

# Foundation of Data Science

## Course overview

w1 - Introduction of data science concepts

w2 - The impact of data today.

w3 - You career as a data professional.

w4 - Data applications and workflow.

w5 - Course 1 end-of-course project.

## Week-02 (The impact of data today)

\* **Data cleaning:** The process of formatting data and removing unwanted material.

\* **Personally Identifiable Information (PII):**

Information that permits the identity of an individual to be inferred by either direct or indirect means.

Examples: Biometric records,

Usernames, & social security number

\* A person's background, experiences, and beliefs leads to **bias**.

\* To eliminate bias: The sample set should be representative to the greater population.

\* Data Privacy:

means preserving a data subject's information and activity any time a data transaction occurs.

○ Privacy rights for the person whose data is being collected:

- ① Protection from unauthorized use.
- ② Freedom from inappropriate use of their data.
- ③ Ability to give consent to data collection.
- ④ Legal right to access the data.

\* Data protection in Action:

- ① Incorporating access permission to ensure that only the people who are supposed to access that information can do so.
- ② Anonymization of the data that means remove PII from the information.
  - Procedure: Blanking, Hashing, Masking.

③ Data aggregation is another way for protecting privacy. In this approach, the multiple data input is grouped by their similarities. For example, store a group of members that contains members whose age are between 25-34 instead of saving each member and their exact age.

### \* Five principles for data team building :

- |                    |                   |
|--------------------|-------------------|
| ① Adaptability.    | ② Activation.     |
| ③ Standardization. | ④ Accountability. |
| ⑤ Business impact. |                   |

### \* Data team :

RACI is a matrix that companies organize roles and responsibilities for individuals or teams to ensure work get done efficiently.

1. Responsible who are directly responsible for the task.

2. Accountable Those who approve the task done by responsible person. often a manager / team lead.

3. Consulted: This role apply to those to offer input on a task. There should be a clear and open line of 2-way communication between those assigned to responsible and consulted. An example of the role may be subject matter expert.

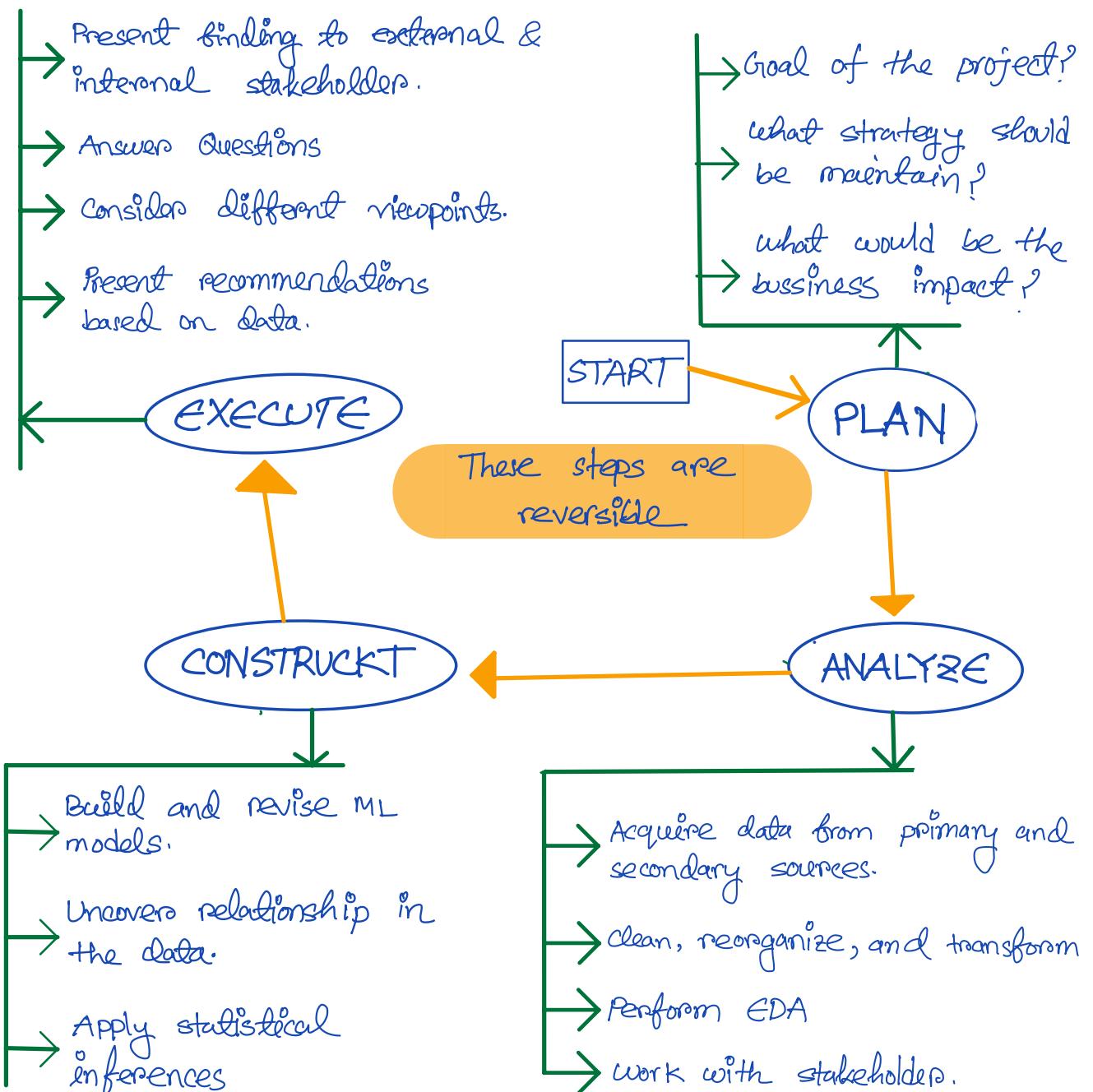
4. Informed: Those in this role need to be kept aware of progress & concern of those working on a project. The senior management holds this position. They only needs the high levels of insights.

