NOTES FROM INDUSTRY

# Analytics Lifecycle Management

7 stages from Problem Formulation to Solution Sunset

Ying Li · Follow

Published in Towards Data Science · 7 min read · May 18, 2021

Adding machine intelligence into our business workflows has become norm now, and there are increasingly more data-drive predictive analytics being developed and integrated into existing business operations to assist decision making, improve efficiency, reduce risks and enhance employee experience.

Nevertheless, with the proliferation of analytics and AI models produced, we are facing the challenge on efficiently managing the analytics lifecycle to ensure that the models yield solid business insights leading to optimal decisions, identified opportunities and righteous actions. It is a multifaceted and complex task.

Below is my view of managing an analytics' entire lifecycle, which provides a step-by-step guidance through the process of formulating a business problem, developing and deploying the analytics, and eventually to

sunsetting the analytics. Note that while the process is presented in a sequential fashion, it is inherently an iterative process which can be used to produce repeatable and reliable predictive result.
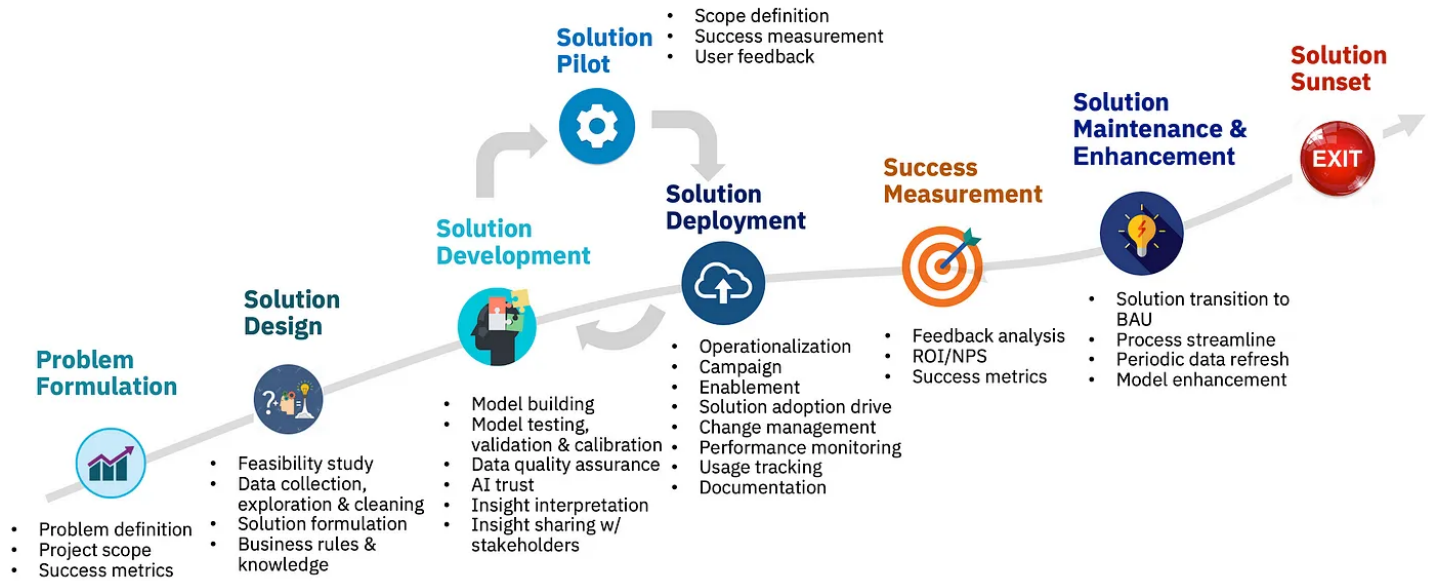


## A. Problem Formulation

Any analytics or AI solution starts with a business problem or a business use case. Only when the business need is clearly understood, can an analytics solution then be designed and developed with a projected business outcome. This is also the stage where a business sponsor needs to be secured who is committed to use the developed analytics in the proposed use case, should it meet the original design goal. Often times a project falls into the pitfall that its analytics output is not adopted by any potential users thus wasting all the efforts. Consequently, involving business sponsors and co-creating solutions with them at the early stage of a project is very critical.

Another two aspects of formulating the problem include the definition of project scope and success metrics. This is the time to ask critical questions such as: What are in scope for this project? Does it include all employee population, or only specific business units, geos, countries, segments? How do we define success? What are the success criteria to be measured? What is the current state, and what will be the future state? What will be the experience of the target users in the future state? Getting clarification to these questions is critical to the Solution Design stage.

