Written Instructions for Instructional Webpage

Sumit Sah

1. Title of Your Instructional Video or Webpage

How to Perform Basic Data Analysis in Python

2. Defining Your Specific User/Audience

Who are you imagining will use your instructions? These instructions are intended for beginners who want to learn how to perform simple data analysis using Python for schoolwork, personal projects, or introductory data science practice.

What do they know or not know already? They may know how to install basic software and open files, but they likely do not know how Python libraries work, how to import data, how to clean datasets, or how to interpret common Python errors.

What is their level of expertise? Beginner-level Python users with minimal familiarity with data analysis.

What is the setting in which they will use the instructions? These instructions will be used in calm, non-emergency environments such as at home, in a computer lab, or in a classroom.

3. Introduction

Why might a user want to learn this task? Python is one of the most widely used languages for data analysis, and learning the basics helps users complete academic projects, explore personal data, and begin developing data science skills.

Advance Organizer (one sentence): In this guide, the user will learn how to install Python libraries, load a dataset, inspect and clean the data, run simple analysis, create visualizations, and save the results.

Materials or Equipment Needed:

- A computer with Python 3.8 or newer installed
- Internet access (recommended for installing libraries)
- Jupyter Notebook, Google Colab, or a Python IDE (VS Code, PyCharm, etc.)
- A CSV dataset saved on the computer
- Python libraries: `pandas`, `numpy`, `matplotlib`

Warnings or Important Information:

- Installing Python or libraries may require administrator permissions.
- The dataset should not include sensitive or personal information.
- File names must be spelled correctly to avoid "FileNotFoundError" issues.
- Save work frequently to avoid losing progress.

4. Numbered Step-by-Step Instructions

Step 1: Install the Required Python Libraries. Open a terminal or command prompt and install the needed libraries:

```
pip install pandas numpy matplotlib
```

Step 2: Open a Python Notebook or Script. Launch Jupyter Notebook, Google Colab, or a Python IDE. (For beginners who do not have an IDE installed, VS Code is a free option.)

Step 3: Import the Data Analysis Libraries. At the top of your file, enter:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

Step 4: Load Your CSV Dataset. Place your CSV file in an accessible location and load it:

```
df = pd.read_csv("dataset.csv")
```

Step 5: Inspect the Dataset. Preview the data using:

```
df.head()
df.info()
df.describe()
```

Step 6: Clean Missing or Duplicate Data. Remove incomplete or repeated entries:

```
df = df.dropna()
df = df.drop_duplicates()
```

Step 7: Perform Basic Data Analysis. Run basic calculations such as averages or correlations:

```
df["column_name"].mean()
df.corr()
np.mean(df["column_name"]) % added based on peer suggestion
```

Step 8: Visualize the Data. Create a basic plot:

```
plt.plot(df["column_name"])
plt.show()
```

Step 9: Save the Cleaned or Analyzed Dataset. Export your file:

```
df.to_csv("cleaned_data.csv", index=False)
```

5. Troubleshooting or FAQ

Problem 1: "pip is not recognized." Cause: Python was not added to PATH. **Fix:** Reinstall Python and select "Add Python to PATH," then restart the terminal.

Problem 2: "FileNotFoundError" when loading the CSV. Cause: File name or path does not match. **Fix:** Check the spelling, confirm the folder location, or use a full file path.

Problem 3: "ModuleNotFoundError: No module named 'pandas'." Cause: Required libraries were not installed. **Fix:** Install the libraries again:

```
pip install pandas numpy matplotlib
```

Problem 4: Plot does not appear. Cause: Some environments require an explicit display command. **Fix:** Make sure to include:

```
plt.show()
```

Problem 5: Syntax or Indentation Errors. Cause: Missing quotes, parentheses, or incorrect spacing. **Fix:** Check for typos or inconsistent indentation.