schemas
- Expand Tables
- Open a new Query tab
- Run the below queries in order. 1_create_tables.txt, 2_convert_state_to_abbriation.txt, 3_churn_age_group.txt, 4_joining_the_two_tables.txt
- After refresh the Tables to see the 2 new added tables.

A2: Installing Dashboard File

I will be providing the Tableau file along with my dataset files in a zipped folder called "D211_Files.zip". The two dataset files are for section A1 and A4. You will need to get it to the VM and then just double click on the file. Is it already export to the version of Tableau that is on the VM version 2021.4.

Steps to install Dashboard:

- Extract the contents of D211_Dashboard.twbx as it's in a zipped format, using the right-click context menu and selecting "Extract All..." to the C:\Users\Public\Downloads folder. This should have been done in section A1.
- Navigate to C:\Users\Public\Downloads\D211_Files\D211_Files\ and double left click the D211_Dashboard.twbx file.
- Once the D21_Dashboard.twbx opens click on Data Source tab on the bottom left.
- When the file explorer screen pops up select go to the C:\Users\Public\Downloads\D211_Files\D211_Files\s folder.
- select D211_Files folder again. You should now see the 2 dataset csv files.
- Click on the Churn_clean.csv file then click open.
- Click on the acs2017_census_tract_data_unix.csv file then click open.

- Click on the white space in Tableau to refresh and see the connections.
- You will now see that the 2 connections are connected and they should be join with a left join already.
- Then click on the Dashboard1 tab.
- The dashboard should now load into Tableau, ready for interaction and evaluation.

A3: Navigation Guide for "My Dashboard"

Opening the Dashboard:

- Double-click on the provided Tableau Packaged Workbook (D211_Dashboard.twbx) file to open it in Tableau.

Exploring the Dashboard:

- The dashboard is composed of several interactive elements: a map (Churn_Map), a bar graph (Income_Bar_Graph), a line graph (Tenure_Line_Graph), and a scatter plot (MonthlyCharge_ScatterPlot).
- At the top right, you will see key metrics such as Churn Cost and Turn Rate.

Using Filters:

- AGE and States Abbr Filters: On the left side, you can filter the entire dashboard by age groups and state abbreviations. Simply check or uncheck the boxes to include or exclude data in the visualizations.
- Churn Filter: Below the age and state filters, you can filter by churn status. Select either "Yes" or "No" to view the data for customers who have churned or not.

Interacting with the Churn Map:

- To Zoom/Pan: Use your mouse scroll wheel or the provided map tools to zoom in and out. Click and drag to pan across the map.
- Viewing Details: Hover over any point on the map to see more information about that specific location.

Analyzing the Income Bar Graph:

- Each bar represents a different income bracket. The colors distinguish between churned (orange) and retained (blue) customers within each bracket.
- Hover over any bar to get the exact number of customers in each category.

Observing the Tenure Line Graph:

- This graph displays the churn rate over different tenure bins. A higher line indicates a higher churn rate at that tenure.
- Hover over any point on the line to see the churn rate for that specific tenure.

Examining the Monthly Charge Scatter Plot:

- The scatter plot shows the relationship between monthly charges and tenure, with orange dots representing churned customers and blue dots representing current customers.
- You can identify patterns such as clusters of churn at certain monthly charge levels or tenures by observing the concentration of orange dots.

Interpreting Key Metrics:

- In the top right corner, you have important KPIs such as Churn Cost and Turn Rate. These give you a quick snapshot of the financial impact of churn and the rate at which customers are leaving.

Exporting Data or Images:

- If you need a snapshot of any graph or the entire dashboard, you can export it by going to 'Worksheet' (for individual views) or 'Dashboard' (for the whole dashboard) at the top menu, and then selecting 'Export' followed by 'Image' or 'Data'.

A4: SQL and Other codes support my Dashboard

Step 1

-- Creating churn_data table and importing data from CSV file.

CREATE TABLE public.churn_data (CaseOrder INTEGER, Customer_id VARCHAR(255), Interaction VARCHAR(255), UID VARCHAR(255), City VARCHAR(255), State VARCHAR(255), County VARCHAR(255), Zip INTEGER, Lat NUMERIC(10, 6), Lng NUMERIC(10, 6), Population INTEGER, Area VARCHAR(255), TimeZone VARCHAR(255), Job VARCHAR(255), Children INTEGER, Age INTEGER, Income NUMERIC(15, 2), Marital VARCHAR(255), Gender VARCHAR(255), Churn VARCHAR(255), Outage_sec_perweek NUMERIC(10, 2), Email INTEGER, Contacts INTEGER, Yearly_equip_failure INTEGER, Techie VARCHAR(255), Contract VARCHAR(255), Port_modem VARCHAR(255), Tablet VARCHAR(255), InternetService VARCHAR(255), Phone VARCHAR(255), Multiple VARCHAR(255), OnlineSecurity VARCHAR(255), OnlineBackup VARCHAR(255), DeviceProtection VARCHAR(255), TechSupport VARCHAR(255), StreamingTV VARCHAR(255), StreamingMovies VARCHAR(255), PaperlessBilling VARCHAR(255), PaymentMethod VARCHAR(255), Tenure NUMERIC(10, 2), MonthlyCharge NUMERIC(10, 2), Bandwidth_GB_Year NUMERIC(15, 2), Item1 INTEGER, Item2 INTEGER, Item3 INTEGER, Item4 INTEGER, Item5 INTEGER, Item6 INTEGER, Item7 INTEGER, Item8 INTEGER);

