Gabriela Masak

D598 – Analytics Programming

Task 1: Program Planning

9/17/2024

Program Plan: Financial Data Analysis

**Part 1: Program Flowchart**

A diagram of a company

Description automatically generated

**Part 2: Program Psuedocode**

1. Start
2. Import necessary libraries
3. Import the data file into a data frame from “D598 Data Set.xlsx” with appropriate column data types and index column (business id)
4. Identify and remove duplicate rows:

for row in dataset:

for otherRows in dataset:

if business\_id of row = business\_id of otherRow:

delete otherRow

1. Use groupby function to group businesses by “Business State”, aggregating results with mean, median, max, and min in subsequent columns, assigning results to businessByState
2. Filter updated original dataset to find businesses with negative debt-to-equity ratios:

for row in updatedDataset:

if debt-to-equity ratio < 0:

add business id to negativeRatio list

print negativeRatio list

1. Calculate the debt-to-income ratio for each business:

for row in updatedDataset:

ratio = divide “long-term debt”/”revenue”

add ratio to ratioList list

1. Concatenate list as new column on updated original dataset
2. End

**Part 3: Explanation of Relationship Between Flowchart and Psuedocode**

Like recipes for baking a cake or an instruction manual for building a piece of furniture, having an organized set of guidance is imperative for developers writing a code. This guidance can be either visual or written, as a flowchart or pseudocode respectively. While they convey the same information, they cater to different preferences and learning styles. Flowcharts provide a visual representation of the process, making it easier to understand the overall structure and flow of the code. On the other hand, pseudocode offers a written, step-by-step approach that can be more detailed and precise, helping developers to translate the logic directly into actual code. Both methods are essential tools in a developer’s toolkit, ensuring clarity and efficiency in the coding process.

The flowchart in Figure 1 captures in the flow of data as it in uploaded and read into the program from the “D598 Data Set.xlsx” file, then follows a series of iterations and calculations that result in three different tables for viewing and further analysis. First, the data types and an index column are defined, before the data is read into the program from the Excel file and stored for future manipulation and calculations. The duplicate rows are removed in a series of iterations. These iterations are visualized in the flowchart to the right, where the arrows and shapes create the loops. First the data enters the first loop with the first row, focusing on the business ID, and subsequently compares the business ID to all other rows. The comparison is visualized in the flowchart with a diamond shape, indicating that a decision is being made. Within the inner loop, this decision will remove rows with a matching business ID or move on to the next row if the IDs do not show a duplicate row. The remaining rows are stored as an updated dataset. This dataset then takes three paths. It is grouped by state, with mean, median, minimum, and maximum values calculated for related quantitative fields, whose results are stored and displayed. This is visualized by the rectangles and arrows to the left of the “call updated dataset” rectangle. The arrows and rectangles to the right of this rectangle show the steps taken to iteratively identify the businesses with a negative debt-to-equity ratio. The third path below the rectangle represent the series of steps taken to calculate the debt-to-income ratio of the businesses and then concatenate the results with a new column on the original updated dataset, whose output is visualized by a rhombus in the flowchart, ultimately leading to the “stop” circle where the program finalizes.

The pseudocode in Part 2 of this document offers a blueprint for program development, representing the various algorithms that manipulate the data and define the program. Independent of the syntax rules that define programming languages like Python or R, the pseudocode is readable and understood by humans, with or without large technical or programming skills, much like the flowchart. Following the guidelines of PEP8, as the program will utilize Python code, the pseudocode lays out the logic and framework of the program. First, necessary libraries are imported. Pandas is a popular Python library used to analyze data such as descriptive statistics mean, median, minimum, and maximum. Various lists and variable are instantiated. The Excel file is loaded into the program, wherein it undergoes two nested loops. The outer loop tracks the initial row for comparison, with the inner row selecting the row it will be compared with, deleting rows that contain matching business ID numbers before continuing. The loops are exited when all rows have been compared. With all duplicated removed, the updated dataset is called, along with a grouping function from the Pandas library, segregating the businesses by state, and calculating the mean, median, minimum, and maximum of the long-term debt, equity, debt-to-equity ratios, liabilities, revenue, and profit margin by state, and ultimately displaying and storing the data before continuing. The previously updated dataset is then filtered to identify businesses with negative debt-to-equity ratios, a list of which is then displayed. Finally, the program again calls the previously updated dataset to divide the long-term debt by the revenue for each row, assigning these values to a list, which is then concatenated as a new column to the original updated dataset. The resulting dataset is output and the program concludes.

While different in form, pseudocode and flowcharts pose a similar solution to the same issue and both serve as organizational tools that aid in program development, understanding, and communication. The biggest difference lies in the ability of flowcharts to branch out in multiple directions and areas, offering an omnidirectional view, whereas pseudocode is unidirectional. Pseudocode is written, flowing from top-to-bottom focusing on algorithms and flow of data, flowcharts is a visual aid, focusing on the flow of data and processes as well as highlighting decisions. Together, they ensure that the developer has a clear roadmap to follow, enhancing both the development process and the communication of ideas among team members.