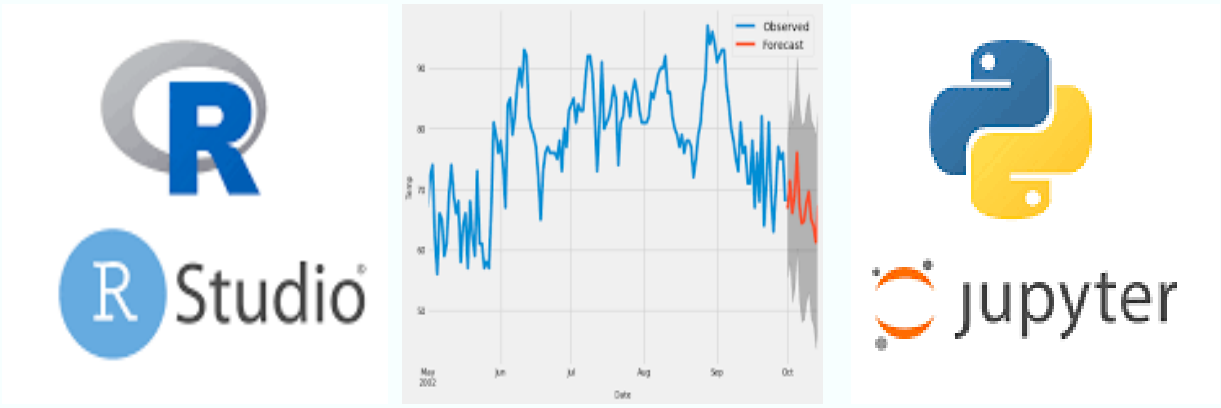
Information Brochure

PONDICHERRY UNIVERSITY
(A Central University)
Organizes

EXECUTIVE DEVELOPMENT PROGRAM
ON

DATA SCIENCE with R & PYTHON
A Statistical Analytics' Training Program
(During 2nd, 3rd, 4th and 5th Weekends of May 2020)



Hosted by

DEPARTMENT OF STATISTICS

Contact Details:
DR. TIRUPATHI RAO PADI
Professor & Training Coordinator
DEPARTMENT OF STATISTICS
Ramanujan School of Mathematical Sciences,
Pondicherry University, Puducherry - 605014, India
Phones: 0413 – 2654829 (Office), 9486492241, 9629862241 (Mobile)
Emails: puedpstpds@gmail.com, traopadipu@gmail.com, traopadi.sta@pondiuni.edu.in

1. ELIGIBILITY FOR PARTICIPATION:

Participants of the following categories are considered for getting the Statistical Training on Data Science and Analytics

- ***Young Faculty Members*** of the Colleges/Universities/Higher Learning Institutions
- ***Research Scholars*** who are pursuing Ph.D. programs,
- ***Employees of IT/Data Analytics***/ and other related fields
- ***Post graduate students of any discipline*** who are having subjects Mathematics/ Statistics/ Computer Science/Any Engineering course at their graduation level.

2. COURSE CONTENTS:

The contents consist of four modules. Each module is for two days with timings 8.30 am to 6.30 PM (10 hours) / 9.00 am to 7.00 PM (10 hours) in each day training so as total of 20 hours per module.

- Modules 1&2 are on R programming and Modules 3&4 are on Python Programming.
- ***Module-1***: Basic Data Science Techniques with R Programming
- ***Module-2***: Advanced Data Science & Data Analytics with R programming
- ***Module-3***: Essential Data Science Tools with Python
- ***Module-4***: Enhanced Techniques of Data Science & Data Analytics with Python

3. TIME AND DAYS OF TRAINING:

Training program is intended to conduct with four modules on four weekends in the month of May 2020 as follows.

Module No.	*Date & Day*	*Timings*	*Topic*
Module-1	09.05.2020 (Saturday)	8.30 AM to 6.30 PM	**Basic Data Science Techniques with R**
	10.05.2020 (Sunday)	9.00 AM to 7.00 PM	
Module-2	16.05.2020 (Saturday)	8.30 AM to 6.30 PM	**Advanced Data Science & Data Analytics with R**
	17.05.2020 (Sunday)	9.00 AM to 7.00 PM	
Module-3	23.05.2020 (Saturday)	8.30 AM to 6.30 PM	**Essential Data Science Tools with Python**
	24.05.2020 (Sunday)	9.00 AM to 7.00 PM	
Module-4	30.05.2020 (Saturday)	8.30 AM to 6.30 PM	**Enhanced Data Science & Data Analytics with Python**
	31.05.2020 (Sunday)	9.00 AM to 7.00 PM	

4. FEES & TRAINING CHARGES:

Course fees are based on the modules. Each module has different cost. The participants may choose the course modules as per their requirement. They have a choice of selecting the number of course modules. They have the option of getting training for one/two/three/all four modules by paying the respective mentioned fees.

