



DICE
ANALYTICS

Data Science and Machine Learning



<https://www.facebook.com/diceanalytics/>



<https://pk.linkedin.com/company/diceanalytics>

OUR DATA SCIENCE TEAM



Anees Akhter



Ahmed Niazi



Ahmad Raza Khan

COURSE DETAILS

Course Outline	Available here
Assignments	02
Quizzes	02
Projects	03
Presentation	01
Final Exam	01

What is Data?



TYPES OF DATA (STRUCTURE-WISE)

Structured Data



Semi-structured Data



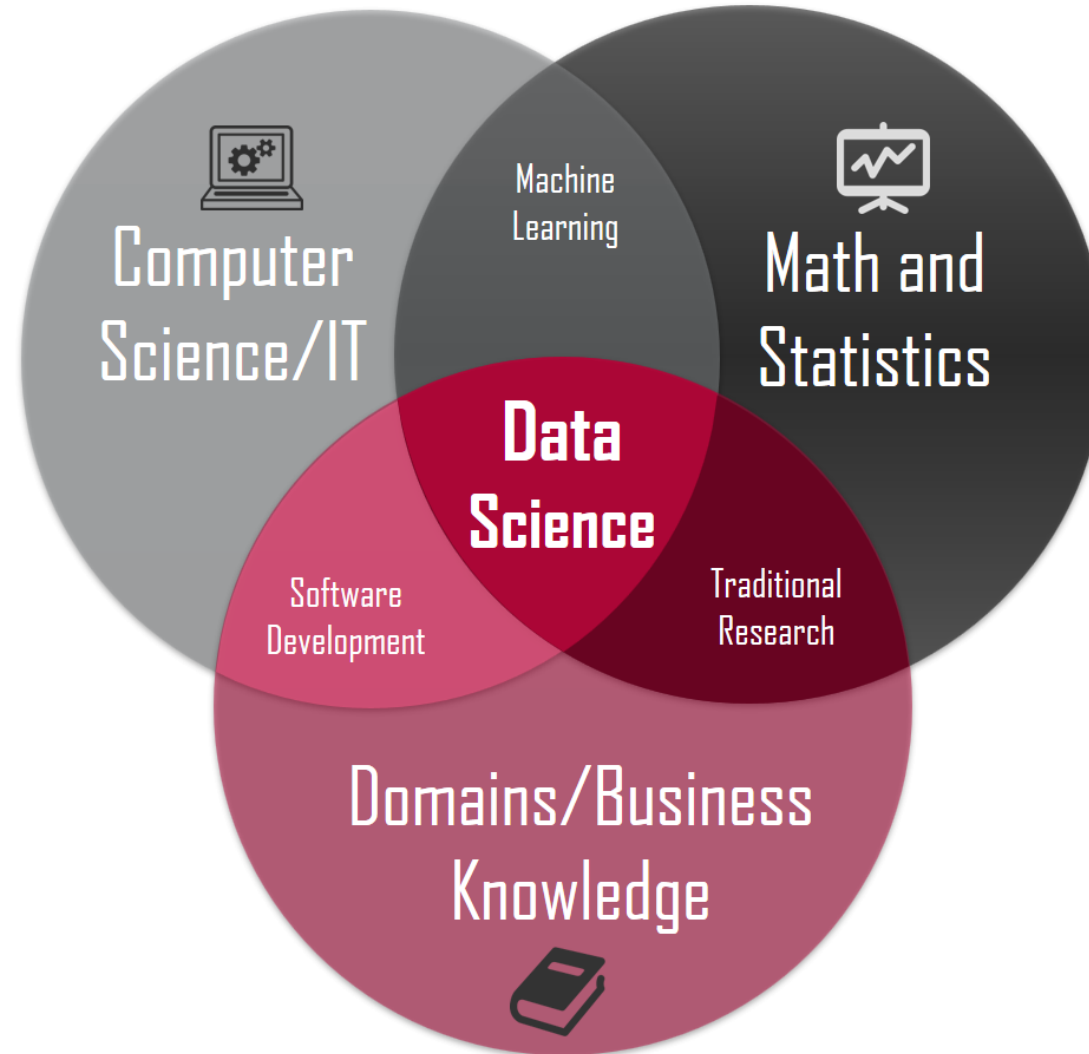
Unstructured Data



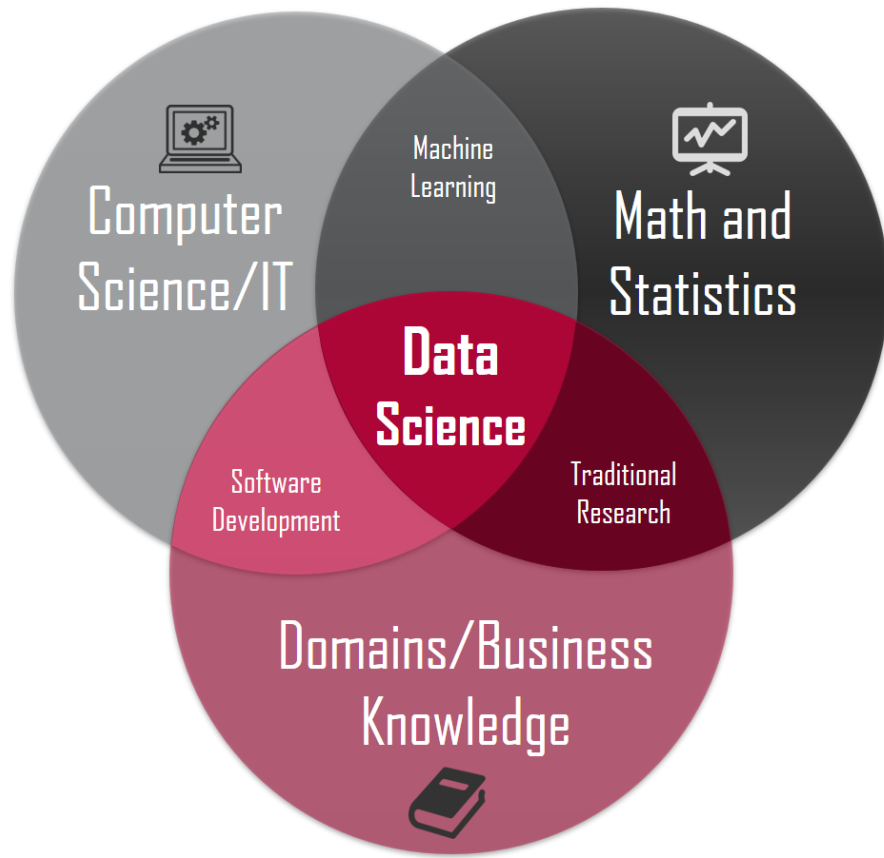
What is Data Science ?