ALTER TABLE public.churn_data OWNER TO postgres;

-- Import the Churn_data CSV into the Churn_data table created. Please place the files in the \Public\Downloads\ folder in the VM.

COPY public.churn_data FROM 'C:\Users\Public\Downloads\D211_Files\D211_Files\churn_clean.csv' WITH (FORMAT csv, DELIMITER ',', HEADER true);

-- Create acs2017_census_data table and importing data from CSV file.

CREATE TABLE public.acs2017_census_data (TractId BIGINT, State VARCHAR(255), County VARCHAR(255), TotalPop INTEGER, Men INTEGER, Women INTEGER, Hispanic NUMERIC(5, 2), White NUMERIC(5, 2), Black NUMERIC(5, 2), Native NUMERIC(5, 2), Asian NUMERIC(5, 2), Pacific NUMERIC(5, 2), VotingAgeCitizen INTEGER, Income NUMERIC(10, 2), IncomeErr NUMERIC(10, 2), IncomePerCap NUMERIC(10, 2), IncomePerCapErr NUMERIC(10, 2), Poverty NUMERIC(5, 2), ChildPoverty NUMERIC(5, 2), Professional NUMERIC(5, 2), Service NUMERIC(5, 2), Office NUMERIC(5, 2), Construction NUMERIC(5, 2), Production NUMERIC(5, 2), Drive NUMERIC(5, 2), Carpool NUMERIC(5, 2), Transit NUMERIC(5, 2), Walk NUMERIC(5, 2), OtherTransp NUMERIC(5, 2), WorkAtHome NUMERIC(5, 2), MeanCommute NUMERIC(5, 2), Employed INTEGER, PrivateWork NUMERIC(5, 2), PublicWork NUMERIC(5, 2), SelfEmployed NUMERIC(5, 2), FamilyWork NUMERIC(5, 2), Unemployment NUMERIC(5, 2));

ALTER TABLE public.acs2017_census_data OWNER TO postgres;

-- Please place the files in the \Public\Downloads\ folder in the VM.

COPY public.acs2017_census_data FROM 'C:\Users\Public\Downloads\D211_Files\D211_Files\acs2017_census_tract_data_unix.csv' WITH (FORMAT csv, DELIMITER ',', HEADER true);

-- referential integrity

-- create states table

CREATE TABLE public.states (state_id SERIAL PRIMARY KEY, state_abbreviation VARCHAR(2) UNIQUE, state_name VARCHAR(255));

-- Add a state_id column

ALTER TABLE public.churn_data ADD COLUMN state_id INT; ALTER TABLE public.acs2017_census_data ADD COLUMN state_id INT;

-- Establish foreign key relationships:

ALTER TABLE public.churn_data ADD CONSTRAINT fk_state_id FOREIGN KEY (state_id) REFERENCES public.states(state_id) ON DELETE RESTRICT ON UPDATE CASCADE;

ALTER TABLE public.acs2017_census_data ADD CONSTRAINT fk_state_id FOREIGN KEY (state_id) REFERENCES public.states(state_id) ON DELETE RESTRICT ON UPDATE CASCADE;

Step 2

-- Convert State to Abbriation on the acs2017_census_data table.

UPDATE public.acs2017_census_data SET State = CASE State WHEN 'Alabama' THEN 'AL' WHEN 'Alaska' THEN 'AK' WHEN 'Arizona' THEN 'AZ' WHEN 'Arkansas' THEN 'AR' WHEN 'California' THEN 'CA' WHEN

