**Basic prerequisite to enter in Data Science Domain**

Data Science is very old and vast domain which include Big data segment,

Analytics (machine learning and deep learning), Data Visualization, Data cleaning, ETL etc.

To understand your area of interest it is very necessary to know deeper about

things which are used to solve a business problem.

Here we will explain steps to enter in analytics fields :

**Step 1 : you are from mathematics background its good else this is your task to brush up your mathematics part :**

* Probability -
* Conditional Probabilities
* Events, outcome, independent etc.
* Bayes theorem.
* Possion distribution, Bernoulli’s distribution, Binomial distribution.

* Statistic -
* Learn central tendency
* Learn central limit theorem (Mean , Median , Mode, std, variance)
* Learn Normal distribution ( Normalization).
* Learn skewness, kutosis, standardization.
* Learn and implement statistics operation z-score, min-max scale etc)
* Learn about hypothesis testing technique like ( t-test, chi-square test, z-test, f-test, anova test ) and P- value concept.
* Learn and understand null hypothesis concept.
* Linear Algebra -
* Learn about scaler quantity and their rank.
* Learn and implement Vectors and understand concept of vector representation.
* Learn about Linear Trasnformation, Linear Combination, Vector in higher dimension
* Learn about importance of vectorization in computationally high operation.

* Matrix Algebra -
* Learn and Understand what Matrix is and why it is important.
* Learn and implement matrix operation ( multiplication , addition, element wise multiplication, singularity, determinant etc)
* Learn matrix dimension and their rank.

* Calculus -
* Learn and implement about differencial calculus, partial derivatives , chain rule, power rule and other concept of calculus.
* Learn and implement basic integration, integration of function like (log, e and other important concept)
* Imp function -
* Learn about monolithic function ( log) , focus more on exceptional cases like (log , e , sin , cos and other important function in maths about their behaviour and property)
* Good to know -
* Learn about number system ( Real, Complex, integer etc.)
* Learn about number series like (Taylor, fourier, fibonacci etc which were there in 1st year of college)
* Learn about Arithmetic mean (A.M.), Harmonic Mean(H.M.) and its variants.
* Learn basic mensuration part ( triangle, rect etc) and their angle variation using trigonometric functions
* Learn basic about circle, parabola, hyperbola, eclipse and other shape which could be inter related at later stage.

Other than this if anything you feel important it is well and good to study always.

**Step 2 - A programming background to implement operation**

* If you are from a programming background it is a plus point but if not, then you need to focus over any programming language ( C++, Python, R, Java etc) in which any one you can be proficient and expert.
* Example : choosing python over other language , which include key concept such as :-
  + - * basic indentation and formation of python
      * declaration of variables and their scope
      * declaration of function and their behaviour and scope.
      * Conditional statement ( if else etc)
      * looping concept ( for, while, for each etc)
      * list , dictionary , set etc are important concept.
      * String operation and manipulation.
      * Zipping, collection etc.
      * Classes, oops, self etc and other concepts of python.

* After completion of basic core of any language , need to focus on machine learning based field like ( pandas , numpy, matplotlib , seaborn, scikit, statesmodel , scipy etc)
* In pandas need to focus more on ( filtering , dataframe creation, series, stacking, melting, reading, pivot table etc.)
* In Numpy need to focus more over all mathematics operation including matrix manipualtion.
* In matplotlib and seaborn focus more on data visualization technique like all plots and their importance. Where it is useful and why. Suitability of graph gives a better understanding.
* Scikit, statesmodel and scipy etc libraries are high level this you plan after you start algorithm implentation. This is last segment you will do in your learning.

