

CONSULTING & ANALYTICS CLUB  
IIT GUWAHATI

## Summer Analytics 2019 Schedule:

The entire course is distributed in 4 weeks. Every week will be distributed in two phases Phase I (4-5 Days) & Phase II (2-3 days).

Phase I : It will consist of learning various concepts of Data Analytics where most of the theory part will be covered, basically Tue-Fri of each week is allotted for this phase.

Phase II : An assignment/quiz will be uploaded every Saturday morning which would help you to apply all the concepts learnt within the theory part during Phase I and you will have 3 days to implement and finish your work.

### **Week 1 :**

#### *Phase I*

We'll be starting with most important tools for Analytics:

Python : This week will comprise of learning basic coding skills in python required for data science.

EDA : In statistics, **exploratory data analysis (EDA)** is an approach to analyzing data sets to summarize their main characteristics, often with visual methods(graphs).

Note - This week you have two extra days 25th and 26th May so Phase I will be of 7 days

Breakdown:

For this week you need to audit two courses the method to do this is described at the end.

*PS: Do make notes while learning and it's always a better practice to try implementing the syntax of the code yourself in a jupyter notebook for better understanding*

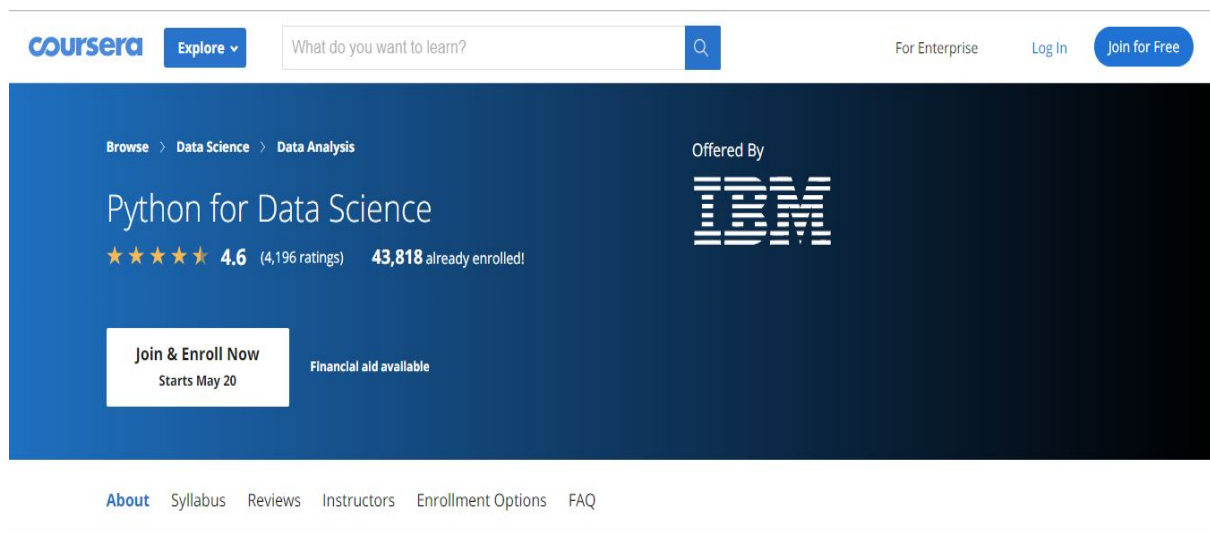
Day 1	Python Basics (Coursera) <i>*complete week 1</i>
Day 2	Python Data Structures(Coursera) <i>*complete week 2</i>
Day 3	i) EDA-Examining Distributions(Stanford) <i>*upto Measure of Spread-range,IQR &amp;Outliers</i> ii) Python Programming Fundamentals(Coursera) <i>*upto loops</i>
Day 4	i) EDA-Examining Distributions(Stanford) <i>*complete</i> ii) Python Programming Fundamentals(Coursera) <i>*complete</i>
Day 5	i) EDA- Examining

