Lab #7

Instructions

With lab #7, your group will write a python program to demonstrate different methods for handling missing value. In this task you will modify the missing values and report what changes in the distribution after the modifications are made.

Submission Requirements

- Python .py file with no markdown language. Only Python Code and comments.
- Group Assignment: Only 1 assignment needs to be submitted for each group.
- You should seek to reduce redundancy in your code.
- Your code should exercise good judgment for displaying output to the user.

Task

- File: http://drd.ba.ttu.edu/isqs3358/Labs/Lab7/team_missing_data.csv
- Handle missing values for the following columns:
  - points_scored
  - points_allowed
  - opposing_team
- Report the following column metrics after handling the missing values:
  - Mean
  - Median
  - Variance
- Use each of the following methods to handle the missing values:
  - Round 1:
    - Drop all NA Records
  - Round 2:
    - Numerics fill with 0
    - Textual with “-default-”
  - Round 3:
    - Numerics fill with Column average
    - Textual with “-default-”
  - Round 4
    - Numerics fill with team_id based average
    - Textual with distribution sampling

Notes and Hints

- You will need to refresh the dataframe after each round.

Example

```
-------------- Task #1 DropNA Method ---------------
Means
points_scored     60.368421
points_allowed    59.578947
dtype: float64
Median
points_scored     59.0
points_allowed    57.0
dtype: float64
Variance
points_scored     143.578947
points_allowed    147.812865
dtype: float64
mean by opposing team
                          points_allowed  points_scored
opposing_team
Blazing Suckerfish              62.000000      59.000000
Garlic Stakers                  65.600000      66.600000
Monsters                        45.000000      61.500000
The Mystfallen                  70.333333      61.666667
The Puddleglums                 51.200000      51.200000
Windrunner International        61.666667      63.666667
-------------- Task #2 FillNA with 0 Method ---------------
Means
points_scored     55.066667
points_allowed    57.500000
dtype: float64
Median
points_scored     58.5
points_allowed    61.5
dtype: float64
Variance
points_scored     513.995402
points_allowed    533.155172
dtype: float64
mean by opposing team
                          points_allowed  points_scored
opposing_team
-default-                       60.666667      70.333333
Blazing Suckerfish              62.000000      59.000000
Garlic Stakers                  54.666667      62.833333
Monsters                        30.000000      54.000000
The Mystfallen                  72.000000      46.250000
The Puddleglums                 55.833333      42.666667
Windrunner International        64.500000      47.750000
```

Figure 1: Displays outputs for Round #1 & Round #2

```
Means
points_scored     61.185185
points_allowed    63.888889
dtype: float64
Median
points_scored     61.092593
points_allowed    62.500000
dtype: float64
Variance
points_scored     165.450830
points_allowed    153.126437
dtype: float64
mean by opposing team
                          points_allowed  points_scored
opposing_team
-default-                       71.314815      70.333333
Blazing Suckerfish              62.000000      59.000000
Garlic Stakers                  65.314815      62.833333
Monsters                        51.296296      54.000000
The Mystfallen                  72.000000      61.546296
The Puddleglums                 55.833333      52.864198
Windrunner International        64.500000      63.046296
-------------- Task #4 Fill with Team based Column Averages Method ---------------
Means
points_scored     61.185185
points_allowed    63.888889
dtype: float64
Median
points_scored     61.0
points_allowed    62.0
dtype: float64
Variance
points_scored     184.541311
points_allowed    170.794872
dtype: float64
mean by opposing team
                          points_allowed  points_scored
opposing_team
Blazing Suckerfish              62.000000      59.000000
Garlic Stakers                  65.600000      62.833333
Monsters                        45.000000      54.000000
The Mystfallen                  72.000000      61.666667
The Puddleglums                 55.833333      51.200000
Windrunner International        64.500000      63.666667
/tmp/ipykernel_16755/2760147846.py:58: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df_teams_merge['points_scored'][df_teams_merge['points_scored'].isna()] = df_teams_merge['avg_points_scored']
/tmp/ipykernel_16755/2760147846.py:59: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df_teams_merge['points_allowed'][df_teams_merge['points_allowed'].isna()] = df_teams_merge['avg_points_allowed']
```

Figure 2: Displays outputs for Round #3 & Round #4