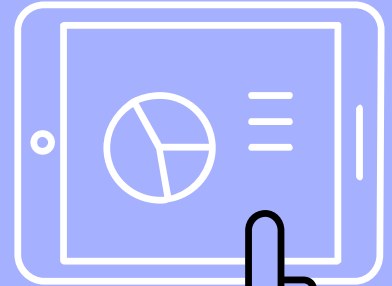
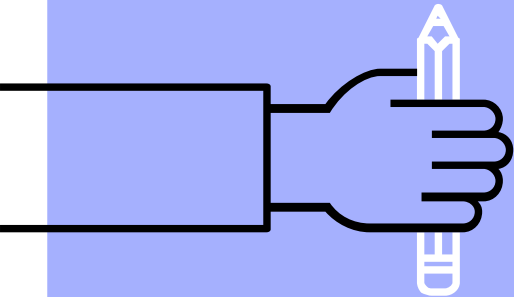
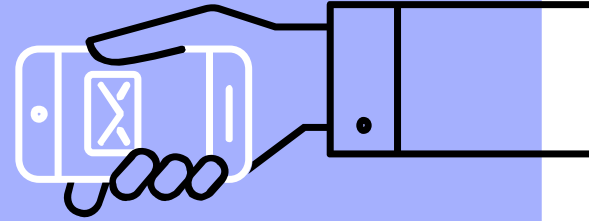
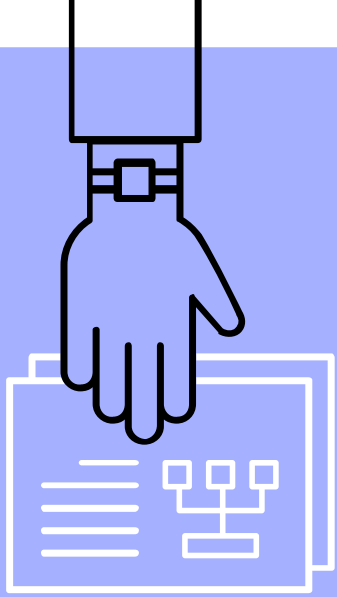
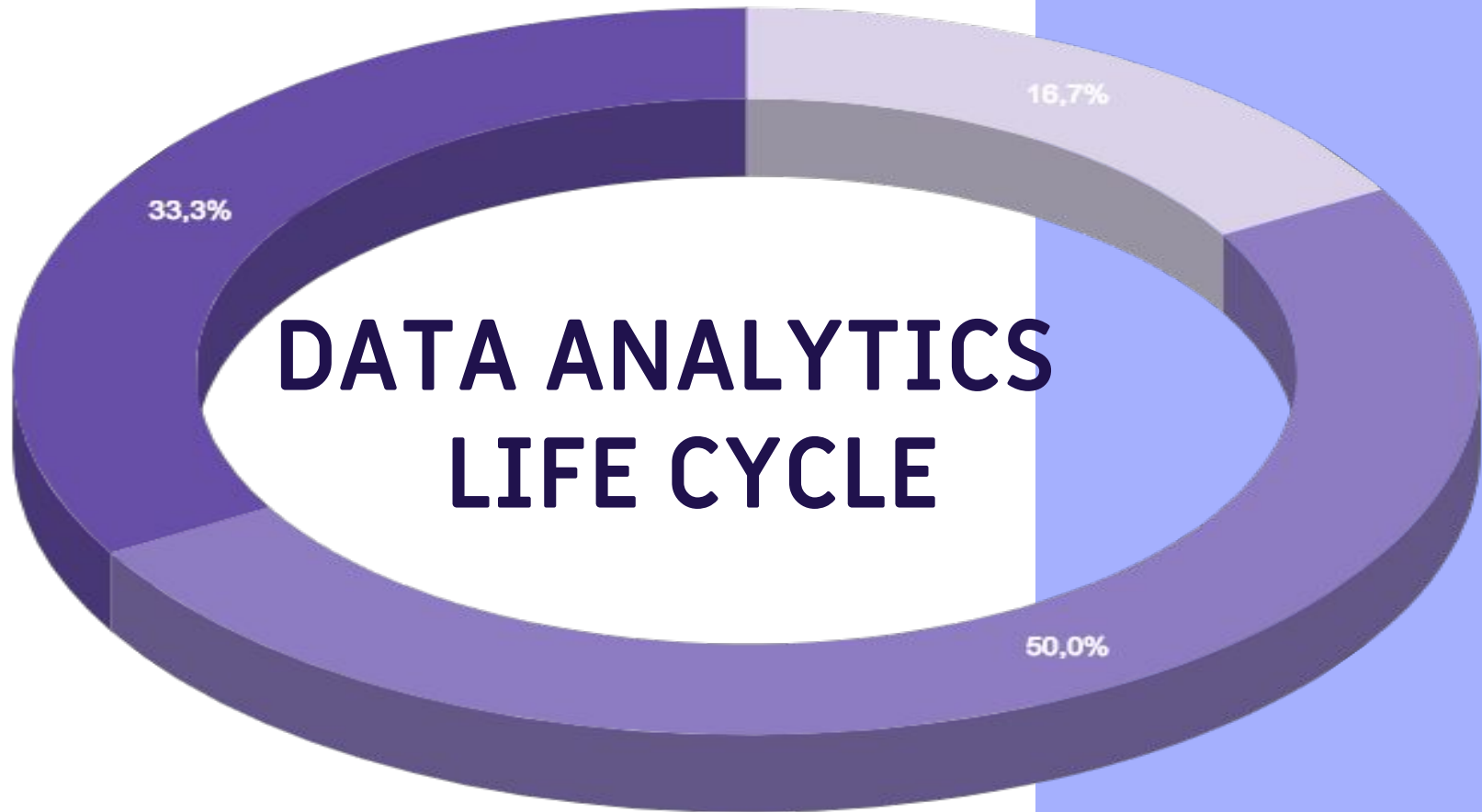