Multi Discipline



We all start from Zero



IT Guys

Non-IT Guys

Math's & Stats



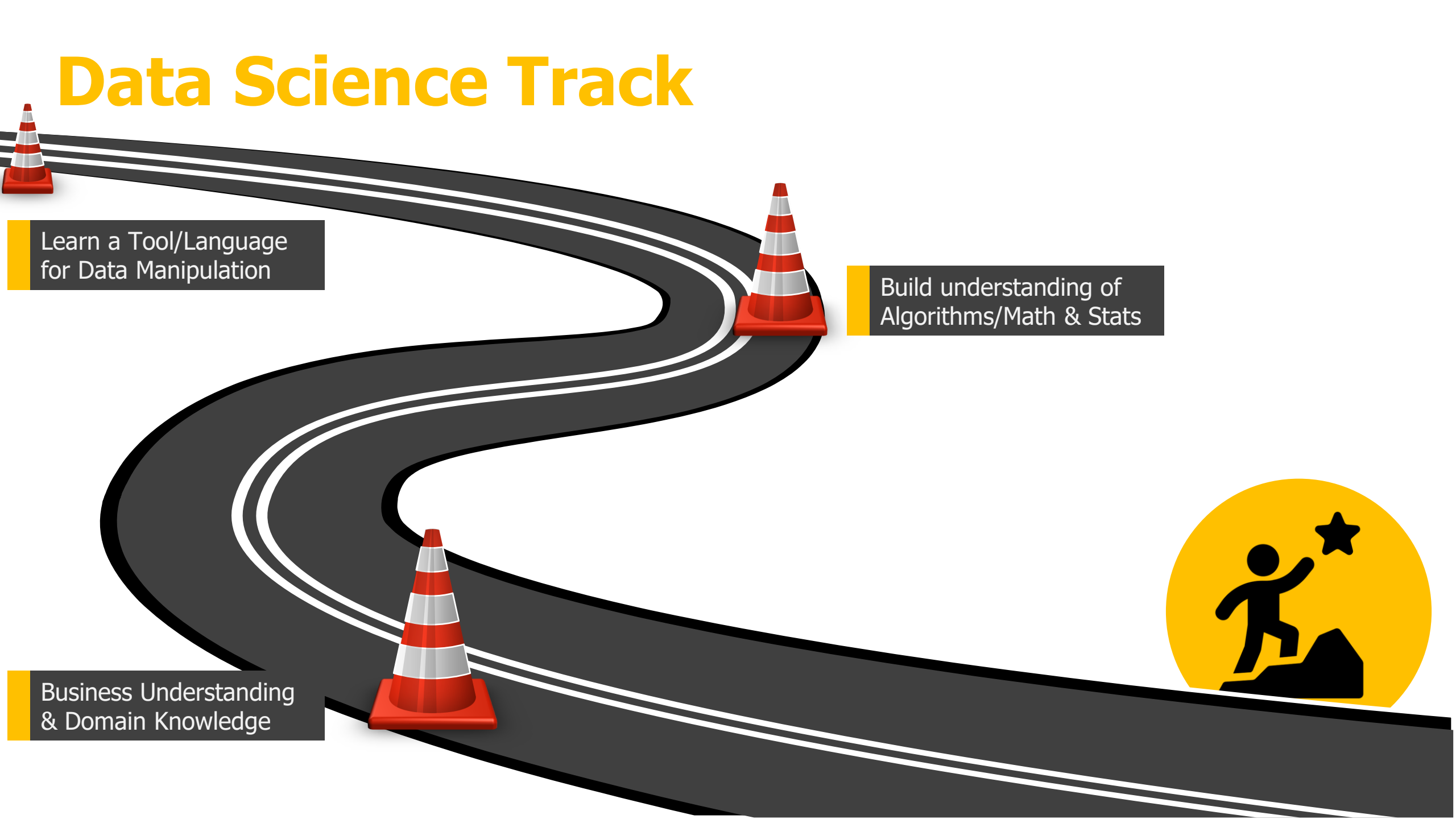
Business Knowledge



IT & Computers



Data Science Track



Learn a Tool/Language
for Data Manipulation

Build understanding of
Algorithms/Math & Stats

Business Understanding
& Domain Knowledge





Download Microsoft Lobe



<https://www.lobe.ai/>

Machine Learning Made Easy





Tour



Documentation



Community



Examples

New Project



Create a new project
to get started.





Untitled

 Label

☒ Train

 Use

All Images 0

One 0

5 images per label
needed to start training.

Label

Import



To start training your model,
import and label some images.