**B. Solution Design**

Once the problem is well formulated, the next step is to design the solution which includes

- Understand the current landscape or existing work in this area, especially those within the company so as to avoid any duplicate effort and potentially build upon existing solutions. In case similar solutions are found in external market, then the question of "buy or build" might need to be discussed with business stakeholders. This step shall also include the feasibility study as well as the sizing of efforts if required by business sponsors

- Identify the data needed for the solution and explore their availability. This step may entail brainstorming sessions with data engineers, data scientists, sponsored users, and business stakeholders. Having people with business domain knowledge to chime in the discussions would be invaluable here. On the other hand, often times data collection turns out to be the most time-consuming step, especially when the data is coming from many different sources with a mixture of database tables, spread

sheets, flat files and other heterogeneous formats. The development cycle can be much shortened when the majority data can be automatically pulled from a centralized data repository

- Clean, consolidate and explore the data. At this step, the team will use various tools to search for relationships, trends and patterns to gain deeper understanding of the data, as well as identifying salient features that will help address the business problem from an analytical perspective. While examining the data, the team may find the need to add, delete or combine features to create more precisely focused models

- Understand the need of applying additional business rules. If there is indeed such need, then they should be taken into account during this design stage

## C. Solution Development

At this stage, numerous analytical and machine learning modeling algorithms are applied to the data to find the best representation of the relationships in the data which will help answer the business questions or address the business needs identified in the first stage. Extensive model building, testing, validation and calibration will be conducted accordingly. Another key area of this development stage is the data quality assurance which ensures the quality of the data at various stages of the development based on both business and analytics requirement.

AI trust has become an increasingly important aspect for any machine learning based solutions. At the end of day, if the end users do not trust the insights, prediction or recommendation produced by the AI solution, they

will simply ignore the data. So, how to increase the data transparency, demonstrate the fairness and robustness of the AI output, as well as make the AI decision explainable, become very critical to drive the solution adoption. There are a few tools developed by IBM Research to foster AI trust, including AI Factsheet 360 (https://aifs360.mybluemix.net/), AI Fairness 360 (http://aif360.mybluemix.net), and AI Explainability 360 (http://aix360-dev.mybluemix.net). A lot of good information can be found there.

The last step of this stage is to interpret the insights in a business context, share and validate them with business stakeholders. The information gathered from AI Trust assessment would be very helpful during such review.

**D. Solution Deployment**

This is the stage where we take the developed solution and derived insights, and put them into action using repeatable and automated processes. Campaigns and enablement sessions to target user groups are good ways to raise the awareness of solution which ultimately drives the solution adoption. On the other hand, integrating the solution into existing operational flows is another great and in fact, preferred way to drive its use and achieve business outcome. Sometimes, certain level of change management would be necessary to include the analytics results into a prior process, and a strong support and endorsement from the business sponsors and stakeholders would make this process a bit easier and faster.

Note that there might be a need for a Solution Pilot stage right after the

development, to validate the solution in a relatively small scope setting before scaling it up. Once the team is confirmed that the pilot outcome indeed meets the predefined success criteria along with business stakeholders' approval, then it can move onto the solution deployment stage.

After the solution is deployed and in use, its performance needs to be monitored, usage to be tracked and feedback to be collected, so as to support any potential solution adjustment and enhancement in the future. On the other hand, if indeed severe defects or serious performance issues are identified, then the team may need to go back to the solution development stage.

### E. Success Measurement

Once the solution is successfully deployed and its usage been continuously tracked, it is time to measure success. The success metrics defined during the Problem Formulation stage would be measured and reported at this stage. Other metrics could include the ROI in terms of cost saving or revenue generation, as well as NPS based on feedback.

### F. Solution Maintenance & Enhancement

Once the solution is in a steady state, it can be transitioned into a BAU mode. It is also the time to streamline/automate the process as much as possible to make it more efficient with minimal maintenance or human intervention. Periodic model refreshes will also happen during this stage.

Moreover, with continuous monitoring and measurement of the model performance based on standardized metrics, the team should continuously assess the validity and effectiveness of the solution to examine whether it still meets the current business needs. If a gap is identified, then the model needs to be enhanced and recalibrated. This may bring the cycle back to solution development, or even to solution design in case that the business needs have evolved over time, existing data source is no longer valid, or new data becomes available. This ultimately makes the lifecycle management an iterative process.

**G. Solution Sunset**

In the event that the original business need is no longer valid, a new solution has been developed with enhanced capability, or the clients have moved on, the existing solution will no longer be needed, and it is time to retire the model.

Having reviewed the end-to-end analytics lifecycle, I do want to point out a few things that could potentially complicate, slow down or even derail this whole process.

- The needed data sources might be scattered across the entire organization, require very manual collection and consolidation, or with very strict access due to its privacy and sensitivity, or even confidential thus cannot be shared

- The built of AI trust in the data insights is lagging behind and the target users have concerns about the validity and ethics in the insights

- The business might be slow to leverage the insights in the target business use case, or the target users do not value such insights and choose to ignore them for any decision-makings

- The integration of analytics into existing workflows might require a lot of change management touching upon different user personas

To avoid or mitigate some of the above pitfalls, below are some suggestions:

- Build a very clear use case with estimated ROI

- Gain very strong support from the business sponsors and stakeholders

- Co-create solution with sponsored users through design thinking sessions and keep them in the loop for the entire analytics lifecycle

- Develop the solution in an agile way

Good luck, and enjoy the management of the analytics lifecycle!