**Step 3: Understand about basic machine learning concept**

* Learn what is machine learning and why it is used ? When it is developed and what is role in our day to day life and business domain.
* Learn about basic type of algorithm in machine learning like ( supervise , unsupervise and clustering) .
* Learn about loading data in dataframe and train test split.
* First very basic regression technique is Linear Regression. It is a very basic model for prediction. Understand properly and try to implement using basic mathematics.
* Thereafter learn about logistic classification which is very basic and important classification method. Understand properly and implement using basic mathematics.
* Learn about Feature enginnering , Feature selection, feature generation, anamoly detection, Data cleaning, Data Wrangling etc related to data.
* Learn Concept of overfitting and underfitting.
* Learn concept of bias and variannce.
* Learn and implement regularization technique ( L1 and L2 )
* Learn about other Regression algorithm such as :
  + - * + Linear Regression variants ( Multivariate, Polynomial, Lasso, Ridge)
        + Gradient descent technique for cost optimization.
        + SVR ( support vector regression)
        + KNN (K Nearest Neighbour for regression)
        + Decision Tree ( decision Tree for regression)
        + Random Forest regression.
        + XGBoost for regression.
* Learn about other Classification algorithm such as :
  + - * + Logistic Regression variants ( sigmoid, softmax)
        + Gradient descent technique for cost optimization.
        + SVM ( support vector classification)
        + KNN (K Nearest Neighbour for classification)
        + Decision Tree ( decision Tree for classification)
        + Random Forest .
        + Light Gradient boosting. (LGBM)
* Learn about Clustering algorithm such as :
* Agglomarative clustering
* K-means clustering
* K-mode clustering
* Learn about Time series and Implement with various method:
* Seasonality, trend, cycle, pattern, noise, white noise.
* Acf, pacf, stationarity, adfuller test, holt winter method, box cox method
* Arima model and variants ( Sarima, Auto arima, multivariate arima).
* ETS ( expoential time series, EWMA)
* Arch and garch model for volatility.
* Mape and smape for accuracy

* Learn about visualization using seaborn and matplotlib and try to visualize data and get insight.
* For regression need to focus on R2 value , RMSE, MSE, cost vs iteration graph, Beta values and graph.
* For classification need to focus over confusion matrix, precision , recall, F1-score, accuracy, ROC-AUC curve etc
* For Time series need to focus on seasonality, trends, random noise, white noise, previous patterns, cylce, correlation and pattern , Mape, smape etc.
* For validity need train and test set divison. And for setting hyper parameter need dev set.
* Use any method like grid cv or k-fold to get correct accuracy.

**Step 4 – Test your skill set in Data set**

* very first step to start with kaggle.com and create a account.
* Download any suitable small dataset and start practicing your skill.
* Try to use pandas and numpy as much as possible in your test scanario and check result manually
* try using data visualization using matplotlib and seaborn.
* Try it untill you get enough confidance and good understading over all this concept
* If you are underconfident about your programming skill take any course based on your programming skills example ( a complete understanding of python ) . there are various vendor like – udemy, courseera , pluralsight or datacamp.
* Start with Andrew Ng course ( Machine leaning by Andrew Ng – standford university)
* once you start course, start with implemanting in your dataset and any programming langugage , remember try not to use any library and implement using purely maths concept.
* Download and read as much as reasearch paper on particular algorithm you are working on.
* Read more article in medium, watch video in youtube and other sites to get more clearance about particular algorithm you are working on.
* Make a goal to list down all important algorithm and learn in sequence as andrew tells.
* once you done with all this need to take participate in any ongoing competiton in either Anayticsvidya.com or kaggle.com.
* Check your level and recheck your solution and approach by taking refrence from experts solutions.
* Try to achieve atleast in top 15-20% in all population. Practice and learn standard of code from other kernels.
* Try to automate things through your programming.

**Step 5 – An important link of courses and documents:**

* <https://www.coursera.org/learn/machine-learning>– coursera machine learning.
* <https://www.udemy.com/complete-python-bootcamp/>- udemy python learning.
* [https://machinelearningmastery.com](https://machinelearningmastery.com/)- for blogs and concept.
* <https://www.analyticsvidhya.com/>- for blogs and competitions.
* <https://www.kaggle.com/>- for learning and competiiton.
* <https://medium.com/>- for reading blogs
* <https://www.3blue1brown.com/about/> -3 blue 1 brown for maths
* <https://www.khanacademy.org/> - for mathematics calculus.
* <https://github.com/yug95/MachineLearning> – github for machine learning papers.

For any other clarification reach to data analytics team they will surely help you out.

\*\*\*\*\*\*\*\*\*\* Happy Learning \*\*\*\*\*\*\*\*\*\*\*\*