

# Getting Started with Python (Google Colab) and R

Prepared by

**Aditya Korekallu Srinivasa**

Scientist (Senior Scale)

ICAR-Indian Agricultural Research Institute  
New Delhi

April 24, 2025

# Contents

<b>1</b>	<b>Python Basics with Google Colab</b>	<b>2</b>
1.1	What is Google Colab? . . . . .	2
1.2	How to Start . . . . .	2
1.3	Importing Libraries . . . . .	2
1.4	Importing Data (Local Excel File) . . . . .	2
1.5	Viewing Data . . . . .	2
1.6	Selecting and Manipulating Columns . . . . .	3
1.7	Selecting Rows . . . . .	3
1.8	Adding and Removing Rows . . . . .	3
1.9	Handling Missing Data . . . . .	3
1.10	Basic Summary and Grouping . . . . .	3
1.11	Basic Plotting . . . . .	3
1.12	Basic Regression Example . . . . .	4
1.13	Saving Data . . . . .	4
1.14	Useful Tips . . . . .	4
<b>2</b>	<b>R Basics for Data Analysis</b>	<b>5</b>
2.1	Getting Started with R . . . . .	5
2.2	Importing Data (Local Excel File) . . . . .	5
2.3	Viewing Data . . . . .	5
2.4	Selecting and Manipulating Columns . . . . .	5
2.5	Selecting Rows . . . . .	5
2.6	Adding and Removing Rows . . . . .	6
2.7	Handling Missing Data . . . . .	6
2.8	Basic Summary and Grouping . . . . .	6
2.9	Basic Plotting . . . . .	6
2.10	Basic Regression Example . . . . .	6
2.11	Saving Data . . . . .	6
2.12	Useful Tips . . . . .	6

# 1 Python Basics with Google Colab

## 1.1 What is Google Colab?

Google Colab (<https://colab.research.google.com>) is a free, cloud-based Jupyter notebook environment. You can write and run Python code directly in your browser, with no setup required. It is ideal for data analysis, machine learning, and reproducible research.

## 1.2 How to Start

1. Go to <https://colab.research.google.com>
2. Sign in with your Google account.
3. Click New Notebook to start.
4. Each cell can contain code or text (Markdown).

## 1.3 Importing Libraries

You need to import packages for data analysis:

```
1 import numpy as np
2 import pandas as pd
3 import matplotlib.pyplot as plt
4 import seaborn as sns
```

## 1.4 Importing Data (Local Excel File)

To use a local Excel file (e.g., Exercise-2\_Table6\_1\_Gujerati.xls) in Colab:

**Step 1: Upload the file from your computer:**

```
1 from google.colab import files
2 uploaded = files.upload() # Use the dialog to select Exercise-2
   _Table6_1_Gujerati.xls
```

**Step 2: Read the Excel file:**

```
1 df = pd.read_excel('Exercise-2_Table6_1_Gujerati.xls')
```

## 1.5 Viewing Data

```
1 df.head()      # First 5 rows
2 df.tail()      # Last 5 rows
3 df.info()      # Data types and non-null counts
4 df.describe()  # Summary statistics
```

## 1.6 Selecting and Manipulating Columns

```
1 df['column_name']          # Select column
2 df[['col1', 'col2']]      # Multiple columns
3 df['new_col'] = df['col1'] * 100 # Create new column
4 df['col1'] = df['col1'] + 1    # Modify column
5 df.drop('new_col', axis=1, inplace=True) # Delete column
```

## 1.7 Selecting Rows

```
1 df.loc[0]          # Select row by index
2 df.iloc[1:3]       # Select rows by position
3 df[df['col1'] > 10] # Conditional selection
```

## 1.8 Adding and Removing Rows

```
1 # Add a row
2 df.loc[len(df)] = [val1, val2, ...]
3 # Remove a row
4 df = df.drop(1) # Remove row with index 1
```