	<p>Relationships(Stanford)</p> <p><i>* upto Scatter plots</i></p> <p>ii)Working with data in python(Coursera)</p> <p><i>* upto pandas</i></p>
Day 6	<p>i)Working with data in python(Coursera)</p> <p><i>* complete</i></p> <p>ii) EDA- Examining Relationships(Stanford)</p> <p><i>*complete</i></p>
Day 7	<p>i) Python Matplotlib (for making plots in python). Follow this link: <a href="https://www.youtube.com/watch?v=yZTBMdPOww">https://www.youtube.com/watch?v=yZTBMdPOww</a></p>

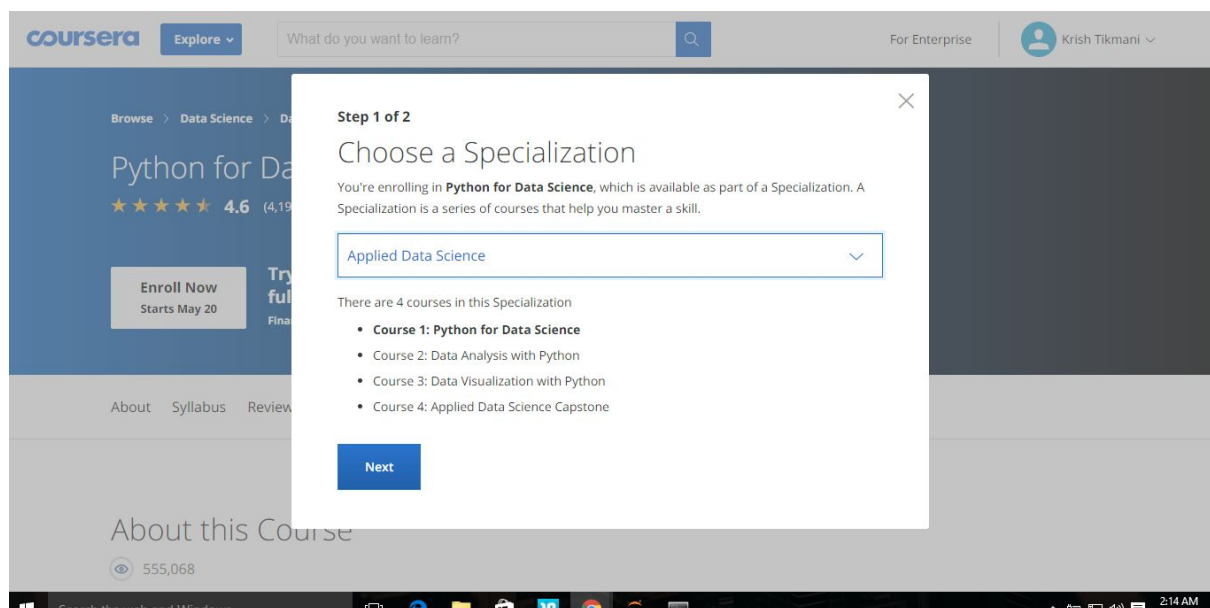
Python For Data Science - visit the link below and follow the steps

<https://www.coursera.org/learn/python-for-applied-data-science?specialization=ibm-data-science-professional-certificate>

