Sangy Academy

www.sangyacademy.cQom

🧠 **Course Title: Introduction to Data Analysis with R**

Duration: 3 Months
Schedule: 2 lessons/week (2 hours per lesson)
Total Lessons: 24
Prerequisite: Basic proficiency in Python

🔍 Course Objectives

By the end of this course, participants will be able to:

- Use R and RStudio effectively for data analysis tasks

- Import, clean, manipulate, and visualize data using tidyverse packages

- Perform exploratory data analysis (EDA)

- Understand and apply basic statistical methods in R

- Communicate data insights using reports and dashboards

📚 Course Outline

🔹 Module 1: Getting Started with R (Weeks 1–2)

Lesson 1 – Introduction to R and RStudio

- Why R for data analysis

- R vs Python for data tasks

- Installing R and RStudio

- Using RStudio interface

Lesson 2 – R Language Basics

- Vectors, data types, and indexing

- Control structures: if, for, while

- Writing simple functions

Lesson 3 – Working with R Scripts and Projects

- R scripts and markdown

- RStudio projects and file management

- Basic input/output and using the console

Lesson 4 – Introduction to Packages and Help System

- Installing and loading packages

- Using `help()` and `?` for documentation

- Introduction to `tidyverse`

◆ **Module 2: Data Manipulation with tidyverse (Weeks 3–5)**

Lesson 5 – Data Frames and tibbles

- Creating and viewing data frames

- Tibbles vs data frames

- `readr` and `readxl` for importing data

Lesson 6 – Data Cleaning with `dplyr`

- `filter()`, `select()`, `mutate()`, `arrange()`, `rename()`

- Chaining with %>% (pipe)

Lesson 7 – Grouped Summaries with `group_by()` and `summarise()`

- Aggregating data

- Handling missing data (`na.rm`, `is.na`)

Lesson 8 – Data Reshaping with `tidyr`

- `pivot_longer()` and `pivot_wider()`

- `separate()` and `unite()` for string manipulation

◆ **Module 3: Data Visualization with ggplot2 (Weeks 6–7)**

Lesson 9 – Grammar of Graphics and Introduction to `ggplot2`

- Aesthetic mappings, layers, and themes

- Scatter plots and bar plots

Lesson 10 – Advanced ggplot2 Visualizations

- Histograms, boxplots, violin plots

- Customizing labels, titles, colors, legends

- Faceting (`facet_wrap`, `facet_grid`)

Lesson 11 – Time Series and Line Plots

- Plotting over time

- Formatting time-based data with `lubridate`

Lesson 12 – Combining Plots and Saving Output

- `patchwork` or `cowplot` for multiple plots

- Exporting plots to image or PDF

◆ **Module 4: Exploratory Data Analysis (EDA) (Weeks 8–9)**

Lesson 13 – EDA Concepts and Strategies

- Identifying types of variables

- Asking good analytical questions

Lesson 14 – Descriptive Statistics in R

- Mean, median, standard deviation

- Frequency tables and proportions

Lesson 15 – Outliers, Distributions, and Data Quality

- Detecting outliers

- Skewness, kurtosis, normality checks

Lesson 16 – Practical EDA Case Study

- Walkthrough of real-world dataset

- Cleaning, exploring, visualizing

◆ **Module 5: Basic Statistics for Data Analysis (Weeks 10–11)**

Lesson 17 – Hypothesis Testing in R

- t-tests, chi-square tests
- Using `t.test()` and `chisq.test()`

Lesson 18 – Correlation and Linear Regression

- Pearson and Spearman correlation
- Simple linear regression with `lm()`

Lesson 19 – ANOVA and Categorical Data Analysis

- One-way ANOVA
- Using `aov()` and interpreting results

Lesson 20 – Introduction to Sampling and Bootstrapping

- Random sampling
- Confidence intervals using `boot` or `infer`

◆ **Module 6: Reporting and Reproducibility (Weeks 12)**

Lesson 21 – R Markdown Basics

- Creating R Markdown documents
- Embedding code and plots in reports

Lesson 22 – Parameterized Reports and Templates

- Dynamic reporting
- Sharing insights with stakeholders

Lesson 23 – Project Presentation Preparation

- Group/individual project work
- Creating final reports and visualizations

Lesson 24 – Final Presentations and Course Wrap-up

- Present analysis projects
- Q&A, review, next steps in learning R

🛠️ Tools Used

- R & RStudio

- tidyverse (dplyr, tidyr, ggplot2, readr)

- lubridate, infer, broom

- R Markdown, knitr

🧪 Capstone Project

By the end of the course, each participant will complete a data analysis project using a real-world dataset, including:

- Data wrangling

- EDA and visualization

- Statistical analysis

- Reporting via R Markdown