No. Module	*Date & Day*	*Topic*	*Fees (Incl. GST)*	*Last Date for apply*
Module-1	09.05.2020 (Saturday)	**Basic Data Science Techniques with R**	**Rs. 4,000/-**	**07.05.2020**
	10.05.2020 (Sunday)			
Module-2	16.05.2020 (Saturday)	**Advanced Data Science & Data Analytics with R**	**Rs. 5,000/-**	**14.05.2020**
	17.05.2020 (Sunday)			
Module-3	23.05.2020 (Saturday)	**Essential Data Science Tools with Python**	**Rs. 5,000/-**	**21.05.2020**
	24.05.2020 (Sunday)			
Module-4	30.05.2020 (Saturday)	**Enhanced Data Science & Data Analytics with Python**	**Rs. 6,000/-**	**28.05.2020**
	31.05.2020 (Sunday)			

.

5. **MODE OF PAYMENT:**

The Selected candidates after their applications can pay the charges on line through net banking, NEFT/IMPS in the name of ***The Coordinator, STP-DSA*** payable at Indian Bank, Pondicherry University Branch, R.V. Nagar, Kalapet, Puducherry 605014. The savings bank account number is **6867639902**, Indian Bank, Pondicherry University Branch, IFSC: IDIB000P152.

6. **RESOURCE PERSONS:**

Resource persons for these training programs are from the faculty members of ***Pondicherry University*** and the ***Experts from Statistical Training, Analytics and Research Consulting Group (STAR Con. Group)***. This group consists of both free lance and permanent consultants of different agencies like Universities, Industries, Consultancy Firms and Corporate Sectors of Data Analytics. All are having very good data handling experience with both open source and proprietary softwares such as R, Python, SAS, SCI LAB, MATLAB, SPSS, MINITAB, etc. The course coordinator is ***Dr. Tirupathi Rao Padi, Professor, Department of Statistics, Ramanujan School of Mathematical Sciences, Pondicherry University***. He is having 32 years of teaching experience in the courses of Mathematical Statistics, Applied Statistics, Operational Research and Statistical Computing for both under graduate and post graduate students of different Indian universities.

7. **ORIENTATION OF THE TRAINING:**

Each module is having a separate training schedule. Participants have to register separately for each module by paying fees separately. Training program for each module will be for 20 hours schedule on Saturday and Sunday as per the mentioned timings. All the theoretical concepts of statistics will be trained by statistics faculty as live demonstration on computers. All the participants should have laptops with compatible operating system to work with latest versions of open source software R and Python. Simultaneous and live practice will be happened along with class room teaching. Participants will be supplied data sets on different platforms and working practice will be provided on the spot. Participants are suggested to attend with their own data sets pertaining to their specific objective of the study.

8. **COURSE OBJECTIVES:**

The prime objective of these courses is to prepare the data using community for handling multiple and heterogeneous tasks of Data Science. All the resource persons of the courses belong to the Statistics domain. They are having good understanding on the statistical techniques that are using for data science and data analytics. This training is more focused and keeping the specific interests of target groups such as graduate and post graduate students, Research Scholars, Faculty Members of different disciplines, data practitioners, statistical data consultants and many more categories of the similar professions. The latest requirements like Data Acquisition, Data Cleaning, Data Formatting, Database Administration, Database Management, Data Security, Data Updating, Data Analysis, Statistical Data Modeling, Predictive Modeling and Forecasting, Data Mining, Data Visualization, Business Intelligence, etc makes the data science more vital for handling the current needs of data consulting activities.

9. **VENUE OF THE TRAINING:**

Department of Statistics, Ramanujan School of Mathematical Sciences,

Pondicherry University, R.V. Nagar, Kalapet, Puducherry (UT), India-605014.

10. **ISSUING OF CERTIFICATES**:

Certificate of partipation along with result/grade in online exam after completion of course module will be issued to the successful candidates in "Statistical Training Program on Data Science & Analytics" duly signed by the competent authorities of *Executive Development Programme* - Pondicherry University.

