DATA SCIENCE LIFECYCLE

Data Science is Empirical, Experimental & Exploratory by nature.

→ When you start with the problem in a dataset, you often don't know what you're gonna find. You might not even know what you are looking for.

↳ That's why we need a Lifecycle that is Flexible and Exploratory.

Q What is a Lifecycle?

A Series of steps you take when you approach a challenge.

As a Data Scientist you need a lightweight approach to discovering insights so you can have structure while still being flexible enough to adapt new ideas.

Follow this P. E. E. P Process ....

Preparation

- → clarify the Problem
- → Research & find Relevant Data
- → Data Understanding
- → Setting Goals
- → Create Analysis Plan

You need to come up with questions to answer.

EXPERIMENTATION

- This is where you get down & dirty with the data.
  - → Start Coding
  - → Create Visualizations
  - → Create Models
  - → Explore the Data
- All these steps are iterative, you might find yourself looping through this part.
- You are going to Understand, Explore, Analyze and Make Sense of the Data.

- Then you're gonna do more problem framing, Generate Hypotheses.
- Then you might have to design experiments to test hypotheses.
- Then you're going to build out features & extract some complexity from that raw data.
- Then you are going to build some candidate models
- Then you are gonna tune some hyper parameters.
- Then more analysis on that.

ENGINEERING

- Start thinking about how to set up a PIPELINE and a PROCESS for Cleaning & Extracting Data.

  → Organize Your Code

  → Make Project Reproducable

  → Think about Presentation

  → Deployment of Model

- Once you have gone through messy work of clean data & experiments, it's a good idea to go back & make sure your code is neat, organized, well documented & version controlled.

PRESENTATION

- → Draw Conclusions
- → State Problem Solutions
- → Create an Executive Summary
- → Practice & Prepare a Story

Here you WRAP UP the work you've done.

→ So to work your way through all of these parts of project is iterative.

→ You gotta move through PEEP framework in a fluid way. This fluidity is a key to success in Data Science.

→ It's key to success in making progress on ambiguous questions.