Share on your Social Media

Data Science with R Tutorial

Published On: August 9, 2024

Data Science with R Tutorial

R is an interactive environment for data science work, not merely a programming language. Explore in detail in this Data Science with R tutorial.

[Download Data Science with R Tutorial PDF](#)

Related Courses at SLA

Introduction to Data Science with R Tutorial

The most in-demand profession today is data science. This is due to the urgent requirement for data analysis and insight construction.

Several crucial technologies are needed to churn the raw data to accomplish this. R is one of the computer languages that offers a sophisticated

Related Posts

Tableau Developer Salary in Chennai

Published On: October 12, 2024

Introduction A Tableau Developer designs, develops, and maintains dashboards and visualizations using Tableau software. Key...

environment for information processing, visualization, and research.

We cover the following in this Data Science with R Tutorial:

- Overview of Data Science with R
- Applications of R in Data Science
- R Libraries for Data Science
- Data Visualization with R

Overview of Data Science with R

R is a programming language that lets users explore, model, and visualize data through its objects, operators, and functions. R is a data analysis tool.

R is used in data science to handle, store, and analyze data. It can be applied to statistical modeling and data analysis.

[Data Science with R Interview Questions](#)

Features of R for Data Science

Below are the key features of R for data science:

- R is a statistical programming language and environment that supports statistical computing and graphics.
- It contains a lot of characteristics that are helpful for representation and statistical analysis.
- It offers a lot of user-friendly packages for carrying out activities.
- Among the well-known R IDEs are Rstudio, RKward, R Commander, and so forth.
- Numerous libraries and packages exist, such as ggplot2, caret, etc.
- In data science, it is primarily utilized for intricate data analysis.

Applications of R in Data Science

Leading Organizations Using R for Data Science:

Social network analytics: It makes use of R to build connections between users and acquire insights on their behavior.

Carrying out various analytical tasks: R is used by the Google Flu Trends project to examine trends and patterns in searches related to the flu.

Analytical solutions: R is also used by IBM to create a range of analytical solutions. IBM Watson, an open computing platform, has made use of R.

Charting components: Uber uses the R package shiny. R was used in the development of Shiny, an interactive online application that embeds interactive visuals.

R Libraries for Data Science

R libraries most frequently used in data science

Dplyr: The dplyr package is used for data analysis and wrangling tasks. We utilize this package to make several R functions for the DataFrame easier to use.

These five features are actually the foundation of Dplyr. Both local data frames and distant database tables are supported. It may be necessary for you to:

- Choose certain data columns.
- Sort your data by filtering out particular rows.
- Sort your data into rows according to order.
- Modify your data frame so that it has additional columns.
- In some way, summarize sections of your data.

Ggplot2: The most well-known visualization library in R is called ggplot2. It offers an attractive collection of interactive graphics.

- The ggplot2 library implements a "grammar of graphics."
- By articulating the connections between the characteristics of data and their graphical representation, this method provides us with a logical means of creating visualizations.

Esquisse: This package has enabled R users to access Tableau's most crucial functionality. You may quickly complete your visualization by simply dragging and dropping.

This improves upon ggplot2. It enables us to create histograms, bar graphs, curves, and scatter plots. The graph can then be exported, or we can obtain the code that generated it.

Mlr: This library nearly possesses every significant and practical algorithm needed to do machine learning jobs. It is the extensible framework for survival analysis, regression, clustering, and multi-classification.

Tidyr: We utilize the Tidyr program to clean and organize our data.

When every variable is a column and every row is an observation, we say that the data is tidy.

Shiny: Shiny is a useful tool for sharing your content with those in your immediate vicinity and facilitating their visual exploration. It is the greatest ally of a data scientist.

Caret: The abbreviation Caret stands for both classification and regression training. You can model intricate regression and classification issues with this function.

E1071: The E1071 package is widely used to implement several types of miscellaneous functions, including clustering, Fourier transform, naive bayes, and support vector machines.

Data Science with R Syllabus PDF

Data Visualization with R Programming

The R programming language is intended for use in scientific research, graphical data analysis, and statistical computation. It is typically chosen for data visualization since, because of its packages, it provides flexibility and little coding.

Types of Data Visualization

R provides a variety of visualizations, some of which are as follows:

Bar Plot

Bar plots come in two varieties: horizontal and vertical, and they show data points as bars of varying lengths that correspond to the value of the corresponding data item.

Plotting of continuous and categorical variables is their usual use. Bar plots that are horizontal or vertical can be obtained by changing the horizontal option to true or false, accordingly.

Example: Horizontal Bar Plot with R

barplot(airquality$Ozone,

   main = 'Ozone Concenteration in air',

   xlab = 'ozone levels', horiz = TRUE)

Example 2: Vertical Bar Plot with R

barplot(airquality$Ozone, main = 'Ozone Concenteration in air',

   xlab = 'ozone levels', col ='blue', horiz = FALSE)

The following possibilities are shown by bar plots:

- To conduct a comparative analysis of the data set's several data categories.
- To examine how a variable has changed over a period of months or years.

Histogram

A histogram and a bar chart are similar in that they both utilize bars of different heights to show the distribution of data.

On the other hand, values in a histogram are arranged into successive intervals known as bins. These adjustable-sized bins are used in histograms to organize and display continuous values.

> **Example:**
>
> *data(airquality)*
>
> *hist(airquality$Temp, main ="La Guardia Airport's*
>
> *Maximum Temperature(Daily)",*
>
> *xlab ="Temperature(Fahrenheit)",*
>
> *xlim = c(50, 125), col ="yellow",*
>
> *freq = TRUE)*

The following situations involve the use of histograms:

- To confirm that the data are distributed equally and symmetrically.
- To find values that deviate from expectations.