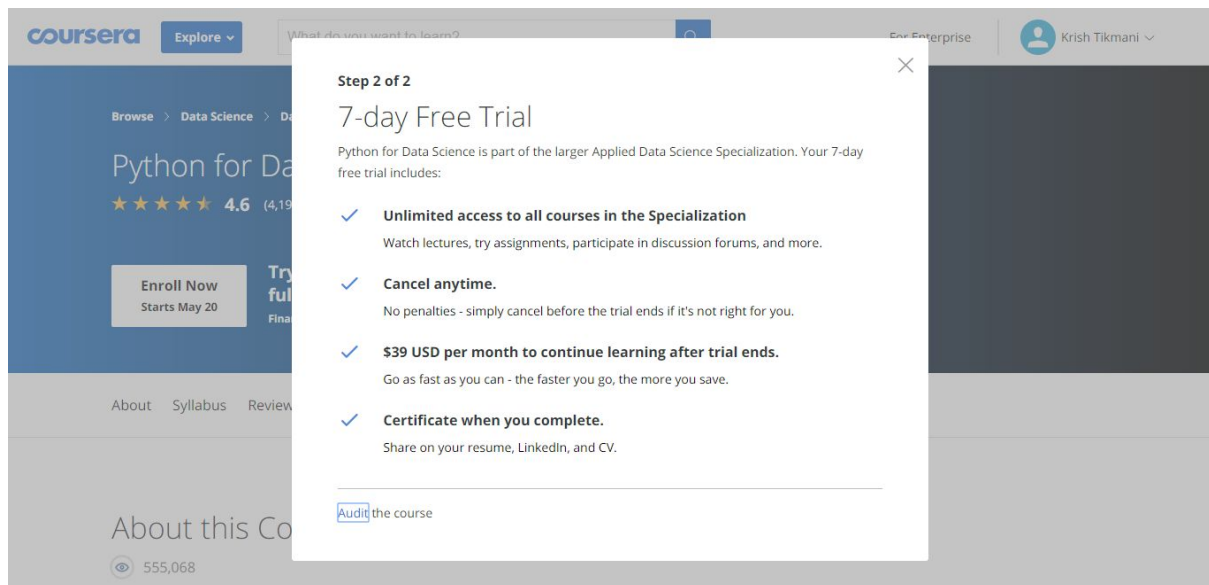
## Step 1 - Click on Join and Enroll now and create your account



## Step 2 - Click on next button



## Step 3 - Click on Audit the course

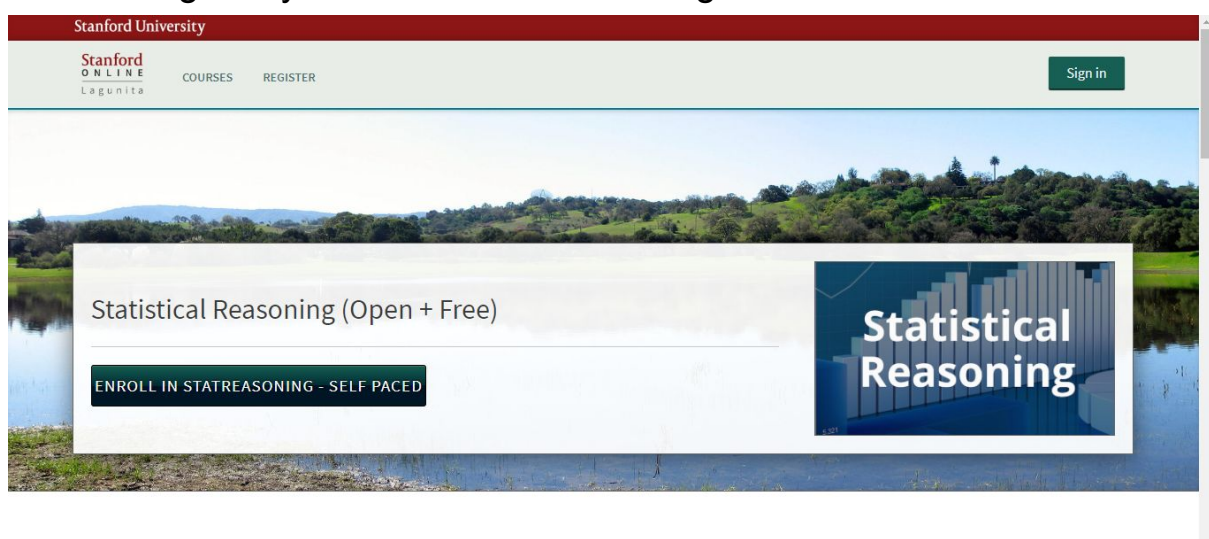


Finally you are good to go with this python course.

Statistics for data analytics - visit the link below

<https://lagunita.stanford.edu/courses/OLI/StatReasoning/Open/about>

Enroll & register yourself to this course to get started.



