Advanced Data Acquisition PA

Vincent Taylor

Western Governors University

Course Number: D211

Part I – Data Dashboards – Section A

1. Copy of both datasets used in this analysis. select CTRL + click
   1. [Churn](https://access.wgu.edu/ASP3/aap/content/f9tjr8djg83jd8c3sdf8.zip) – Link to WGU
   2. [Kaggle Telecom Data](https://www.kaggle.com/code/bhartiprasad17/customer-churn-prediction/data) – Link to Kaggle website for csv file
2. **Dashboard Installation Instructions (A2)**
   1. Take the d211.zip file and move the file to ‘C:\Users\Public\Downloads’
   2. From here right-click and select extract all, this will create a new folder.
   3. Go back to the desktop and click on pgAdmin.
   4. Once this has been opened from the left most panel select Servers > Databases > Churn.
   5. Right-click on the Churn database and select query tool.
   6. Once this has been selected there will be a folder in the top right select it and navigate to ‘C:\Users\Public\Downloads’ and select the cleaning\_data\_code.txt file. (*Note if you cannot find it in this folder change the format from sql to All Files*)
   7. Once text file has loaded into the query editor, from your keyboard select Fn + F5 this will execute the query. (*Note you may also select the Play button and this will execute the code.*)
   8. Once the query has been executed successfully exit out of PgAdmin and from your desktop open Tableau 2021.4
   9. From the top left select File > Open > Navigate to ‘C:\Users\Public\Downloads’ > and select the d211.tbwx file.
   10. Once you have opened this file you will be asked for a username and password, please use the following (*Note this case sensitive*): **Username**: postgres / **Password**: Passw0rd!
   11. Congratulations! You should now have a working dashboard and can move to A3 for Dashboard installation instructions.
3. **Navigating the Dashboard (A3)**
   1. From here everything should be included on the tab d211\_Dashboard. This contains key KPI metrics on our consumers compared to an industry competitor. We have both our customers which contains 10,000 unique customer responses and our competitor data has about 7000 unique customer responses. To segment our dashboard by each company we can use the True/False filter with our company representing True and the competitor representing False. We also have demographic information on our customers such as the percentage of male/female/Nonbinary customers. All of the dashboard worksheets can also be filtered by clicking anywhere on the dashboard. There is also the average monthly charge based on payment method type as well as tenure. Below is a dictionary on payment types as well.
      1. 
4. **SQL Code(A4)**

|  |
| --- |
| 1. DROP TABLE IF EXISTS competitor; 2. CREATE TABLE IF NOT EXISTS competitor( 3. customerID text PRIMARY KEY, 4. gender text, 5. SeniorCitizen int, 6. Partner text, 7. Dependents text, 8. tenure int, 9. PhoneService text, 10. MultipleLines text, 11. InternetService text, 12. OnlineSecurity text, 13. OnlineBackup text, 14. DeviceProtection text, 15. TechSupport text, 16. StreamingTV text, 17. StreamingMovies text, 18. Contract text, 19. PaperlessBilling text, 20. PaymentMethod text, 21. MonthlyCharges numeric, 22. TotalCharges text, 23. Churn text 24. ); 25. copy competitor 26. FROM 'C:\Users\Public\Downloads\kaggle.csv' 27. DELIMITER ',' 28. CSV HEADER; 29. UPDATE competitor 30. SET "contract" = 31. CASE 32. WHEN "contract" = 'Month-to-month' THEN '1' 33. WHEN "contract" IN ('1 year', 'One year') THEN '2' 34. WHEN "contract" = 'Two year' THEN '3' 35. ELSE "contract" 36. END, 37. "paymentmethod" = 38. CASE 39. WHEN "paymentmethod" = 'Bank transfer (automatic)' THEN '1' 40. WHEN "paymentmethod" = 'Credit card (automatic)' THEN '2' 41. WHEN "paymentmethod" = 'Electronic check' THEN '3' 42. WHEN "paymentmethod" = 'Mailed check' THEN '4' 43. ELSE "paymentmethod" 44. END; 45. ALTER TABLE competitor 46. RENAME COLUMN SeniorCitizen TO senior\_citizen; 47. ALTER TABLE competitor 48. RENAME COLUMN partner TO married; 49. ALTER TABLE competitor 50. RENAME COLUMN PhoneService TO phone; 51. ALTER TABLE competitor 52. RENAME COLUMN MultipleLines TO multiple\_lines; 53. ALTER TABLE competitor 54. RENAME COLUMN InternetService TO internet\_service; 55. ALTER TABLE competitor 56. RENAME COLUMN OnlineSecurity TO online\_security; 57. ALTER TABLE competitor 58. RENAME COLUMN OnlineBackup TO online\_backup; 59. ALTER TABLE competitor 60. RENAME COLUMN DeviceProtection TO device\_protection; 61. ALTER TABLE competitor 62. RENAME COLUMN TechSupport TO tech\_support; 63. ALTER TABLE competitor 64. RENAME COLUMN StreamingTV TO streaming\_tv; 65. ALTER TABLE competitor 66. RENAME COLUMN StreamingMovies TO streaming\_movies; 67. ALTER TABLE competitor 68. RENAME COLUMN Contract TO contract\_id; 69. ALTER TABLE competitor 70. RENAME COLUMN Customerid TO customer\_id; 71. ALTER TABLE competitor 72. RENAME COLUMN PaperlessBilling TO paperless\_billing; 73. ALTER TABLE competitor 74. RENAME COLUMN PaymentMethod TO payment\_id; 75. ALTER TABLE competitor 76. RENAME COLUMN MonthlyCharges TO monthly\_charge; 77. ALTER TABLE competitor 78. RENAME COLUMN TotalCharges TO total\_charge; 79. ALTER TABLE competitor 80. ADD COLUMN company bool DEFAULT false; 81. ALTER TABLE customer 82. ADD COLUMN company bool DEFAULT true; 83. ALTER TABLE competitor 84. ALTER Column contract\_id TYPE Numeric 85. USING contract\_id::numeric; 86. ALTER TABLE competitor 87. ALTER Column payment\_id TYPE Numeric 88. USING payment\_id::numeric; 89. CREATE TABLE combined\_df AS 90. SELECT 91. subquery.customer\_id, 92. subquery.gender, 93. subquery.churn, 94. subquery.round\_tenure, 95. subquery.monthly\_charge, 96. subquery.payment\_id, 97. subquery.contract\_id, 98. subquery.company 99. FROM ( 100. SELECT 101. customer\_id, 102. gender, 103. churn, 104. round(tenure) as round\_tenure, 105. monthly\_charge, 106. payment\_id, 107. contract\_id, 108. company 109. FROM customer 111. UNION ALL 113. SELECT 114. customer\_id, 115. gender, 116. churn, 117. round(tenure) as round\_tenure, 118. monthly\_charge, 119. payment\_id, 120. contract\_id, 121. company 122. FROM competitor 123. ) subquery; 124. UPDATE customer SET company = true; |