11. **DEPARTMENT OF STATISTICS:**

Statistics department was established in the year 2006 as a separate entity form department of Mathematics. Department is having the sanctioned strength of 9 Faculty members. Currently it consists of 2 professors, 1 Associate Professor, and 4 Assistant professors. 2 positions for associate professor are vacant and the recruitment process was initiated. Faculty members are working on different areas of research namely Applied Probability, Distribution Theory, Stochastic Modeling, Optimization Techniques, Applied Operational Research, Multivariate Data Analysis, Biostatistics, Statistical Inference, Reliability theory, Statistical Quality Control, Survival Analysis, Applied Statistics, Sampling Theory, Designs of Experiments, etc. The venue of training programs is at department of Statistics, Pondicherry University (A Central University), Puducherry - 605014, India. The department is having the entire necessary infrastructure to provide effective statistical training programs.

12. **ABOUT EXECUTIVE DEVELOPMENT PROGRAM:**

Pondicherry University was established in 1988 under the act of parliament, Govt. of India. Executive Development Program of Pondicherry University with new guidelines was introduced from the Academic Year 2019-20. The objective of this program is to have industry university interaction to identify the needs of the industry so as possible remedial services shall be extended from the university. It is an imitative of the Pondicherry University to extend the community out reaching with different stake holders of the university. The activities like conduct of short term training programs, skill development workshops, knowledge sharing conferences, extending university intellectual interaction with outside users in applied domains are the targeted with different categories of people of university. As a part of this our university is coordinating the programs pertaining to academic, research, capacity building, consultancy, skill enhancement, updating the curriculum as per the societal needs, etc.

13. **Schedules of the programme:** the total program is having four separate schedules for each separate course module.

SCHEDULE OF MODULE-1

BASIC DATA SCIENCE & ANALYTICS WITH R		
Day /Session/ Time	**Course Contents**	**Details of the contents**
Day-1 (1st Saturday): 8.30 AM to 6.30 PM		
08.30 AM to 09.00 AM: Registration & Inauguration		
Day-I/ 08.30 AM - 09.00 AM	Registration &Inauguration	Declaration on Orientation of Workshop/ Skilled Training / Capacity Building by the coordinator and introduction of resource persons
Day-I/ Session-1 09.00AM to 11.00AM	Basic Statistics	1. Measures of Central Tendency 2. Measures of Dispersion, Skewness &Kurtosis 3. Ordinal/Positional Measures 4. Relative/Ratio Measures 5. Frequency Tables 6. Correlations and Regression 7. Association Measures
11.00 AM to 11.15 AM: Morning Tea Break		
Day-I/ Session-2 11.15AM to 1.15PM	Explanatory Data Analysis - EDA	1. Data Validation and Quality 2. Data Cleaning &Analyzing 3. Transformation to explore data 4. Patterns and models 5. Process & Visualize the data 6. Bar plot, Box plot, Correlation Plot
01.15 PM. to 02.00PM: Lunch Break		
Day-I/ Session-3 02.00 PM to 04.00PM	Introduction to R-I	1. R/ R-Studio as a statistical Software and Language 2. Functions in R 3. Packages in R 4. Data frames 5. Qualitative and Quantitative data Measures 6. Fundamental of the R Language 7. Basic commands in R Programming 8. Functions in R 9. Data preparations with R
04.00 PM to 04.15PM: Evening Tea Break		
Day-I Session-4 04.15PM to 05.45PM	Introduction to R-I	1. Importing Data into R 2. Exploring your dataset 3. Basic operations with a Data Frame 4. Filtering a Data Frame 5. Building Data frames 6. Merging Data Frames 7. Subscripts with Arrays & Lists 8. Plots and their interpretations
Day-I Session-5 05.45PM to 06.30PM	Participant's Interactions	Discussions & Case study

