



FREE AI Classes in Every City

April 25 – May 28 2020 | FREE 5-Day Machine Learning Immersion in 80+ Cities



## DAY 1: FREE AI Class in Every City 2022

### INTRODUCTION TO ARTIFICIAL INTELLIGENCE

- Introduction to Data Scientists Network and AI Invasion
- Artificial Intelligence, Machine Learning, and Data Science.
- Machine Learning Use Cases/Applications
- Introduction to platforms (Anaconda/Jupyter Notebook) and Google Colab
- Brief warmup with python

All Datasets to be used can be found [here](#)

Code:

[https://colab.research.google.com/drive/1O4CZOTzBTEFoTX2LAhfkw4GFhGCQSGF  
?usp=sharing](https://colab.research.google.com/drive/1O4CZOTzBTEFoTX2LAhfkw4GFhGCQSGF?usp=sharing)

### Introduction to Data Scientists Network and AI Invasion



A hand reaches towards a glowing blue network of connections forming a sphere. Inside the sphere is the DSN logo. Below the sphere, the text "DSN Innovations: *Community, Product and Partnership model*" is displayed, followed by the subtitle "Powering AI innovations for the 2 billion people in emerging market".

## About DSN

---

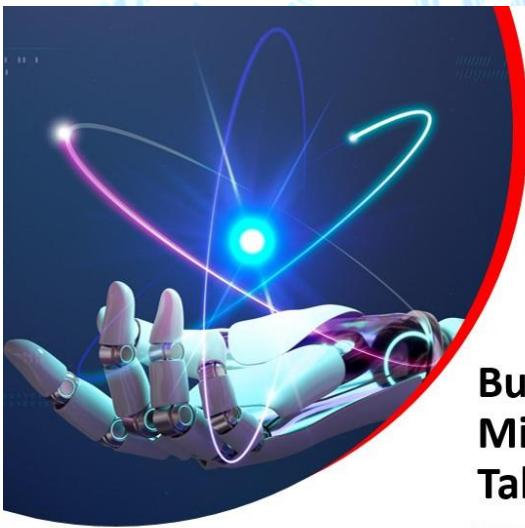


Our commitment is to raise 1 million AI talents and build AI solutions that improve the quality of life and wellbeing of 2 billion people in emerging market.



Our renewed focus will see us deliver on our vision through a renewed focus on 3 core areas;

- *Community for learning and research*
- *Product development for social impact*
- *Partnerships for solution delivery*



**Building one  
Million African AI  
Talents in 10 Years**

---



## DSN– Evolution (2016 to 2021)

---

Learning and Community



Research and Social Good Applications



Corporate support and Solution deployment



Start-ups/ventures, AI- vertical partnership and AI for National Development




---

**DSN won the best Academic Poster at the 21st edition of the global Economic and Computation Conference**

---

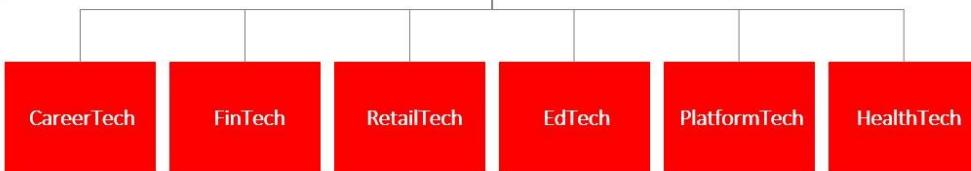


## DSN is the only African finalist at the 2021 XPrize competition on building advanced AI algorithm to address COVID19



### Dual business model for global impact through DSN AI Innovations

Building cross-country AI learning and research communities and developing AI solutions that enhance the quality of life and wellbeing of 2 billion people in emerging market

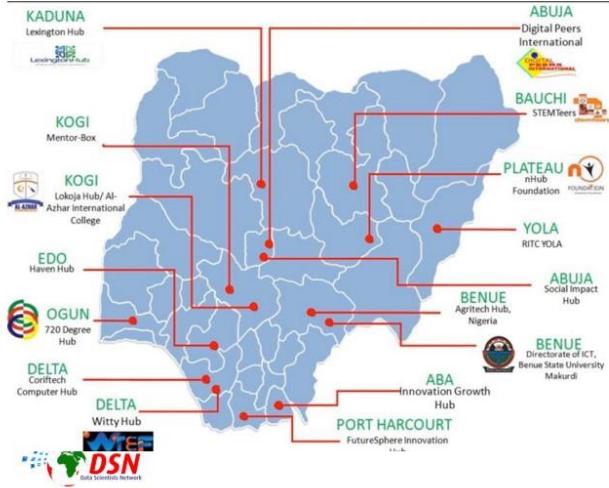


## Employability-oriented trainings for students-

Pre-NYSC , Higher Institutions and Post-NYSC

### AI Everywhere 6-weeks Class

*Introductory Python Programming and Machine learning*



1,000+ students on the #AI\_in\_Cities program in **17 Cities**



## Employability-oriented trainings for students-

Pre-NYSC , Higher Institutions and Post-NYSC

### AI Everywhere 6-weeks Class

*Introductory Python Programming and Machine learning*



## We also run the biggest pan Nigeria AI learning platform with 5-day introductory machine learning class in Cities

As part of our plan to democratize the learning of Artificial Intelligence and Data Science, Data Science Nigeria kicks off a one-week (30 hours) of introductory Machine Learning in 30 Cities.





**AI+ INVASION 2019**

**30 HOURS OF INTRO MACHINE LEARNING IN 30 CITIES**

MARCH 18 TO APRIL 17, 2019

Coming to a city nearest to you....

- Each city session will run for a maximum of one week and will focus on Python programming and Introduction to Machine Learning
- AI+ Knowledge Box will be distributed at these meet-ups
- AI+ Club will also be formally set up

For more information:

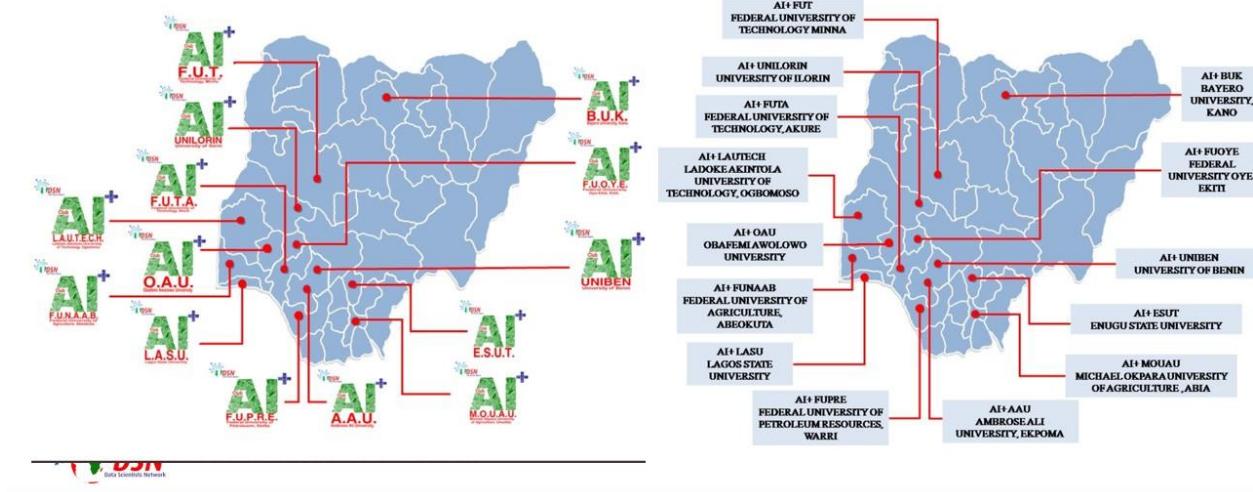
- DataScienceNigeria
- DataScienceNigeria.org
- https://poco.ai/cup
- @www.datascienceinigeria.org
- info@datascienceinigeria.org

The more information on the other events, please click [www.datascienceinigeria.org/2019plans](http://www.datascienceinigeria.org/2019plans)



## 18 universities and 23 cities

To Sustain The Amazing Impacts that are made via our Cohort Trainings, we have the AI+ Clubs on Campuses and Cities. This runs continuously year-round as Meet-Ups.



## AI+ Community

---



## AI+ Community

---





## Artificial Intelligence Bootcamp – the biggest and most competitive inter- campus learning platform

Driving industry-relevant skills by exposing students to industry data e.g.  
Hackathon with max.ng

**Can you HACK the future?**

**CHALLENGE**

Practical and Bold Machine Learning Applications/Solutions that can add immediate value to a company like

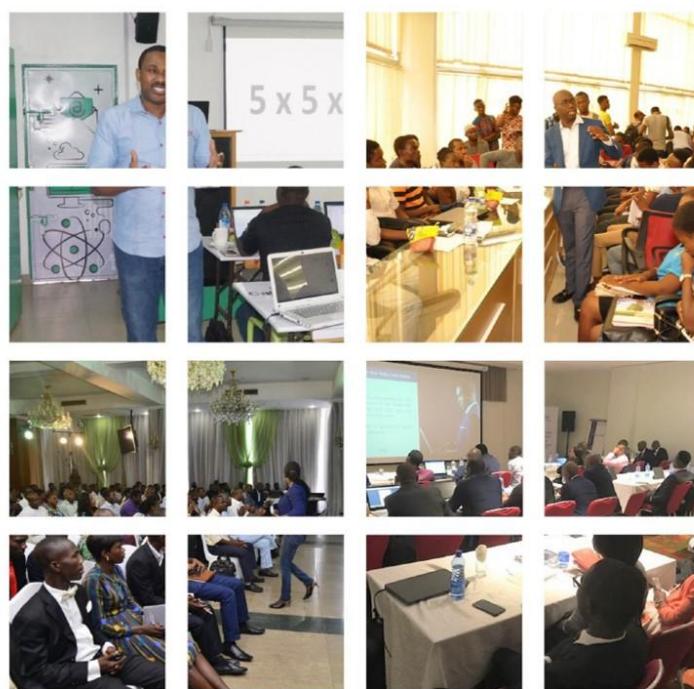
**max.ng**

Reward: cash prize and an interview with SeqHub for an Opportunity for a 3 month contract position

**DSN** **max.ng** **SeqHub**



### Companies We Have Trained



### Driving industry level knowledge level -AI for Executive



A one-day masterclass to demystify the whole concept of Artificial Intelligence for executives, business leaders and strategic decision-makers.





