Adrian Alvarez

9/9/2018

*Objective:*

Write an original ETL code that combines files and identifies cases that have a low blood pressure (for their age) for longer than 15 minutes.

*Attached Files:*

Original code file in Python – ETL\_Alvarez.py

Final report in csv file - FinalReport.csv

*Notes:*

I created multiple functions to allow for a more scalable program. For example, if we were provided with a “high” systolic blood pressure limit then the BPFromAge function can be modified to include that. Ideally the functions in this code will be separated into a structure below with each script being its own file, this allows scripts to be run on multiple projects. Code was also extensively commented on.

* Main directory
  + Cleaning Scripts
    - Sorting
    - Check for null
    - Return database
  + Inspecting Scripts
    - Shows top 5 rows
    - Shows bottom 5 rows
    - Gives extra info
      * Columns, types, memory usage
    - Gives simple state
      * Mean, Avg, Std…
  + Application Scripts
    - Vitals thresholds
    - Statistical algorithms
  + Project Specific Scripts
    - Loads databases
    - Calls from other scripts

*Plots:*

Helps me see what I’m working with.

|  |  |
| --- | --- |
| Person: 123  Date: 1/5/2016 | Person: 123  Date: 2/13/2016 |
| Person: 234  Date: 1/6/2016 | Person: 234  Date: 2/17/2016 |
| Person: 234  Date: 9/8/2015 | Person: 345  Date: 5/23/2015 |
| Person: 345  Date: 8/11/2015 | Person: 456  Date: 10/16/2017 |
| Person: 456  Date: 5/21/2015 | Person: 456  Date: 6/8/2015 |
| Person: 456  Date: 8/15/2015 | Person: 456  Date: 7/7/2017 |