

Phase 2: Innovation

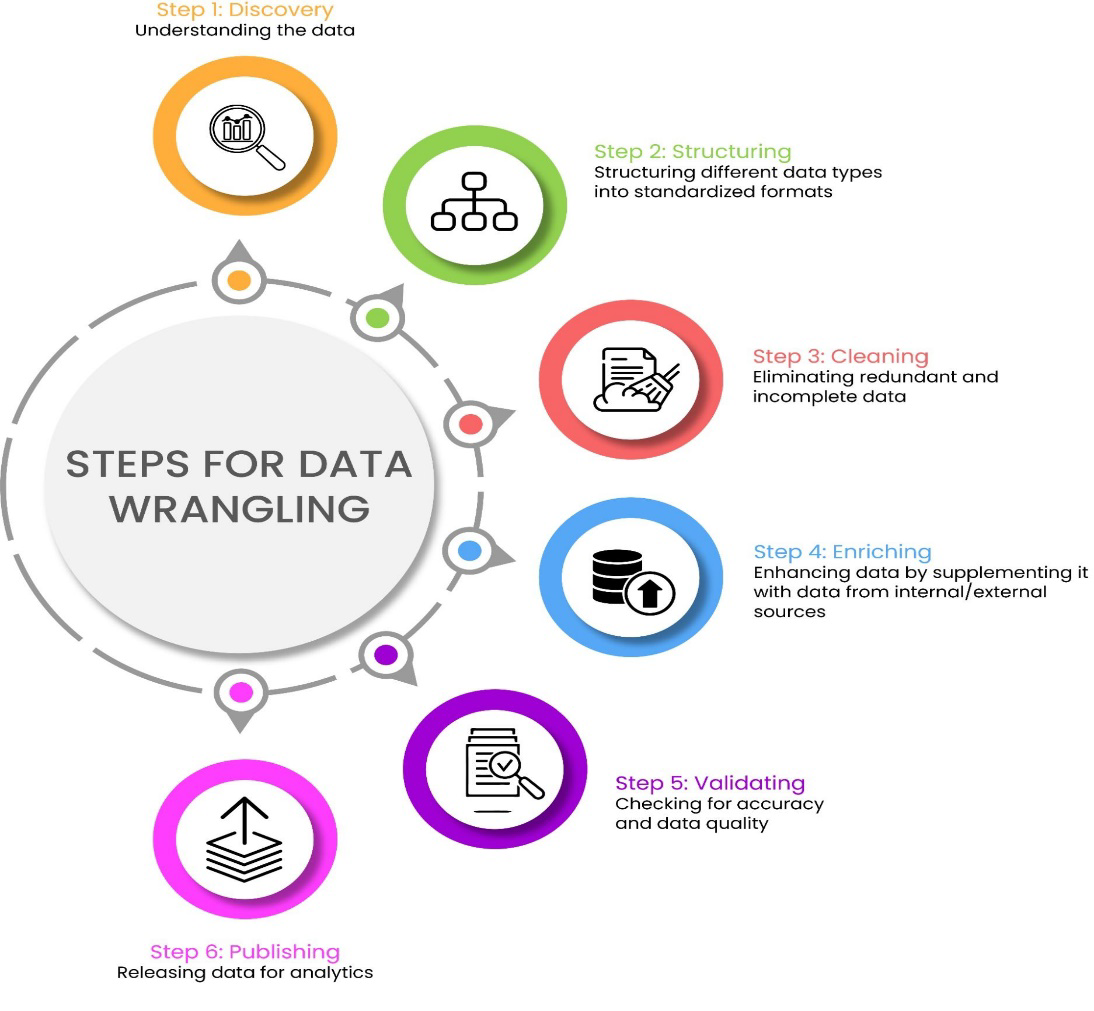
In this section you need to put your design into innovation to solve the problem. Create a doc around it and share the same for assessment.

MODULE 3:

DATA WRANGLING:

data wrangling, is a process that removes the risk of losing valuable insights from refined data. It ensures that data is in a reliable state before it is analyzed. The entire process is focused on cleaning the data and maintaining a specific format before uploading it to a database. It can be a time-consuming process when conducted manually. This article will discuss data wrangling in [Python](https://www.turing.com/kb/why-is-python-best-suited-for-competitive-coding), lay out the steps to write files with Python CSV writer, and much more.