AI+ Club for School- AI trainings Primary and Secondary School Students

---

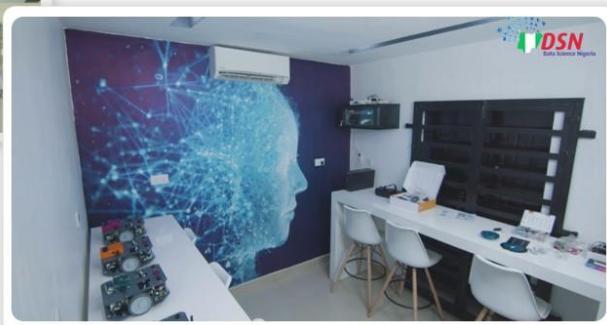
## AI+ Club FOR SCHOOLS

Preparing students to thrive in the  
future of work with Artificial Intelligence  
& Python Programming



## DSN has a dedicated AI Research Lab Hub in Yaba, Lagos

---



## Thank You

### Data Scientists Network Foundation

-  Website: [www.datasciencenigeria.org](http://www.datasciencenigeria.org)
-  Twitter: [@Datasciencenig](https://twitter.com/Datasciencenig)
-  Instagram: [@Datasciencenigeria](https://www.instagram.com/datasciencenigeria/)
-  Facebook: [facebook.com/datasciencenig](https://facebook.com/datasciencenig)
-  YouTube: <https:// goo.gl/Vcjyp>
-  Annual Report: <https://bit.ly/DSNAnnualreport-2020-2021>



## JOIN OUR COMMUNITY:

<https://www.datasciencenigeria.org/ai-communities/>

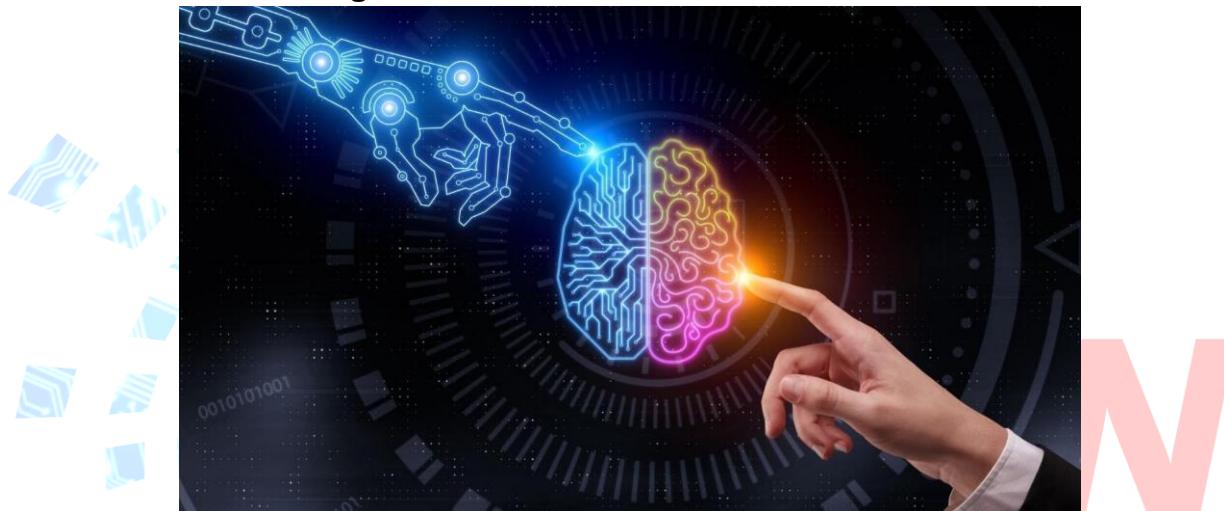


## Artificial Intelligence, Machine Learning, and Data Science

Artificial intelligence (AI) is undoubtedly the hottest buzzword today and for all the good reasons. Over the past decade, AI has truly become a factor of production, with the potential to introduce new sources of growth and change the way work is done across industries.

The function and popularity of Artificial Intelligence are soaring by the day. Artificial intelligence is the ability of a system or a program to think and learn from experience. AI has significantly evolved over the past few years and has found its applications in almost every sector.

### What is Artificial Intelligence?



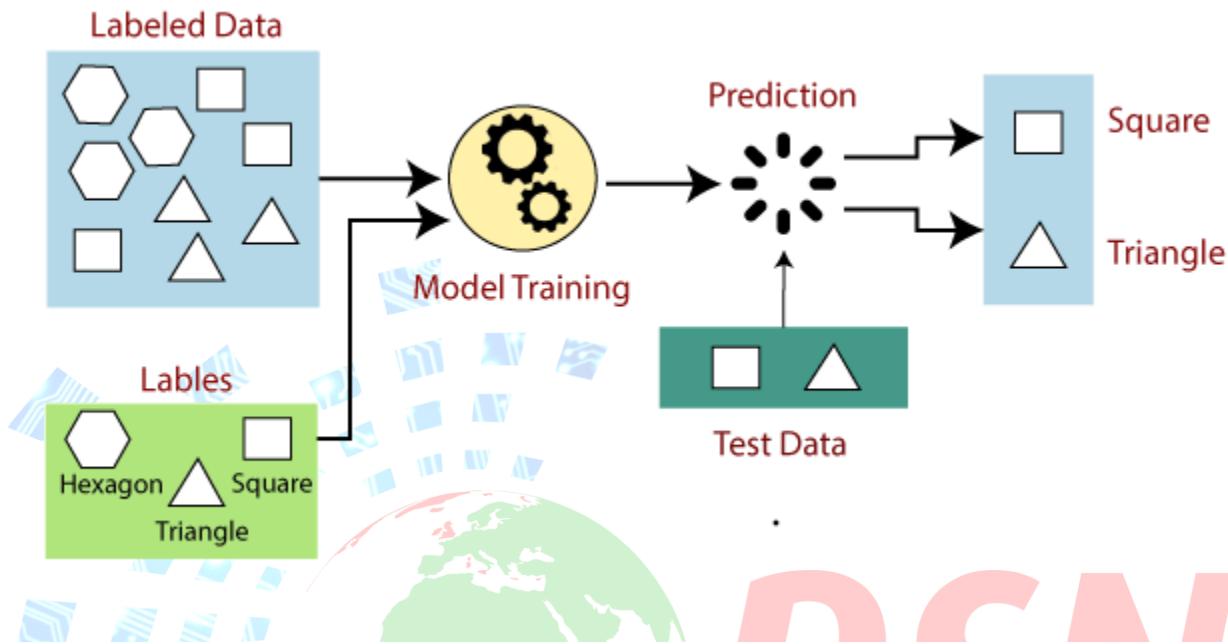
Artificial intelligence is a technique for having a computer or machines think intelligently in the same way that a human brain does. AI is achieved through evaluating the cognitive process and researching the patterns of the human brain. AI is a combination of Machine Learning techniques and Deep Learning.

### What Is Machine Learning?

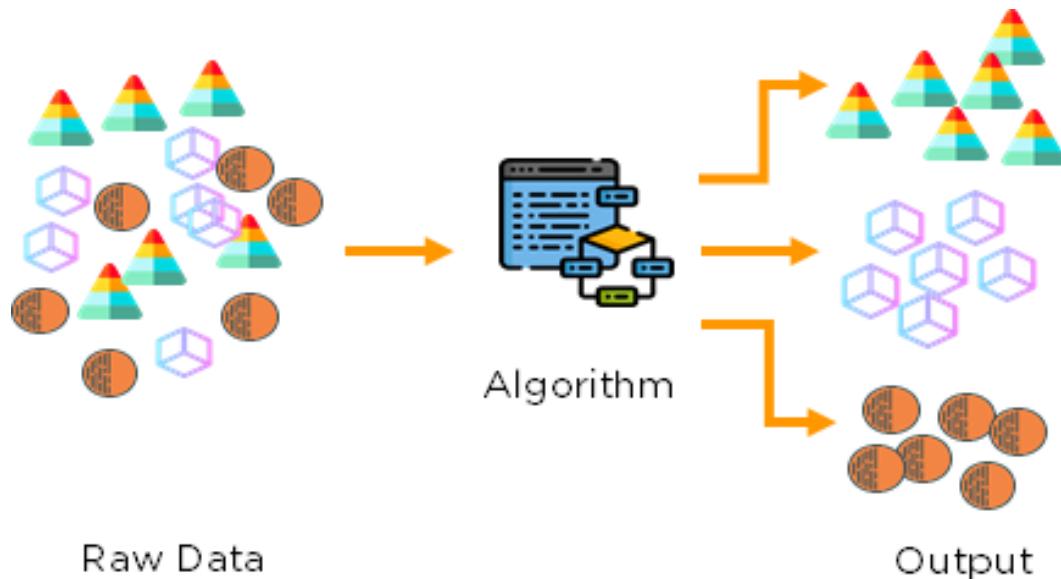
Machine learning is the subset of artificial intelligence that involves the study and use of algorithms and statistical models in computer systems to perform specific tasks without human interaction. Machine learning models rely on patterns and inference instead of manual human instruction. Almost all tasks that can be completed with a data-defined pattern or set of rules can be done with machine learning. Machine learning algorithms can be subdivided into supervised and unsupervised learning.

**Supervised learning:** In supervised learning, you train your model on a labelled dataset which means we have both raw input data as well as its results. We split our

data into a training dataset and a test dataset where the training dataset is used to train our network whereas the test dataset acts as new data for predicting results or to see the accuracy of our model. Hence, in supervised learning, our model learns from seen results the same as a teacher teaches his students because the teacher already knows the results. Accuracy is what we achieve in supervised learning as model perfection is usually high.

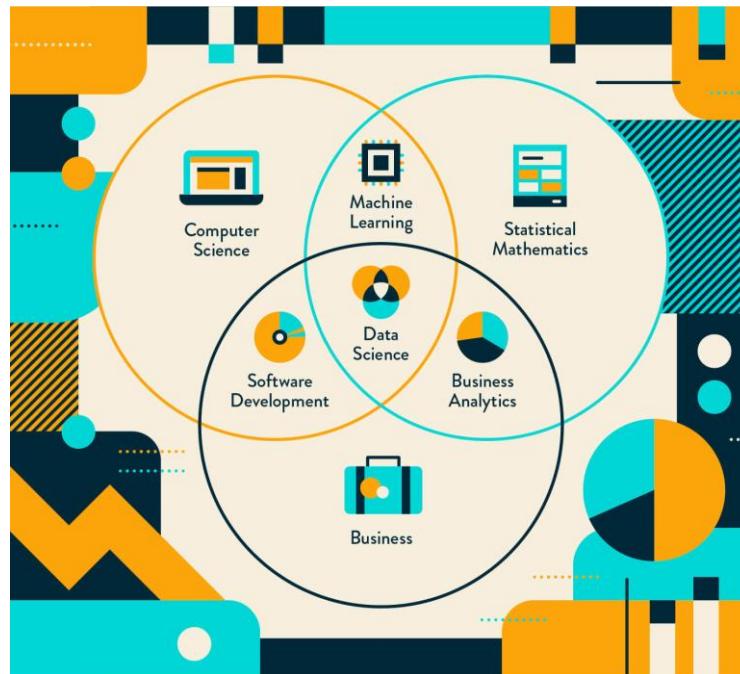


**Unsupervised learning:** In unsupervised learning, the information used to train is neither classified nor labelled in the dataset. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabelled data. The main task of unsupervised learning is to find patterns in the data. Once a model learns to develop patterns, it can easily predict patterns for any new dataset in the form of clusters. The system doesn't figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabeled data.



### What is Data Science?

Data science is an [interdisciplinary](#) field that uses scientific methods, processes, algorithms and systems to extract [knowledge](#) and insights from structured and [unstructured data](#), and apply knowledge and actionable insights from data across a broad range of application domains. Data science is related to [data\\_mining](#), [machine learning](#) and [big data](#).



### Artificial Intelligence Applications

The integration of AI tools with various software and devices is digitizing the world. We will see the applications of Artificial Intelligence in robotics, healthcare, defence, and many other sectors below:

- **AI in Robotics**

The field of robotics involves the designing and creation of automated machines or robots in such a way that they possess the ability to perform tasks on their own. Nowadays, robots are becoming more and more advanced and efficient in accomplishing tasks. This is due to the Artificial Intelligence tools and techniques that are specially designed for the field of robotics. Advanced robots consist of sensors, high-definition cameras, voice recognition devices, etc.

These robots are capable of learning from their past mistakes and experience and can adjust the algorithms according to the environment. Artificial Intelligence is an extremely useful tool for robotic applications. When AI is combined with advanced devices, it can help in optimizations. It is helpful in enhancing the complex manufacturing process in industries such as aerospace.