\* Edge computing: A way of distributing computational task over a bunch of nearby processor.

\* Data stewardship: The practice of an organization that ensure the data is accessible, useable, and safe.

## Week-04 (Data application & workflow)

\* PACE framework for big data analysis :



## \* Key considerations when crafting a message :

- ① **Purpose** : The reason why the communication is taking place.
- ② **Receiver** : the audience who receiving the information.
- ③ **Sender** : The person responsible for crafting the message/communication.

## \* Seven traps for effective communication :

- ① **Speak the language of the audience** : Identify the needs of your audience.

### key takeaways :

- (i) Breakdown technical concepts into simpler terms.
- (ii) Use shorter sentence.
- (iii) Use direct language and avoid unnecessary details.
- (iv) Pay attention to diverse background.

- ② **Invite Questions and welcome feedback** : Everyone can use feedback whether it is positive reinforcement or area for improvement.

- key takeaways** : Analyze feedback. If not then set-up follow up meeting to understand him the project goal.

3. Be the connection to data: You are your team's direct connection to the insights your data provide.

key takeaways:

- ① Focus on the objectives to help others better understand your data process.
- ② Tell the story with compelling and cohesive narrative.
- ③ Respond to the question in a timely manner.
- ④ Find opportunity to address stakeholder questions.
- ⑤ Demonstrate your value to the team.

4. Let your visualization help tell the story: visualizations are one of the best way to communicate idea, especially when dealing with big-data.

key takeaways:

- ① Be sure that your visuals tell the story within the data. Design visualization for inclusivity.
- ② Use labels and text to clarify, not clutter.
- ③ Use fonts that are easy to read.
- ④ Use high-contrast, shading, and other customization to communicate the message clearly.

5. Build Positive Professional relationships: This will build credibility and influence in the workplace and allow continuous growing.

key takeaways :

- ① Focus on what matters to the audience.
- ② Invite feedback and discussion.
- ③ Be a trusted SME who communicate clearly and inclusively.

6. Identify assumption about data: The background, experiences, beliefs, and worldviews of people can influence the information containing the data. The data professional need to identify bias from the data. If not than the data might lead to false information.

key takeaways :

- ① Is there anything I am taking for granted.
- ② Can I determine if the assumptions is correct?
- ③ Can I identify the bias?

7. Identify Limitation of the Data: Have to encounter limitation of data that can impact the analysis. And it should be addressed before the analysis process.

## key takeaways :

- ① Is the data complete? Are there missing values?
- ② Are the dataset formatted correctly?
- ③ Is the sample sufficient to conduct the analysis of an entire population?
- ④ Does the data contains any PII? How should we protect them?

## \* Elements of a projects proposal :

- ① **Project title:** Title of the project should be top of the document.
- ② **Project Objective:** One to three statements, the explanation of what the project trying to achieve.
- ③ **Milestones:** ~grouping of tasks within a project, breaking the work needed to be smaller, manageable goals. It also include the scheduling of the project that needs to be completed within the projects.
- ④ **Tasks:** Details the work needs to completed within a milestone.

5. **Outcome**: ~ are the completed actions or result that allow a project to continue.
6. **Deliverables**: ~ are item that can be shared among team members. These are the end product of some portion of the work.
7. **Stakeholders**: The individuals or group who are directly involved the project and have a vast interest in the success of a project.
8. **Estimated time**: At the beginning of a project, the estimated time should be complete as the milestone.

## Week-05 (Course-1 end-of-course project)

### \* Start Your Project :

1. Gather information about the business problem or question to be answered.
2. Respond to key questions posed in the PACE strategy document.
3. Create a project proposal for cross functional team members.

### \* Course-1 end-of-course deliverables :

1. Complete course-1 PACE strategy document.
2. Create project proposal.

### \* TikTok workplace scenario :

○ **Project background :** TikTok data team is in the earliest stage of the claims classification project. The following tasks are needed before the team can begin the data analysis process :

1. Organize project task into milestones.
2. Classify tasks using PACE workflow.
3. Identify relevant stakeholders.