



**Proposal for
“Aspiration”
an
Experiential Workshop
On
Data Analytics using R**

By

**AUROPATH
(Auro Service Trust)**

**Society House, Ho. 11, St. Martin Street
PUDUCHERRY
9826268541**



About the Workshop:-

The era of big data drastically changed the requirements for extracting meaning from business data. In the world of relational databases, administrators easily generated reports on data contents for business use, but these provided little or no broad business intelligence. For that, they employed data warehouses, but data warehouses generally cannot handle the scale of big data cost-effectively. While data warehouses are certainly a relevant form of data analytics, the term data analytics is slowly acquiring a specific subtext related to the challenge of analyzing data of massive volume, variety, and velocity.

Today the field of data analytics is growing quickly, driven by intense market demand for systems that tolerate the intense requirements of big data, as well as people who have the skills needed for manipulating data queries and translating results.

The significance of Data and its effective analysis is unavoidable to have a good piece of research and to make a fruitful decision. From students to scholars, individual researchers to research institutes throughout the world come across this data analysis part to carry out their research.

Statistics and statistical analysis plays a crucial role in research and analyzing data. The aim of this workshop is to provide hands on training in understanding the concepts in data science and statistical analysis using R programming which is one of the most widely used open source package in the field of Data Science across the world. This workshop will help the participants in applying proper research methodology and improving their analysis by applying various statistical techniques using R programming.

About Auropath :-

A unit of auroservice, works in close collaboration with SAS to help colleges, universities and centres of higher education to improve their working and performance at every level and in every field.

AUROPATH assists higher education institutions in the accreditation and recognition of National Assessment and Accreditation Council (NAAC) and National Board of Accreditation (NBA).

Some of the post-NAAC/NBA consultancy programmes are,

- Curricular development aspects
- Teaching, learning and evaluation
- Research, consultancy and extension
- Infrastructure and learning resources
- Student support and progression
- Governance, leadership and management
- Innovation and best practices
- Institutional social responsibility
- Innovative technology development
- Faculty development
- Character and citizenship building

Potential Beneficiary:-

Academicians, Faculties, Current and Prospective Research Scholars, PG Students and UG Students from all disciplines. A maximum of 60 candidates will be selected for the training programme on First-Come-First-Serve basis.

Objectives of the Workshop:-

- Get familiar with R programming language;
- Understand statistical analysis using R programming;
- Explore how to statistically analyze research data using R programming
- Explore how to represent data using basic graphs in R.
- How to prepare presentation, document, web pages, blogs in R

**Delegates Fee:-**

A nominal fee of ~~Rs 2000/-~~ ^{Rs 500/-} covering the cost of research material, lunch and refreshments' for ~~three~~ ^{two} days shall be payable per participant. In case of residential participants the institute will facilitate accommodation as per the requirement of participants only. The cost of which will be done by the participants. Fees once paid will not be refunded. However, a substitute delegate is acceptable.

Resource Persons:-

Meenakshi Ramesh Patel, ICT Mentor, Auropath, Sri Aurobindo Society, Pondicherry.
M. Tech(CT), Specialization in Data Analytics.

Program Schedule**❖ Day 1**

- Introduction to R
- Intermediate R
- Introduction to Tidyverse
- Importing Data in R and Cleaning data
- Exploratory Data Analysis

❖ Day 2

- Data Manipulation in R with dplyr
- Data Visualization with ggplot2
- Correlation and Regression
- Reporting with R Markdown

Unique Features of the program:-

Basics of Data Analysis and Visualization will be covered through R- programming. Hands on experience on R. E-material will be provided.

Prerequisites:-

- Participants should come with their Laptops fully charged(Along with charger).
- Participants should download R Gui and R Studio in their machines. [Instructions for downloading will be shared]



Short introduction to the topics covered:-

Introduction to R

In this introduction to R, you will learn the basics of this beautiful open source language, including factors, lists and data frames. With the knowledge gained in this workshop, you will be ready to undertake your first very own data analysis.