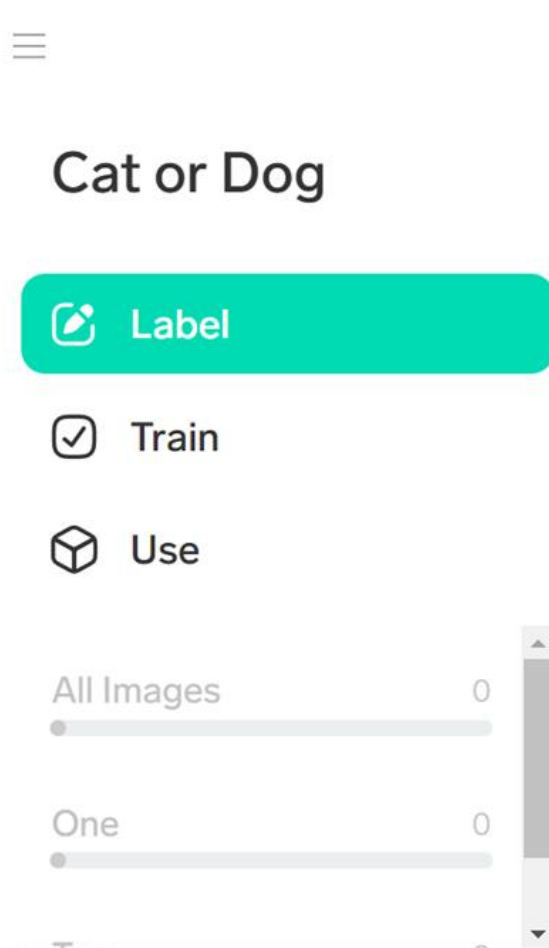
 Learn More

 Watch Tour

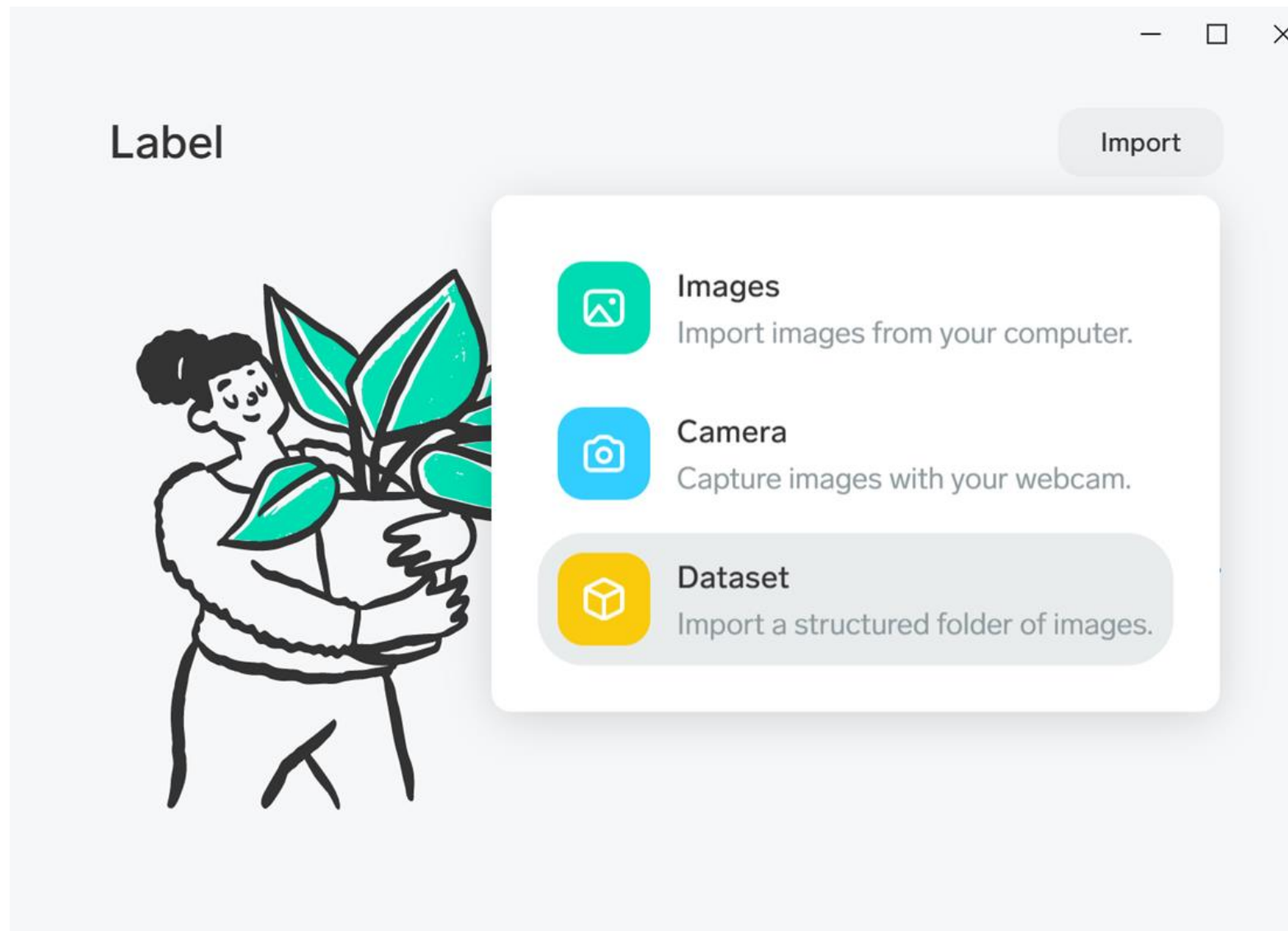
PROBLEM STATEMENT

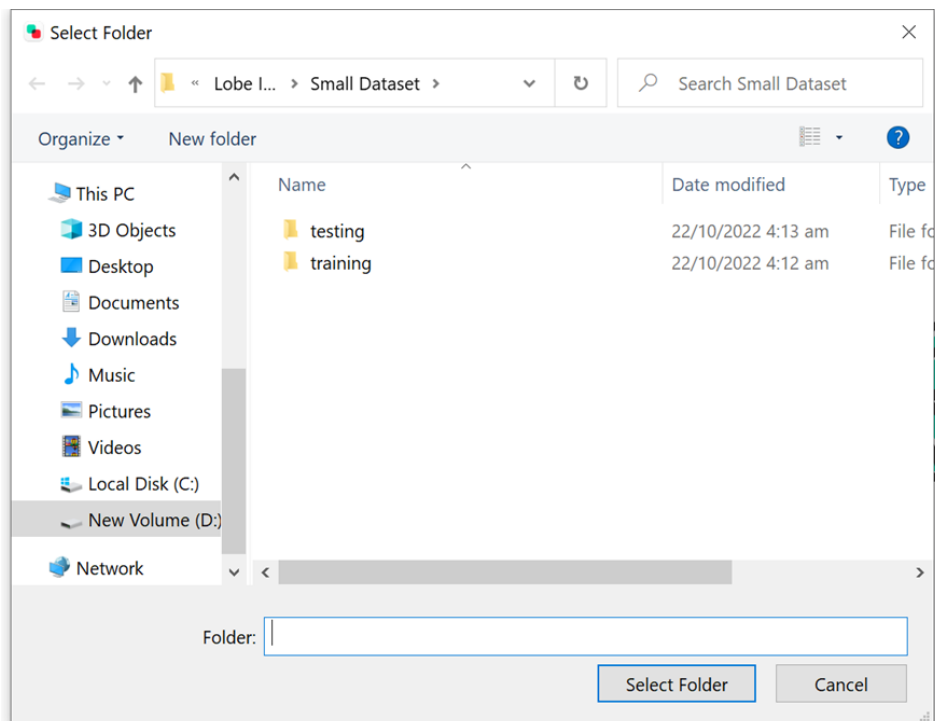
To identify if an image is of a cat or a dog.

DATA INGESTION



5 images per label
needed to start training.





To sta
import

Le

Label Images

Would you like Lobe to use the folder names
to automatically label your images?



☒ **Label Using Folder Name**
Use folder names to label your images.

☐ **Label Manually**
Use Lobe to manually label your images.

Cancel

Import

MODEL TRAINING



Cat or Dog

 Label

☐ Training...

 Use

cats 72

dogs 72

Lobe is starting to train
your model...

All Images

View

Import

cats 72



MODEL RESULTS



Cat or Dog

 Label

☒ Train

 Use

cats 72



dogs 72



97% of your images are
predicted correctly,
3% incorrectly.

