Lab #6

Instructions

With lab #6, your group will write a python program that allows you to merge 2 different dataframes and manipulate the data to fix missing values. Pay special attention to the comment about correlation, it is a hint to where you can find your data to fill in missing values.

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 outputs should match the supplied examples for column order.

Task

- Files:
  - hr_data.csv: http://drd.ba.ttu.edu/isqs3358/Labs/Lab6/hr_data.csv
  - sales_data.csv: http://drd.ba.ttu.edu/isqs3358/Labs/Lab6/sales_data.csv
- Please join these datasets into a single dataset.
- Using the information that columns with missing data are highly correlated with the position (in the full dataset this is especially true), please address the missing values in the dataset.
  - Print to console what your method and reasoning is for this. Please do the best possible job in fixing these values.
- Create a new column called total benefits:
  - total_benefits = Salary + Benefits
- Create a new column identifying if the employee's total_benefits is high or lower than the average total_benefits for all salaries (assume ties as "high")
  - name column: employee_comp
- Add a new column identifying if the employee's total_benefits is higher or lower than the average total_benefits within the employee's title (assume ties as "high")
  - name column: position_comp
- Create the following reports:
  - all_data.csv – contains….yes you guessed it, all data.
  - employee_high_benefit.csv – contains employee info for employees that make higher than the average total_benefit.
  - positiondata_high.csv – contains employees that make higher than the position average total_benefit

Notes and Hints

- Again, notice the correlation. A group based aggregation will be helpful here.
- Check the example files for delimiter and column order.

Example

- Can be found in lab_6_example.zip found in this lab on blackboard.