Intermediate R

The intermediate R is the logical next stop on your journey in the R programming language. In this R workshop you will learn about conditional statements, loops and functions to power your own R scripts. Next, you can make your R code more efficient and readable using the apply functions. Finally, the utilities chapter gets you up to speed with regular expressions in the R programming language, data structure manipulations and times and dates. This R workshop will allow you to learn R and take the next step in advancing your overall knowledge and capabilities while programming in R.

Introduction to Tidyverse

This is an introduction to the programming language R, focused on a powerful set of tools known as the "tidyverse". In this part you'll learn the intertwined processes of data manipulation and visualization through the tools dplyr and ggplot2. You'll learn to manipulate data by filtering, sorting and summarizing a real dataset in order to answer exploratory questions. You'll then learn to turn this processed data into informative line plots, bar plots, histograms, and more with the ggplot2 package. This gives a taste both of the value of exploratory data analysis and the power of tidyverse tools. This is a suitable introduction for people who have no previous experience in R and are interested in learning to perform data analysis.

Importing Data in R and Cleaning data

Importing data into R to start your analyses—it should be the easiest step. Unfortunately, this is almost never the case. Data come in all sorts of formats, ranging from CSV and text files and statistical software files to databases and HTML data. Knowing which approach to use is key to getting started with the actual analysis. In this course, you will get started with learning how to read CSV and text files in R. You will then cover the readr and data.table packages to easily and efficiently import flat file data. After that you will learn how to read XLS files in R using readxl and gdata.

Running exciting analyses on interesting datasets is the dream of every data scientist. But first, some importing and cleaning must be done.

Exploratory Data Analysis

Scientists seek to answer questions using rigorous methods and careful observations. These observations—collected from the likes of field notes, surveys, and experiments—form the backbone of a statistical investigation and are called data. Statistics is the study of how best to collect, analyze, and draw conclusions from data. It is helpful to put statistics in the context of a general process of investigation: 1) identify a question or problem; 2) collect relevant data on the topic; 3) analyze the data; and 4) form a conclusion. In this part, you'll focus on the first two steps of the process.



Data Manipulation in R with dplyr

In this interactive session, you will learn how to perform sophisticated dplyr techniques to carry out your data manipulation with R. First you will master the five verbs of R data manipulation with dplyr: select, mutate, filter, arrange and summarise. Next, you will learn how you can chain your dplyr operations using the pipe operator of the magrittr package. In the final section, the focus is on practicing how to subset your data using the group_by function, and how you can access data stored outside of R in a database. All said and done, you will be familiar with data manipulation tools and techniques that will allow you to efficiently manipulate data.

Data Visualization with ggplot2

The ability to produce meaningful and beautiful data visualizations is an essential part of your skill set as a data scientist. This part, the first R data visualization session in the series, introduces you to the principles of good visualizations and the grammar of graphics plotting concepts implemented in the ggplot2 package. ggplot2 has become the go-to tool for flexible and professional plots in R. Here, we'll examine the first three essential layers for making a plot - Data, Aesthetics and Geometries. By the end of the session you will be able to make complex exploratory plots.

When your dataset is represented as a table or a database, it's difficult to observe much about it beyond its size and the types of variables it contains. In this session, you'll learn how to use graphical and numerical techniques to begin uncovering the structure of your data. Which variables suggest interesting relationships? Which observations are unusual? By the end of the session, you'll be able to answer these questions and more, while generating graphics that are both insightful and beautiful.

Correlation and Regression

Ultimately, data analysis is about understanding relationships among variables. Exploring data with multiple variables requires new, more complex tools, but enables a richer set of comparisons. In this session, you will learn how to describe relationships between two numerical quantities. You will characterize these relationships graphically, in the form of summary statistics, and through simple linear regression models.

Reporting with R Markdown

Learn how to write a data report quickly and effectively with the R Markdown package, and share your results with your friends, colleagues or the rest of the world. Learn how you can author your own R Markdown reports, and how to automate the reporting process so that you have your own reproducible reports. By the end of the interactive data analysis reporting sessions, you will be able to generate reports straight from your R code, documenting your work — and its results — as an HTML, PDF, slideshow or Microsoft Word document.