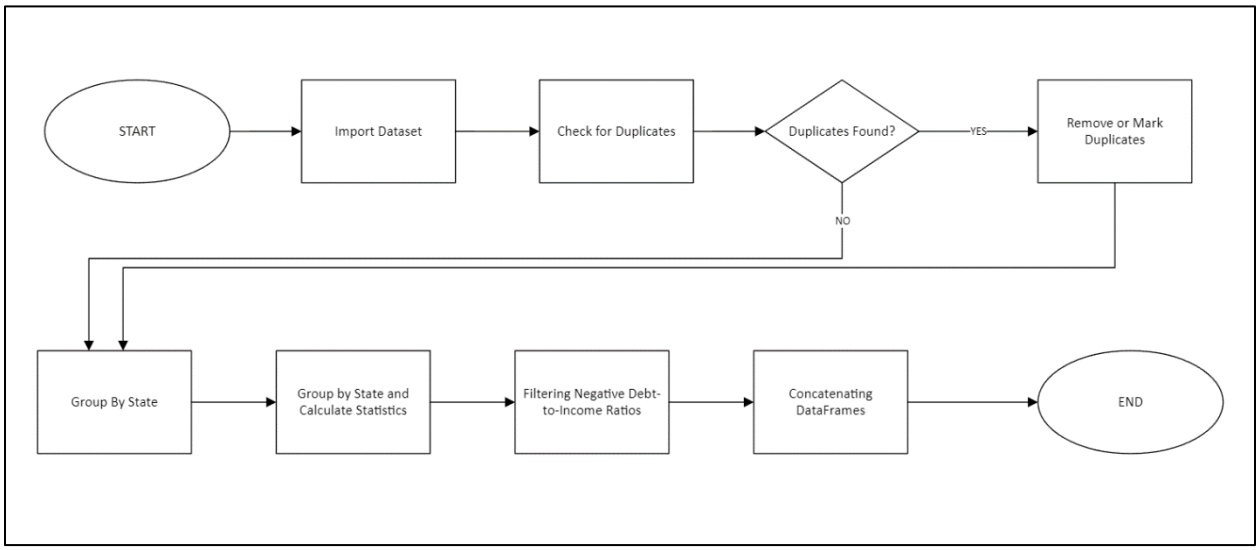
Robert Pavlik
D598 Analytics Programming
Task 1

A. Flowchart



B. Pseudocode

```
//START
START

//STEP 1 : Import the Dataset
IMPORT dataset AS ds

//STEP 2: Check for Duplicates
IF ds contains duplicate rows THEN

    //STEP 3: Remove or Mark Duplicates
    REMOVE or MARK duplicate rows in ds

END IF

//STEP 4: Group Data by State
grouped_ds = GROUP ds BY 'State'

//STEP 5: Calculate Descriptive Statistics by State
Stats_ds = Calculate mean, median, min, max FOR all numeric columns in
grouped_ds

//STEP 6: Filtering Negative Debt-to-Equity Ratios
Negative_debt_equity_ds = FILTER ds WHERE 'Debt to Equity' < 0

//STEP 7: Calculate Debt-to-Income Ratio for Each Business
debt_income_ds = CALCULATE 'Debt-to-Income' as 'Long-term Debt' / 'Total
Revenue' FOR all businesses in ds

//STEP 8: Concatenate DataFrames
final_ds = CONCATENATE ds WITH debt_income_ds

//END
END
```

C.

1. The flowchart and the pseudocode follow a sequential and structured approach to perform data analysis tasks, ensuring that all key operations are completed efficiently. The steps can be broken down as follows:

    - Data Input and Validation: The program begins by importing the dataset into a data frame. Both the flowchart and pseudocode indicate that checking for duplicates is an essential step before proceeding with data analysis, as duplicates can skew results. The decision-making process to either remove or mark duplicates reflects an important data-cleaning step.

    - Data Grouping and Aggregation: After ensuring data integrity, the program groups the data by state, allowing for specific statistical analyses. This grouping is essential for calculating state-level summaries and descriptive statics, which help in analyzing the performance of businesses in different regions.

    - Filtering: The program filters out companies with negative debt-to-equity ratios, as these companies are considered to have unfavorable financial standing, based on the fund's rebalancing strategy. This step isolates businesses that may require closer scrutiny or exclusion from further investment

    - New Metric Calculation: The pseudocode then proceeds to calculate the debt-to-income ratio for each business, a crucial metric in assessing a company's financial health. By calculating this metric, the program provides an additional layer of analysis, allowing the fund managers to understand businesses' debt burdens relative to their revenues.

    - Data Consolidation: Lastly, the pseudocode concatenates the original data with the newly calculated debt-to-income ratio, ensuring that the final dataset contains all the necessary information for decision-making

2. The flowchart and pseudocode are closely aligned, as they represent the same process in different formats, both outlining the logical steps to be performed in sequence. Here's how they match up:

    - Start to Import Dataset: Both the flowchart and the pseudocode begin with importing the dataset into a data frame. This is the first necessary action to work with the data.

    - Checking for Duplicates: After importing, both the flowchart and pseudocode include a step to check for duplicates in the dataset. The decision diamond in the flowchart corresponds to the conditional logic (IF Statement) in the pseudocode, where duplicates are either removed or marked if found.

    - Data Grouping and Statistics Calculation: In the flowchart, the step labeled "Group by State" and "Group by State and Calculate Statics" mirrors the GROUP BY operation in the pseudocode, which creates a new data frame containing descriptive statistics for all numeric variables by state.

    - Filtering Negative Debt to Equity Ratios: The next step in the flowchart shows the filtering of negative debt-to-equity ratios, which corresponds to the FILTER statement in the pseudocode that identifies businesses with a negative ratio.

    - Debt-to-Income Calculation: In both the flowchart and the pseudocode, calculating the debt-to-income ratio is a separate step. In the pseudocode, this step involves using the formula Long-term Debt / Total Revenue for each business, which aligns with the flowchart's representation.

    - Concatenating Data Frames: Finally, both the flowchart and pseudocode include the steps of concatenating the new debt-to-income ratio data with the original dataset. This ensures that all relevant information is available in the final data frame, preparing it for the final analysis.

The Flowchart provides a high-level overview of the logical steps involved, while the pseudocode breaks these steps down into detailed, executable instructions. Both tools follow the same sequence, ensuring consistency in the approach to solving the data analysis problem.