AIT 580 – Assignment 10 – Python with Pandas

Due Sunday October 27, 2019

Submission to BOTH your **Github** and **Blackboard**

- **Getting Ready**:
  - Target Data: Employee Attrition Dataset (`EmployeeAttrition.csv`)
  - Clone/pull this repository (if you haven't done so): https://github.com/myeong/AIT580
  - Copy the "`Python_assignment_10`" folder to your own Git folder that you made for Assignment 3.
  - Acquire data from the `data` folder.
- **Tool**
  - You need to use Jupyter Notebook, Rodeo, or Pycharm for this assignment. You can use either your own IDE on your PC *or* the one inside Virtual Box VM.
- **Procedure**
  - Choose an IDE
    - (If you're using the tool in VM) You can run
      - Rhodeo: First turn off 3D Acceleration option under Display Setting of Virtual Box, otherwise you will see black screen when you try to open Rhodeo. Then double click on the Rhodeo icon at the Desktop.
      - Jupyter Notebook: Open web browser Chrome (circle icon) on the left side launch bar. This will open Chrome browser screen. Type localhost:8888 in the URL and press enter. This will open JupterLab tool. Click on Python 2 and new window will be open for writing python code.
      - PyCharm: Click on the PC icon on left side launch bar to open PyCharm IDE.
    - If you are using your own PC (not VM), you can install Jupyter Notebook through installing Anaconda (https://www.anaconda.com/distribution/).
  - Create a Python file for your assignment
    - [If you're using Rhodeo or PyCharm] Copy and paste `Assignment-10-Python-Pandas.py` and rename the copied file to `YourLastName_GNumber.py`
    - [If you're using Jupyter Notebook] Copy and paste `Assignment-10-Python-Pandas.ipynb` and rename the copied file to `YourLastName_GNumber.ipynb` (Note that Jupyter Notebook uses a different file type than other Python IDEs).
  - Open this file in your Python IDE (i.e., either Rhodeo, Jupyter, or PyCharm).
    - If you're using Rhodeo or PyCharm, you can open the file directly.
    - If you're using Jupyter Notebook, you need to go to the command line, go to the assignment folder, and type "`jupyter notebook`" and hit return. Then, you can see a list of files in that folder. Click on the "`ipynb`" file that you just

created.
- o Load `EmployeeAttrition.csv` file using the Python script that is already written in the assignment file.
  - ▪ Make sure to add the complete path of `EmployeeAttrition.csv` file inside source file. For example, if you store `EmployeeAttrition.csv` in "data" folder then your complete path should be `[your_git_path]/data/EmployeeAttrition.csv`
- o Run the source code
  - ▪ If you're using Rhodeo, you can click the "`Run Script`" button.
  - ▪ If you're using Jupyter Notebook, scripts are executed block by block. You can put your cursor in the first block, and click the "`Run`" button. Or, you can run the focused block by hitting "Shift + Enter" keys (shortcut).
- o If you want to compile and run single line/statement in your code, press `Run` button inside RStudio window on top.
- o If everything goes well, you will see output in black color in the console below your code. If your code has any error, it will be written in another color.
- o Write your remaining code in the same source file to answer all the questions below.

Practice Scripts:
Please see the practice Python scripts for your reference. This will help you answer the assignment questions. These files are in the same folder. These are NOT part of your assignment but just for your reference. Data needed for these scripts are all available under the "`data`" folder.
- • ecommerce_pandas_practice.py
- • imdb_pandas_practice.py
- • olympics_pandas_practice.py

Submission Guideline: Submission for this assignment is two ways.
1. After writing scripts on your Python file (either `.py` or `.ipynb`), commit and push it to your own Git repository that you made for Assignment 3 (you shouldn't use other repos. If the instructor or TA cannot see your repository, it is regarded as no-submission). Make sure to use "`print`" or "`head`" statement for answering each question, so TA can run your code directly without modifying the code. If class TA does not see printed output when running the Python code, it will be regarded as "question not answered," which will lead to the deduction in points.
2. After pushing your assignment to your Github repository, submit the screenshot of your Github repository page that shows your uploaded files through Blackboard.
3. If needed, report the answer to each question as a comment (using # sign).

Assignment Questions:

Write Python code using Pandas to answer these questions:
1. Find the number of entries/rows and columns in the data.
2. What is the average Monthly Income
3. What is the highest amount of HourlyRate
4. What is the Department, JobRole, MaritalStatus, and OverTime of EmployeeNumber "10"?

5. What is the Employee ID of highest MonthlyIncome paid employee?
6. What is the Average (mean) DailyRate for all Employees Group By Age whose age is greater than 58. (hint: use groupby function)
7. How many unique EducationField are there?
8. What are the top 5 most common JobRole?
9. How many JobRoles represented by less than 100 employees?
10. What is the correlation between Education and JobSatisfaction?