Sample: <https://www.youtube.com/watch?v=XuGJqajHAHo>

- **AI in Defence**

Defence is one of the most crucial sectors where Artificial Intelligence is contributing to nation-building. Security systems can be vulnerable to attacks by hackers, these attackers would do so to sell private information which can prove to be detrimental to any country. This is where the involvement of Artificial Intelligence proves to be of great use. The analysis of large amounts of data becomes easy with the help of Artificial Intelligence. Tools powered by AI can help find suspicious activity over the system and keep track of the security of an organisation/nation database. Any alterations in the database by an unknown source are immediately tracked down for action.

Sample: <https://www.analyticsinsight.net/role-of-artificial-intelligence-in-defence/>

- **AI in Transport**

Artificial Intelligence in the transport industry has completely changed the era of travelling. As the competition in the travel industry is high, there is a need to analyze all the factors that influence the travel business. These factors are price, seasons, festivals, the number of travellers, etc.

With the help of predictive analytics, the software can analyze data related to these factors that impact the cost of transport. Tools powered by AI can help perform predictive analytics efficiently on the data, like predicting the best prices in specific routes. Another application of Artificial Intelligence in transportation is route optimization. Businesses like Uber and Bolt can use Artificial Intelligence in their app to show optimized paths thereby moving consumers faster from point A to B.

Sample: [https://www.youtube.com/watch?v=s\\_Ze2o8ixNM](https://www.youtube.com/watch?v=s_Ze2o8ixNM)

- **AI in Healthcare**

When it comes to healthcare, AI never lags behind. Most healthcare organizations rely on AI-based software for their day-to-day tasks. These tasks vary from patient diagnosis to hospital data management. Since large amount of data are generated by the healthcare industry per day, there emerges a need for AI-based advanced processors that can extract, manipulate, analyze, and draw some meaningful insights from this data.

AI and ML technologies are doing a fabulous job in the healthcare industry. The AI-based algorithms fed into the systems are capable enough to spot patterns much more efficiently than humans. AI-based devices also help measure real-time data such as blood pressure, heartbeat, body temperature, and many more.

Sample: <https://www.youtube.com/watch?v=ii-FfE-7C-k>

- **AI in Marketing**

Today, the marketing industry is revolutionized by the applications of Artificial Intelligence. Various industries such as e-commerce, e-learning, advertising, media, and entertainment use Artificial Intelligence to boost profitability. Suppose, you are searching for a product on Amazon. Along with the product, it will also show you the best sellers, similar products, varieties of the same product, and the 'Recommended for you' list of products.

AI-based algorithms understand the interests of customers and give recommendations to searches made. An example of such is Netflix.

Sample: [https://www.youtube.com/watch?v=FYMjXD3G\\_Y](https://www.youtube.com/watch?v=FYMjXD3G_Y)

- **AI in Automotive Industry**

The invention of self-driving cars has completely changed the world of automobiles. There are various companies developing self-driving cars such as Tesla, Google, Bosch, Nissan, Audi, Volvo, and many more. The self-driving cars are built using a combination of various technologies, and one of the majorly used technologies is Artificial Intelligence. A self-driving car uses sensors, cameras, voice detectors, and many other devices. It analyzes the surroundings by collecting data. The AI-enabled advanced systems used in the self-driving car will find an optimized path to the destination.

With the help of AI, we can address problems such as traffic accidents, respond to natural disasters, etc.

Sample: [https://www.youtube.com/watch?v=VGGHCH0T\\_SQ](https://www.youtube.com/watch?v=VGGHCH0T_SQ)

## Other Use Cases

- **Voice assistants**

This consumer-based use for machine learning applies mostly to smartphones and **smart home** devices. The voice assistants on these devices use machine learning to understand what you say and craft a response. The **machine\_learning\_models** behind voice assistants were trained on human languages and variations in the human voice, because it has to translate what it hears into words and then make an intelligent, on-topic response.

- **Dynamic pricing**

This machine-based pricing strategy is most known in the **travel industry**. Flights, hotels, and other travel bookings usually have a dynamic pricing strategy behind them.

Consumers know that the sooner they book their trip the better, but they may not realize that the actual price changes are made via machine learning.

- **Email filtering**

This is a classic use of machine learning. Email inboxes also have a spam inbox, where your email provider automatically filters unwanted spam emails.

- **Product recommendations**

Amazon and other online retailers often list “recommended products” for each consumer individually. These recommendations are based on past purchases, browsing history, and any other behavioral information they have about consumers. This is a great way for online retailers to provide extra value or **upsells** to their customers using machine learning.

- **Personalized marketing**

Marketing is becoming more personal as technologies like machine learning gain more ground in the enterprise. Now that much of marketing is online, marketers can use characteristic and behavioral data to segment the market.

- **Process automation**

There are many processes in the enterprise that are much more efficient when done using machine learning. These include analyses such as risk assessments, demand forecasting, **customer churn prediction**, and others. Machine learning for process automation alleviates the timeliness issue for enterprises. Machine learning can even help with customer loyalty analyses like **sentiment analysis**.

- **Fraud detection**

Banks use machine learning for fraud detection to keep their consumers safe, but this can also be valuable to companies that handle credit card transactions. Fraud **detection** can save money on disputes and chargebacks, and machine learning models can be trained to flag transactions that appear fraudulent based on certain characteristics.

## Introduction To Platforms

### How to Install and Setup Anaconda

- Anaconda: <https://www.anaconda.com/products/individual>

- Pandas: <https://pandas.pydata.org/>
- Jupyter Notebook Installation without Anaconda: <https://jupyter.org/>

## Introduction to Google Colab

Colab, or "Colaboratory", allows you to write and execute Python in your browser, with

- Zero configuration required
- Access to GPUs free of charge
- Easy sharing



Whether you're a student, a data scientist or an AI researcher, Colab can make your work easier. Visit the link below for how to get started

<https://colab.research.google.com/>

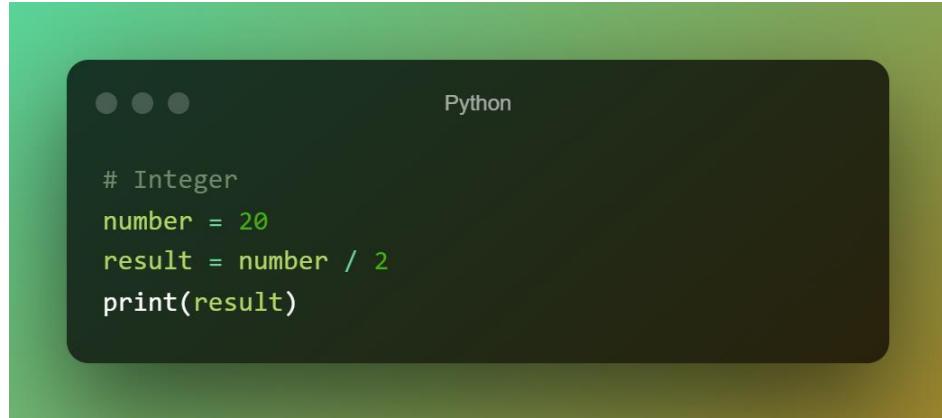
## Brief Warmup With Python

Python is a widely-used general-purpose, high-level programming language. It is a programming language that lets you work quickly and integrate systems more efficiently.

## Data Types In Python And Their Implementation

- Integer: Integers are numerical (number) values that are positive or negative whole numbers without decimals

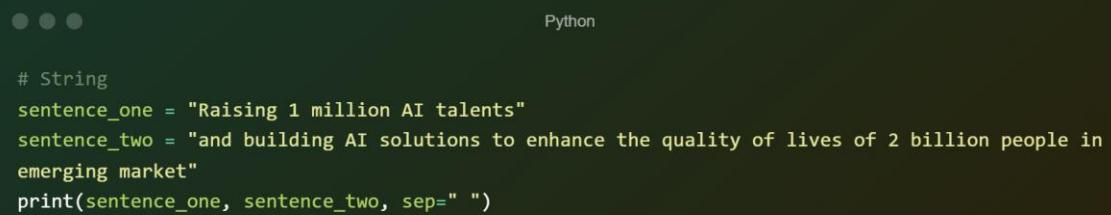
A large, semi-transparent watermark for the Data Scientists Network (DSN) is positioned on the right side of the slide. It features the letters "DSN" in a large, bold, red font. Below "DSN", the words "Data Scientists Network" are written in a smaller, blue, sans-serif font. The background of the watermark is a light blue color.



```
Python

# Integer
number = 20
result = number / 2
print(result)
```

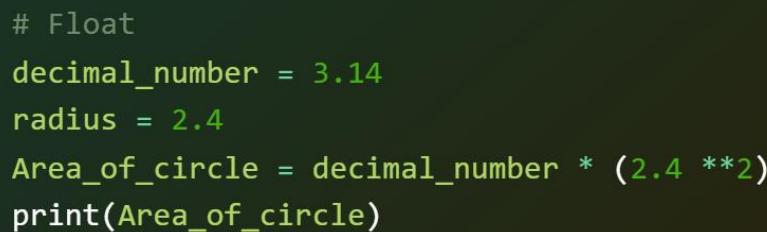
- String: Characters or texts are known as strings in Python. A string object must be inside a single or double quote. Otherwise, Python will throw an error.



```
Python

# String
sentence_one = "Raising 1 million AI talents"
sentence_two = "and building AI solutions to enhance the quality of lives of 2 billion people in
emerging market"
print(sentence_one, sentence_two, sep=" ")
```

- Float: Float (floating point) numbers are numbers with a decimal point.



```
Python

# Float
decimal_number = 3.14
radius = 2.4
Area_of_circle = decimal_number * (2.4 **2)
print(Area_of_circle)
```

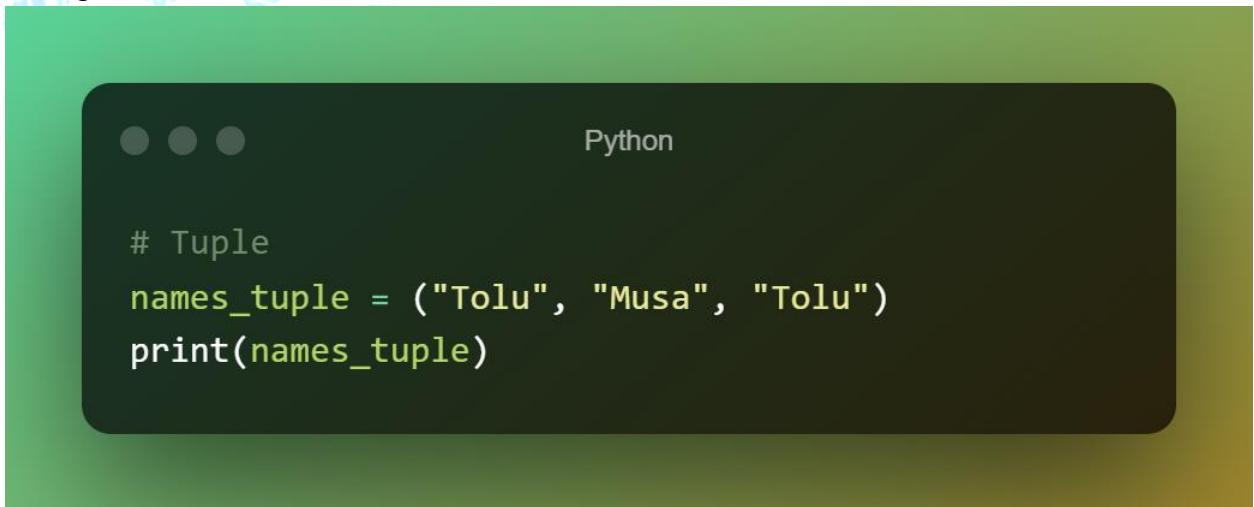
- List: Lists are used to store multiple items in a single variable.



```
Python

# List
names = ["Tolu", "Uche", "Musa"]
names.append("John")
print(names)
```

- Tuple: A tuple in Python is similar to a list. The difference between the two is that we cannot change the elements of a tuple once it is assigned whereas we can change the elements of a list.



```
Python

# Tuple
names_tuple = ("Tolu", "Musa", "Tolu")
print(names_tuple)
```

- Sets: A Sets in Python is similar to a list. The difference between the two is that set does not allow for duplicate elements whereas duplicate elements are allowed in a list.

