**Task 1: Program Planning**

Joseph Emmanuel Cayetano

Western Governors University

D598: Analytics Programming Task 1

September 27, 2024

**A. Create a flowchart for a program to perform the required task**

A diagram of a process

Description automatically generated

**B. Write pseudocode for a program to perform the required task**

A screenshot of a computer code

Description automatically generated

**C1. Describe the logic behind the flowchart and pseudocode**

As for the logic behind the flowchart, it consists of tasks that will be required for analyzing the data from the “D598 Data Set.” The data is about the performance of 150 U.S. companies from the recent quarter. The first step will be to import pandas and NumPy libraries, as these libraries are necessary for performing data analysis. The second step is to read the “D598 Data Set” and store it into a data frame. This is the data to be analyzed. The third step is to check whether the data contains duplicate rows. If it does, duplicate rows need to be removed to ensure each row in the data frame is unique. Otherwise, proceed to the next step. The fourth step is to group the data by state, and calculate descriptive statistics such as mean, median, min, and max for each numeric variable. This will help uncover trends across different states. The fifth step is to store the grouped data into a new data frame. The sixth step is to filter the data to identify businesses that have negative “debt to equity” ratios. The seventh step is to store the filtered data into a new data frame. The eight step will be to concatenate the filtered data frame with the original data frame. The final step is to display the combined data frame.

As for the logic behind the pseudocode, it consists of step-by-step statements for analyzing the data from the “D598 Data Set.” Pseudocode provides more detailed steps on how the data analysis will be performed “without conforming to specific syntax rules” (Pi, 2022, par. 2). The first step is to import pandas and NumPy libraries, as both python libraries are necessary for data manipulation. The second step is to create a variable called ‘df,’ and this will serve as a storage for the main data. The third step is to load the data from the ‘D598 Data Set.xlsx,’ and store it inside the ‘df’ variable. The variable ‘df’ will be the original data frame in the data analysis. The fourth step is to check for duplicate rows in the data. If there are duplicates, remove them to prevent redundant analysis. Otherwise, proceed to the next step. The fifth step is to group the data by state and calculate descriptive statistics for all numeric columns in the data. This helps in understanding the business performance for each state. The sixth step is to create a variable called ‘grouped\_df’ to store the grouped data. The seventh step is to filter the data to identify businesses with negative ‘Debt to Equity.” The eight step will be to calculate ‘Debt to Income’ ratio for all businesses in the data. This can be done by dividing long-term debt by revenue. The ninth step is to create a variable called ‘df2’ to store the data that contains the calculated ‘Debt to Income’ ratio. The tenth step is to concatenate the new data frame ‘df2’ with the original data frame ‘df.’ The original data frame will now contain the ‘Debt to Income’ ratio for all businesses. The final step is to print the data from the combined data frame.

**C2. Explain the alignment between flowchart and pseudocode**

Both the flowchart and pseudocode align with each other in performing data analysis using the “D598 Data Set.” As for Importing the Pandas and NumPy libraries, the flowchart uses a processing node to represent the import, while the pseudocode describes the import process in a line of text. As for loading the data, the flowchart uses an input node to represent it because the data will be used to conduct data analysis, while the pseudocode describes the loading process in detail by also mentioning that the data will be stored inside the ‘df’ variable. As for checking duplicate rows, the flowchart uses a decision node to represent “yes or no”, while the pseudocode uses an if statement and an indentation. As for grouping the data by state and calculating descriptive statistics, the flowchart uses a processing node to represent the grouping and calculation operation, while the pseudocode describes the operation in a line of text. As for creating a variable a variable containing the grouped data, the flowchart uses a processing node to represent it, while the pseudocode describes the operation in a line of text. As for filtering the data to identify businesses with negative ‘Debt to Equity’ ratios, the flowchart uses a processing node to represent the process of filtering for negative ratios, while the pseudocode describes the filtering process using text. As for calculating ‘Debt to Income’ ratios, the flowchart uses a processing node to represent the process of calculating the ratios, while the pseudocode describes the calculation process in detail. As for creating a variable containing ‘Debt to Income’ ratios, the flowchart uses a processing node to represent the creation of a new variable, while the pseudocode represents the operation textually. As for concatenating two data frames, the flowchart uses a processing node to represent the concatenation, while the pseudocode details it in plain language. As for display the final data from the combined data frame, the flowchart uses an output node to represent the output of the data, while the pseudocode details it textually. Overall, the flowchart visually represents the data analysis steps, while the pseudocode describes the same steps in text.

**References**

Pi, R. (2022, October 25). *What is pseudocode?*. FutureLearn. https://www.futurelearn.com/info/courses/block-to-text-basedprogramming/0/steps/39492