**DVC with Git - Step-by-Step Guide**

**Step-by-Step Guide for DVC Setup and Data Versioning**

**1. Initial Setup**

- Install Visual Studio Code or any preferred IDE.

- Install Git for Windows: https://git-scm.com/downloads

- Create a folder on your Desktop and name it 'DVC'.

- Paste the dataset file 'Salary\_Data.csv' into this folder.

**2. Initialize Git and DVC**

Open terminal in the DVC folder:

git init

pip install dvc

python -m dvc init

**3. Track Dataset with DVC**

python -m dvc add Salary\_Data.csv

git add Salary\_Data.csv.dvc .gitignore

git commit -m "Track the dataset with DVC"

**4. Create and Run alterv1.py**

import pandas as pd

data = pd.read\_csv("Salary\_Data.csv")

data["YearsExperience"] += 1

data.to\_csv("Salary\_Data.csv", index=False)

Run in terminal:

python alterv1.py

python -m dvc add Salary\_Data.csv

git add Salary\_Data.csv.dvc

git commit -m "v1: Added 1 year to YearsExperience"

**5. Create and Run alterv2.py**

import pandas as pd

data = pd.read\_csv("Salary\_Data.csv")

data["YearsExperience"] += 2

data.to\_csv("Salary\_Data.csv", index=False)

Run in terminal:

python alterv2.py

python -m dvc add Salary\_Data.csv

git add Salary\_Data.csv.dvc

git commit -m "v2: Added 2 years to YearsExperience"

**6. Create and Run alterv3.py**

import pandas as pd

data = pd.read\_csv("Salary\_Data.csv")

data["YearsExperience"] += 3

data.to\_csv("Salary\_Data.csv", index=False)

Run in terminal:

python alterv3.py

python -m dvc add Salary\_Data.csv

git add Salary\_Data.csv.dvc

git commit -m "v3: Added 3 years to YearsExperience"

**Data Version Control Commands Summary** Show all Git commits: git log --oneline

Checkout a specific version: git checkout <commit\_id> Pull data from DVC storage: dvc pull

Return to the latest (master) state: git checkout master