Data Analytics Life Cycle

DS with Python
Lecture 6





DISCOVERY

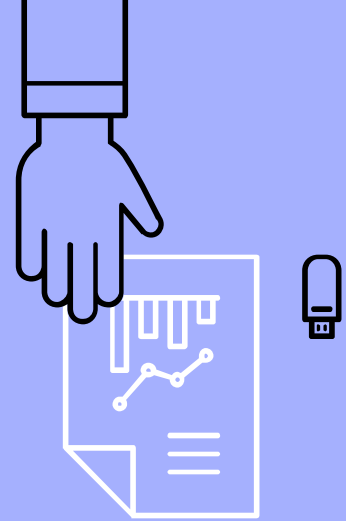
PHASE I



**DO I HAVE
ENOUGH
INFORMATION
TO DRAFT AN
ANALYTIC
PLAN AND
SHARE FOR
PEER
REVIEW?**

- 1. Learning the Business Domain**
- 2. Resources**
- 3. Framing The Problem**
- 4. Identifying Key Stakeholders**
- 5. Interviewing the Analytics Sponsor**

STEP BY STEP PROCESS



Common Questions that are helpful to ask during Discovery Phase

- 1. What business problem is the team trying to solve?**
- 2. What is the desired outcome of the project?**
- 3. What industry issues may impact the analysis?**
- 4. What timelines need to be considered?**
- 5. Who could provide insight into the project?**

Common Questions that are helpful to ask during Discovery Phase

- 1. Who has the final decision making authority on the project?**
- 2. How will the focus & scope of the problem change if the following dimensions change?**

Time

People

Risk

Resources

Size

DATA PREPARATION

PHASE II



**DO I HAVE
ENOUGH
GOOD
QUALITY
DATA TO
START
BUILDING THE
MODEL?**

- 1. Preparing the Analytic Sandbox**
- 2. Performing ETLT**
- 3. Learning About the Data (Data available/collected)**
- 4. Data Conditioning**
- 5. Survey And Visualize**

Common Questions that are helpful to ask during Preparation Phase

- 1. What are the data sources?**
- 2. What are the target fields/table columns?**
- 3. How clean is the data?**
- 4. How inconsistent are the contents/data types?**
- 5. Look for Systematic errors caused by auto-data feeds.**

Common Questions that are helpful to ask during Preparation Phase

- 1. Review Data Sources to ensure calculations remain consistent within columns or tables.**
- 2. Assess the granularity of the data (range of values, level of aggregation)?**
- 3. Does the data represent the population of interest ?**
- 4. Are measurements in sync like datetime/timestamp?**

Common Questions that are helpful to ask during Preparation Phase

- 1. Is names standardized/data normalized?**
- 2. Are scales consistent? If not, by how much.**
- 3. Are State or Country names, used in full/abbreviated or ISD codes?**
- 4. Are units using Metric System or else (same with currency, temperature, date formats etc)**

MODEL PLANNING

PHASE III



**DO I HAVE A
GOOD IDEA
ABOUT THE
TYPE OF
MODEL TO
TRY? CAN I
REFINE THE
ANALYTIC
PLAN?**

1. **Assess the Structure of the Datasets**
2. **Accept or Reject the existing hypothesis**
3. **Determines if situation warrants a single model or a series of models (Model Selection).**
4. **Data Exploration & Variable Selection**

MODEL BUILDING

PHASE IV



**IS THE MODEL
ROBUST
ENOUGH?
HAVE WE
FAILED FOR
SURE?**

1. **Build a Robust Model that meets all objectives.**
2. **Ensure the model appears valid and accurate on the test data**
3. **Ensure the model output/behaviour make sense to domain experts.**

That is, the model is giving answers that make sense in this context.



