

Introduction to Big Data: Environment Set Up

Anaconda Users

If you already have Anaconda installed, open the Anaconda Prompt and run:

```
conda create -n YourEnvName -c conda-forge python=3.12 pandas numpy matplotlib seaborn pymysql
```

Replace YourEnvName with the name you want for your environment. Do not use quotes.

Non-Anaconda Users

If you are working offline without Anaconda, check if Python is already installed on your PC. Open Command Prompt (CMD) and run:

```
python --version
```

If the installed version is 3.12, you can proceed to create a virtual environment.

If it is a different version (e.g., 3.10 or 3.11), you have two options:

Option A: Use the existing Python version to create a virtual environment (minor differences with 3.12 are acceptable).

Option B: Install Python 3.12 alongside the existing version for full compatibility. During installation, choose “Install for me only” and “Add Python 3.12 to PATH”. Then create the environment using Python 3.12 explicitly:

```
py -3.12 -m venv YourEnvName
```

After creating and activating your environment, install the required libraries:

```
pip install pandas numpy matplotlib seaborn pymysql
```

Tip: Using CMD is recommended over PowerShell or other terminals, because it handles pip and local package installations reliably.

I strongly recommend using Command Prompt instead of PowerShell or other terminals, because it handles pip and local Python package installations more reliably.