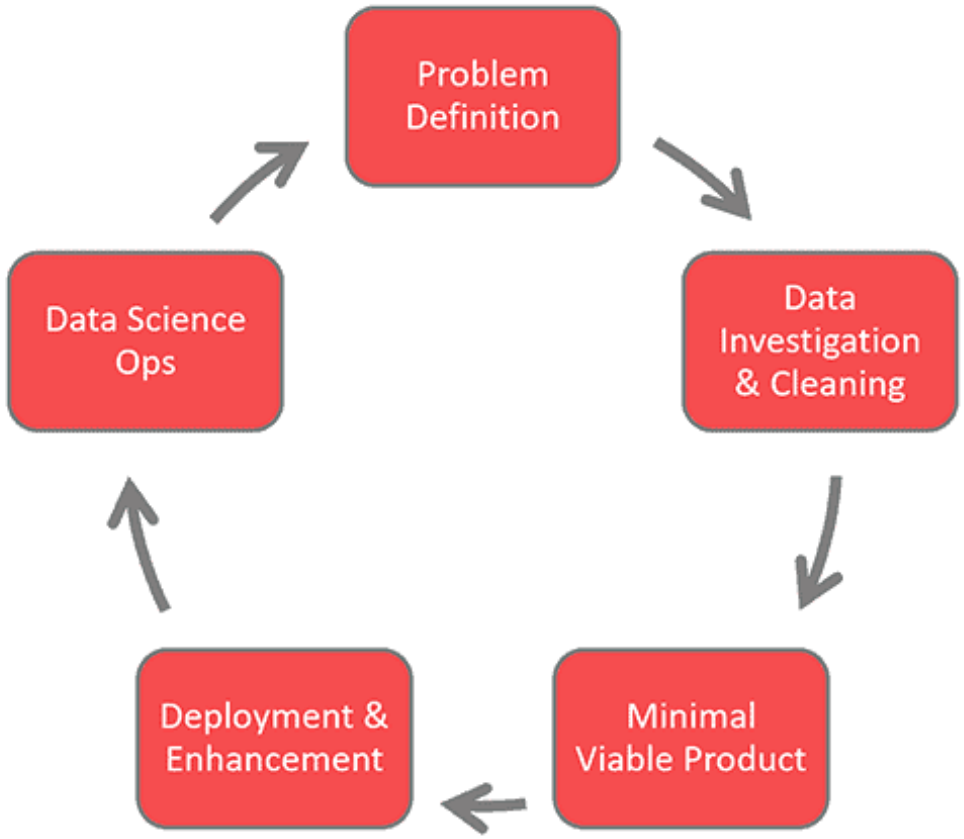
Data Science Life Cycle

By : L Vinay Rajiv Reddy

What is a life cycle ?

- Iterative set of steps
- while delivering a data science product /project.
- Each project has specific requirements to do with the data.

Generalized life cycle



1) Problem Definition

- Just like any other IT life cycle
- here also it starts with “why” and “what”

- **What’s Unique ? :**
- **Identify the type of problem being solved**
- **Weather it could be solved by ML ?**

2) Data Investigation and Cleaning

- Once data sourcing is done :
- Document the data quality
- Clean the data
- Combine various data sets to create new views
- Load the data into the target location (often to a cloud platform)
- Visualize the data
- Present initial findings to stakeholders and solicit feedback

- **What's Unique ? :**
- what is going on and can help reframe the business problem.

3) Minimal Viable Model :

- Don’t build full pledged product, rather, get out something of value and receive feedback.

- **How do you build the Minimal Viable Model?**
- 1) “Lab” Validation.
- 2) “Wild” Validation.

- **What’s Unique ? :**

- Does the model perform better than baseline performance?
- Is it worthwhile to deploy the model in production?
- Are there any unintended consequences?

4) Deployment and Enhancements

- **Deployment** : getting your model output in a Tableau dashboard.

- Enhancement : MVP may not lead as expected after loading complete product, necessary steps had to be done.

5) Data Science Ops

- Usually the life cycle terminates at deployment itself.
- But as the product mature into mainstream operations, companies need to take a stronger product focus that includes plans to maintain the deployed systems long-term.

- **It includes :**
- **Software Management**
- **Model and Data Management**
- **On-going Stakeholder Management**