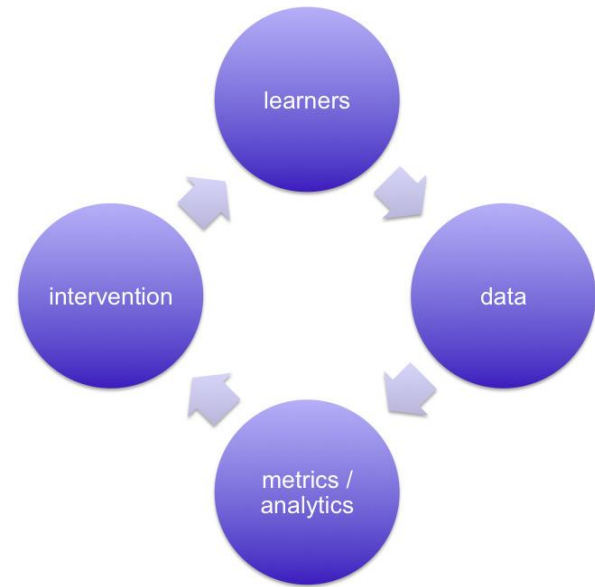
**IS THE MODEL
ROBUST
ENOUGH?
HAVE WE
FAILED FOR
SURE?**

1. Does the parameter values of the fitted model make sense to the domain context.
2. Is the model sufficiently accurate to meet the goal.
3. Does the model avoid intolerable mistakes?
4. Are False Positives more serious or False Negatives?



**IS THE MODEL
ROBUST
ENOUGH?
HAVE WE
FAILED FOR
SURE?**

1. Are more data or more inputs needed?
2. Do any of the inputs need to be transformed or eliminated?
3. Will the kind of model chosen support the runtime requirements?
4. Is a different form of model required to address the business problem ?



**Once the Data Science Team can
evaluate either if the model is
sufficiently robust to solve the
problem or if the team has failed, it
can move to the next phase in Data
Analytics LifeCycle**