Analytics Lifecycle  Analytics  Editors Pick  Notes From Industry

# Written by Ying Li

40 Followers · Writer for Towards Data Science

Ying Li is the Global Head of People Analytics at PepsiCo. She leads team developing advanced analytics solutions to support leaders in key decision-making

**More from Ying Li and Towards Data Science**

Ying Li in IBM Data Science in Practice

## Understanding Skills Scarcity in An External Market—Part 1

What is Skills Scarcity? How to measure it? and Where to use it?

Jun 16, 2021 👏 4

Steve Hedden in Towards Data Science

## How to Implement Graph RAG Using Knowledge Graphs and...

A Step-by-Step Tutorial on Implementing Retrieval-Augmented Generation (RAG),...

Sep 6 👏 1.2K 💬 13





Ahmed Besbes in Towards Data Science

## What Nobody Tells You About RAGs

A deep dive into why RAG doesn't always work as expected: an overview of the...

✦ Aug 23 👏 1.7K 💬 25
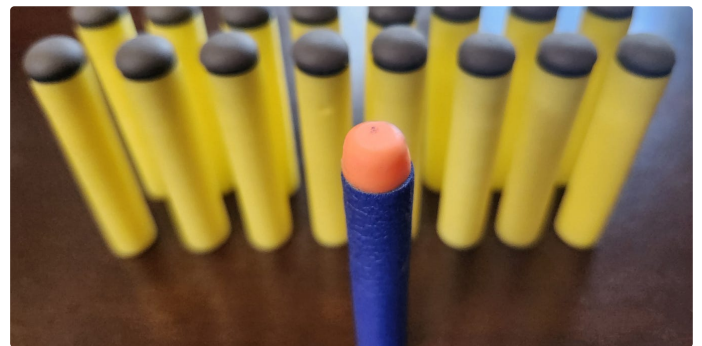
Ying Li

## Proactively Sourcing Candidates in Talent Acquisition

What is proactive sourcing? How to accomplish it with quality? and What's the...

May 3

# Recommended from Medium



Aakash Gupta

### What the Product Sense Interview is Evaluating

I found Meta's 5 criteria for the product sense interview. They're surprising.

May 31   👏 25   💬 1



Henry Ha (Here & Now) in GoPenAI

### The Cornerstone of Success: Data Science Project Documentation

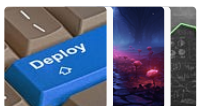A part of the Data Science & AI project planning and management Series

⭐ Apr 17   👏 2

# Lists

### The New Chatbots: ChatGPT, Bard, and Beyond

12 stories · 467 saves



### Business

41 stories · 143 saves



### Predictive Modeling w/ Python
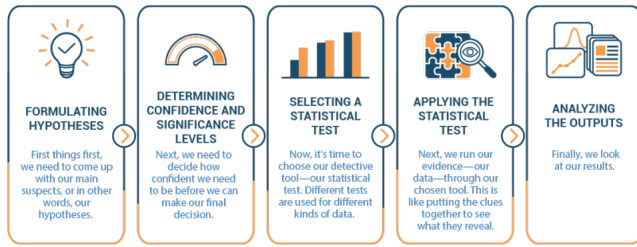
20 stories · 1561 saves



### ChatGPT prompts

48 stories · 2039 saves

## HYPOTHESIS TESTING

**FORMULATING HYPOTHESES** — First things first, we need to come up with our main suspects, or in other words, our hypotheses.

**DETERMINING CONFIDENCE AND SIGNIFICANCE LEVELS** — Next, we need to decide how confident we need to be before we can make our final decision.

**SELECTING A STATISTICAL TEST** — Now, it's time to choose our detective tool—our statistical test. Different tests are used for different kinds of data.

**APPLYING THE STATISTICAL TEST** — Next, we run our evidence—our data—through our chosen tool. This is like putting the clues together to see what they reveal.

**ANALYZING THE OUTPUTS** — Finally, we look at our results.

Muchid Ariyanto

## A Simple Guide to Understanding Hypothesis Testing: The Power of...

Hypothesis testing is crucial in any data-driven company, especially when teams hav...

Sep 16  👏 1

Alexander Nguyen in Level Up Coding

## The resume that got a software engineer a $300,000 job at...

1-page. Well-formatted.

⭐ Jun 1  👏 23K  💬 455



Mohammed Azarudeen Bilal

## 99 Most Intimidating Business Analyst Interview Questions

Unlock the Answers You Probably Didn't Know About

⭐ Aug 22  👏 293  💬 2



Bhavik Patel in Product Coalition

## Analysing Tech Layoffs: Which Roles Were Hit Hardest?

In the aftermath of tech layoffs, it would appear that engineering was hit hardest, bu...

⭐ Sep 17  👏 819  💬 16

See more recommendations