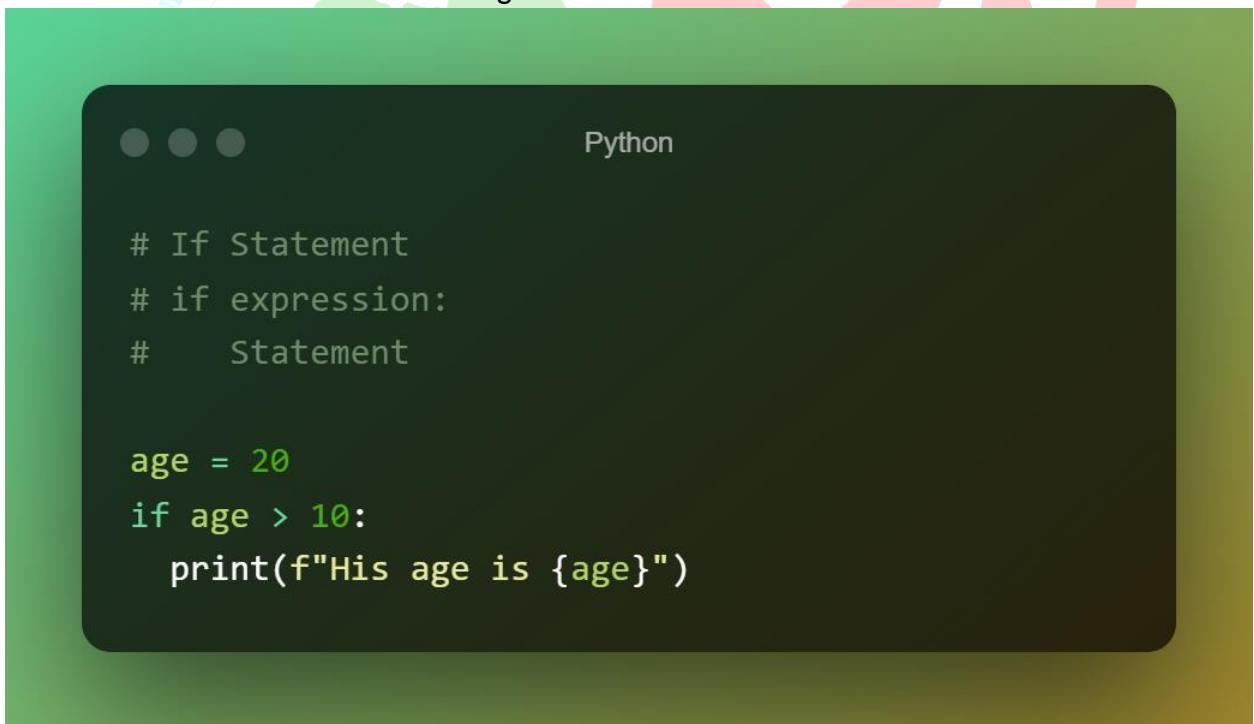


A screenshot of a dark-themed Python code editor. At the top right, it says "Python". Below that is a snippet of Python code:

```
# Set
names_set = {"Ade", "Bayo", "Ali", "Ngozi"}
names_set.add("Baki")
names_set.add("Ade")
print(names_set)
```

## Conditional Statement In Python

- The If Statement: The If statement is the most fundamental decision-making statement, in which the code is executed based on whether it meets the specified condition. It has a code body that only executes if the condition in the if statement is true. The statement can be a single line or a block of code.



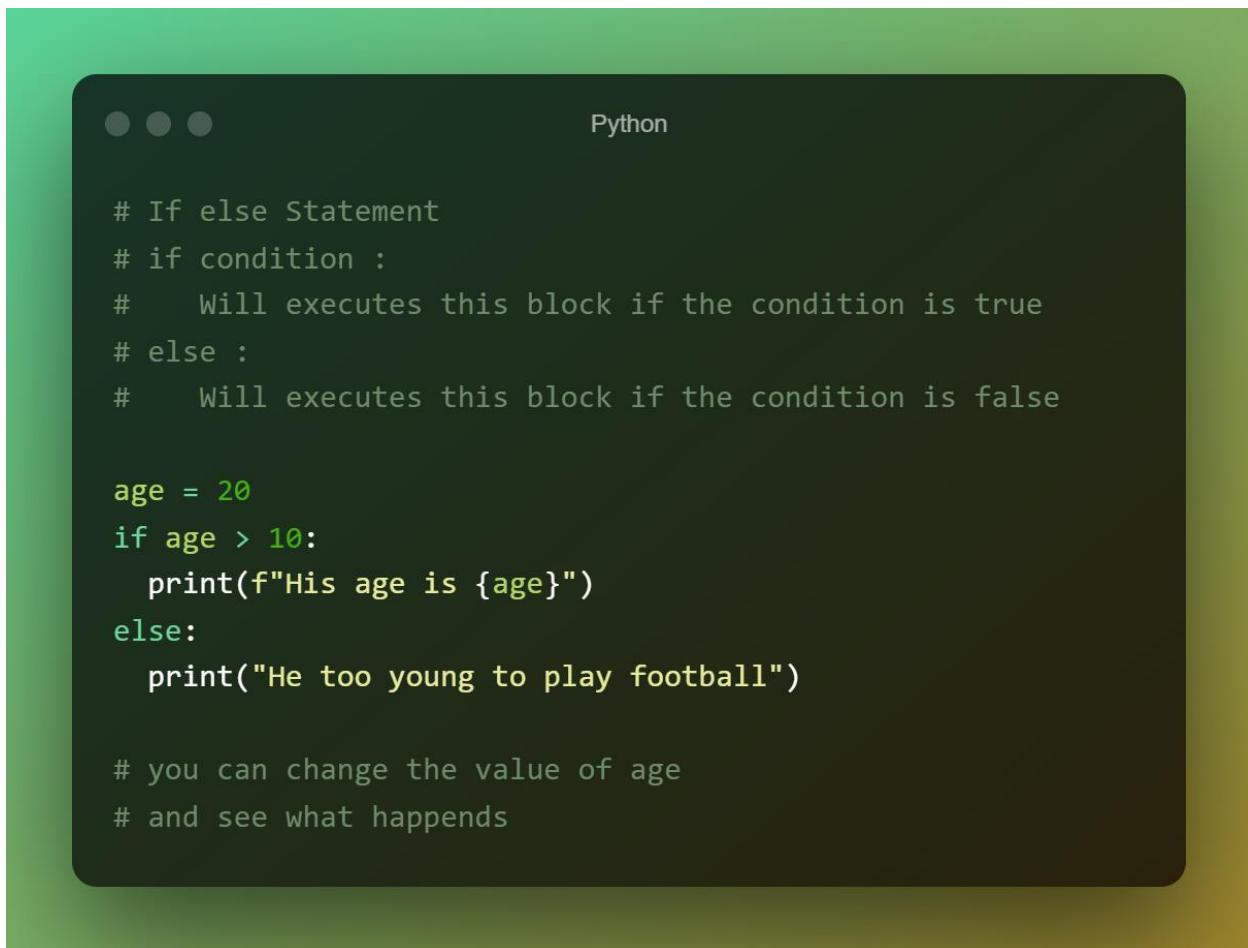
A screenshot of a dark-themed Python code editor. At the top right, it says "Python". Below that is a snippet of Python code:

```
# If Statement
# if expression:
#     Statement

age = 20
if age > 10:
    print(f"His age is {age}")
```

- The If else statement: This statement is used when both the true and false parts of a given condition are specified to be executed. When the condition is true, the

statement inside the if block is executed; if the condition is false, the statement outside the if block is executed.



A screenshot of a dark-themed Python code editor. The code demonstrates an if-else statement:

```
# If else Statement
# if condition :
#     Will executes this block if the condition is true
# else :
#     Will executes this block if the condition is false

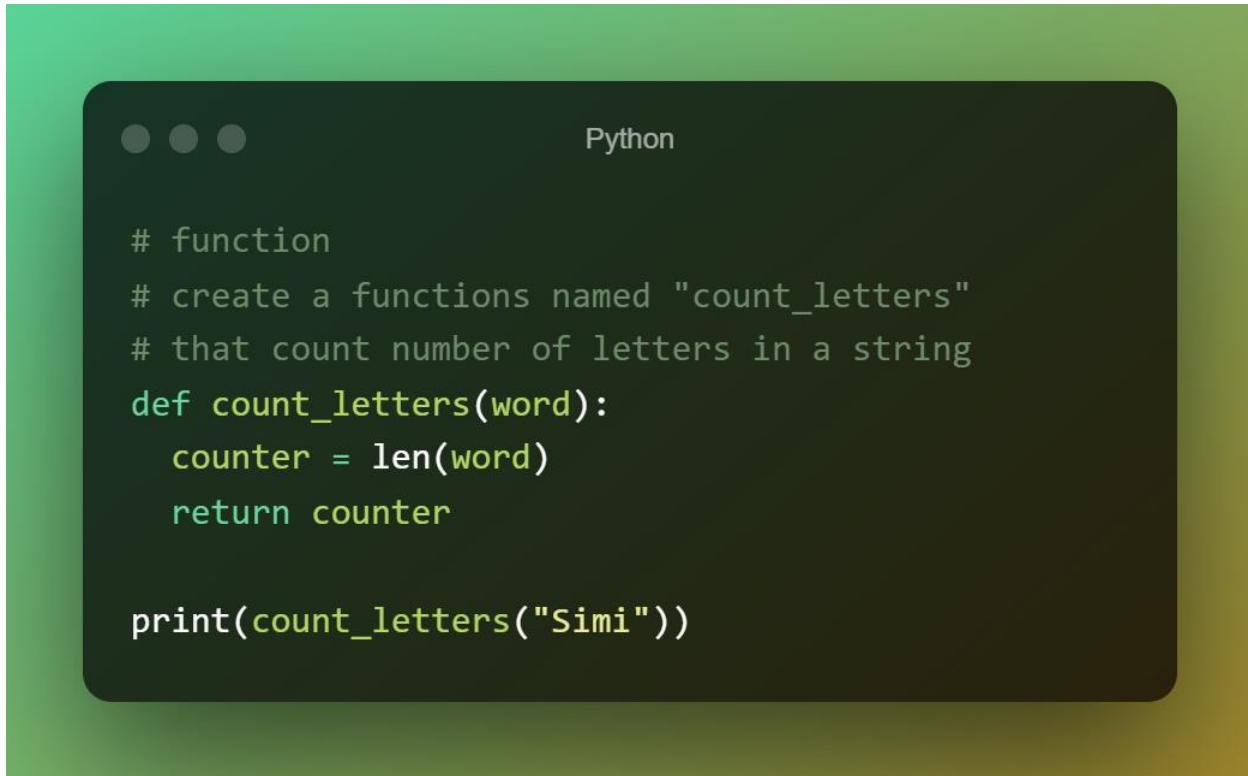
age = 20
if age > 10:
    print(f"His age is {age}")
else:
    print("He too young to play football")

# you can change the value of age
# and see what happens
```

## Functions In Python

A function is a block of code that only executes when called. You can pass data, known as parameters, into a function and the function can return data as a result.