**DATA WRANGLING STEPS**



It is impossible to select a single data science skill set that is the most important for business professionals as the insights are only as good as the data that governs them. Thus, it is vital to shape raw data into a usable form. This is where data wrangling helps as it cleans and enriches the data.

The exact tasks required in data wrangling depend on the type of transformations implemented to make a dataset usable. Below are the steps involved:

Step 1: Discovery

The first step involves discovering what type of information is contained in a data source and deciding if it has some value. Discovery is a small process in the entire data analysis operation and is a good way to explore the data. It is an opportunity to understand the data in-depth. Wrangling divides the data according to consistency and quality.

Step 2: Structuring

Standardizing the data format for the different data types makes it usable for various processes. Raw data does not have structure as, in most cases, it comes in different formats and there is no predefined structure. It needs to be restructured so data analysts can use the same for further analysis.

Step 3: Cleaning

Cleaning eliminates redundant and incomplete data that could skew the entire data analysis process. In this phase, high-quality analysis takes place. Data containing Null Values is changed either to an empty string or zero. This is followed by standardizing the format to clean the data to a higher quality. The goal is to get pure data that is ready for the final analysis.

Step 4: Enriching

Enriching entails deciding if there is enough data or if we need to seek out additional interior or third-party sources. Here, the data is transformed into a new kind of data, going from being cleaned to its formatted version. The ideal way to achieve this is to upscale, downsample, and, finally, predict the data.

Step 5: Validating

Validating deals with conducting tests to expose data quality and consistency issues. Data quality rules are used to evaluate the quality of datasets. Both consistency and quality are verified once data is processed. These parameters help establish a strong base upon which to tackle security issues. The tests are conducted along with multiple dimensions that adhere to the constraints of the syntax.

Step 6: Publishing

Publishing is all about making wrangled data available to stakeholders in downstream projects. It is the last step of the data wrangling process. The final form of the data is matched to its format which is then used for analytics.

MODULE4:

## What is an Artificial Neural Network?

These are computational models and inspire by the human brain. Many of the recent advancements have been made in the field of Artificial Intelligence, including Voice Recognition, Image Recognition Robotics using it. They are the biologically inspired simulations performed on the computer to perform certain specific tasks like -

* Clustering
* Classification
* Pattern Recognition

**In general -** It is a biologically inspired network of artificial neurons configured to perform specific tasks. These biological methods of computing are known as the next major advancement in the Computing Industry.

### What is a Neural Network?

The term ‘Neural’ has origin from the human (animal) nervous system’s basic functional unit ‘neuron’ or nerve cells present in the brain and other parts of the human (animal) body. A neural network is a group of algorithms that certify the underlying relationship in a set of data similar to the human brain. The neural network helps to change the input so that the network gives the best result without redesigning the output procedure. You can also learn more about [ONNX](https://www.xenonstack.com/blog/onnx/) in this insight.

**The advantages of are listed below**:

* A neural network can perform tasks that a linear program cannot.
* When an element of the neural network fails, its parallel nature can continue without any problem.
* A neural network learns and reprogramming is not necessary.
* It can be implemented in any application.
* It can be performed without any problem.

**The disadvantages of are described below:**

* The neural network needs training to operate.
* The architecture of a neural network is different from the architecture of microprocessors. Therefore, emulation is necessary.
* Requires high processing time for large neural networks

**How do neural networks work?**

Think of each individual node as its own [linear regression](https://www.ibm.com/topics/linear-regression) model, composed of input data, weights, a bias (or threshold), and an output. The formula would look something like this:

∑ wixi + bias = w1x1 + w2x2 + w3x3 + bias

OUTPUT = f(x) = 1 if ∑w1x1 + b>= 0; 0 if ∑w1x1 + b < 0

Once an input layer is determined, weights are assigned. These weights help determine the importance of any given variable, with larger ones contributing more significantly to the output compared to other inputs. All inputs are then multiplied by their respective weights and then summed. Afterward, the output is passed through an activation function, which determines the output. If that output exceeds a given threshold, it “fires” (or activates) the node, passing data to the next layer in the network. This results in the output of one node becoming in the input of the next node. This process of passing data from one layer to the next layer defines this neural network as a feed forward network.

