PDS Mini project walkthrough

Here's a high-level representation of the approach followed by the student in 18 steps:

1. Load Data
   - Read the 'bank.csv' file.
   - Clean the data by removing quotes.
   - Load the cleaned data into a DataFrame using pandas with ';' as the separator.

2. Inspect Data Types
   - Check the data types of all columns in the DataFrame.

3. Check for Wrong Data Types
   - Verify if any data types are incorrectly identified and convert them if necessary (not needed in this case).

4. Missing Values Check
   - Calculate the sum of missing values (NaN) for each column.

5. Duplicate Records Check
   - Calculate the sum of duplicate rows in the DataFrame.

6. Average Balance Calculation
   - Calculate the average balance for customers who are subscribed ('yes') and not subscribed ('no').
   - Visualize the average balances using a bar plot.

7. Education Distribution Visualization
   - Create a pie plot to show the distribution of education levels.

8. Feature Engineering - Season Creation
   - Define a function to map months to seasons.
   - Apply the function to the 'month' column to create a new 'season' column.

9. Class Distribution Visualization
   - Create count plots for the distribution of the 'season' and 'y' (subscription status) variables.

10. 'Pdays' Feature Analysis
    - Create a bar plot to analyze the effect of 'pdays' on subscription status.

11. Replace Specific 'Pdays' Values
    - Replace '-1' values in 'pdays' with NaN.

12. Reanalyze 'Pdays' Feature
    - Create a bar plot to analyze the effect of 'pdays' on subscription status after replacement.

13. Negative Balance Subscription Check
    - Create a count plot to check if customers with a negative balance have subscribed.

14. Maximum Balance by Job Type
    - Use a pivot table to find the maximum balance for each job type.

15. Age and Balance Relationship Analysis
    - Create a scatter plot to analyze the relationship between age, balance, and subscription status.

16. Balance Distribution Visualization
    - Create violin and box plots to show the distribution of balance for each subscription class.

17. Proportion of Defaulters Visualization
    - Create a pie plot to show the proportion of defaulters and non-defaulters.

18. Pdays' Distribution Visualization
    - Create box and strip plots to visualize the distribution of 'pdays' with respect to subscription status.

Note: The pseudocode above is a simplified representation of the steps taken by the student and does not include specific details such as function parameters or data manipulation techniques. It is intended to provide an overview of the student's analytical approach.

****