Syntax of a function statement in python



```

# function
# create a functions named "count_letters"
# that count number of letters in a string
def count_letters(word):
    counter = len(word)
    return counter

print(count_letters("Simi"))

```

## References

1. *Mitchell, Tom* (1997). [Machine Learning](#). New York: McGraw Hill. [ISBN 0-07-042807-7](#). [OCLC 36417892](#).
2. Machine Learning Use Cases: <https://algorithmia.com/blog/machine-learning-use-cases>
3. ["1. Introduction: What Is Data Science? - Doing Data Science \[Book\]"](https://www.oreilly.com/programming/topics/introduction-to-data-science/). [www.oreilly.com](https://www.oreilly.com). Retrieved 3 April 2020.
4. *Top Applications of Artificial Intelligence*: [Top Applications of Artificial Intelligence\(AI\) in 2021 \(intellipaat.com\)](https://intellipaat.com/)
5. Artificial Intelligence Tutorial for Beginners: [Artificial Intelligence Tutorial for Beginners \[Updated 2020\] \(simplilearn.com\)](https://www.simplilearn.com/tutorials/artificial-intelligence-tutorial-for-beginners)

## Activity

- List other industries where AI can be utilised
- Write a python function called “two\_sum” that takes in two arguments(integer) and calculates the sum of two number
- Write an if-else statement that print “This is a prime number” if the number is prime, or print out “This is an Ordinary number” if the number is not prime



## DAY 2: FREE AI Class in Every City 2022

### ANALYZING AND UNDERSTANDING DATA

- Introduction to the AI Invasion Project
- The machine learning development workflow
- Data gathering, cleaning and preparation
- Exploratory data analysis

Code:

<https://colab.research.google.com/drive/1lkpE0gTB7oR5ngoHOF2taPtyaxCkRwEe?usp=sharing>

### AI Invasion Project

The AI Invasion project is the last project you must complete before graduating from this program. The goal of this project is to reinforce what you've learned, put some of the abilities you've gained into practice, and have a project to show potential employers in your portfolio.

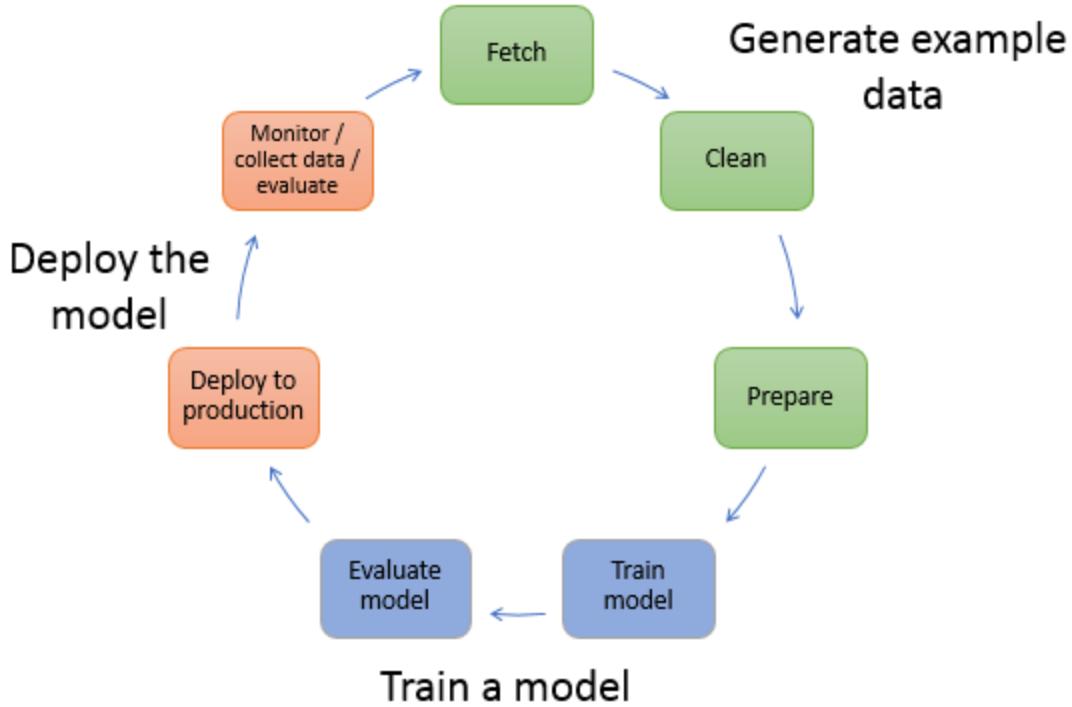
Skills needed to complete this project include

1. Knowledge of Python programming
2. Understanding of data preprocessing
3. Familiarity with some machine learning libraries like Pandas, Scikit-learn, e.t.c.
4. A basic understanding of the Zindi platform and how to submit predictions

Do not worry, you will learn these skills and more in this course. The project will focus on building a machine learning model that can predict the prices of used cars

### Machine Learning Development Workflow

Machine learning workflows define the steps initiated during a particular machine learning project. The typical phases include data collection, data pre-processing, data segmentation, model training and refinement, evaluation, and deployment to production. Kindly note that machine learning workflows vary by project.



## Data Collection

Data collection is one of the most important stages of machine learning workflows. During data collection, you are defining the potential usefulness and accuracy of your project with the quality of the data you collect.

To collect data, you need to identify your sources and aggregate data from those sources into a single dataset. This could mean streaming data from the Internet, downloading open-source data sets, or constructing a data lake from assorted files, logs, or media.

## Data Pre-Processing

Once your data is collected, you need to pre-process it. Pre-processing involves cleaning, verifying, and formatting data into a usable dataset. If you are collecting data from a single source, this may be a relatively straightforward process. However, if you are aggregating several sources you need to make sure that data formats match, that data is equally reliable, and remove any potential duplicates.

## Data Segmentation

This phase involves breaking processed data into three datasets—training, validating, and testing:

- Training set— used to initially train the algorithm and teach it how to process information. This set defines model classifications through parameters.
- Validation set— used to estimate the accuracy of the model. This dataset is used to finetune model parameters.
- Test set— used to assess the accuracy and performance of the models. This set is meant to expose any issues in the model.

## Training And Refinement

Once you have datasets, you are ready to train your model. This involves feeding your training set to your algorithm so that it can learn appropriate parameters and features used in classification.

Once training is complete, you can then refine the model using your validation dataset. This may involve modifying or discarding variables and includes a process of tweaking model-specific settings (hyperparameters) until an acceptable accuracy level is reached.

## Machine Learning Evaluation

Finally, after an acceptable set of hyperparameters is found and your model accuracy is optimized you can test your model. Testing uses your test dataset and is meant to verify that your models are using accurate features. Based on the feedback you receive you may return to training the model to improve accuracy, adjust output settings, or deploy the model as needed.

## Machine Learning Deployment.

Deployment of machine learning models, or simply, putting models into production, means making your models available to other systems within the organization or the web so that they can receive data and return their predictions.

## Data Gathering, Cleaning And Preparation

### Open Source Dataset

These are lists of some of the publicly available datasets you can work with as a data scientist

- AFRIFASHION40000: AFRIFASHION40000 is an openly available dataset of African fashion images generated using Generative Adversarial Networks (GANs) and created by Data Science Nigeria. Link: [https://bit.ly/DSN\\_AFRIFASHION40000](https://bit.ly/DSN_AFRIFASHION40000)
- Data.gov: It consists of a variety of datasets from US Government agencies. Domains include Education, Climate, Food, Chronic disease and so much more. Link: <https://www.data.gov/>
- UCI Machine Learning Repository: This site consists of datasets hosted by the University of California, Irvine. It has a collection of about 400+ datasets aimed toward the Machine Learning community. Link: <http://archive.ics.uci.edu/ml/index.php>
- Google Public Datasets: Google has hosted tons of datasets on Google Public Datasets which is basically their Cloud Platform. You can browse through their dataset collection using BigQuery. The first 1 Terabyte of queries you make are basically free. Link: <https://cloud.google.com/bigquery/public-data/>
- Datasets on Github: It hosts tons of awesome datasets. This GitHub boasts a variety of datasets such as Climate Data, Time Series data, Plane crash data etc. Feel free to dig in. Link: <https://github.com/awesomedata/awesome-public-datasets>
- For more datasets, you can check the link below: <https://medium.com/analytics-vidhya/top-100-open-source-datasets-for-data-science-cd5a8d67cc3d>

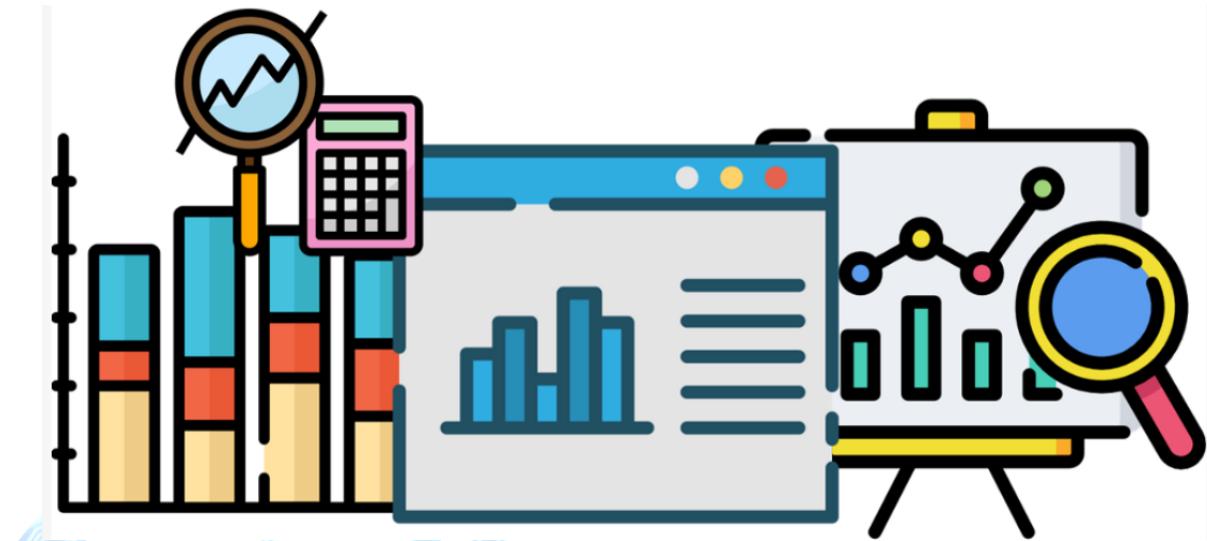
## The Data Cleaning Process

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. When combining multiple data sources, there are many opportunities for data to be duplicated or mislabeled. Most raw data, whether text, images, video – often even data stored in spreadsheets – is improperly formatted, incomplete, or downright dirty and needs to be properly cleaned and structured before you begin your analysis. Here are some steps to make sure your data is clean and ready to go.

- Remove irrelevant data
- Deduplicate your data
- Fix structural errors
- Deal with missing data
- Filter out data outliers
- Validate your data

## Exploratory Data Analysis

Exploratory Data Analysis refers to the critical process of performing initial investigations on data so as to discover patterns, spot anomalies, test hypotheses and check assumptions with the help of summary statistics and graphical representations.



The importance of performing EDA are:

- Helps identify errors in data sets.
- Gives a better understanding of the data set.
- Helps detect outliers or anomalous events.
- Helps understand data set variables and the relationship among them

## References

*Data Cleaning Steps & Process to Prep Your Data for Success.* (2021, June 3).

