PROJECT MILESTONE 2

PACE: Plan Stage

- How can I best prepare to understand and organize the provided information?

  I will begin by exploring the structure of the provided dataset and consider reviewing the provided Data Dictionary.

- What follow-along and self-review codebooks will help me perform this work?

  I will use the follow-along notebooks from Milestone 2 to help me perform this work.

- What are some additional activities a resourceful learner would perform before starting to code?

  I would recall the Python documentation to repeat some crucial aspects briefly.

PACE: Analyze Stage

- Will the available information be sufficient to achieve the goal based on my intuition and the analysis of the variables?

  Before starting the analysis of the given dataset, I thought it would be enough. But, as I "dive" into analysis, I discovered some insights for additional information from the stakeholders.

- How would I build summary dataframe statistics and assess the min and max range of the data?

  By using the powerful pandas and NumPy libraries.

- Do the averages of any of the data variables look unusual? Can I describe the interval data?

  The median values (not the mean values because of the outliers produced by mean calculations) of all the count variables all far smaller than their max values and standard deviations.
  The interval data for every count variable is extremely large.

PACE: Construct Stage

The Construct stage does not apply to this workflow. The PACE framework can be adapted to fit the specific requirements of any project.

PACE: Execute Stage

- Given my current knowledge of the data, what would I initially recommend to my client to investigate further prior to performing EDA analysis?

  I would recommend to check the extremely big anomalies (outliers) in values of every count variable in the dataset.

- What data initially presents as containing anomalies?

  All count variables contain anomalies.

- What additional types of data could strengthen this dataset?

  Maybe some additional information about the level of engagement could strengthen the dataset.