COMMUNICATE RESULTS

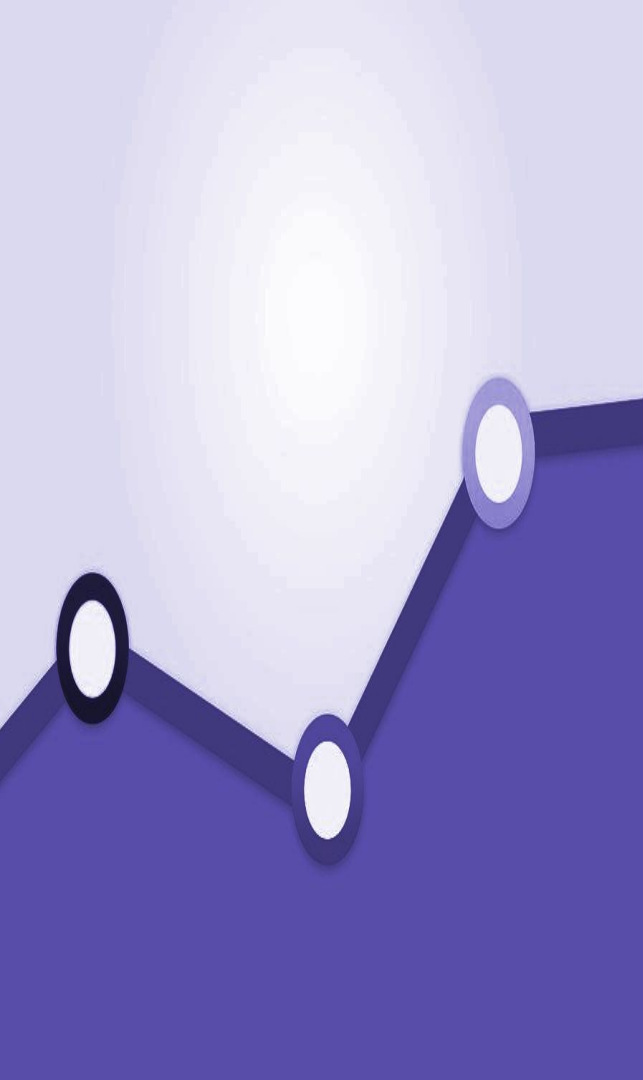
PHASE V



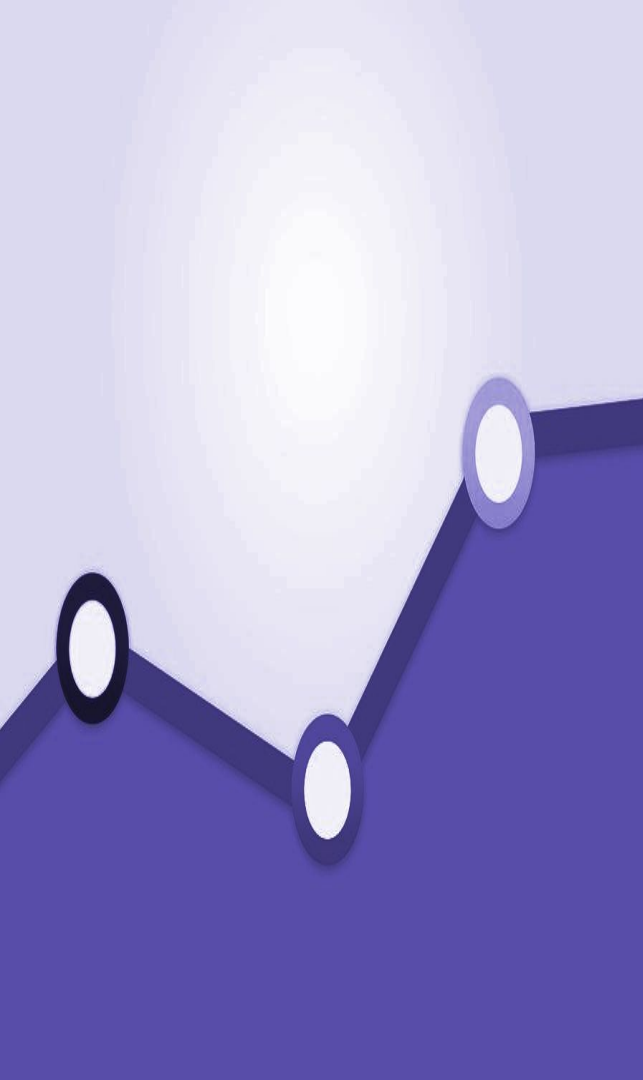
**IS THE MODEL
ROBUST
ENOUGH?
HAVE WE
FAILED ??**

1. When conducting this assessment determine if the results are statistically significant and valid.
2. This phase deal with sharing your findings with the stakeholder as it is.
3. The stakeholders must understand how the model affects their processes.

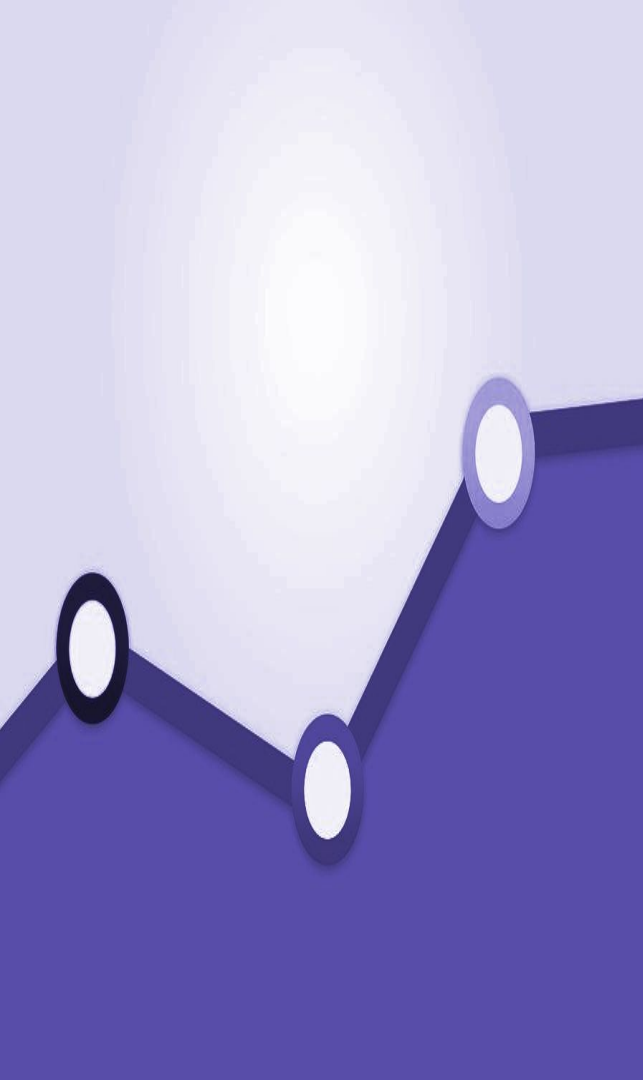
OPERATIONALIZE



Finally the team communicates the benefits of the project and sets up a pilot project to deploy the work in a controlled way before broadening the work to a full ecosystem.



The risk can be managed more effectively and the team can learn by undertaking a small scope, pilot deployment before a wide-scale rollout.



This approach enables the team to learn about the performance and related constraints of the model in a production environment on a small scale and make adjustments before a full deployment.

THANKS!

Any questions?

You can find me at:
ankita.sinha8118@gmail.com

