**Business Understanding**

* **What is the goal & objectives?**
* **Seeking clarification - What is the goal?**
  + Establishing a clearly defined question starts with understanding the GOAL of the person who is asking the question.
  + Once the goal is clarified, the next piece of the puzzle is to figure out the objectives that are in support of the goal.
* **Supporting the goal. figuring out the supported goal objectives.**
  + By breaking down the objectives, structured discussions can take place where priorities can be identified in a way that can lead to organizing and planning on how to tackle the problem.
* **Getting stakeholders "buy-in" and support...**
  + Depending on the problem, different stakeholders will need to be engaged in the discussion to help determine requirements and clarify questions.
  + Applying the concepts
  + Pilot project kickoff
* **What is the sponsor's involvement? it is critical**
  + Set overall direction
  + Remained engaged and provided guidance.
  + Ensured necessary support, where needed.
* **Identify the business requirements.**
  + The correct approach depends on the business requirements for the model.

**Analytics Approach**

* **Which analytic approach to pick?**
  + Selecting the right analytic approach depends on the question being asked.
  + The approach involves seeking clarification from the person who is asking the question, so as to be able to pick the most appropriate path or approach.
  + Once the problem to be addressed is defined, the appropriate analytic approach for the problem is selected in the context of the business requirements.
  + Once a strong understanding of the question is established, the analytic approach can be selected. This means identifying what type of patterns will be needed to address the question most effectively.
* **Pick analytic approach based on type of question.**
  + **Descriptive**
    - Current status
    - If the question is to show relationships, a descriptive approach maybe be required.
    - This would be one that would look at clusters of similar activities based on events and preferences.
  + **Diagnostic (Statistical Analytics)**
    - What happened?
    - Why is this happening?
    - Statistical analysis applies to problems that require counts.
    - if the question requires a yes/ no answer, then a classification approach to predicting a response would be suitable.
  + **Predictive (Forecasting)**
    - What if these trends continue?
    - What will happen next?
    - If the question is to determine probabilities of an action, then a predictive model might be used.
    - Decision-tree classification model
      * Categorical or classified outcome
      * Explicit decision path
    - To predict an outcome
  + **Prescriptive**
    - How do we solve it?
    - …
  + In the case where the question is to learn about human behavior, then an appropriate response would be to use Clustering Association approaches.
  + …
  + A decision tree classification model was used to identify the combination of conditions leading to each patient's outcome.
  + In this approach, examining the variables in each of the nodes along each path to a leaf, led to a respective threshold value. This means the decision tree classifier provides both the predicted outcome, as well as the likelihood of that outcome, based on the proportion at the dominant outcome, yes or no, in each group.
  + A decision tree classification model is easy for non-data scientists to understand and apply, to score new patients for their risk of readmission.
  + Categorical outcome
  + Explicit “decision path” showing conditions leading to high risk
  + Likelihood of classified outcome
  + Easy to understand and apply
* **Will Machine learning be utilized?**
  + Machine Learning is a field of study that gives computers the ability to learn without being explicitly programmed.
  + Machine Learning can be used to identify relationships and trends in data that might otherwise not be accessible or identified.
  + In the case where the question is to learn about human behavior, then an appropriate response would be to use Clustering Association approaches.

**Data Requirements**

* **Data Requirements - Cooking with Data.**
  + If you don’t have the right ingredients, then your success will be compromised. Each step is critical in making the goal.
    - Which ingredients are required,
    - How to source or the collect them,
    - How to understand or work with them,
    - How to prepare the data to meet the desired outcome.
  + Building on the understanding of the problem at hand, and then using the analytical approach selected, the Data Scientist is ready to get started.
  + It's vital to define the data requirements for decision-tree classification.
* **Selecting the Cohort**
* **Defining the Data**
  + This includes identifying the necessary
    - data content,
    - Formats
    - and sources for initial data collection.
    - Representation suitable for decision tree classifier
  + Then the content, format, and representations of the data needed for decision tree classification were defined.
    - One record per patient with columns representing variables (depending variables and predictors)
    - Content covering all aspects of each patient’s clinical history
      * Transactional format
      * Transformation required

**Data Collection**

* **Data Requirements - Cooking with Data.**