## 1.9 Handling Missing Data

```
1 df.isnull().sum()          # Count missing values
2 df.fillna(0)               # Replace missing with 0
3 df.dropna()                # Remove rows with missing data
```

## 1.10 Basic Summary and Grouping

```
1 df['col1'].mean()          # Mean
2 df['col1'].sum()           # Sum
3 df.groupby('col2')['col1'].mean() # Grouped mean
```

## 1.11 Basic Plotting

```
1 df['col1'].hist()
2 plt.show()
3 sns.boxplot(x='col2', y='col1', data=df)
4 plt.show()
```

## 1.12 Basic Regression Example

```
1 import statsmodels.api as sm
2 X = df[['col1', 'col2']] # Replace with actual column names
3 y = df['dependent_var']
4 X = sm.add_constant(X)
5 model = sm.OLS(y, X).fit()
6 print(model.summary())
```

## 1.13 Saving Data

```
1 df.to_csv('mydata.csv', index=False)
2 from google.colab import files
3 files.download('mydata.csv')
```

## 1.14 Useful Tips

- Use Tab for auto-completion.
- Use Shift+Enter to run a cell.
- Use ? after a function for help, e.g., df.head?
- Use Markdown cells for notes and equations.

## 2 R Basics for Data Analysis

### 2.1 Getting Started with R

R is a free software environment for statistical computing and graphics. You can use R via:

- RStudio Desktop (recommended)
- R command line
- R in the cloud (e.g., RStudio Cloud, Google Colab with R kernel)

### 2.2 Importing Data (Local Excel File)

First, install and load the `readxl` package if not already installed:

```
1 install.packages("readxl") # Run once if needed
2 library(readxl)
```

Read the Excel file (assume it is in your working directory):

```
1 df <- read_excel("Exercise-2_Table6_1_Gujerati.xls")
```

If you want to select the file interactively:

```
1 df <- read_excel(file.choose())
```

### 2.3 Viewing Data

```
1 head(df)           # First 6 rows
2 tail(df)           # Last 6 rows
3 str(df)            # Structure and types
4 summary(df)        # Summary statistics
```

### 2.4 Selecting and Manipulating Columns

```
1 df$col1             # Select column
2 df[c("col1", "col2")] # Select multiple columns
3 df$new_col <- df$col1 * 100 # Create new column
4 df$col1 <- df$col1 + 1    # Modify column
5 df$new_col <- NULL       # Delete column
```

### 2.5 Selecting Rows

```
1 df[1, ]            # First row
2 df[2:3, ]          # Rows 2 and 3
3 subset(df, col1 > 10) # Conditional selection
```

## 2.6 Adding and Removing Rows

```
1 # Add a row
2 df <- rbind(df, data.frame(col1=val1, col2=val2, ...))
3 # Remove a row
4 df <- df[-2, ] # Remove second row
```

## 2.7 Handling Missing Data

```
1 is.na(df)           # Check for NAs
2 sum(is.na(df))      # Count NAs
3 df$col1[2] <- NA    # Introduce NA
4 df[is.na(df)] <- 0  # Replace NAs with 0
5 df <- na.omit(df)   # Remove rows with NAs
```

## 2.8 Basic Summary and Grouping

```
1 mean(df$col1)       # Mean
2 sum(df$col1)        # Sum
3 aggregate(col1 ~ col2, data=df, mean) # Grouped mean
```

## 2.9 Basic Plotting

```
1 hist(df$col1)
2 boxplot(col1 ~ col2, data=df)
```

## 2.10 Basic Regression Example

```
1 model <- lm(dependent_var ~ col1 + col2, data=df)
2 summary(model)
```

## 2.11 Saving Data

```
1 write.csv(df, "mydata.csv", row.names=FALSE)
```

## 2.12 Useful Tips

- Use Tab for auto-completion in RStudio.
- Use Ctrl+Enter to run a line or selection.
- Use ?function for help, e.g., ?mean
- Use R Markdown for reproducible reports.