

# Essential Softwares for Data Science and Analytics Foundation Course

## Module 2: Data Importation in Python

Dr. Austin Brown

Kennesaw State University

# Data Importation

- ▶ **Objective:** Learn how to import necessary packages and data (eg. CSV and XLSX files) in Google Colab.
- ▶ **Importance:** Importing data is a fundamental step in data analysis and data science workflows.

# Importing Necessary Packages

- ▶ **Using pandas for DataFrame operations:**

- ▶ Import pandas as pd:

```
import pandas as pd
```

- ▶ **Using openpyxl for XLSX support in pandas (if needed):**

- ▶ Install openpyxl and import it:

```
!pip install openpyxl
```

- ▶ **Using read\_excel from pandas for reading Excel files:**

- ▶ Import read\_excel:

```
from pandas import read_excel
```

# Sample Usage

## ► Putting it all together:

```
!pip install openpyxl  
import pandas as pd  
from pandas import read_excel
```

# Importing Data in Google Colab using URL

## ► Importing CSV Files:

- Use `pandas.read_csv()` function.
- Example:

```
# Import pandas library
import pandas as pd
# Load CSV file into a DataFrame
url = 'https://path/to/your/file.csv'
mydata = pd.read_csv(url)
# 'url' is the link to the CSV file
```

## ► Importing XLSX Files:

- Use `pandas.read_excel()` function.
- Example:

```
import pandas as pd
url = 'https://path/to/your/file.xlsx'
mydata = pd.read_excel(url, sheet_name='Sheet1')
# 'sheet_name' specifies the sheet to read
```

# Manual File Importation

- ▶ Checking the Working Directory in Google Colab

- ▶ **Using !pwd to print the working directory:**

- ▶ Example:

```
!pwd
```

- ▶ Explanation:

- ▶ !pwd is a shell command that prints the current working directory in Colab.
  - ▶ Use this to confirm the location where files should be uploaded.

# Uploading and Importing Files in Google Colab

## ► **Download Data:**

- Download the StudentGPA CSV and XLSX files from D2L.

## ► **Manually Uploading Files:**

- Use the file upload button in the Colab interface to upload your CSV/XLSX file.

```
# Run and Select file on your PC
from google.colab import files
uploaded = files.upload()
```

## ► **Explanation:**

- `files.upload()` opens a file upload dialog.
- then you can navigate and select needed files.

# Uploading and Importing CSV Files in Google Colab

## ▶ Using `pandas.read_csv()`:

- ▶ Example to import a CSV file after uploading:

```
import pandas as pd
# Uploaded file is named 'StudentGPA.csv'
mydata = pd.read_csv('C:/PROJECT/StudentGPA.csv')
```

## ▶ Explanation:

- ▶ `files.upload()` opens a file upload dialog.
- ▶ `pd.read_csv()` reads the uploaded CSV file into a DataFrame.
- ▶ `head()` displays the first few rows of the DataFrame.



# Uploading and Importing XLSX Files in Google Colab

## ▶ **Alternative Manual Uploading Files:**

- ▶ Use the file upload button in the Colab interface to upload your XLSX file to the current working directory.

## ▶ **Using `pandas.read_excel()`:**

- ▶ Example to import an XLSX file after uploading:

```
import pandas as pd
mydata_xlsx = pd.read_excel('StudentGPA.xlsx'
, sheet_name='Sheet1')
# Display the first few rows of the XLSX data
mydata_xlsx.head()
```

## ▶ **Explanation:**

- ▶ `pd.read_excel()` reads the uploaded Excel file into a DataFrame.
- ▶ `head()` displays the first few rows of the DataFrame.

## Explore Imported Data

```
# Display the first few rows of the CSV data  
mydata.head()
```

	Student ID	Sex	Sleep	alcohol	exercise	TV	Study	S
0	1	Female	7.0	0.0	3.0	120.0	29	2
1	2	Male	6.0	5.0	2.0	35.0	31	3
2	3	Male	6.0	10.0	NaN	30.0	48	1
3	4	Female	6.0	0.0	2.0	30.0	28	2
4	5	Female	9.5	0.0	1.0	30.0	48	1