MonkeyLearn. Retrieved April 11, 2022, from <https://monkeylearn.com/blog/data-cleaning-steps/>

Guide, S. (2021, November 11). *What is Exploratory Data Analysis? Steps and Market Analysis.* Simplilearn. Retrieved April 6, 2022, from <https://www.simplilearn.com/tutorials/data-analytics-tutorial/exploratory-data-analysis>

*Guide To Data Cleaning: Definition, Benefits, Components, And How To Clean Your Data.* (n.d.). Tableau. Retrieved April 6, 2022, from  
<https://www.tableau.com/learn/articles/what-is-data-cleaning>

Patil, P. (2018, March 23). *What is Exploratory Data Analysis? | by Prasad Patil*. Towards Data Science. Retrieved April 6, 2022, from  
<https://towardsdatascience.com/exploratory-data-analysis-8fc1cb20fd15>

### Activity

- Using Pandas download and load a dataset from one of the open-source datasets
- Perform EDA on the titanic dataset. You can download the dataset using  
`pd.read_csv(https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv)`





## DAY 3: FREE AI Class in Every City 2022

### SUPERVISED LEARNING: INTRODUCTION TO REGRESSION ANALYSIS

- What is Linear Regression
- How Linear Regression Works
- Steps involved in working on a regression problems
- Other machine learning algorithms

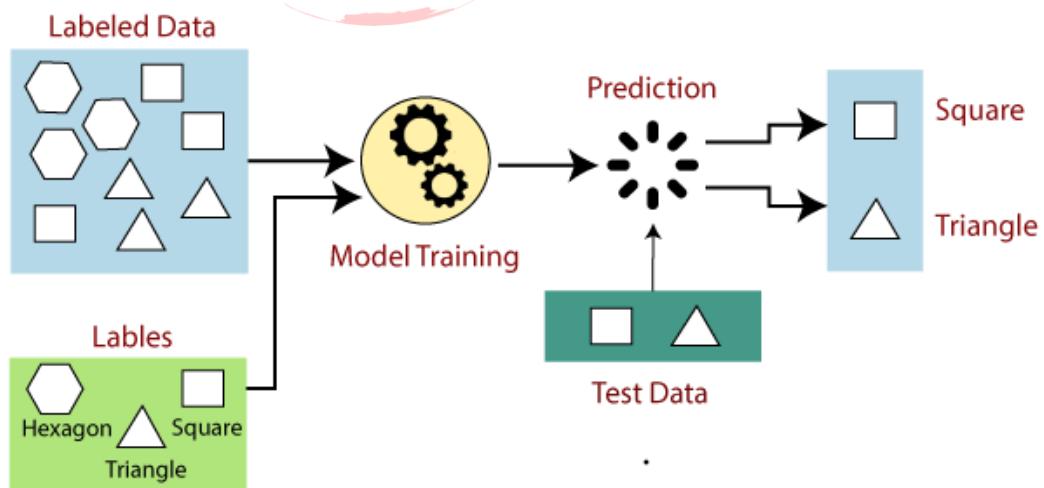
Dataset here: Download [AI Invasion In-Class Dataset.xlsx](#) here

Code:

<https://colab.research.google.com/drive/1cd4qOzrZ5o2ettmgfJ9RGMg133eSHtd2?usp=sharing>

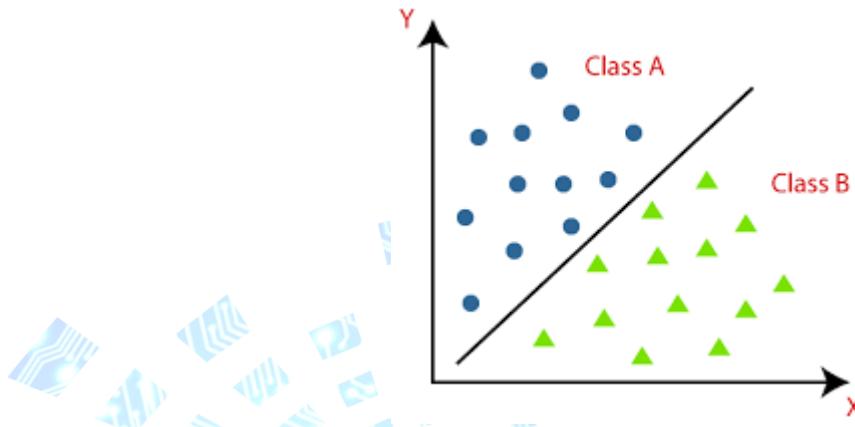
### Supervised Learning

Supervised Learning can be subdivided into Regression and Classification. Regression and Classification are both parts of a class of Machine Learning Algorithms called “Supervised Learning Algorithms”. Remember that Machine Learning is all about generating future occurrences/predictions of an event while making reference to past occurrences of the event in question. And for this particular class of Machine Learning(Supervised learning), we have labelled data with known examples with which the algorithm is trained to predict unknown data.



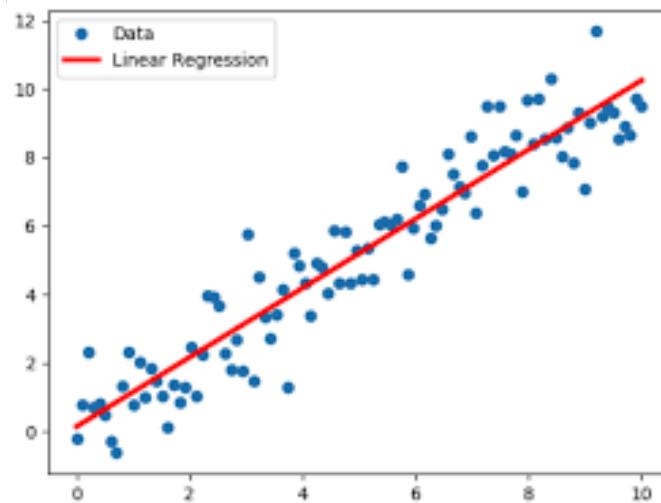
**Classification:** A classification classifier (model) is the type of model in which the output variable (i.e the Label) is Discrete. For example:

- Predict if the patient has cancer or not. [Label: Cancer and Not cancer (2)]
- Predict if an employee will leave or stay [Label: Leave and Stay (2)]
- Predict if an email is spam or not. [Label: Spam and Not Spam (2)]



**Regression:** A Regression model is the type of model in which the output variable is continuous. For example:

- Predict the price of a house in an area. [Label: Price of a house ]
- Predict how much rainfall will occur in an area. [Label: Rainfall measurement ]



Similarities:

- Both belong to the class of Machine Learning Algorithms called “Supervised Learning”.

- Both model a problem by learning a mapping function/relationship from the input(X) to the output(Y), using known examples.

Differences:

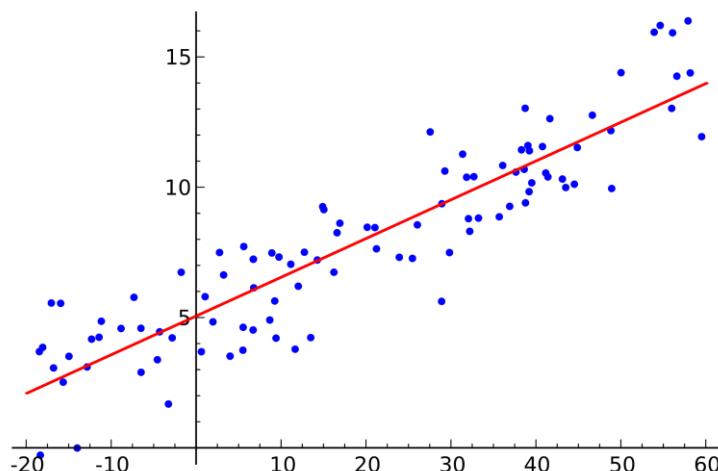
Property	Regression Algorithms	Classification Algorithms
Output/Target Variable	Predicts a continuous real value	Predicts a class/discrete values
Learning function	A line of best fit	A decision boundary
Evaluation metrics/Loss functions	Mean Squared Error, Mean Absolute Error, Root Mean Squared Error	Accuracy, F1 score, Area under the curve

Regression is a predictive statistical process where the model attempts to find the important relationship between dependent and independent variables. The goal of a regression algorithm is to predict a continuous number such as sales, income, and test scores.

First, we will have a quick introduction to building models in Python, and what better way to start than one of the very basic models, linear regression? Linear regression will be the first algorithm used and you will also learn more complex regression models.

## What is Linear Regression?

Linear Regression is considered the most natural learning algorithm for modelling data, primarily because it is easy to interpret and models efficiently for most natural problems. It belongs to the family of “Linear Models/Predictors” in machine learning (one of the most useful hypothesis spaces)

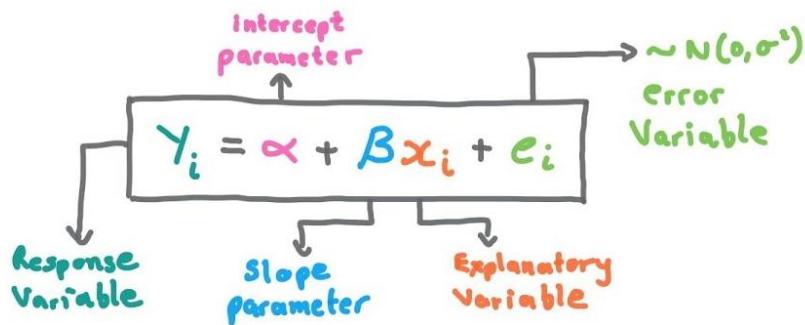


## The Mathematical Theory Of Linear Regression

Mathematically, the linear regression model can be defined by a dependent variable 'Y', also called the regressand and an independent variable(or set of independent variables) 'X', also called the regressor(s), and a sample space 'n'.

### Regression Model

- model response & explanatory variables
- model bivariate data points



### Steps Involved In Working On Regression Problems

Step#1 Importing the required libraries

Step#2 Loading the dataset

Step#3 Clean the dataset

- Remove irrelevant data
- Deduplicate your data
- Fix structural errors
- Deal with missing data
- Filter out data outliers
- Validate your data

Step#4 Perform data segmentation

Step#5 Load your data into the Linear Regression model i.e Train your model

Step#6 Make predictions

Step#7 Evaluate your model.

## Other Machine Learning Algorithms

- Decision Tree

- Support Vector Machine
- XGBoost

## References

Brownlee, J. (2017, December 11). *Difference Between Classification and Regression in Machine Learning*. Machine Learning Mastery. Retrieved April 6, 2022, from <https://machinelearningmastery.com/classification-versus-regression-in-machine-learning/>

*Introduction to Machine Learning for Beginners | by Ayush Pant.* (n.d.). Towards Data Science. Retrieved April 6, 2022, from <https://towardsdatascience.com/introduction-to-machine-learning-for-beginners-eed6024fdb08>

Singh, R. (n.d.). *Step-by-Step Regression Analysis. What is Regression Analysis? | by Great Learning*. Medium. Retrieved April 6, 2022, from <https://medium.com/@mygreatlearning/step-by-step-regression-analysis-f7e3e3ebf296>

## Activity

- Build a machine learning algorithm



## DAY 4: FREE AI Class in Every City 2022

### SUPERVISED LEARNING: INTRODUCTION TO REGRESSION ANALYSIS

- Building a prediction model
- Evaluating machine learning models
- Exporting predictions of machine learning models

Code: [https://colab.research.google.com/drive/1MAvq-FUA15ebYqlxpa\\_ze0VqRDWJqJ5s?usp=sharing](https://colab.research.google.com/drive/1MAvq-FUA15ebYqlxpa_ze0VqRDWJqJ5s?usp=sharing)

#### Building A Prediction Model

##### Class Activity

Build a Regression Machine Learning Model using Mama Tee restaurant dataset.