Box Plot

A boxplot is used to visually represent the given data's statistical summary. Data points such as the lowest and highest, the median, the first and third quartiles, and the interquartile range are displayed in a boxplot.

> **Example**
>
> *data(airquality)*
>
> *boxplot(airquality$Wind, main = "Average wind speed*
>
> *at La Guardia Airport",*
>
> *xlab = "Miles per hour", ylab = "Wind",*
>
> *col = "orange", border = "brown",*
>
> *horizontal = TRUE, notch = TRUE)*

The following code can also be used to create multiple box plots at once:

boxplot(airquality[, 0:4],

   main ='Box Plots for Air Quality Parameters')

Uses for box plots include:

- To provide a thorough statistical explanation of the data using a visual aid.
- To determine which outlier points fall outside the data's interquartile range.

Scatter Plot

Numerous points on a Cartesian plane make up a scatter plot. Every point indicates the value that two parameters have taken, making it simple to see how they relate to one another.

> **Example**
>
> *data(airquality)*
>
> *plot(airquality$Ozone, airquality$Month,*
>
>   *main ="Scatterplot Example",*
>
>   *xlab ="Ozone Concentration in parts per billion",*
>
>   *ylab =" Month of observation ", pch = 19)*

The following situations involve the usage of scatter plots:

- To demonstrate whether two sets of bivariate data are associated.
- To gauge the direction and strength of this kind of interaction.

Heat Map

A heatmap is a graphical data representation that uses colors to show the matrix's value. Heatmaps are plotted using the heatmap() function.

Syntax: heatmap(data)

Parameters: data: Setting parameters information

It displays matrix data, including row and column values.

Return: A heatmap is created by this function.

> **Example**
>
> *data <- matrix(rnorm(50, 0, 5), nrow = 5, ncol = 5)*
>
> *colnames(data) <- pasteo("col", 1:5)*
>
> *rownames(data) <- pasteo("row", 1:5)*
>
> *heatmap(data)*

Data Science with R Training

Map Visualization in R

Here, we're utilizing the R programming language and the maps package to visualize and show geographic maps.

install.packages("maps")

Example

data <- read.csv("worldcities.csv")

df <- data.frame(data)

library(maps)

map(database = "world")

points(x = df$lat[1:500], y = df$lng[1:500], col = "Red")

3D Graphs in R

The preps() function, which creates 3D surfaces in perspective view, will be used in this instance. Drawing perspective plots of a surface over the x–y plane is possible using this function.

Syntax: persp(x, y, z)

Parameter: The arguments that this function takes are x, y, and z, where x and y are vectors that

indicate the location along the x- and y-axes, respectively. The height of the surface in the matrix z will be the z-axis.

Return Value: When projecting 3D coordinates (x, y, z) into the 2D plane using homogeneous 4D coordinates (x, y, z, t), persp() yields the viewing transformation matrix.

Example

```
cone <- function(x, y){

sqrt(x ^ 2 + y ^ 2)

}

# prepare variables.

x <- y <- seq(-1, 1, length = 30)

z <- outer(x, y, cone)

persp(x, y, z,

main="Perspective Plot of a Cone",

zlab = "Height",

theta = 30, phi = 15,

col = "orange", shade = 0.4)
```

Data Science with R Skills Salary

Benefits of R Data Visualization

Compared to alternative data visualization technologies, R offers the following benefits:

- R provides a large number of visualization libraries and a wealth of online documentation on how to use them.
- R also provides multipanel charts and 3D models for data display.
- R makes it simple to alter the axes, fonts,

legends, annotations, and labels in our data visualizations.

Application Areas

- The data's analytical conclusions.
- Health monitoring equipment to detect variations in blood pressure, cholesterol, and other parameters.
- To find recurrent themes and patterns in marketing and customer data.
- Meteorologists evaluate global weather patterns.
- Traffic monitoring and trip time estimation are two uses of visualization seen in real-time maps and geo-positioning systems.

Conclusion

We hope you have gained basic knowledge of R and data science with this Data Science with R tutorial. Fine-tune your skills with our **data science with R training in Chennai**.

Share on your Social Media

Softlogic Academy

Softlogic Systems

KK Nagar [Corporate Office]

Navigation

About Us

Blog Posts

Careers

Contact

Placement Training

Corporate Training

Hire With Us

Job Seekers

No.10, PT Rajan Salai, K.K. Nagar, Chennai – 600 078.
Landmark: Karnataka Bank Building
Phone: +91 86818 84318
Email: enquiry@softlogicsys.in
Map: Google Maps Link

OMR

No. E1-A10, RTS Food Street
92, Rajiv Gandhi Salai (OMR), Navalur, Chennai - 600 130.
Landmark: Adj. to AGS Cinemas
Phone: +91 89256 88858
Email: info@softlogicsys.in
Map: Google Maps Link

Courses

- Python
- Software Testing
- Full Stack Developer
- Java
- Power BI
- Clinical SAS
- Data Science
- Embedded
- Cloud Computing
- Hardware and Networking
- VBA Macros
- Mobile App Development
- DevOps

- SLA's Recently Placed Students
- Reviews
- Sitemap

Important Links

- Disclaimer
- Privacy Policy
- Terms and Conditions

Social Media Links

Review Sources

- Google
- Trustpilot
- Glassdoor
- Mouthshut
- Sulekha
- Justdial
- Ambitionbox
- Indeed
- Software Suggest
- Sitejabber