All Images

View

Import

cats 72

Show Less



MODEL TESTING



Cat or Dog

 Label

☒ Train

 Use

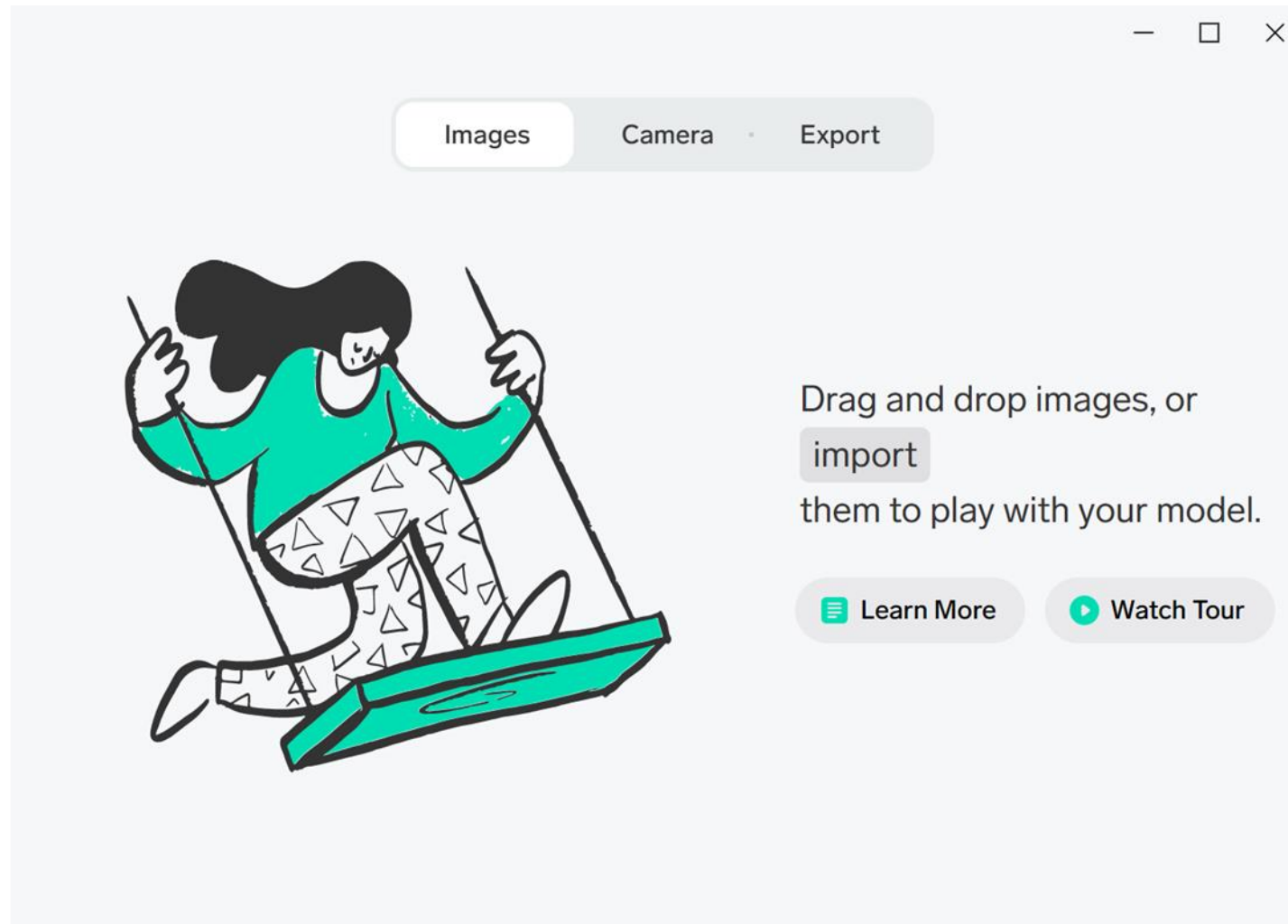
cats 94%

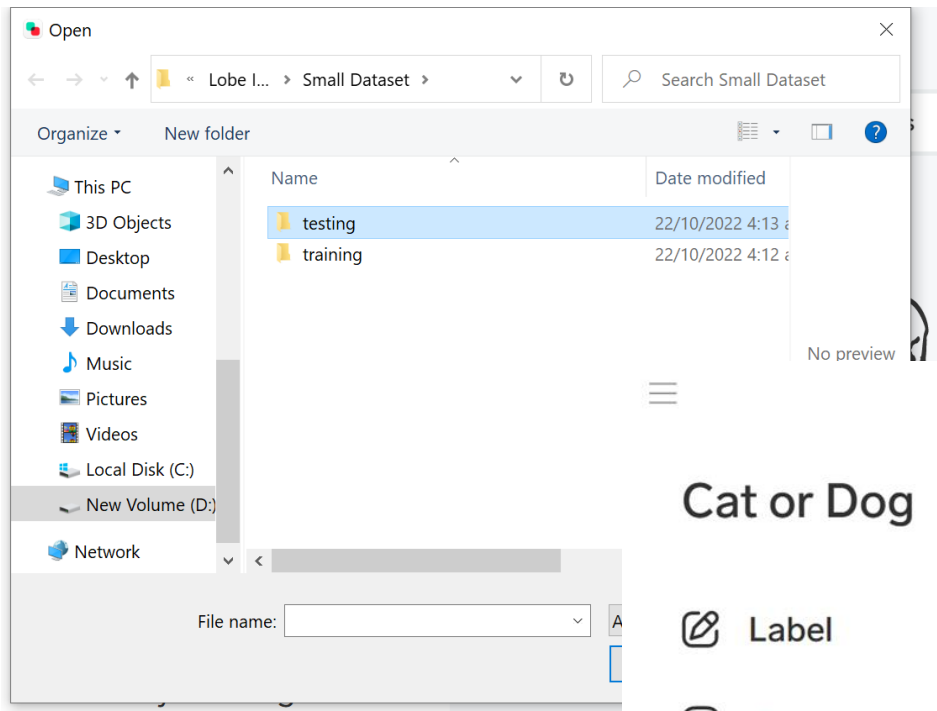


dogs 100%



97% of your images are
predicted correctly,
3% incorrectly.





