**Data Analysis Methodology**

1. Find the **dimensions** of the dataset
2. Look for **Column names inconsistency** (use df.columns). If yes, fix them and make all the names consistent.
3. Data Cleaning
   1. Find if there are any **wrong data types** used in the columns. If yes fix them.
   2. Look for the jargon data like the ‘?’, ‘.’ Etc. Try to fill them with .nan.
   3. Find **missing data**
   4. Fix missing data
4. If it is a classification problem, find if it’s a **balanced or imbalanced dataset** (use df[‘column\_name’].value\_counts())
5. Exploratory Data Analysis (**EDA)**
   1. Plot Scatter plot