*Phase II*

This week's assignment consists of two parts Assignment-1(a) & Assignment -1(b) based on whatever that is covered in lectures of week 1.

The link to Assignment file is:

[https://drive.google.com/drive/folders/1mh5YGIHEJZBw\\_Q1ARQkAhiZ9Wq3v34oC?usp=sharing](https://drive.google.com/drive/folders/1mh5YGIHEJZBw_Q1ARQkAhiZ9Wq3v34oC?usp=sharing)

Please follow the instructions as given in above folder to complete the assignment.

*Submit your Assignment by following instructions below :*

Instructions for uploading assignment:

- 1) Create a Github account
- 2) Create a new Repo and name it Assignment\_Week1. Write some description about the project.
- 3) Upload all the files of the assignment that is shared with you (ipynb replaced with your new updated one)
- 4) Yayy! You're done. Now share the link of the Github repo in the Google Form below

Link for submission :

[https://docs.google.com/forms/d/e/1FAIpQLSdHqUNygPBfO9yizK-o7aB5dvsOc8L6bjJ33gl5\\_-76yaA58A/viewform?usp=sf\\_link](https://docs.google.com/forms/d/e/1FAIpQLSdHqUNygPBfO9yizK-o7aB5dvsOc8L6bjJ33gl5_-76yaA58A/viewform?usp=sf_link)

If you are new to Github you can take help from link below:

<https://www.youtube.com/watch?v=73l5dRucCds>

(Note: You can have multiple Submissions the one with best marks will be selected)

P.S. If you feel any doubt you can post it on facebook page or feel free to contact us.

*Sample Solution : Click on the link below to get the solutions to first Assignment-*

[https://drive.google.com/drive/folders/1Sh\\_VPHNdf8CJvr-0INOWNtdoCJCCym0\\_?usp=sharing](https://drive.google.com/drive/folders/1Sh_VPHNdf8CJvr-0INOWNtdoCJCCym0_?usp=sharing)

## **Week 2:**

### **Phase I**

**Data Analysis in Python Using Pandas** : In this week we will enhance our skill and learn more tricks for playing with a Dataset.

**Introduction to Probability:** Probability theory is the mathematical foundation of statistical inference which is indispensable for analyzing data affected by chance, and thus essential for data scientists. I know it may seem a bit cliché at starting but trust me probability is one of the most important foundation used in Machine Learning algorithms.

**Inferential Statistics:** This is the most important concept, you'll be applying these throughout your career and it will help you answer many questions in various interviews.

The main purpose of inferential statistics is to:

- A. Summarize data in a useful and informative manner.
- B. Estimate a population characteristic based on a sample.
- C. Determine if the data adequately represents the population.

*\*\*This week too you will have to audit two courses, the link and steps to do are given at the end.\*\**

*Some Instructions:*

- a) For the python part this week it is highly advised to duplicate everything after watching the video on your own for better grasp.*
- b) For the courses of Cousera just watch the video lectures (Speed 1.5x) and it is not required to do it's assignments and quizzes.*
- c) Making notes is highly advised. It will help you grasp things better and save you time when you need to go through concepts.*

Breakdown:

DAY 1	<p>(i)Data Analysis in Python using Pandas: visit the link- <a href="https://www.youtube.com/watch?v=yzIMircGU5I&amp;list=PL5-da3qGB5ICCsgW1MxlZ0Hq8LL5U3u9y&amp;index=1">https://www.youtube.com/watch?v=yzIMircGU5I&amp;list=PL5-da3qGB5ICCsgW1MxlZ0Hq8LL5U3u9y&amp;index=1</a> (upto video 8 watch in 2x speed)</p> <p>(ii)Producing Data: Sampling (Stanford) *complete</p> <p>(iii)Producing Data: Designing Studies (Stanford) *complete</p>
-------	--