Cat or Dog

 Label

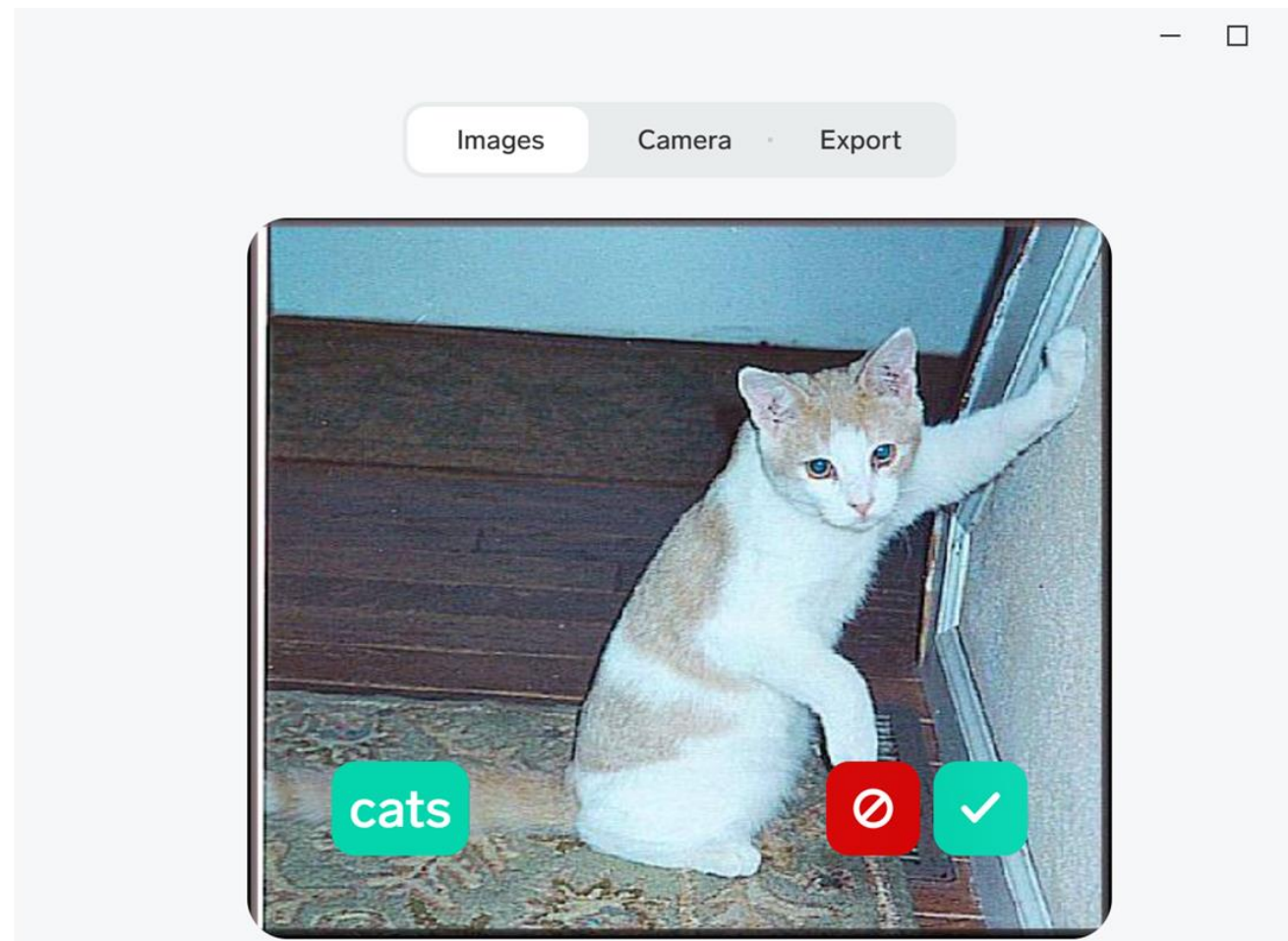
☒ Train

 Use

cats 94%

dogs 100%

97% of your images are
predicted correctly,
3% incorrectly.



Download Complete Dataset (optional)

<https://www.microsoft.com/en-us/download/details.aspx?id=54765>

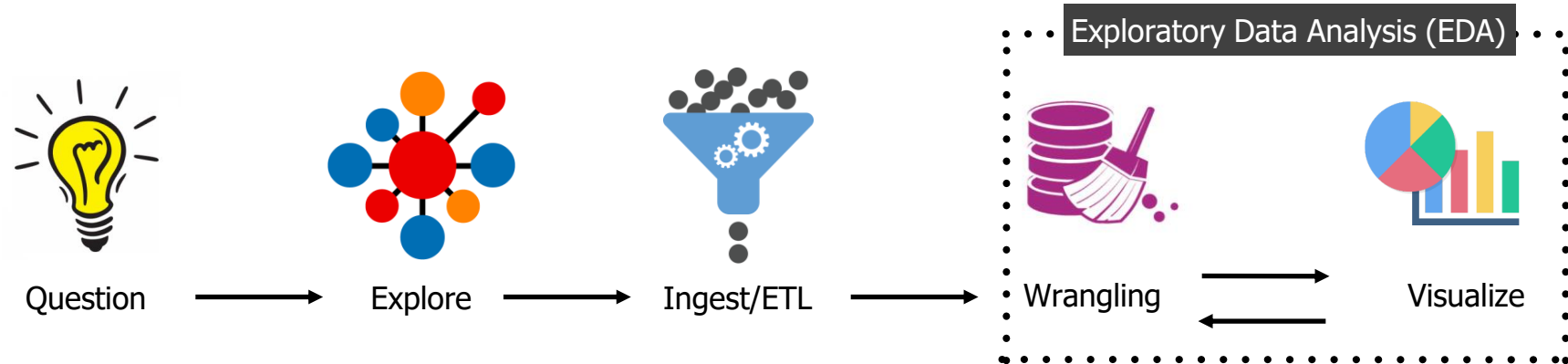
Kaggle Competition (optional)

<https://www.kaggle.com/c/dogs-vs-cats>

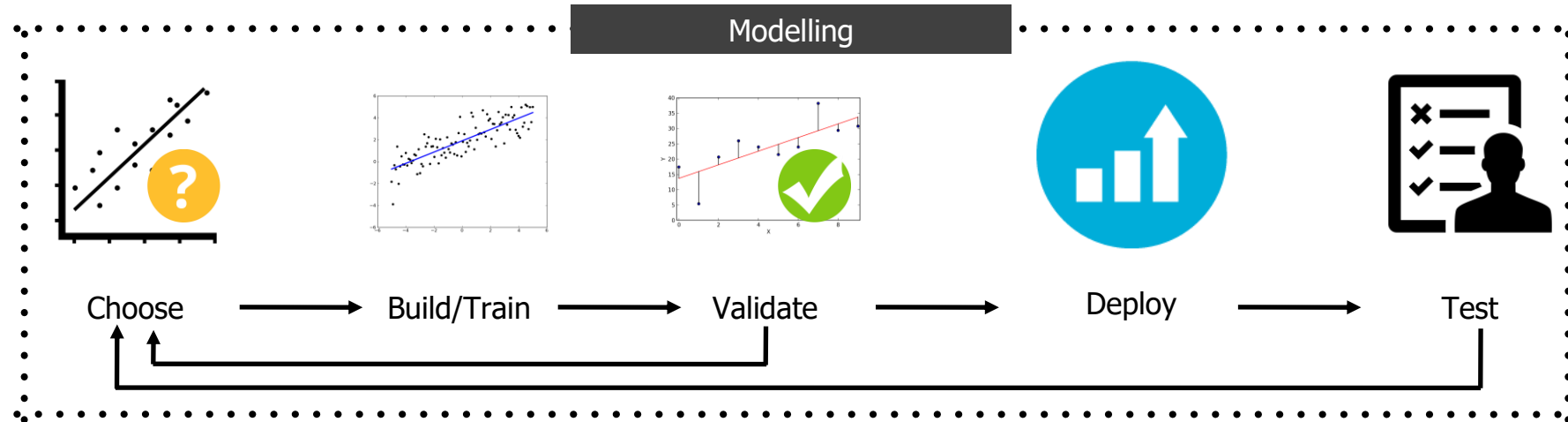
Now Challenge Your Imagination!



1.



