Course: Introduction to Data Science
Start Date: March 2021
Coordinates: Flexible
Instructor: Dr. Raghunathan Ramakrishnan (ramakrishnan@tifrh.res.in)

Syllabus:

1) **Data Science:** Big Data, Facets of data (structured/unstructured data)

2) **Toolboxes:** Python libraries, SCIKIT-Learn, PANDAS

3) **Statistics:** Distributions, Outlier, Skewness, Pearson's/Spearman's/Kendall's coefficient, Kernel density

4) **Statistical Inference:** Hypothesis testing, Confidence Intervals

5) **Supervised Machine Learning:** What is machine learning? Learning curves, Support Vector Machines, Random Forest

6) **Regression:** Linear Regression, Logistic Regression

7) **Unsupervised Machine Learning:** Clustering, Case studies

8) **Big Data concepts:** Handling large data, Hadoop, Spark, NoSQL, Graph databases, Natural language processing, MapReduce

Required Text
1. *Introduction to Data Science. A Python Approach to Concepts, Techniques and Applications,* Laura Igual, Santi Segu, Springer (2017).
2. *Introducing Data Science,* Davy Cielen, Arno D. B. Meysman, Mohamed Ali, Manning (2016).

Evaluation Method: Assignment (50%), Presentations (50%)