DAY 2	<p>(i)Data Analysis in Python using Pandas: visit the link-  <a href="https://www.youtube.com/watch?v=YPItfQ87qjM&amp;list=PL5-da3qGB5ICCsgW1MxlZ0Hq8LL5U3u9y&amp;index=9">https://www.youtube.com/watch?v=YPItfQ87qjM&amp;list=PL5-da3qGB5ICCsgW1MxlZ0Hq8LL5U3u9y&amp;index=9</a>  <i>(videos 9-16 watch in 2x speed)</i></p> <p>(ii)Introduction to probability: Week 3 (Coursera)  <i>*complete</i>  <i>(you don't need to do week 1 &amp; week 2)</i></p>
DAY 3	<p>(i)Data Analysis in Python using Pandas: visit the link-  <a href="https://www.youtube.com/watch?v=OYZNk7Z9s6I&amp;list=PL5-da3qGB5ICCsgW1MxlZ0Hq8LL5U3u9y&amp;index=17">https://www.youtube.com/watch?v=OYZNk7Z9s6I&amp;list=PL5-da3qGB5ICCsgW1MxlZ0Hq8LL5U3u9y&amp;index=17</a>  <i>(videos 17-21 watch in 2x speed)</i></p> <p>(ii)Introduction to probability: Week 4 (Coursera)  <i>*complete</i></p>

DAY 4	<p>(i)Data Analysis in Python using Pandas: visit the link- <a href="https://www.youtube.com/watch?v=0s_1IsROgDc&amp;list=PL5-da3qGB5ICCsgW1MxlZ0Hq8LL5U3u9y&amp;index=24">https://www.youtube.com/watch?v=0s_1IsROgDc&amp;list=PL5-da3qGB5ICCsgW1MxlZ0Hq8LL5U3u9y&amp;index=24</a> (videos 24-30 watch in 2x speed)</p> <p>(ii)Inferential Statistics (Coursera)- Week 1 <i>*upto CLT &amp; Sampling</i></p>
DAY 5	<p>(i)Inferential Statistics (Coursera)- Week 1 <i>*upto Confidence Intervals</i></p> <p>(ii)Inferential Statistics (Coursera)- Week 2 <i>*complete i.e. Hypothesis Testing &amp; Significance</i></p>

**For Introduction To Probability & Data – visit the link below and follow the same steps as done earlier for Python for Data Science.**

<https://www.coursera.org/learn/probability-intro>

**For Inferential Statistics – visit the link below and follow the same steps as done above.**

<https://www.coursera.org/learn/inferential-statistics-intro>

\*\*\*This is all for the Phase I part of Week 2, in the upcoming week we'll be continuing with Inferential Statistics and start with Machine Learning, So be motivated guys push yourself and complete this Week something bigger is waiting for you in the upcoming one :) :) \*\*\*

## **Phase II**

The link to Assignment 2 is given below:

[https://drive.google.com/drive/folders/1mh5YGIHEJZBw\\_Q1ARQkAhiZ9Wq3v34oC?usp=sharing](https://drive.google.com/drive/folders/1mh5YGIHEJZBw_Q1ARQkAhiZ9Wq3v34oC?usp=sharing)

This assignment contains 3 parts, you have to complete first 2 parts and follow the same procedure as done in Assignment 1 for submission.

The link for submission is :

[https://docs.google.com/forms/d/e/1FAIpQLSdY7yRIKrnkbDqtt516ozHyljmcej0Hrl89xfV0DTa5z3Wzvw/viewform?usp=sf\\_link](https://docs.google.com/forms/d/e/1FAIpQLSdY7yRIKrnkbDqtt516ozHyljmcej0Hrl89xfV0DTa5z3Wzvw/viewform?usp=sf_link)

Part 3 of the assignment is for your practice and its a good application based problem so try doing it.

Duration Given For Assignment : 2 Days

Deadline:11th June 2019 23:59

## **Week 3:**

### **Phase I**

As mentioned earlier this week we will be finishing Inferential Statistics and start with Machine learning.

**Machine Learning:** The field of study that gives computers the ability to learn without being explicitly programmed.

Or

If performance of a program at task T as measured by P improves with experience E it is called Machine Learning.