The objective of the regression task is to predict the amount of tip (gratuity in Nigeria naira) given to a food server based on total\_bill, gender, smoker (whether they smoke in the party or not), day (day of the week for the party), time (time of the day whether for lunch or dinner), and size (size of the party) in Mama Tee restaurant.

Dataset: Download [Tips.csv dataset here](#)

#### Evaluating Machine Learning Models

Model evaluation is very important in data science. It helps you to understand the performance of your model and makes it easy to present your model to other people. There are many different evaluation metrics out there but only some of them are suitable to be used for regression. This lesson will cover the different metrics for the regression model and the difference between them.

By the end of this lesson, you will understand which evaluation metrics to apply to your future regression model.

There are 3 main metrics for model evaluation in regression:

1. R Square/Adjusted R Square
2. Mean Square Error(MSE)/Root Mean Square Error(RMSE)
3. Mean Absolute Error(MAE)

## R Square/Adjusted R Square

R Square is calculated by the sum of the squared of prediction error divided by the total sum of the square which replaces the calculated prediction with mean. R Square value is between 0 to 1 and a bigger value indicates a better fit between prediction and actual value.

R Square is a good measure to determine how well the model fits the dependent variables. However, it does not take into consideration of overfitting problem.

$$R^2 = 1 - \frac{SS_{Regression}}{SS_{Total}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

## Mean Square Error(MSE)/Root Mean Square Error(RMSE)

MSE is calculated by the sum of the square of prediction error which is real output minus predicted output and then divide by the number of data points. It gives you an absolute number on how much your predicted results deviate from the actual number. You cannot interpret many insights from one single result but it gives you a real number to compare against other model results and help you select the best regression model.

Root Mean Square Error(RMSE) is the square root of MSE. It is used more commonly than MSE because firstly sometimes MSE value can be too big to compare easily. Secondly, MSE is calculated by the square of error, and thus square root brings it back to the same level of prediction error and makes it easier for interpretation.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

## Mean Absolute Error (MAE)

Mean Absolute Error(MAE) is similar to Mean Square Error(MSE). However, instead of the sum of the square of error in MSE, MAE is taking the sum of the absolute value of error.

Compare to MSE or RMSE, MAE is a more direct representation of the sum of error terms. MSE gives larger penalization to big prediction errors by squaring them while MAE treats all errors the same.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

### Implementation Of The Model Evaluation Using Sklearn

R Square/Adjusted R Square

```
Python

# R Square/Adjusted R Square
from sklearn.metrics import r2_score

r2_score(y_true, y_pred)
```

Mean Square Error(MSE)/Root Mean Square Error(RMSE)



Python

```
# Mean Square Error(MSE)
from sklearn.metrics import mean_squared_error

mean_squared_error(y_true, y_pred)
```

Mean Absolute Error(MAE)



Python

```
# Mean Absolute Error(MSE)
from sklearn.metrics import mean_absolute_error

mean_absolute_error(y_true, y_pred)
```

## Exporting Predictions Of Machine Learning Models

At last, you're ready to submit some predictions for scoring. You can write your predictions to a CSV file using the `.to_csv()` method on a pandas DataFrame.

```
import pandas as pd

df = pd.DataFrame({'name': ['Tolu', 'Musa', 'Uche'],
                   'mask': ['red', 'purple', 'Blue'],
                   })
df.to_csv('./save_file.csv', index=False)
```

## References

Smith, I. (2020, May 23). *3 Best metrics to evaluate Regression Model? | by Songhao Wu*. Towards Data Science. Retrieved April 6, 2022, from <https://towardsdatascience.com/what-are-the-best-metrics-to-evaluate-your-regression-model-418ca481755b>

## Activity

- Build an XGBoost Model suing the tips.csv dataset





## DAY 5: FREE AI Class in Every City 2022

### REVISION AND AI INVASION COMPETITIONS

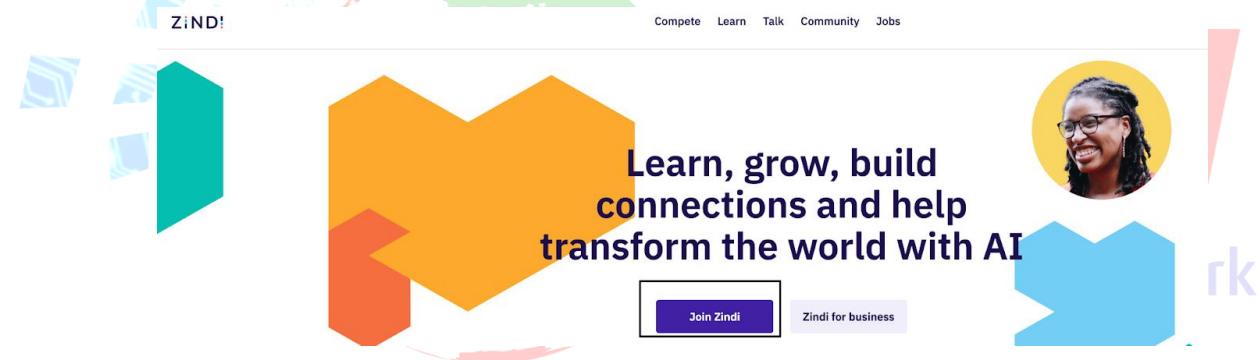
- In Class Exercise: Nigerian Cars Price Prediction
  - Download Dataset [AI Invasion In-Class Dataset.xlsx](#) here
- Introduction to Zindi and how to make a submission
- Going further with deep learning

### Introduction To Zindi And How To Make A Submission

#### How To Sign Up On Zindi

Steps:

1. Visit the site: <https://zindi.africa/>
2. Click on 'Join Zindi'



3. Fill in the necessary details as required and choose a recognizable 'username'.

## Sign up

Username

First Name  Last Name   
We will not show your name publicly.

Gender  (optional)

Country  City   
(optional)

Email

Password

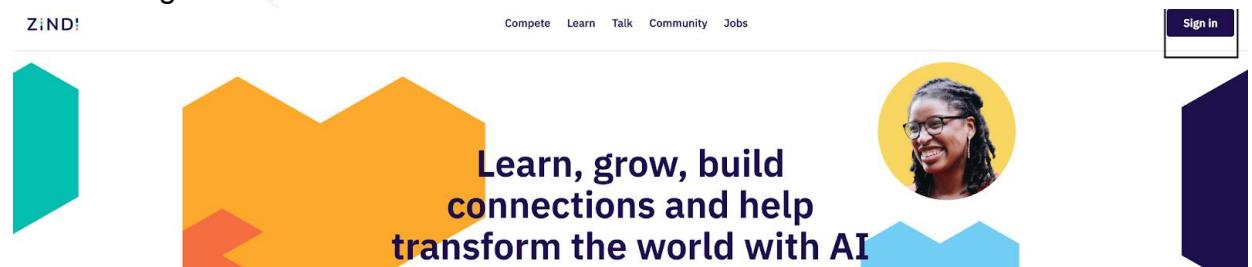
By submitting, you acknowledge that you have read and understood, and agree to the Zindi Terms of Use.

[Sign Up](#)

[Sign in](#)

This site is protected by reCAPTCHA and the Google [Privacy Policy](#) and [Terms of Service](#) apply.

4. Click on the sign-up bottom
5. An email will be sent to your mailbox for account confirmation.
6. Next, re-visit the site: <https://zindi.africa/>
7. Click on sign-in.



8. Enter your “username” and Password”

## Sign in

Username or email

Password

This site is protected by reCAPTCHA and the Google [Privacy Policy](#) and [Terms of Service](#) apply.

[Sign In](#)

[Sign up](#) [Forgot password?](#)

## How To Join A Competition On Zindi

1. On signing in, the following webpage is displayed.
2. Click on the “Compete Tab”
3. Zindi provides different categories of competition:
  - Prize competition: you win prize money if you are among the top 3 winners of a particular competition.
  - You win points: Points increase your ranking among other data scientists on the platform.
  - You gain knowledge: Knowledge competitions are where you can learn and increase your skillset.
4. Navigate to the hackathon of choice.
5. The competition page provides some information to help you understand the problem you are going to solve, by reading the problem statement, how you can participate and how to submit your solution on the platform.
6. On The competition page, the following tab can be found;
  1. Info: The info tab contains the problem statements of the competition and a list of organizations that have either provided the dataset or funded the competition. On the left side, you can see a list of vertical tabs that provide more information about the competition.
    - (a) Description

- It provides the problem statement of the competition and a list of organizations or companies that supported the competition, either by providing the datasets or funding the competition.
- We suggest you read the description of this competition in order to understand the problem and machine learning approach you can choose to solve it. If you read the description, you will understand that the objective of this competition is to create a machine learning model to predict which individuals are most likely to have or use a bank account.

### (b) Rules

Description	
Rules	<b>Rules</b>
Prizes	As this is a learning challenge, aside from the rules in the Terms of Use, no other particular rules apply. This challenge is open to all and not restricted to any country.
Evaluation	

- Each competition has its own rules. Breaking the rules can lead to disqualification, so make sure to carefully read and understand all the rules of the competition.

### (c) Prizes

- Now, this is the best part: the prize section provides details about the prize money that will be provided for the first, second and third place winners of the competition. But remember not all competition provides prize money for its winners; in other competitions, you can get Zindi points or gain knowledge.

### (d) Evaluation

- Each competition has its own evaluation metric that will be used to evaluate your results and rank you on the leaderboard. It also shows how you should prepare your submission file before uploading your file on the platform.
- For this competition, the evaluation metric will be the percentage of survey respondents for whom you predict the binary 'bank account' classification incorrectly.

### (e) Timeline

- This section provides information about the start date of the competition and the end date and time of the competition. If you submit your solution after the deadline you will receive a score but it will not reflect on the leaderboard. Make sure to submit before the deadline if you want to be considered for a prize.
- This competition has been reopened as a knowledge challenge, this means it will not close.

## 2. Data

- The data tab contains a description of the dataset you are going to use for this competition. On the right side of the page, you can see a list of links to download the dataset and other important files. You will download:
  - VariableDefinition.csv — This file contains a definition of each variable in the train and test data.
  - SubmissionFile.csv — The file contains a sample of how the submission file should look like.
  - Test.csv — This is a test data file you will use for prediction and save your results in the submission file.
  - Train.csv - This is a train data file that contains both the independent variable and the target one. You will use this dataset to train your model.
  - These files may differ between competitions. Also, keep in mind that you must join the competition in order to have access to the data files.

