D598 Task 3

Aaron Pacheco

1. The following is an explanation of how my code from Task 2 functions.

Step 1: The code first imports the pandas library, which allows us to create and manipulate dataframes. Then it imports a .csv of the original data and formats it as a dataframe.

Step 2: The code cleans the data by identifying and dropping any rows that are exact duplicates of previous rows.

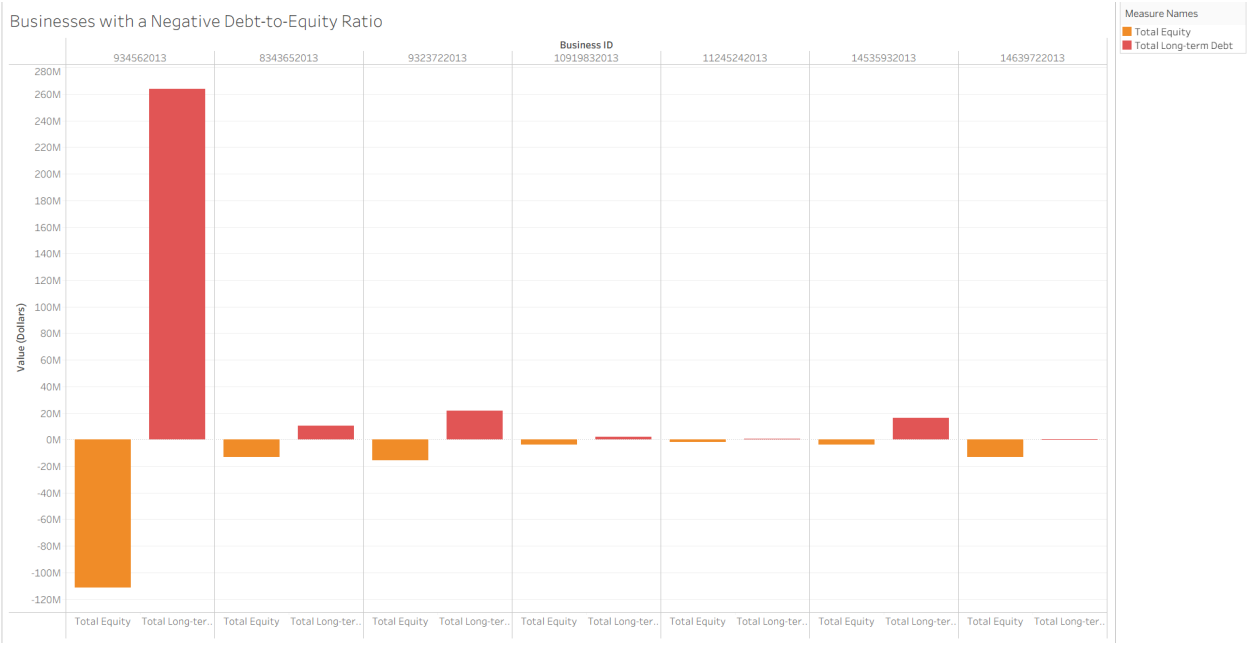
Step 3: The code then defines the columns which we want to calculate statistics of as "cols_to_agg". Then, the code creates a new dataframe, where it groups all of the original data by "business state" and calculates each state's mean, median, minimum and maximum for each of the "cols_to_agg". Finally, it prints this new dataframe.

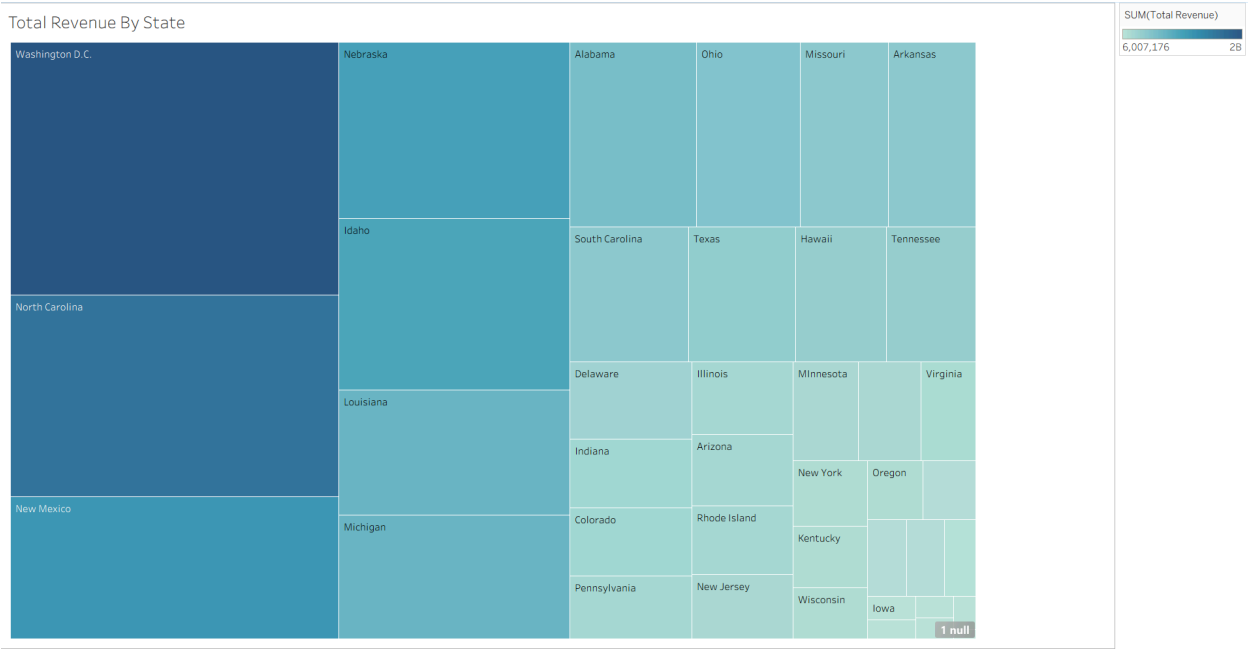
Step 4: The code filters the original data for only businesses with a negative debt-to-equity ratio and places these businesses in a new dataframe. Then it prints that new dataframe.

