**1. IDENTIFY THE PROBLEM**

* What is the business or product objective?
* What are the goals and criteria for success?
* What would the ideal dataset have?

**2. ACQUIRE THE DATA**

* What is the right/ideal dataset?
* Time sensitivity
* Possible supplementary data
* How is the data hosted: Local | Remote
* What are the most appropriate tools to work with data?
* Preprocess: Excel | Python | R
* Analysis: Python, R
* Database: Plaintext (CSV) | SQL | NoSQL
* Visualization: MatPlotlib | R | Tableau | Gephi

**3. PARSE THE DATA**

* Is there documentation for the data?
* What are your observations from Exploratory Data Analysis
* What is the Data Quality?
* Missingness
* Sparsity
* Errors / Impossible Values
* Inconsistent Coding

**4. MINE THE DATA**

* What is the sampling methodology? (Random | Representative of population)
* What needs to formatted, cleaned, sliced and combined?
* What are the necessary derived/computed columns for the new data?
  + Averages
  + Deviations / Absolute Differences

**5. REFINE THE DATA**

* Are there any trends or outliers?
* What are the descriptive statistics for the key variables?
  + Central Tendency
  + Variability
* Do you need to transform the data?
  + Make into Normal Distribution
  + Scale to a common mean, min, max

**6. BUILD A DATA MODEL**

* What is the appropriate model for the data?
  + Supervised vs. Unsupervised
  + Classification vs. Regression
* How is the initial performance of the model?
* How can you refine the model based on the initial performance?

**7. PRESENT THE RESULTS**

* How would you summarize the findings in a narrative/story?
* What are the limitations and assumptions of your analysis?
* What are follow-up problems and questions for future analysis?