Day-2 (1st Sunday): 9.00 AM to 7.00 PM		
Day-II/ Session-1 09.00AM to 11.00AM	Statistical Modeling – Basic Concept	1. What is linear regression? 2. Why Linear Regression? 3. Bivariate data 4. Scatter plot 5. Measures of association – Covariance – Correlation coefficient 6. Simple linear regression – Fitting a regression line – Interval estimation & Prediction – Basic tests 7. What is Collinearity? 8. Concept and why Dummy Variables?
11.00 AM to 11.15 AM: Morning Tea Break		
Day-II/ Session-2 11.15AM to 01.15PM	Data Mining, Cleaning/ Wrangling	1. Databases handling 2. Fetch and retrieve the data 3. Connect various databases into R 4. Missing Value Mechanisms & Patterns 5. When can be Missing Values Ignored? 6. List-Wise & Pair-Wise Deletion 7. Missing Value Imputation Methods 8. Outliers
01.15 PM to 02.00PM: Lunch Break		
Day-II/ Session-3 02.00 PM to 04.00PM	Statistical Linear modeling with R Practicum	1. Statistical model building – Linear regression model with Business Problem 2. Validation of the model – Linear Regression 3. Interpretation of the summary on Linear Regression
04.00 PM to 04.15PM: Evening Tea Break		
Day-II Session-4 04.15 PM to 05.15PM	Statistical Logistic modeling with R Practicum	1. Statistical model building – Logistic Regression model with Business Problem 2. Validation of the model – Logistic Regression 3. Interpretation of the summary on Logistic Regression
Day-II, Session-5, 5.15PM to 6.00 PM	Interaction	1. Discussion on the Data sets of the participants, Statistical Data Analysis Planning
Day-II, Session-6, 6.00 PM to 6.30 PM	Examination & Screening Test	1. Online examination, evaluations, and declaration of results and grades
06.30 PM to 7.00 PM: Valedictory, Issue of Certificates & Feedback		

SCHEDULE OF MODULE-2:

ADVANCED DATA SCIENCE & DATA ANALYTICS WITH R		
Day /Session/ Time	**Course Contents**	**Details of the contents**
Day-1 (2nd Saturday): 8.30 AM to 6.30 PM		
08.30 AM to 09.00 AM: Registration & Inauguration		
Day-I/ 08.30 AM - 09.00 AM	Registration &Inauguration	Declaration on Orientation of Workshop/ Skilled Training / Capacity Building by the coordinator and introduction of resource persons
Day-I/Session-1	Basic Statistics	1. Measures of Central Tendency

09.00AM to 11.00AM		2. Measures of Dispersion, Skewness& Kurtosis 3. Ordinal/Positional Measures 4. Relative/Ratio Measures 5. Frequency Tables 6. Correlations and Regression 7. Association Measures
11.00 AM to 11.15 AM: Morning Tea Break		
Day-I/Session-2 11.15AM to 1.15PM	Data Science	1. Introduction: What is Data Science? - Current landscape of perspectives - Skill sets needed 2. Data Science hype – Why now? 3. Supervised learning 4. Unsupervised Learning 5. Topics covered in Supervised learning 6. Topics covered in unsupervised learning
01.15 PM. to 02.00PM: Lunch Break		
Day-I/Session-3 02.00 PM to 04.00PM	Requirements of Good Machine learning System and ML in R – I	1. Data preparations with R for ML 2. Analysing data 3. Patterns identification 4. Prediction 5. Conclusion
04.00 PM to 04.15PM: Evening Tea Break		
Day-I/ Session-4 04.15PM to 05.45PM	ML in R – II	1. Classification 2. Regression
Day-I/Session-5 5.45 PM to 6.30 PM	Interaction of Participants	1. Discussions, Case studies, data sets of participants, Statistical Analysis Planning, etc.
Day-2 (2nd Sunday): 9.00 AM to 7.00 PM		
Day-II/Session-1 09.00AM to 11.00AM	ML in R – III Un-Supervised Learning - I	1. Concepts in Unsupervised learning 2. Unstructured data 3. Why Linear Regression? 4. Logistic regression 5. Clustering
11.00 AM to 11.15 AM: Morning Tea Break		
Day-II/Session-2 11.15AM to 01.15PM	ML in R – IV Un-Supervised Learning-II	1. Text Mining 2. Sentimental analysis 3. Targeted Marketing
01.15 PM to 02.00PM: Lunch Break		
Day-II/Session-3 02.00 PM to 04.00PM	Data Analytics with R-1	1. Test Dataset 2. Validation dataset 3. Training Dataset
04.00 PM to 04.15PM: Evening Tea Break		
Day-II/ Session-4 04.15 PM to 05.15PM	Data Analytics with R-2	1. Statistical model – Logistic Regression model with Business Problem 2. Validation of the model – Logistic Regression 3. Interpretation of the summary on Logistic Regression
Day-II/ Session-5 5.15PM to 6.00 PM	Interaction	1. Discussion on the Data sets of the participants, Statistical Data Analysis Planning
Day-II/ Session-6 6.00 PM to 6.30 PM	Examination & Screening Test	1. Online examination, evaluations, and declaration of results and grades
06.30 PM to 7.00 PM: Valedictory, Issue of Certificates & Feedback		