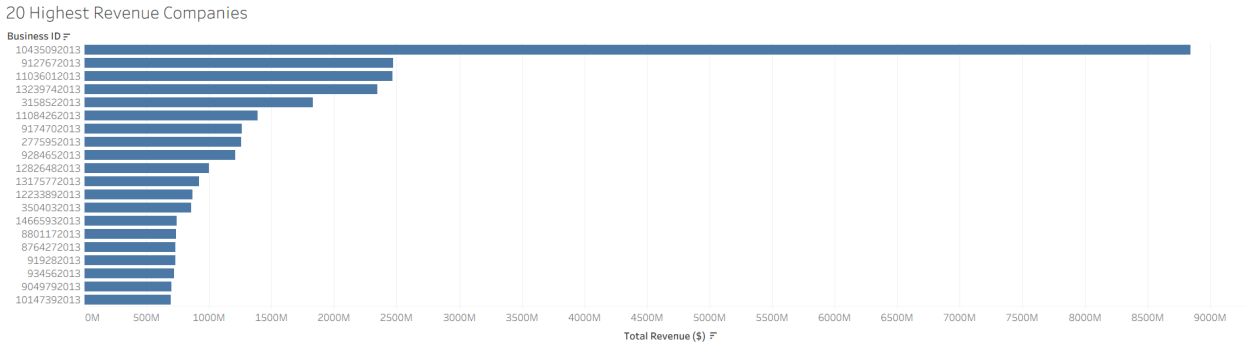
Step 5: The code creates a new dataframe including only Business IDs and a debt-to-income ratio, which it calculates by dividing "long-term debt" by "total revenue". There is also a failsafe included to ensure that this calculation is only performed when "total revenue" isn't 0. If it is zero, then the code instead returns an empty cell.

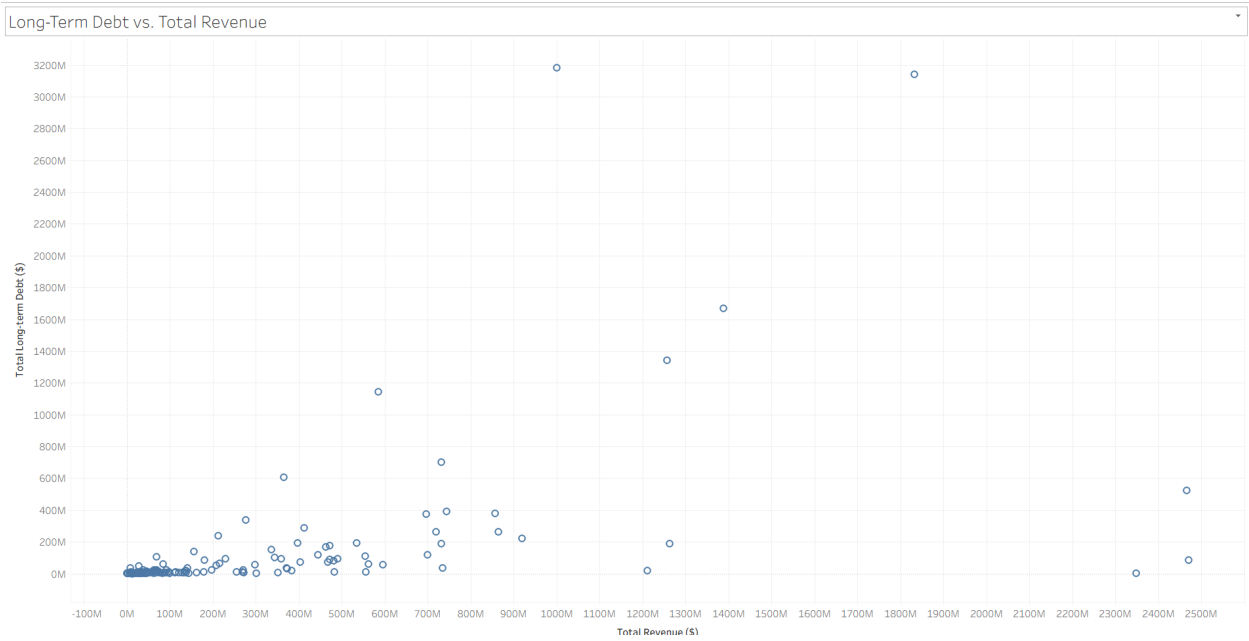
Step 6: The code merges the new dataframe from step 5 with our original data frame, such that the "debt-to-income" ratios that we just calculated were added to the original dataframe, aligned by the business IDs. Then it prints this new complete dataframe.