*\*\*For this week you will have to audit one course from edX i.e. Machine Learning with Python : A practical introduction by IBM. The link and method to enroll is given at the end.\*\**

*This is the most interesting and important week of this course where you will enjoy every bit of whatever you learn.*

*Instructions:*

*(i) In the course given by edX there are Lab sessions after every topic covered, it is highly suggested to go through the code in that as it will help you understand the implementation of the topic.*

*(ii) Making notes is highly advised. It will help you grasp things better and save you time when you need to go through concepts.*

Breakdown:

DAY 1	<p>(i)Intro of Machine Learning, visit link- <a href="https://www.youtube.com/watch?v=ukzFI9rgwfU">https://www.youtube.com/watch?v=ukzFI9rgwfU</a></p> <p>(ii)Inferential Statistics (Coursera)-Week 3 <i>*complete</i></p>
DAY 2	<p>(i)Machine learning with Python (edX) – Module 1 <i>*complete</i></p> <p>(ii)Inferential Statistics (Coursera)-Week 4 <i>*Inference for Proportions (i.e. till Hypothesis test for comparing two proportions)</i></p>

DAY 3	(i)Machine Learning with Python (edX) – Module 2 <i>*upto Simple Linear Regression</i>  (ii)Inferential Statistics (Coursera) – Week 4 <i>*complete</i>
DAY 4	(i)Machine Learning with Python (edX) – Module 2 <i>*complete</i>
DAY 5	(i)Machine Learning with Python (edX) – Module 3 <i>*upto Decision Trees</i>

**For Machine Learning with Python course by edX visit the link and follow the steps as given below :**

<https://www.edx.org/course/machine-learning-with-python>

**Step 1 – Click on enroll now**

The screenshot shows the edX website interface for the course 'Machine Learning with Python: A Practical Introduction' by IBM. The page layout includes a navigation bar with 'edX' logo and links to 'Courses', 'Programs & Degrees', 'Schools & Partners', and 'edX for Business'. A search bar and a user profile 'krishn\_tikmani' are also visible. The main content area features a yellow banner with a Python logo and books, followed by the course title and a description: 'Machine Learning can be an incredibly beneficial tool to uncover hidden insights and predict future trends. This Machine Learning with Python course will give you all the tools you need to get started with supervised and unsupervised learning.' Below this is the IBM logo. To the right, there is a green 'Enroll Now' button, a checkbox for email notifications, and a note that the course is part of a 'Professional Certificate Program'. At the bottom, there is an 'About this course' section with a detailed description and a table of course details: Length (5 weeks), Effort (4 to 6 hours per week), and Price (FREE, with an option to add a verified certificate).

## Step 2 – Click on Audit This Course

The screenshot shows the edX course page for "Machine Learning with Python: A Practical Introduction". The user is logged in as "krishn\_tikmani". The page displays two options for enrolling in the course:

- Pursue the Verified Track**: Highlight your new knowledge and skills with a verified certificate. Use this valuable credential to improve your job prospects and advance your career, or highlight your certificate in school applications. Benefits of the Verified Track include:
  - Unlimited Course Access**: Learn at your own pace, and access materials anytime to brush up on what you've learned.
  - Graded Assignments**: Build your skills through graded assignments and projects.
  - Easily Sharable**: Add the certificate to your CV or resume, or post it directly on LinkedIn.
  - Support our Mission**: EdX, a non-profit, relies on verified certificates to help fund affordable education to everyone globally.A green button labeled "Pursue the Verified Track (₹2709 INR)" is visible.
- Audit This Course (No Certificate)**: Audit this course for free and have access to course materials and discussions forums. **This track does not include graded assignments, or unlimited course access.** A blue button labeled "Audit This Course" is visible.

Below the options, there is a "Connecting..." status bar.

Finally you are good to go with this course.

(\*\*Congrats guys you have completed about 75% of the course just one more week to go after this so push yourself, you have reached this far so complete it and develop a highly demanded skill in today's world within you :) \*\*)

## Phase II

The link for Assignment 3 is :

[https://drive.google.com/open?id=1mh5YGIHEJZBw\\_Q1ARQkAhiZ9Wq3v34oC](https://drive.google.com/open?id=1mh5YGIHEJZBw_Q1ARQkAhiZ9Wq3v34oC)

This assignment has two parts one for implementation of Simple Linear Regression and other for KNN from scratch.