Let’s break down what one single node might look like using binary values. We can apply this concept to a more tangible example, like whether you should go surfing (Yes: 1, No: 0) decision to go or not to go is our predicted outcome, or y-hat. Let’s assume that there are three factors influencing your decision-making:

1. Are the waves good? (Yes: 1, No: 0)
2. Is the line-up empty? (Yes: 1, No: 0)
3. Has there been a recent shark attack? (Yes: 0, No: 1)

Then, let’s assume the following, giving us the following inputs:

* X1 = 1, since the waves are pumping
* X2 = 0, since the crowds are out
* X3 = 1, since there hasn’t been a recent shark attack

Now, we need to assign some weights to determine importance. Larger weights signify that particular variables are of greater importance to the decision or outcome.

* W1 = 5, since large swells don’t come around often
* W2 = 2, since you’re used to the crowds
* W3 = 4, since you have a fear of sharks

Finally, we’ll also assume a threshold value of 3, which would translate to a bias value of –3. With all the various inputs, we can start to plug in values into the formula to get the desired output.

Y-hat = (1\*5) + (0\*2) + (1\*4) – 3 = 6

If we use the activation function from the beginning of this section, we can determine that the output of this node would be 1, since 6 is greater than 0. In this instance, you would go surfing; but if we adjust the weights or the threshold, we can achieve different outcomes from the model. When we observe one decision, like in the above example, we can see how a neural network could make increasingly complex decisions depending on the output of previous decisions or layers.

In the example above we used perceptron to illustrate some of the mathematics at play here, but neural networks leverage sigmoid neurons, which are distinguished by having values between 0 and 1. Since neural networks behave similarly to decision trees, cascading data from one node to another, having x values between 0 and 1 will reduce the impact of any given change of a single variable on the output of any given node, and subsequently, the output of the neural network.

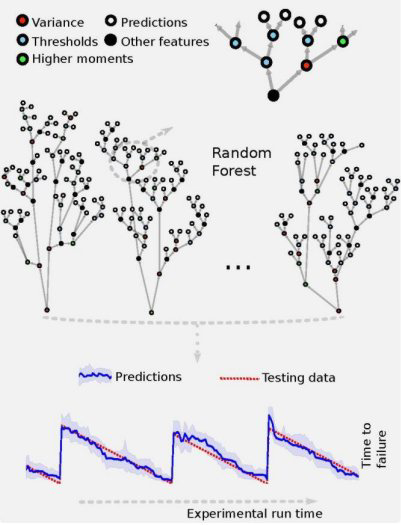
As we start to think about more practical use cases for neural networks, like image recognition or classification, we’ll leverage supervised learning, or labeled datasets, to train the algorithm. As we train the model, we’ll want to evaluate its accuracy using a cost (or loss) function. This is also commonly referred to as the mean squared error (MSE). In the equation below,

* *I*represents the index of the sample,
* y-hat is the predicted outcome,
* y is the actual value, and
* *M* is the number of samples.

𝐶𝑜𝑠𝑡 𝐹𝑢𝑛𝑐𝑡𝑖𝑜𝑛= 𝑀𝑆𝐸=1/2𝑚 ∑129\_(𝑖=1)^𝑚▒(𝑦 ̂^((𝑖) )−𝑦^((𝑖) ) )^2

Ultimately, the goal is to minimize our cost function to ensure correctness of fit for any given observation. As the model adjusts its weights and bias, it uses the cost function and reinforcement learning to reach the point of convergence, or the local minimum. The process in which the algorithm adjusts its weights is through gradient descent, allowing the model to determine the direction to take to reduce errors (or minimize the cost function). With each training example, the parameters of the model adjust to gradually converge at the minimum

**DIAGRAM:**



As illustrated in the above figure, each tree makes its prediction (white leaf node), following a series of decisions (colored nodes) based on features of the acoustic signal during the current window). The RF prediction (blue line) on data it has never seen (testing data) with 90% confidence intervals (blue shaded region). The predictions agree remarkably well with the actual remaining times before failure (red curve).

A particular success of this research was that it identified new signals, previously thought to be low-amplitude noise, that provide forecasting information throughout the earthquake cycle.

**FLOWCHART:**