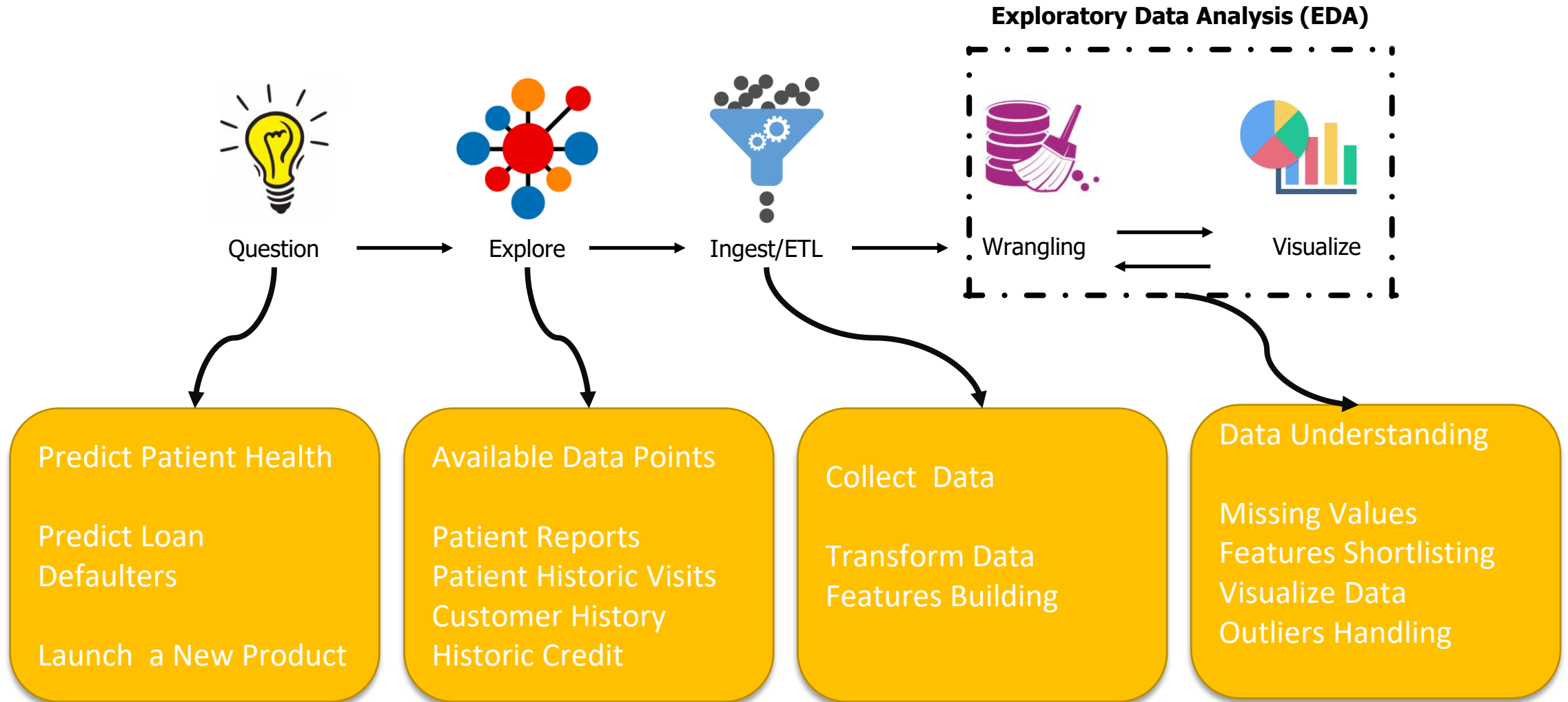
2.



3.

