# Data Management Practices:

1. Data vs. Code
   1. Data stores the data. Do not mess with it, as version control is not available for this.
   2. Data must be present in OneDrive only, and cannot be copied to local/GitHub repository.
   3. Code is just the code, without any data.
   4. Code can exist on GitHub, but try not to keep it on local machine. Version Control will be made available for the Code.
2. Data Folder Name Scheme
   1. Do not edit files inside “01\_raw” folder.
      1. This is where the raw data is stored. If you mess with this folder, your code may/maynot work in other member’s PC.
   2. Feel free to edit files in “02\_processed” folders, but make sure you ask permission of the file owner.
      1. This is where you edit the data, clean it, filter it, publish EDA results as csv, etc.
      2. Always save files with your name to indicate ownership, and prohibit others from making uninformed changes.
      3. You can also save the file with name of the process/analytics/purpose, if multiple people are working on it.
   3. Do not edit files in “03\_final” folder, without discussing and getting approval from the entire team.
      1. This is where the final merged data is stored for analysis.
      2. This data will be cleaned as per the requirements of the analytics to be performed.
   4. Each analytics may require different kinds of data, and the file name should indicate, what analytics will be performed on it.