Python ETAV

(DUE: 5 PM EST on 4/21/2023)

Create a python notebook to extract, transform, analyze, and visualize a dataset of your choosing. Use the sections below as the structure of your notebook. The more documentation, the better!

1. EXTRACT: Choose a dataset and load it into a notebook.
    a. Your source can be from anywhere, a few suggestions including the seaborn datasets (choosing a couple of these might be best, since they are pretty small), datasets from Kaggle, or (**extension**!) datasets from websites that you've scraped.
    b. If your extraction process is more extensive than simply loading a csv into a dataframe, be sure to take note of what you did.

2. TRANSFORM: Clean your dataset.
    a. Import your chosen dataset into a Jupyter Notebook.
    b. Clear out or convert any missing data.
    c. Replace or remove any incorrect data.
    d. Separate data into any number of appropriate data frames (groupbys, logical separations, categorical separations, ect).
    e. **Extension**: Define your own functions to use in .applymap() and map over the entire dataset! Example - RealPython Data Cleaning Article

3. ANALYZE: Examine your dataset to understand its structure and any insights that may be hidden within it.
    a. Find the maximum, minimum, and average of all numerical data.
    b. Group the data categorically and extract aggregate maximum, minimum and average values.
    c. Add calculated columns to your dataframes to demonstrate important relationships between other columns.
    d. Merge any related data onto your main dataframes.
    e. Display any comparative data in a display dataframe.
    f. Interpret the Data - Draw conclusions from observations.
    g. **Extension** - Use simple machine learning to make predictions for a particular attribute - (warning, HARD)!
        i. Linear regression - LinReg with Tensorflow
        ii. Logistic regression - LogReg with SkLearn

4. VISUALIZE: Create a few charts to easily visualize your drawn conclusions.
    a. Plot the data using Pandas built in plotting tool, or use MatPlotLib or Seaborn.
    b. Use at least 2 different kinds of plots, preferably more.
    c. **Extension** - Create an interactive dashboard with Widgets, Voila, or a similar python package for dashboards.