**Task 3: Presentation**

Joseph Emmanuel Cayetano

Western Governors University

D598: Analytics Programming Task 3

September 30, 2024

**A. Explain how the code works for the program you submitted in Task 2**

I will be explaining the code for each step in my data analysis program. I will provide an image for each code for visualization.

A close up of a text

Description automatically generated

In this image, I am using python’s import statement to import two popular libraries for data manipulation, which are pandas and NumPy. I also used an alias for both libraries (pd and np) to easily reference the libraries’ functions without the need for typing the full name of the libraries each time.

A close up of a text

Description automatically generated

Here, I am using the read\_excel function from pandas to read the dataset to be used for our data analysis. The file name of the data set is “D598 Data Set.xlsx.” I am storing the data set into a variable called ‘df.’ This variable will be the original data frame to be used throughout the analysis.

A screenshot of a computer screen

Description automatically generated

In this image, I am using the describe function from pandas to provide basic statistics for each numerical column in the ‘df’ data frame.

A screenshot of a computer code

Description automatically generated

Here, I am using the dtypes function from pandas to display the data types for all the columns in the data. The column ‘Business State’ is the only column that has a string data type. The rest have numerical data types.

A screenshot of a computer

Description automatically generated

In this image, I am using isnull and sum functions to add the null values for each column. There are no null values in the dataset.

A white rectangular object with black text

Description automatically generated

Here, I am using the duplicated function to check for duplicate rows in the data and storing it into the duplicates variable. The data does not contain any duplicate rows.

A screenshot of a computer

Description automatically generated

In this image, I am using the groupby function to group the data by ‘Business State,’ and calculated the mean, median, min, and max for each numeric column by using the agg function. Below is the result of the code.

A close up of a text

Description automatically generated

Here, I am storing the grouped data from above into a new variable called ‘df\_grouped.’ It is now a newly created data frame.

A screenshot of a graph

Description automatically generated

In this image, I filtered the data by selecting only the rows that have negative debt to equity values. Out of 150 rows, only seven were returned, which means only seven businesses have negative debt to equity values in the data set.

A screenshot of a screen

Description automatically generated

Here, I am calculating debt to income ratios for all businesses by dividing long term debt by total revenue and storing the calculation into a new column called ‘Debt to Income Ratio.’ After the calculation, an additional column was added to the original data frame.

A screenshot of a data frame

Description automatically generated

In this image, I have created a new data frame called ‘df2’ that only consists of two columns: business ID and debt to income ratio. The column, debt to income ratio, is a newly created column from the previous step. The data frame ‘df2’ will be combined with the original data frame ‘df’ in the next step.

A screenshot of a computer screen

Description automatically generated

Here, I merged the newly created data frame ‘df2’ with the original data frame ‘df’ using the concat function from pandas. The final data frame now contains the ‘debt to income ratio’ column. I had to drop the ‘debt to income ratio’ that was produced two steps earlier to show that I can merge two different data frames using the concat function.

**B. Provide 4 customized data visualizations**

1st Visualization:

A graph with orange dots

Description automatically generated

2nd Visualization:

A screenshot of a graph

Description automatically generated

3rd Visualization:

A graph with numbers and lines

Description automatically generated

4th Visualization:

A pie chart with different colored circles

Description automatically generated

**C. Explain how customized visualizations in part B were created**

Visualization #1:

A graph with orange dots

Description automatically generated

I was able to create a scatter plot with regression line for ‘total liabilities vs. total revenue’ by using the regplot function from the Seaborn library. Total revenue is how much money the company generates before expenses, while total liabilities are “any debts and obligations that a company owes to another party” (Habiger, 2023, Par. 3). Regarding the parameters of the regplot function, I put the column ‘Total Liabilities’ for the X-axis and the column ‘Total Revenue’ for the Y-axis. I set the value of the ‘scatter’ parameter to ‘True’ to ensure the data points are plotted. I set the color to orange to have an orange regression line and dots. For the title, I put ‘Total Liabilities vs. Total Revenue.’ I set the label for X-axis to ‘Total Liabilities,’ and the label for Y-axis to ‘Total Revenue.’ I have also included a grid by setting the grid function to true.

This visualization suggests that U.S. companies with higher total liabilities may also tend to have higher total revenue, although the relationship between the two variables is not that strong. Some outliers can also be seen from this visualization.

Visualization 2:

A screenshot of a graph

Description automatically generated

I was able to create a bar chart that shows the total revenue for each U.S. state in the data set by using the bar function from matplotlib. Before putting values into the bar function, I sorted the data first by ‘Total Revenue’ in descending order, which means states with higher revenues are on the left, while states with lower revenues are on the right. As for the parameters of the bar function, I put the column ‘Business State’ for the X-axis, and the column ‘Total Revenue’ for the Y-axis. I set the color to orange to have orange bars. For the title, I put ‘Total Revenue by Business State.’ I set the label for X-axis to ‘Business State,’ and the label for Y-axis to ‘Total Revenue.’ I used the xticks function to rotate and resize the font size of the state labels to make it readable.

This visualization suggests that North Carolina has the highest total revenue out of all U.S. states. Also, North Carolina has a significant lead over others.

Visualization #3:

A graph with numbers and lines

Description automatically generated

I was able to create a histogram that shows the distribution of debt-to-income ratio for all U.S. businesses in the data by using the hist function from matplotlib. As for the parameters of the hist function, I put the column ‘Debt to Income Ratio’ for the X-axis, set the number of bins to 4, changed the color to orange, and set ‘alpha’ value to 0.8. The visualization is titled ‘Distribution of Debt to Income Ratio.’ I set the label for X-axis to ‘Debt to Income Ratio,’ and the label for Y-axis to ‘Frequency.’ I’ve also included a grid by setting the grid function to true.

This visualization suggests that most U.S. companies in the dataset have a low debt to income ratio that is between 0 and 1. This means these companies manage their debt relative to their income effectively.

Visualization #4:

A pie chart with different colored circles

Description automatically generated

I was able to create this pie chart that shows the top five states with the highest total liabilities by using the pie function from matplotlib. Before putting values into the pie function, I grouped the data first by ‘Business State’ and calculated the ‘Total Liabilities’ for each state. I also sorted the grouped data to descending. As for the parameters of the pie function, I put the grouped data to be created as a pie, set labels to be state names, display percentages with one decimal place, and rotate the start of the pie chart to 140 degrees. The visualization is titled ‘Top 5 States by Total Liabilities.’

This visualization shows the proportion of total liabilities among the top five states in the U.S. Texas is the state that has highest value of total liabilities.

**References**

Habiger, S. (2023, April 17). *Total liabilities: Definition & Calculation*. FreshBooks. https://www.freshbooks.com/glossary/accounting/total-liabilities?srsltid=AfmBOor1jVx0zUMypOmgXNzdqRZOwXED\_yVJ19QQAQ\_9S\_oFwfOuXrNW