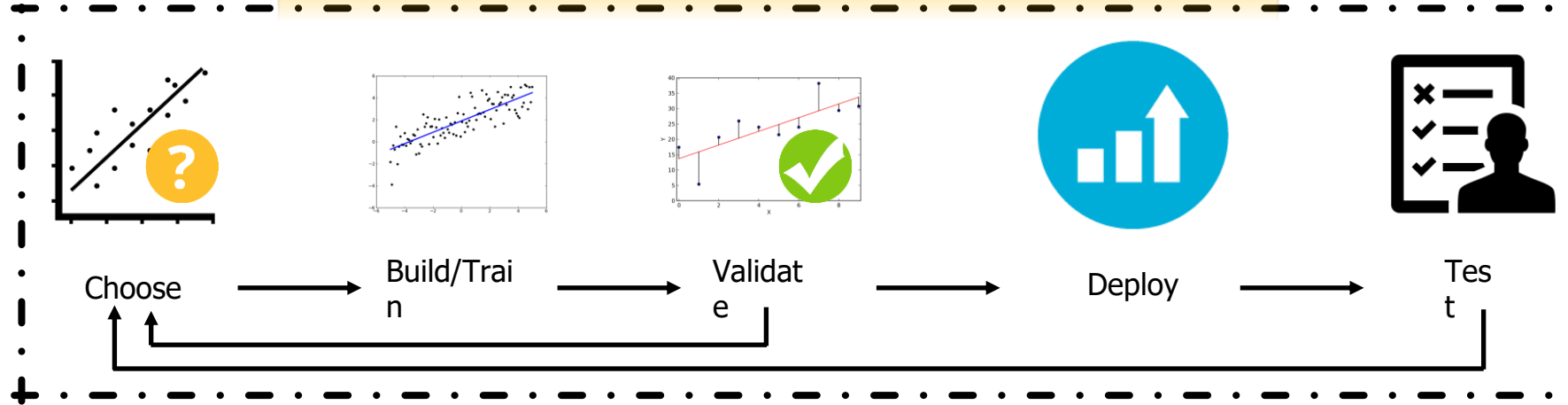


DATA SCIENCE LIFECYCLE



DATA SCIENCE LIFECYCLE

Core Artificial Intelligence / Machine Learning



Cards Color Guess



*Law Enforcement Agency
Training Centers*

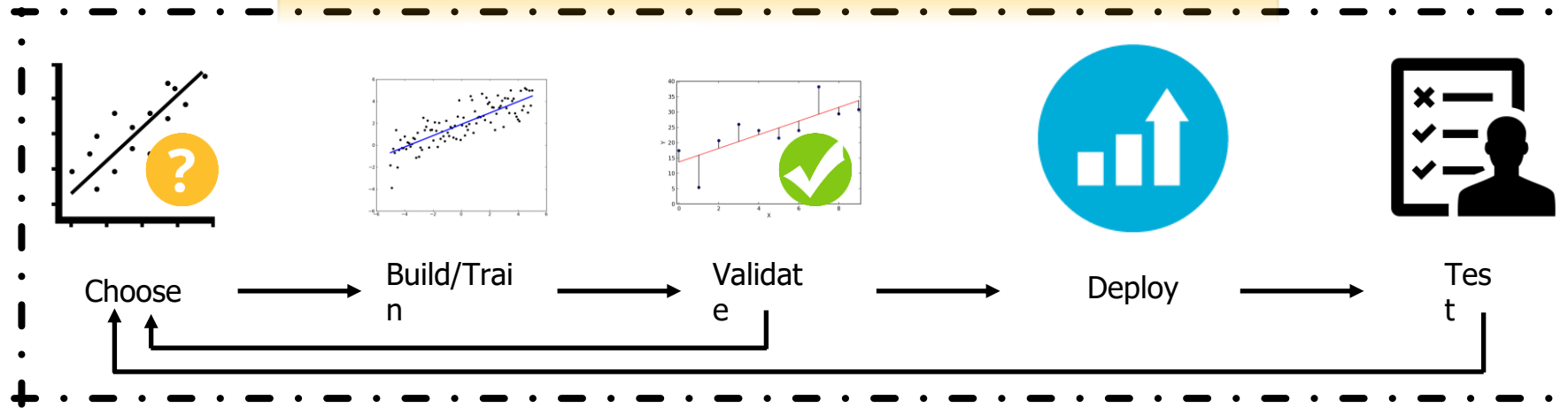


Examination Centers



DATA SCIENCE LIFECYCLE

Core Artificial Intelligence / Machine Learning



Examination Centers



Approach
Number 01



Teacher Used
Few Examples
in Class Room

Teacher Used
Few Examples
in Class Room

Teacher Used
Same
Examples in
Exam Hall

Marks Will Be
High

Approach
Number 02



Teacher Used
Few Examples
in Class Room

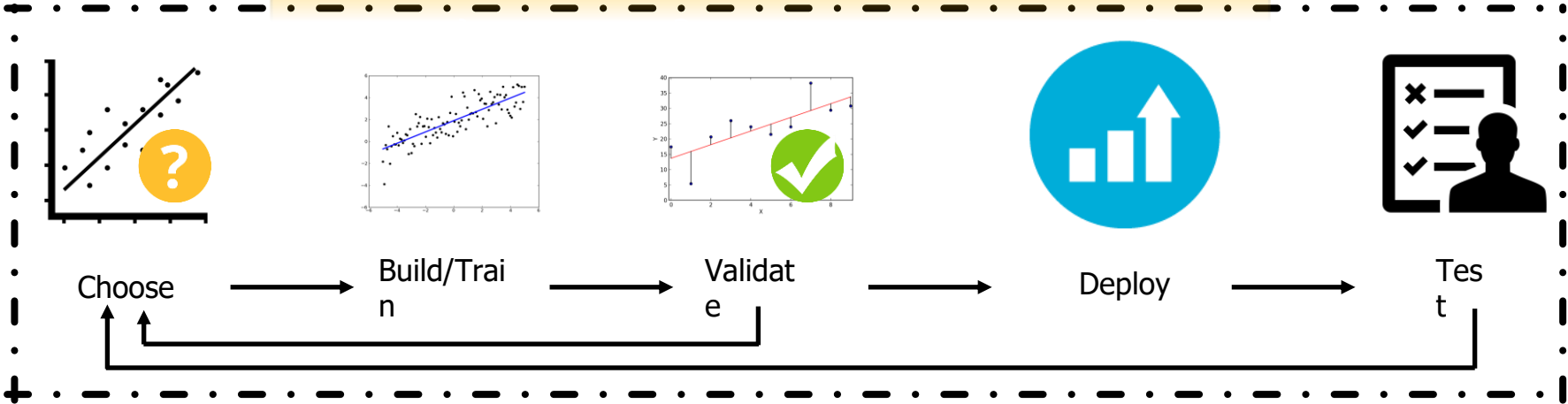
Teacher Used
Few Examples
in Class Room

Teacher Used
Different
Examples in
Exam Hall

Marks Will Be
Relatively Low

DATA SCIENCE LIFECYCLE

Core Artificial Intelligence / Machine Learning



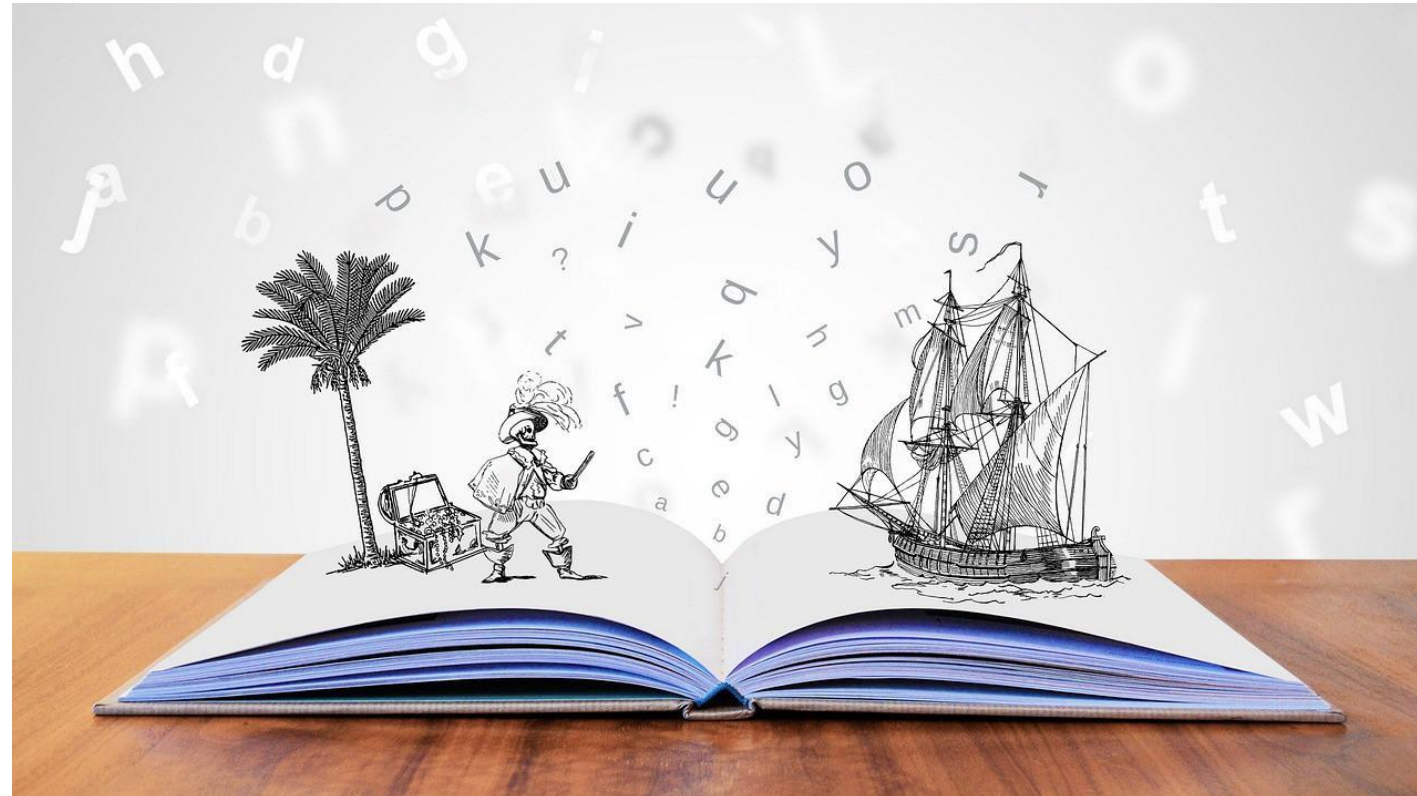
LEA Training



This Soldier Must be Sent back to Training Academy to Retrain Himself for Changing Tactics of Criminals.