'Colorado' THEN 'CO' WHEN 'Connecticut' THEN 'CT' WHEN 'Delaware' THEN 'DE' WHEN 'Florida' THEN 'FL' WHEN 'Georgia' THEN 'GA' WHEN 'Hawaii' THEN 'HI' WHEN 'Idaho' THEN 'ID' WHEN 'Illinois' THEN 'IL' WHEN 'Indiana' THEN 'IN' WHEN 'Iowa' THEN 'IA' WHEN 'Kansas' THEN 'KS' WHEN 'Kentucky' THEN 'KY' WHEN 'Louisiana' THEN 'LA' WHEN 'Maine' THEN 'ME' WHEN 'Maryland' THEN 'MD' WHEN 'Massachusetts' THEN 'MA' WHEN 'Michigan' THEN 'MI' WHEN 'Minnesota' THEN 'MN' WHEN 'Mississippi' THEN 'MS' WHEN 'Missouri' THEN 'MO' WHEN 'Montana' THEN 'MT' WHEN 'Nebraska' THEN 'NE' WHEN 'Nevada' THEN 'NV' WHEN 'New Hampshire' THEN 'NH' WHEN 'New Jersey' THEN 'NJ' WHEN 'New Mexico' THEN 'NM' WHEN 'New York' THEN 'NY' WHEN 'North Carolina' THEN 'NC' WHEN 'North Dakota' THEN 'ND' WHEN 'Ohio' THEN 'OH' WHEN 'Oklahoma' THEN 'OK' WHEN 'Oregon' THEN 'OR' WHEN 'Pennsylvania' THEN 'PA' WHEN 'Rhode Island' THEN 'RI' WHEN 'South Carolina' THEN 'SC' WHEN 'South Dakota' THEN 'SD' WHEN 'Tennessee' THEN 'TN' WHEN 'Texas' THEN 'TX' WHEN 'Utah' THEN 'UT' WHEN 'Vermont' THEN 'VT' WHEN 'Virginia' THEN 'VA' WHEN 'Washington' THEN 'WA' WHEN 'West Virginia' THEN 'WV' WHEN 'Wisconsin' THEN 'WI' WHEN 'Wyoming' THEN 'WY' ELSE State END;

Step 3

ALTER TABLE public.churn_data ADD COLUMN Age_Group VARCHAR(5);

UPDATE public.churn_data SET Age_Group = CASE WHEN Age < 30 THEN '18-29' WHEN Age < 50 THEN '30-49' ELSE '50+' END;

B: Panopto Video

https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=4b16c8b4-b6a1-4cc6-8ea9-b141017289a0

C1: Dashboard Purpose and Function

In the age of data-driven decision-making, the ability to visualize and interact with data is invaluable. The dashboard I've developed serves as a window into the complex mechanics of customer churn and demographics, leveraging data from two distinct yet complementary datasets to provide executives with actionable insights. The dashboard's core purpose is to bring the variables defined in the data dictionary to life. For instance, the data dictionary may outline key demographics and churn metrics. my dashboard visualize these variables into visual stories. Through maps, graphs, and KPIs, it doesn't just display data. It makes it speak to the users, ensuring alignment with the emphasis on understanding customer behavior and its impact on revenue.

C2: Business Intelligence Tool Justification

Tableau was selected as the business intelligence tool for this project due to its robust data visualization capabilities, ease of use, and dynamic interactivity. Its ability to integrate with SQL databases and perform on-the-fly calculations makes it ideal for creating a responsive and informative dashboard that can adapt to the evolving needs of executive stakeholders.

C3: Data Cleaning and Preparation

Data integrity is paramount in analytical operations, especially when the analysis is expected to influence executive decision-making. The process of preparing the datasets for inclusion in the dashboard involved a series of cleaning and normalization steps, detailed as follows:

Data Importation:

- The raw data was initially imported into PostgreSQL. This was done using the COPY command, which allows for bulk insertion of rows from CSV files into the corresponding tables public.churn_data and public.acs2017_census_data. This step necessitated ensuring that the CSV files were properly formatted, with columns matching the table schema and data types compatible with the SQL definitions.

Standardization of State Abbreviations:

- The state fields within the imported datasets featured full state names. To maintain consistency and enable efficient joins with other datasets that might use state abbreviations, a CASE statement was applied to convert full state names into their respective two-letter abbreviations. This ensured uniformity across the dataset and facilitated more accurate geographical analysis on the dashboard.

Age Categorization:

- The age data was categorized into brackets: '18-29', '30-49', and '50+'. This classification allowed for a more granular analysis of the customer base and enabled the dashboard to display insights tailored to specific age demographics, which are often critical in understanding and addressing churn patterns.

Identification and Resolution of Inconsistencies:

- The datasets underwent a thorough examination for any inconsistencies or anomalies, such as duplicate records, outliers, or implausible values that could skew the analysis. Where applicable, records were corrected or removed as per predefined data governance rules to ensure the accuracy of reporting.

Normalization:

- Data normalization included scaling numerical values where necessary and ensuring that categorical data followed a consistent format. For example, monetary values were verified for format consistency, and categorical fields such as 'Yes/No' columns were checked to ensure uniform capitalization.