2.


Businesses with a Negative Debt-to-Equity Ratio
Business ID
934562013
8343652013
9323722013
10919832013
11245242013
14535932013
14639722013
Value (Dollars)
280M
260M
240M
220M
200M
180M
160M
140M
120M
100M
80M
60M
40M
20M
0M
-20M
-40M
-60M
-80M
-100M
-120M
Total Equity
Total Long-ter..
Measure Names
Total Equity
Total Long-term Debt


Total Revenue By State
Washington D.C.
Nebraska
Alabama
Ohio
Missouri
Arkansas
Idaho
South Carolina
Texas
Hawaii
Tennessee
North Carolina
Delaware
Illinois
MInnesota
Virginia
Louisiana
Indiana
Arizona
New York
Oregon
New Mexico
Colorado
Rhode Island
Kentucky
Michigan
Pennsylvania
New Jersey
Wisconsin
Iowa
1 null
SUM(Total Revenue)
6,007,176
2B





3. All customized visualizations were created in Tableau using data printed from the code in Task 2.

The first visualization is a side-by-side bar graph showing all of the companies with a negative debt-to-equity ratio. This visualization shows both the long term debt and the total equity side-by-side. The scale of the graph is chosen to accurately fit all data, and to highlight specifically the exceptional amount of debt and the severely negative equity of the company with Business ID 934562013. Red was chosen to represent the long-term debt because it is a color that will raise alarm over the large amounts of debt these companies have taken on. Orange was chosen to represent the total equity to raise alarm at the fact that these companies are presenting negative equity. The title clearly defines that we are looking specifically at companies with a negative Debt-to-Equity ratio and the key provides insight into what specific data points we are analyzing.

The second visualization is a treemap that shows the portion of the total revenue we are managing comes from each state. A treemap was chosen over something like a pie chart because of the large number of different states we have data for. A pie chart for this data was far too busy and made it nearly impossible to distinguish between many of the individual states. A tree map will more clearly demonstrate how much of the total revenue we manage is held by just a few states, namely Washington DC, North Carolina, and New Mexico. This major difference in revenue is not only demonstrated by the size of each state's box, but also by the color, as each box occupies a color within a gradient of blue, light blue meaning a small portion of total revenue and deep blue representing a very large portion of the total revenue. Blue was chosen since it is a neutral color that demonstrates that all of these states are providing revenue, but some are providing more than others. The scale of the graph accurately portrays just how large of a percentage of the total revenue of our companies is coming from such a small number of states, with nearly 50% of revenue coming from only DC, NC, NM, NE, ID, LA, and MI. The title clearly states that we are looking at the total revenue, broken down by state. The key provides a gradient, showing that the darkest blue represents a revenue of about $2 Billion and the lightest blue represents a revenue of around $6 million.

The third visualization is a bar graph showing the revenues of each of the 20 highest revenue companies in our data. The scale is chosen to accurately show the revenue of all of the companies in the top 20, and to highlight specifically how large of a revenue our largest company, Company 10435092013, has compared to the rest of the companies in our database. Once again, the color blue was chosen as a neutral color, meant to show that this is neutral revenue data. There was no need to emphasize anything in the graph with color, since the sheer size of the largest bar is enough to illustrate the sheer size of the highest revenue company. The title clearly states that we are looking specifically at the 20 highest revenue companies in our data, while our axes clearly define that we are examining revenue numbers for each company.

The fourth visualization is a scatter plot demonstrating the relationship between a company's long-term debt and that company's total revenue. The scale was chosen because it shows both the high concentration of relatively low-debt, low-revenue companies in our data but also shows the slight correlation between an increasing revenue and an increasing amount of long-term debt. Once again, the color blue was chosen as a neutral color, because my goal with this visualization was simply to present a correlation between the two values, not to induce any other particular conclusions. The title and axes

labels clearly define that we are looking at long-term debt data via the y-axis and revenue data via the x-axis.

4. No citations are necessary for this assignment.