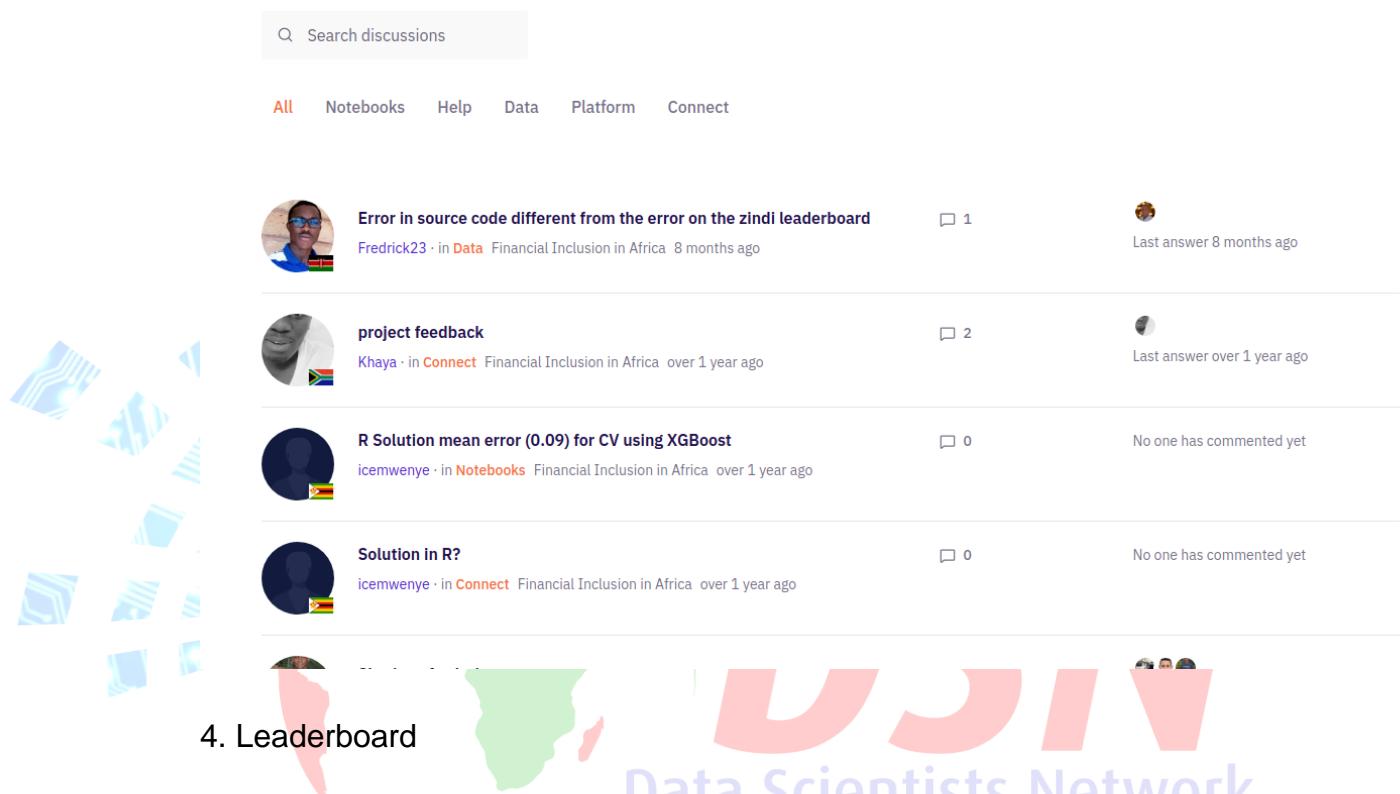
## 3. Discussion

- We're not Liverpool FC fans but we like their slogan: "You will never walk alone". That is the purpose of the discussion page, you don't need to walk alone throughout the competition. If you face any challenge or uncertainty during the competition or you want to ask a question to understand more about the dataset provided, you can

post on the discussion page and other data scientists enrolled in the competition can help you to solve the problem.

- The discussion board is very active and full of knowledgeable and helpful African data scientists willing to assist you.

## Discussions



Search discussions

All Notebooks Help Data Platform Connect

	Error in source code different from the error on the zindi leaderboard	<input type="checkbox"/> 1		Last answer 8 months ago
	project feedback	<input type="checkbox"/> 2		Last answer over 1 year ago
	R Solution mean error (0.09) for CV using XGBoost	<input type="checkbox"/> 0		No one has commented yet
	Solution in R?	<input type="checkbox"/> 0		No one has commented yet

## 4. Leaderboard

- After you have uploaded your submission file, you will appear on the leaderboard. The leaderboard shows your position among all enrolled data scientists in the competition. Your position will depend on your performance after your solution has been evaluated. For this competition, you can submit ten times a day.

## 5. Team

- You don't want to do the competition on your own? That's OK! You can create a team with fellow data scientists enrolled in the competition and work together. The maximum number for a team is 4 members. Remember that sharing code between individuals is not allowed, so if you want to share code with someone else, they must be on the same team as you.



### You are not a part of any team

You don't want to do the competition on your own? That's ok! You can create a team with fellow data scientists enrolled in the competition and work together. [Learn more](#)

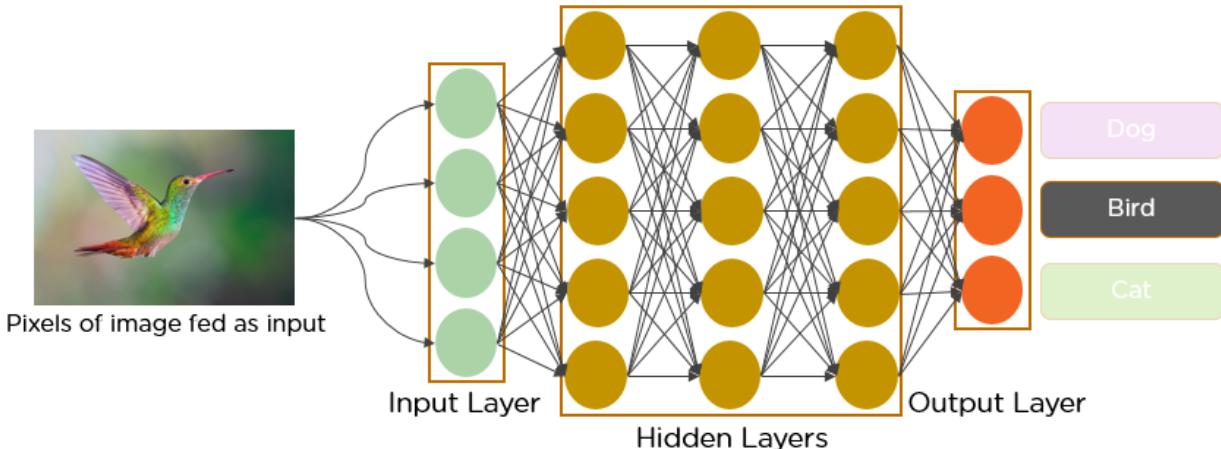
 [Create](#)

## 6. Submission

- The submission page is where you will upload your submission file, by clicking the orange button at the top right side of the page. After you have uploaded your solution, it will be evaluated according to the evaluation metric specified in the competition. Then you will see the score that will define your position on the leaderboard.

## What Is Deep Learning?

Deep learning is a subset of Machine Learning using the approach of neural networks. This branch works with algorithms built with the aim of mimicking the structure and function of the human brain.

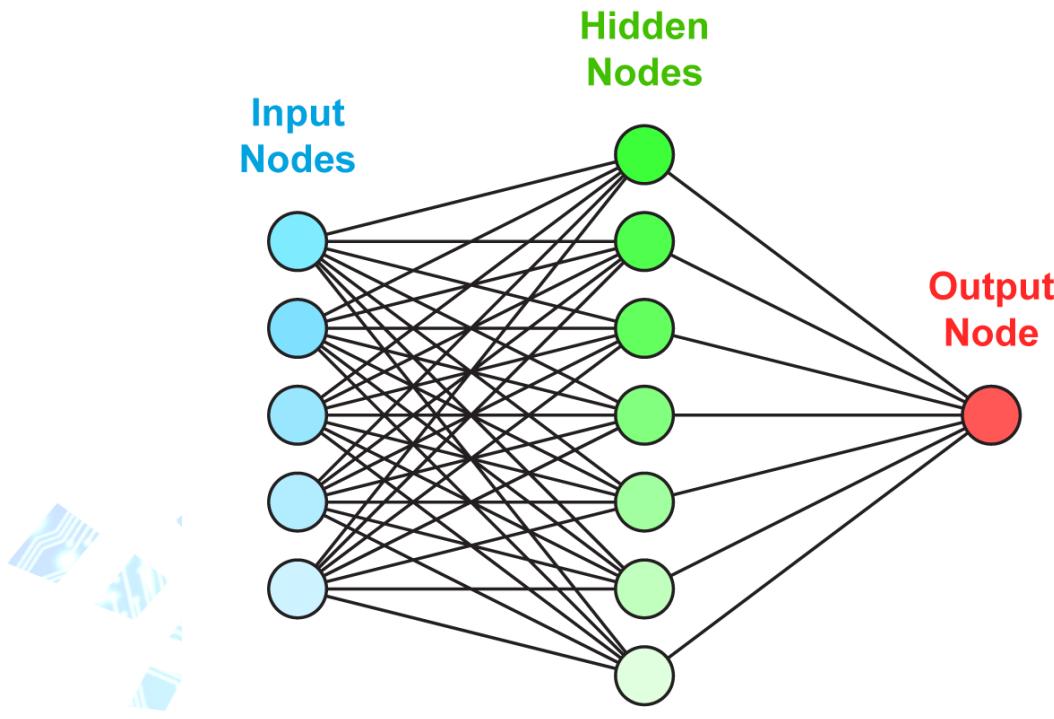


### Key Features of a deep learning algorithm

The three main components of a deep learning network are;

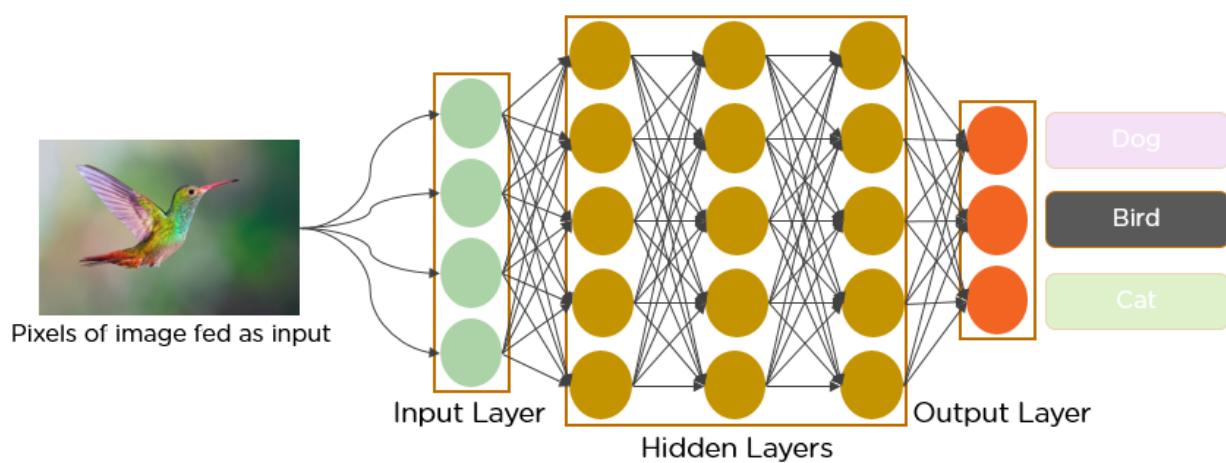
- Input Layer: The input, the first layer carrying information/features about the data.

- Hidden Layers: this can be one or more, connecting the input and output layer..
- Output Layer: where the expected output is obtained.



A node/ neuron; is a single unit in the network that receives information from other neurons. It is the basic building block of a neural network and is also referred to as 'the learning unit'.

The image below shows how a deep learning model works at a glance



Key difference between Deep Learning and machine learning;

	Machine Learning	Deep Learning
<b>Training dataset</b>	Small	Large
<b>Choose features</b>	Yes	No
<b>Number of algorithms</b>	Many	Few
<b>Training time</b>	Short	Long

"Use case of deep learning models"

1. Game-playing and decision making (backgammon, chess, racing)
2. Pattern recognition (radar systems, face identification, object recognition),
3. Sequence recognition (gesture, speech, handwritten text recognition)
4. medical diagnosis
5. Financial applications