Data Validation:

- Post-cleaning, the data was validated by running a series of SELECT statements to ensure that the data transformations were applied correctly. This step was crucial to verify that the data aligned with the expected outcomes and that no errors were introduced during the cleaning process.

C4: Dashboard Creation Steps

The development of the D211_Dashboard in Tableau was a multi-stage process that transformed raw data into a compelling visual narrative for executive decision-making. Here's a breakdown of each step:

Database Connection:

- Initiated the dashboard creation by establishing a live data connection to a PostgreSQL database. This step ensured real-time data availability for the most up-to-date analysis.

Data Joining:

- Executed custom SQL queries within Tableau to seamlessly join the churn_data with acs2017_census_data. Careful attention was paid to match the State fields across both datasets to enable accurate, combined analysis.

Calculated Fields:

- I designed calculated fields to categorize age into meaningful groups and convert state names to abbreviations. For instance, ages were classified into '18-29', '30-49', and '50+' to analyze demographic trends in churn behavior, while state names were converted to abbreviations for uniformity with other data elements.

Iterative Design:

- Through an iterative design process, various visual representations were explored. The Churn Map was crafted to depict customer distribution geographically, while the Tenure Line Graph illustrated churn rate trends over different tenure lengths. Each visualization underwent numerous refinements to ensure clarity and impact.

Interactivity:

- I integrated interactive filters for age and state abbreviations allowed users to tailor the dashboard view to their specific analysis needs. This interactivity empowered users to explore the data from multiple dimensions, drilling down to the most pertinent information.

Visual Refinement:

- I paid meticulous attention to the dashboard's aesthetic aspects, selecting color schemes that accommodate color vision deficiencies and arranging elements for a clean, user-friendly layout.

Performance Indicators:

- I incorporated key performance indicators (KPIs) such as Churn Cost and Turn Rate prominently within the dashboard. These KPIs were calculated using combined data, providing executives with immediate visibility into critical metrics.

Feedback Incorporation:

- Conducted reviews with stakeholders to gather feedback and validate the dashboard's efficacy in conveying the intended insights, leading to further refinements.

C5: Results of Data Analysis

The analysis of customer churn and demographic data yielded significant insights that are instrumental for strategic decision-making. For instance:

- Churn Map Visualization: By geographically mapping churn data, the dashboard highlighted regions with the highest customer attrition rates. This enables executives to pinpoint areas that may require targeted customer service improvements or market analysis.
- Income Bar Graph: The relationship between customer income brackets and churn rates was brought to light, with lower income brackets exhibiting higher churn. This suggests that pricing or service value could be key factors affecting customer retention in these segments.
- Tenure Line Graph: A clear pattern was discerned from the tenure analysis - as customer tenure increases, the likelihood of churn decreases. This underlines the importance of customer loyalty programs and the value of long-term customer relationships.
- Monthly Charge Scatter Plot: An interesting pattern emerged from the scatter plot, showing that higher monthly charges are not always correlated with higher churn rates, indicating a complex interplay between service satisfaction, perceived value, and pricing.

These analytical outcomes enable executives to:

- Prioritize regions for market research and customer satisfaction initiatives.
- Tailor marketing and pricing strategies to address the needs of sensitive income brackets.
- Design loyalty programs that incentivize long-term engagement.
- Reevaluate pricing strategies to ensure competitiveness without compromising on profit margins.

C6: Limitations of Data Analysis

Despite the rich insights gleaned from the dashboard, the analysis is subject to certain limitations:

- Historical Data Constraint: The current analysis is grounded in historical data, which inherently limits the ability to predict future trends. As customer behavior and market dynamics evolve, the analysis may not fully capture emerging patterns.
- Data Entry Quality: The reliability of churn predictions and other insights is directly tied to the accuracy of the data collected. Any inconsistencies in data entry or data collection processes can lead to misinformed conclusions.
- Market Variables Exclusion: External factors such as economic shifts, competitive actions, and changes in consumer preferences are not encapsulated in the datasets. Their omission can lead to an incomplete understanding of the churn phenomenon.

- Lack of Real-time Data: The dashboard does not integrate real-time data streams, which could provide more timely insights into customer behavior and market changes, enabling more agile decision-making.

To surmount these limitations, it is recommended to:

- Incorporate forward-looking analytics and predictive modeling to anticipate future trends.
- Enhance data governance and validation processes to improve data quality.
- Broaden the data scope to include external market intelligence and competitor analysis.
- Invest in capabilities to process and analyze real-time data for up-to-the-minute insights.

Recognizing these limitations is essential for executives to understand the context and the applicability of the data analysis, ensuring that strategic decisions are made with a comprehensive perspective.

D: Source

(Kaggle, 2024): https://www.kaggle.com/datasets/muonneutrino/us-census-demographic-data

E: Source

No external source was used