The assignment for remaining topics taught in this week will be uploaded at the end of the course as it would become hectic for you guys.

The link for submission is :

<https://forms.gle/nGApe3RJH1ATgMKh8>

Duration Given For Assignment : 3 Days

Deadline:20th June 2019 23:59

## **WEEK 4:**

### **Phase I**

We will be completing the Machine Learning course on edX this week. Also we will be covering some of the important algorithms in depth. So for this we have a reading part everyday which has links to the blogs that discusses the algorithm and its implementation in details.

This is the last week of our course, you guys have come this far so make sure you complete this week.

#### *Instructions:*

- (i) For the reading portion make sure you take notes and try to implement the code wherever possible. It is highly advised to google the portion of code that you don't understand.*
- (ii) Watch the Cross validation portion videos at 1.5x speed.*

Breakdown:

DAY 1	<p>(i)Machine Learning with Python (edX) – Module 3 (Classification) <i>*Complete</i></p> <p>(ii)Linear Regression (reading) – visit link: <a href="https://towardsdatascience.com/linear-regression-using-python-b136c91bf0a2">https://towardsdatascience.com/linear-regression-using-python-b136c91bf0a2</a></p>
DAY 2	<p>(i)Machine Learning with Python (edX) - Module 4 (Clustering) <i>*upto K-means Clustering</i></p> <p>(ii)KNN (reading) – visit link: <a href="https://medium.com/machinelearningalgorithms/k-nearest-neighbors-c9823dca611b">https://medium.com/machinelearningalgorithms/k-nearest-neighbors-c9823dca611b</a></p> <p>(iii)Decision Tree (reading) – visit link: <a href="https://medium.com/datadriveninvestor/decision-tree-algorithm-with-hands-on-example-e6c2afb40d38">https://medium.com/datadriveninvestor/decision-tree-algorithm-with-hands-on-example-e6c2afb40d38</a></p>



DAY 3	<p>(i)Machine Learning with Python (edX)</p> <p>–</p> <p>Module 4 (Clustering)</p> <p><i>*upto Hierarchical Clustering</i></p> <p>(ii)Logistic Regression (reading) – visit link:  <a href="https://towardsdatascience.com/building-a-logistic-regression-in-python-301d27367c24">https://towardsdatascience.com/building-a-logistic-regression-in-python-301d27367c24</a></p>
DAY 4	<p>(i)Machine Learning with Python (edX)</p> <p>–</p> <p>Module 4 (Clustering)</p> <p><i>*complete</i></p> <p>(ii)SVM (reading) – visit link:  <a href="https://medium.com/datadriveninvestor/support-vector-machines-ae0ff2375479">https://medium.com/datadriveninvestor/support-vector-machines-ae0ff2375479</a></p>

DAY 5	<p>Cross Validation:</p> <ol style="list-style-type: none"><li>1. <a href="#">Four Types Of Cross Validation</a></li><li>2. <a href="#">Improve Your Model Performance using Cross Validation</a></li><li>3. <a href="#">Selecting the best model in scikit-learn using cross-validation</a></li></ol>
-------	--

(\*\*This marks the end of the theoretical phase of our course. I hope that you guys had great and productive time learning with us :) :). Following this you will have your final assignment which will earn you a certificate for successful completion of course\*\*)

## **Phase II**

The link for the final assignment is given below:

[https://drive.google.com/open?id=1mh5YGIHEJZBw\\_Q1ARQkAhiZ9Wq3v34oC](https://drive.google.com/open?id=1mh5YGIHEJZBw_Q1ARQkAhiZ9Wq3v34oC)

This is the final assignment of this course and it has 1 problem. The descriptions about the problem is well described in the assignment folder.

Link for submission-

<https://forms.gle/Yh6QUPRuJtWE9ZCQ8>

Deadline For Submission - 12th July 2019

[Update Learning Progress Here](#)