Homework #2

Task

In this lab, you will use Pandas to perform the necessary task on the following datasets. As a reminder, this is an individual problem and should be done solely by yourself.

Submission requirements

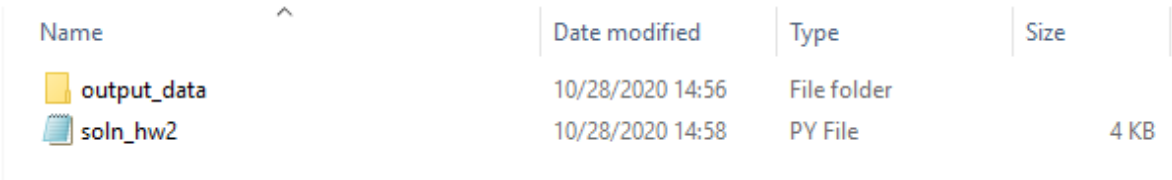
- Python file (.py), you can name this lastname_hw2.py.
- You are required to structure your solution directory as we have done in class.
    - Source data should be pulled direct from the provided web URL.
        - http://drd.ba.ttu.edu/isqs3358/hw2/hr_data.csv
        - http://drd.ba.ttu.edu/isqs3358/hw2/sales_data.csv
        - http://drd.ba.ttu.edu/isqs3358/hw2/item_data.csv
    - Your output should go to the output specified on below in this document:
        - Either output directory or console if using Pandas
    - ***Your code must have url and settings variables at the top of the file. There should be no hard-coded values throughout the homework.***

Source Files:

You have been provided 3 source files.

- http://drd.ba.ttu.edu/isqs3358/hw2/hr_data.csv
- http://drd.ba.ttu.edu/isqs3358/hw2/sales_data.csv
- http://drd.ba.ttu.edu/isqs3358/hw2/item_data.csv

File Structure:



Task:

The data contained in these files represents sales data, HR data and item data related to an employee. There are several fields of data missing, and management would like you to run multiple reports on the data. Please perform the following task. ***NOTE, FOR ALL CSV OUTPUTS, COLUMN ORDER AND DESIGN SHOULD MATCH GIVEN EXAMPLES!***

- 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.
- Management has asked that you produce the following computational columns:
    - Per_Item_Benefit = Benefits / ItemSold
    - Total_Compensation = Salary + Benefits
    - Performance_metrics = (Total_Compensation / ItemSold) /Avg_ItemSold
    - employee_raise_elligible: Defined as
        - "Yes" if
            - Performance_metrics > 238 and Title = Sales Associate 1
            - Performance_metrics > 704 and title = Sales Associate 2
            - Performance_metrics > 938 and title = Sales Associate 3
            - Performance_metrics > 2146 and title = Sales Manager
        - Else "No"
- Compute the average Total_Compensation, Per_Item_Sale, Performance_metrics by Title.
    - Output to file called "title_aggregate.csv"
- Create the following outputs.
    - Employees that will get a raise called "employee_raise.csv"
- Due to the high inflation, the management wants to increase the benefit a little bit. As such compute the following and generate a report named "title_benefit_raises.csv"
    - Create a column to impose the following structure (updated_benefit):
        - Sales Associate 1: 7.5% raise
        - Sales Associate 2: 7% raise
        - Sales Associate 3: 6.5% raise
        - Sales Manager: 6% raise
    - Now create a column (benefit_diff) to show the benefit increase amount for each level.
    - Compute the total_benefit, updated_benefit, benefit_diff by job title for employees getting a raise.
        - ***NOTE, USE SUM, NOT MEAN***
    - Generate the output CSV.