Part II – Demonstration – Section B

1. The Panopto video for my presentation can be found [here](https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=68d1967a-5261-4ae6-8051-b0f701684520&edit=true).

Part III – Report – Section C

1. The purpose of this dashboard was to be able to pull key KPI metrics for our finance team in order to help with a future potential campaign in order to retain more long-term customers. In order to assist the team, they asked for a dashboard that pulled our data and compared it against a dataset from one of our competitors and would be able to be continuously updated for future quarters. I was able to find a similar dataset to our in house data from Kaggle that compared customer tenure as well as monthly charges alongside payment methods in order to complete this assignment.
2. **Justify Tool (C2)**

The business intelligence tool used for this analysis was tableau. I believe that tableau was the best choice for this assignment due to it being a user-friendly app, that also provides immense data visualization tools. Tableau is also a useful tool in terms of being able to help with real-time capabilities. Lastly tableau supports a wide range of various formats in terms of where it can pull data from. Using it in tandem with PgAdmin allows for us to complete two separate parts of the data lifecycle.

1. **Data Preparation (C3)**

Below are the steps used to prepare the data. The customer data was already prepared and due to the way it was prepared I decided to match the external dataset :

* 1. I created a table named "competitor" is created with various customer demographic data representing customer information, such as gender, tenure, contract details, payment method, and churn status.
  2. I then import the external data from a CSV file ('kaggle.csv') into the "competitor" table using the COPY command.
  3. Next I use the UPDATE command to transform the "contract" and "paymentmethod" columns into numeric values to standardize the data.
  4. I then need to rename the columns in order to create consistency with our customer data table I was able to do this using the ALTER TABLE ... RENAME COLUMN commands.
  5. I then create a a new column (Boolean) called company and I set the default value for all of the external data to false.
  6. I then use the UPDATE command to set the "company" column to true for all records in the "customer" table. This indicates that these records belong to the company.
  7. Lastly, I needed to create a new table with the combined data that will be needed for visualization. I combined the data using the UNION ALL operator.
  8. Note at the very end I did use the UPDATE command to set the "company" column to true for all records in the "customer" table.

1. **Steps to create dashboard (C4)**

Average Tenure by Churn:

* Move Churn to Rows
* Move Round Tenure to the text marks
* Right-click and change the measure type from SUM to Average
* Change the visualization to text table.
* Move Company to Filters and right click and select show filter
* Rename title to Tenure by Churn and center it as well as change the font to bold

Average Charge by Payment Method

* Create a new Worksheet
* Move payment id to color as well as rows
* Move Monthly Charge to columns and change the type to AVG
* Under the marks change the chart type to Bar
* Rename the Title to Average Charge by payment method.

Customer Count by Tenure

* Move Gender to the color
* Move comined\_df count to the label mark
* Move the combined\_df count to detail
* Move monthly charge to the filters and right click and select show filter

Dashboard

* Select Tab Create new Dashboard
* Move count of customers by gender to the top left
* Select use as filter
* Move average charge by payment type to the middle
* Select use as filter
* Move tenure by gender below average charge by payment type
* Select use as filter
* Lastly move tenure by churn to the far right and select use as filter

1. **Results of the Analysis (C5)**

One of the stories I wanted to be able to tell is that we have more customers who pay by check than our competitors. Looking at the dataset 23% of our customers pay by check which for comparison would put us in line with our competitors who check represents 23% of their payment method as well, with the key difference of almost 3000 less responses. This would lead me to recommend to senior management that we would need to have our representatives continue to push an electronic method. Another important KPI for senior management in order to figure out which demographic is most likely to churn based on their contract.

1. **Limitations of data analysis (C6)**

A critical limitation is that we are missing offerings such as Phone, Internet service, etc. not having this in our Churn database makes it harder for us to pull actual useful insights against our customer data. In D210 we had that data available which led to much more useful insights compared to what is available in our PgAdmin query.

References

No third-party references were used for this assignment.

References

No third-party code was used for this assignment.