SCHEDULE OF MODULE-3

ESSENTIAL DATA SCIENCE TOOLS WITH PYTHON		
Day /Session/ Time	**Course Contents**	**Details of the contents**
Day-1 (3rdSaturday): 8.30 AM to 6.30 PM		
08.30 AM to 09.00 AM: Registration & Inauguration		
Day-I/ 08.30 AM - 09.00 AM	Registration &Inauguration	Declaration on Orientation of Workshop/ Skilled Training / Capacity Building by the coordinator and introduction of resource persons
Day-I/Session-1 09.00AM to 11.00AM	Basic Statistics	1. Measures of Central Tendency 2. Measures of Dispersion, Skewness & Kurtosis 3. Ordinal/Positional Measures 4. Relative/Ratio Measures 5. Frequency Tables 6. Correlations and Regression 7. Association Measures
11.00 AM to 11.15 AM: Morning Tea Break		
Day-I/Session-2 11.15AM to 1.15PM	Explanatory Data Analysis - EDA	1. Data Validation and Quality 2. Data Cleaning &Analyzing the data 3. Transformation to explore data 4. Patterns and models 5. Process & Visualize the data 6. Bar plot, Box plot, Correlation Plot
01.15 PM. to 02.00PM: Lunch Break		
Day-I/Session-3 02.00 PM to 04.00PM	Introduction to Python	1. Significance and installation of Python 2. Values, variables and statements 3. Conditional executions 4. Iterations like while, nested, for, infinite loops 5. Functions 6. Lists, objects, custom types, 7. Imports and exports of files in Python
04.00 PM to 04.15PM: Evening Tea Break		
Day-I/ Session-4 04.15PM to 05.45PM	Basic Model Building Concepts with Python	1. What & why Linear regression? 2. What & why Logistic Regression? 3. Bivariate data 4. Scatter plot 5. Measures of association Covariance/ Correlation coefficient 6. Simple linear regression – Fitting a regression line – Interval estimation & Prediction – Basic tests 7. What is Collinearity? 8. Concept and why Dummy Variables?
Day-I/Session-5 05.45PM to 06.30PM	Participant's Interactions	1. Discussions & Case study
Day-2 (3rdSunday): 9.00 AM to 7.00 PM		
Day-II/Session-1 09.00AM to 11.00AM	Data Cleaning/ Wrangling	1. Missing Value Mechanisms & Patterns 2. When can be Missing Values Ignored? 3. Case Analysis 4. List-Wise & Pair-Wise Deletion 5. Missing Value Imputation Methods

		6. Outliers
11.00 AM to 11.15 AM: Morning Tea Break		
Day-II/Session-2 11.15AM to 01.15PM	Data handling	1. Databases handling 2. Fetch and retrieve the data 3. Connect various databases into Python
01.15 PM to 02.00PM: Lunch Break		
Day-II/Session-3 02.00 PM to 04.00PM	Statistical Linear modelling with Python Practicum	1. Statistical model building – Linear regression model with Business Problem 2. Validation of the model – Linear Regression Interpretation of the summary on Linear Regression
04.00 PM to 04.15PM: Evening Tea Break		
Day-II; Session-4 04.15 PM to 06.15PM	Statistical Logistic modeling with python Practicum	1. Statistical model building – Logistic Regression model with Business Problem 2. Validation of the model – Logistic Regression Interpretation of the summary on Logistic Regression
Day-II/ Session-5 5.15 PM to 6.00 PM	Interaction	1. Discussion on the Data sets of the participants, Statistical Data Analysis Planning
Day-II/ Session-6 6.00 PM to 6.30 PM	Examination & Screening Test	1. Online examination, evaluations, and declaration of results and grades
06.30 PM to 7.00 PM: Valedictory, Issue of Certificates & Feedback		

SCHEDULE OF MODULE-4

ENHANCED TECHNIQUES OF DATA SCIENCE & DATA ANALYTICS WITH PYTHON		
Day /Session/ Time	**Course Contents**	**Details of the contents**
Day-1 (4th Saturday): 8.30 AM to 6.30 PM		
08.30 AM to 09.00 AM: Registration & Inauguration		
Day-I/ 08.30 AM - 09.00 AM	Registration &Inauguration	Declaration on Orientation of Workshop/ Skilled Training / Capacity Building by the coordinator and introduction of resource persons
Day-I/Session-1 09.00AM to 11.00AM	Python programming	1. Strings 2. Classes 3. Dates and its challenges in conversions 4. Operators 5. Data extraction and cleaning in python
11.00 AM to 11.15 AM: Morning Tea Break		
Day-I/Session-2 11.15AM to 1.15PM	Statistics	1. Normal Distribution, Binominal & Poisson distribution 2. Testing of Hypothesis, Null hypothesis, Alt hypothesis, p-value 3. Z-test, Chi-square test, F-test, t-test 4. Curve Fitting &Principle of Least square 5. S.E, one & two-tailed test, parameter and statistics, sample and population 6. Theory of Estimation