DATA SCIENCE LIFECYCLE



Business Owner



Question

Data Engineer/Data Scientist



Acquire



Ingest/ETL

Data Analyst/Data Scientist



Wrangling

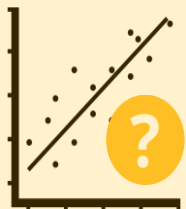
Exploratory
Data Analysis
(EDA)



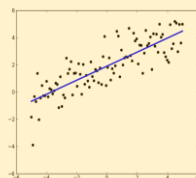
Visualize

Data Scientist/Machine Learning Engineer

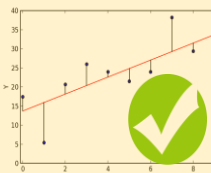
Modelling



Choose



Build/Train



Validate

DevOps/Data Scientist



Deploy

Business Owner/ Data Scientist



Test

Business
Owner/
Data Scientist/
Data Analyst



Comparison of Data Science Skills

	Data Analyst	ML Engineer	Data Engineer	DevOps Engineer	Data Scientist
Programming Tools	●	●	●	●	●
Data Visualization and Communication	●	●	●	●	●
Data Intuition	●	●	●	●	●
Statistics	●	●	●	●	●
Data Wrangling	●	●	●	●	●
Machine Learning	●	●	●	●	●
Software Engineering	●	●	●	●	●
Multivariable Calculus and Linear Algebra	●	●	●	●	●

● Very Important

● Somewhat Important

● Not that Important

DATA SCIENCE - USE CASES



**Churn
Prevention**



**Customer
Lifetime Value**



**Customer
Segmentation**



**Next Best
Action**



**Predictive
Maintenance**



**Product
Propensity**



**Quality
Assurance**



**Risk
Modeling**



**Sentiment
Analysis**



**Up-selling and
Cross-selling**



**Predicting
Buying Behavior**



**Fraud
Detection**



**Healthcare
Diagnosis**



**Content
Recommendation**



**Virtual
Assistance**



**Campaign
Management**



**Volume
Prediction**

Anaconda Installation

Other alternatives

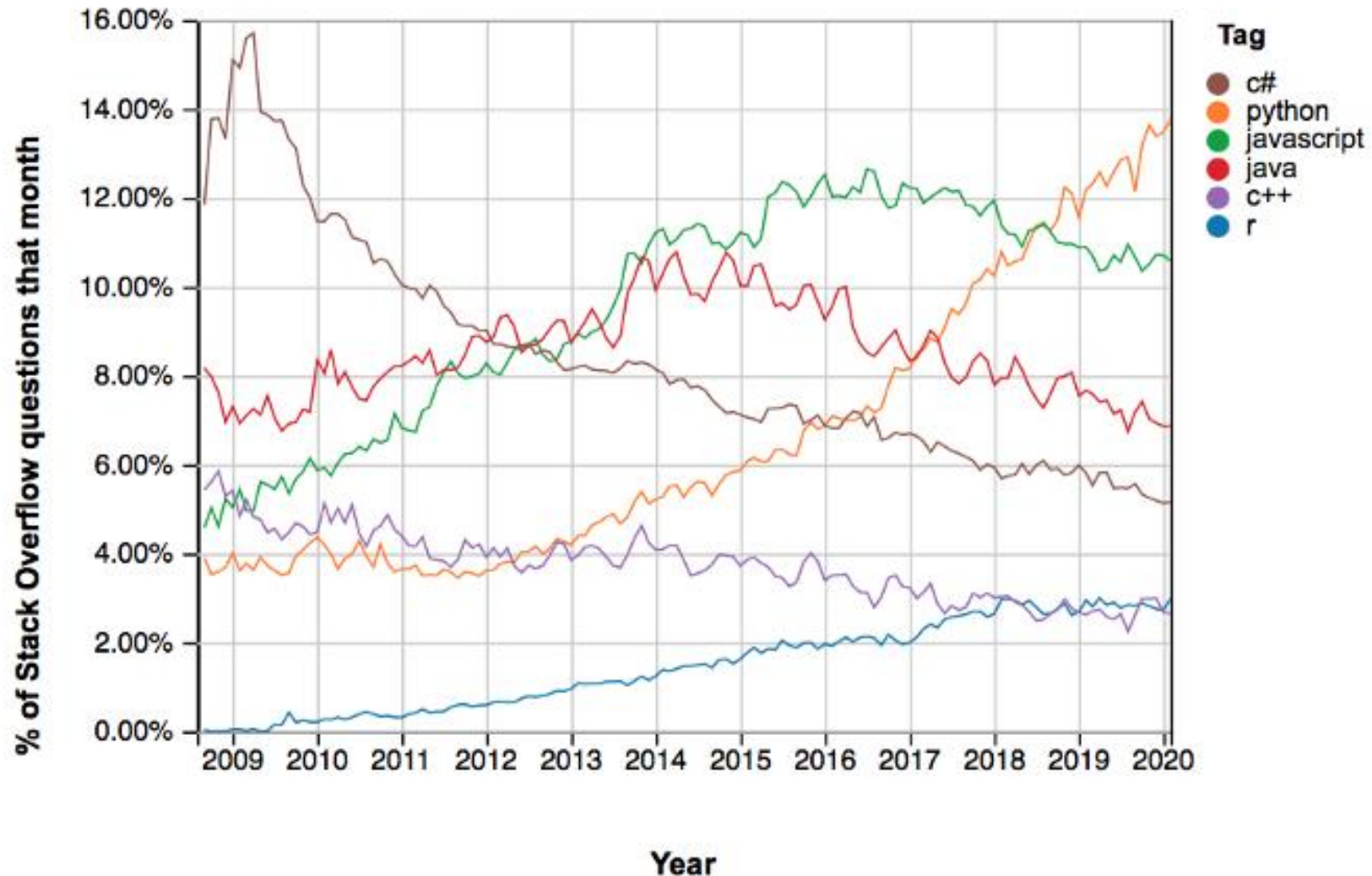
- *Google Colab*
- *Kaggle*
- *CoCalc*

Jupyter Overview

Why Python?



1. Popular



- **Large user community**
- **Well-maintained libraries**
- **Online guidance (Stack Overflow)**

2. Easy

Python Code :

```
print "Hello world"
```

C++ Code :

```
#include <iostream.h>
void main()
{
    cout << "Hello world" << endl;
}
```

Why do people like it ?? 

- **Code is intuitive and expressive (compare C++)**
- **Suited to large quantities of data**

3. Ecosystem

Data science work-flow

