Gabriela Masak

D598 – Analytics Programming

Task 3: Presentation

9/17/2024

Stakeholder Report: Financial Data Analysis

**Part 1: Explanation of Code**

The program, “D598EconomicAnalysis.py” serves to clean, manipulate, and analyze data of 150 businesses in the “D598 Data Set.xlsx” file. This Excel file is loaded into the program and checked for duplicates, which are removed, and the updated data frame is stored for future use. This future use includes grouping the businesses by their defined state and calculating the mean, median, minimum, and maximum descriptive statistical values by state of the long-term debt, equity, debt-to-equity ratios, liabilities, revenues, and profit margins. These values are then displayed in the terminal below. The program then filters the businesses to find those with negative debt-to-equity ratios, storing those that meet the criteria in a new, appropriately named data frame. The terminal then displays the list of business identification numbers with negative debt-to-equity ratios. Finally, the program calculates the debt-to-income ratios of all businesses by dividing the values found in their respective long-term debt and revenue columns. These calculated values are then added to the data frame as a new rightmost column.

**Part 2: Customized Visualizations and Explanation**

Programming with Python allows for data to be extracted, cleaned, manipulated, and analyzed, while also allowing for visualizations to be created. This is typically done using the MatPlotLib library, which offers several functions to create graphs and other visualizations. Four visualizations were created using the data set provided.

A graph of a number of people

Description automatically generated with medium confidence

Figure 1: Histogram of Debt-to-Income Ratios

Figure 1 shows a histogram that shows a skewed distribution of debt-to-income ratios, indicating that most businesses have a low, near-zero ratio. Only two businesses out of 150 have debt-to-income ratios that are greater than 2, indicating outliers. Further economic research would be needed to address whether these businesses are a viable investment. Figure 1 was created with the Python code below, with in-line comments providing explanation of the code:

plt.hist(businessDataUpdated['Debt to Income Ratio'], bins=150, color='skyblue', edgecolor='black')  
  
# Add labels and title  
plt.xlabel('Debt to Income Ratio')  
plt.ylabel('Number of Businesses')  
plt.title('Visualization 1: Distribution of Debt to Income Ratio')  
  
# Show the plot  
plt.show()

A graph with numbers and a bar

Description automatically generated with medium confidence

Figure 2: Histogram of Average Profit Margins by State

Figure 2 shows the distribution of average profit margins by state. The majority of states have a positive profit margin, but there are notable exceptions. A handful of states have a slightly negative profit margin, but there is an outlier. Viriginia has a profit margin of approximately -25. Further economic analysis may show that investing in Virginia businesses may be unfavorable. The Python code used to generate the histogram is below, with in line comments showing where the data is compiled, called to create the histogram, and labels and titles created:

meanProfitByState = businessDataUpdated.groupby('Business State').agg({'Profit Margin': ['mean']})  
  
plt.hist(meanProfitByState, bins=50, color='skyblue', edgecolor='black')  
  
# Add labels and title  
plt.xlabel('Average Profit Margin')  
plt.ylabel('Number of States')  
plt.title('Visualization 2: Distribution of Profit Margins by State')  
  
# Show the plot  
plt.show()

A graph with blue dots

Description automatically generated

Figure 3: Scatterplot of Business Revenue

Figure 3 displays a scatterplot of business revenue. While most businesses have revenues below $100,000,000, there is some presence in the $100,000,000 to $300,000,000 revenue band. One company represents an outlier with almost a billion dollars in revenue. An investment firm may find this information useful and choose to diversify their investment portfolio to include businesses of all sizes instead of mostly smaller businesses. The Python code used to generate the scatterplot is below:

# Prepare Data  
businessID = businessDataUpdated['Business ID']  
revenue = businessDataUpdated['Total Revenue']

# Create scatter plot  
plt.scatter(businessID, revenue)  
  
# Add title and labels  
plt.title('Visualization 3: Business Revenues')  
plt.xlabel('Businesses')  
plt.ylabel('Revenue')  
  
# Show plot  
plt.show()

A colorful pie chart with different states

Description automatically generated

Figure 4: Pie Chart of Revenue Per State

Figure 4 shows the proportion of total revenue per state. Michigan and Washington D.C. represent the largest portions of revenue. These portions are distinct compared to the GDP of the states, where states like California and Texas are known for their large contribution to national GDP, and as such would be seemingly underrepresented. An investment firm may want to explore diversifying into companies of different sizes for different states. The code to create Figure 4 is displayed below:

# defining labels  
states = list(businessDataUpdated['Business State'].unique())  
revenueByState = businessDataUpdated.groupby('Business State').agg({'Total Revenue': ['sum']})  
revenue = revenueByState['Total Revenue']  
revSum = list(revenue['sum'])  
  
# portion covered by each label  
slices = revSum  
  
# color for each label  
color = []  
n = 50  
  
for i in range(n):  
 color.append('#%06X' % randint(0, 0xFFFFFF))  
  
# plotting the pie chart  
plt.pie(slices, labels = states, colors=color, startangle=90, shadow = True, radius = 1.4, autopct = '%1.1f%%')  
  
# showing the plot  
plt.show()