		7. Statistical Inference
01.15 PM. to 02.00PM: Lunch Break		
Day-I/Session-3 02.00 PM to 04.00PM	FORECASTING ANAYTICS	1. Forecasting Analytics – 1 2. Why forecasting 3. Data collection 4. Data Quality 5. Time series components 6. Additive and Multiplicative model
04.00 PM to 04.15PM: Evening Tea Break		
Day-I/ Session-4 04.15PM to 05.45PM	TIME SERIES ANALYSIS	1. Application with Python in Time series Practical Example
Day-I/Session-5 05.45PM to 06.30PM	Participant's Interactions	1. Discussions & Case study
Day-2 (4thSunday): 9.00 AM to 7.00 PM		
Day-II/Session-1 09.00AM to 11.00AM	Python	1. Important Libraries 2. Object creation 3. Data frames 4. Basic statistics with Python commands like mean, histogram 5. Joining, Reshaping, Stack, Grouping 6. Pivot tables, plotting, selection, operations
11.00 AM to 11.15 AM: Morning Tea Break		
Day-II/Session-2 11.15AM to 01.15PM	Data Visualization	1. What is Data visualization? 2. Why Data visualization in Data Science 3. DV in excel 4. DV in Tableau 5. Basic Tableau 6. Example in DV with Excel and tableau
01.15 PM to 02.00PM: Lunch Break		
Day-II/Session-3 02.00 PM to 04.00PM	Practicum ML with Python	1. End to End project 2. Define Problem 3. Prepare Data 4. Evaluate Algorithms 5. Improve Results 6. Present Results
04.00 PM to 04.15PM: Evening Tea Break		
Day-II; Session-4 04.15 PM to 05.30 PM	Practicum KNN with Python	1. k-nearest neighbours 2. practical examples with KNN 3. end to end KNN project with Python
Day-II/ Session-5 5.30 PM to 6.00 PM	Interaction	1. Discussion on the Data sets of the participants, Statistical Data Analysis Planning
Day-II/ Session-6 6.00 PM to 6.30 PM	Examination & Screening Test	1. Online examination, evaluations and declaration of results and grades
06.30 PM to 7.00 PM: Valedictory, Issue of Certificates & Feedback		

APPLICATION FORM

(Click on the following link for online enrolment)

https://docs.google.com/forms/d/1IgR05cn3VwNOLCgtYWJRxxfRpPfSjqIS5BCVNf_MG3A/edit

DATA SCIENCE with R & PYTHON

A STATISTICAL ANALYTICS TRAINING PROGRAM

Activity Under

EXECUTIVE DEVELOPMENT PROGRAM

PONDICHERRY UNIVERSITY - DEPARTMENT OF STATISTICS

Name of the Participant	:				
Gender (Male/Female)	:				
Educational Qualifications	:				
Current Status PG Student/ Research Scholar/ Young Faculty/ IT Employee/Any Other (Specify)	:				
Address of the Current Affiliation	:				
Email Id(s)	:				
Contact Number(s)	:				
Mailing/ Correspondence Address	:				
Preferred Course Module to Get Training (Tick on either one/two/ three/ all course modules)	:	Module-1 with R Programming	Module-2 with R Programming	Module-3 with PYTHON	Module-4 with PYTHON
Dates of Participation	:	9^{th},10^{th} May 2020 Sat.day, Sunday	16^{th},17^{th}May 2020 Sat.day, Sunday	23^{rd},24^{th}May2020 Sat.day, Sunday	30^{th},31^{st} May 2020 Sat.day, Sunday
Mode of Payment (*Online transfer NEFT/ D.D.)	:				
*NEFT/DD in favour of **The Coordinator, STP-DSA**, Account Number: **6867639902**, payable at Indian Bank, Pondicherry University branch, Puducherry, IFSC: IDIB000P152)					
Details of Payment	:				
Amount Paid					
Module-1:R Rs.4,000/-	:				
Module-2:R Rs.5,000/-	:				
Module-3: PYTHON Rs.5,000/-	:				
Module-4:PYTHON Rs.6,000/-	:				

Signature of the Candidate