Milestone 1: Data Exploration, Preprocessing and Visualization

- Use quantitative measure and visualization to verify the data's reasonable condition. For example, how should the missing values be handled? Are the values in a reasonable range, given what you would expect for each variable? Are there obvious outliers? Are variables correlated with each other?
- Ensure consistency in the definitions of fields, units of measurement, time periods, and so on. In this step, new variables are also typically created from existing ones. For example, "duration" can be computed from start and end dates.
- Reduce the data dimension, if necessary: Dimension reduction can involve operations such as eliminating unneeded variables, transforming variables (e.g., turning "money spent" into "spent > $100" vs. "spent ⩽ $100"), and creating new variables. Make sure that you know what each variable means and whether it is sensible to include it in the model.
- Avoid making arbitrary decisions during preprocessing. Also, the objective of each visualization should align with the project objective. Avoid making meaningless plots.
- Submit a milestone 1 report detailing the aforementioned objectives
- Submit the data files and the python notebook