

## Data Assessment – 1

Tasks:

### Day 1: Setup and Data Collection

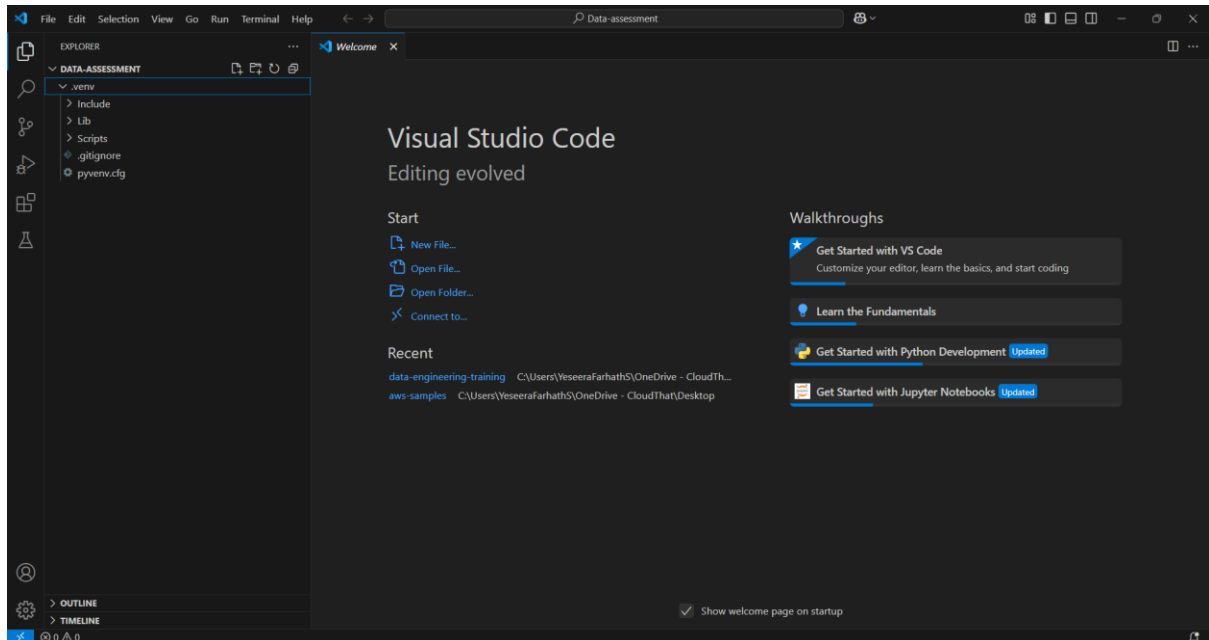
#### 1. Team Setup

- Set up virtual environments for Python dependencies
- Create a new GitHub repository for the project
- Install required Python libraries (NumPy, Pandas, Matplotlib, scikit-learn)

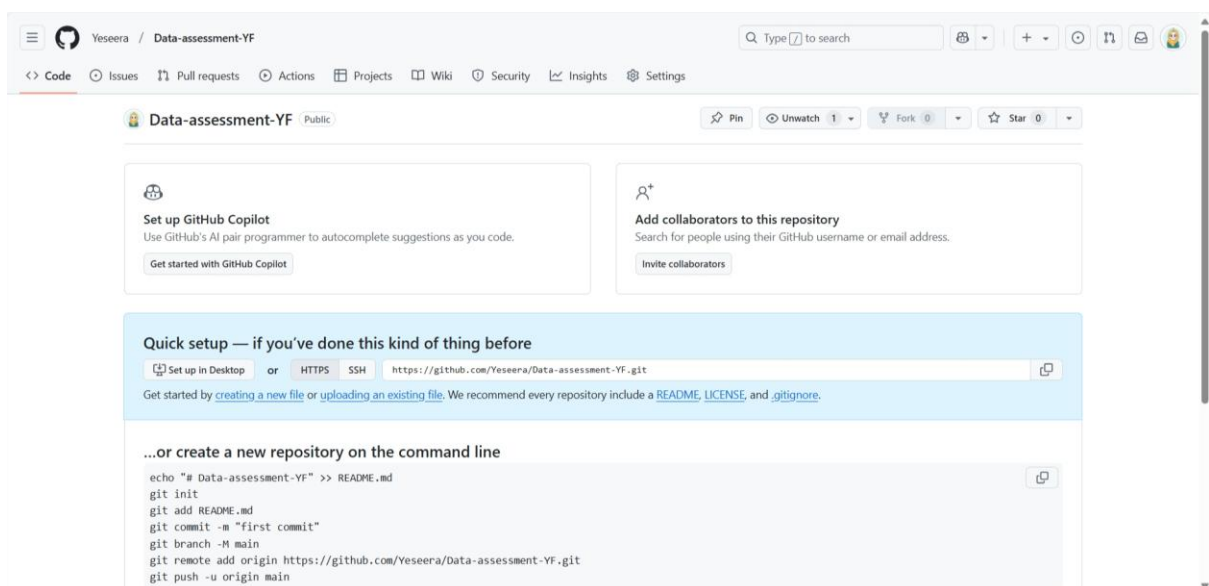
#### 2. **Dataset Selection:** Choose any from one datasets from data.gov.in

#### 3. Environment Setup

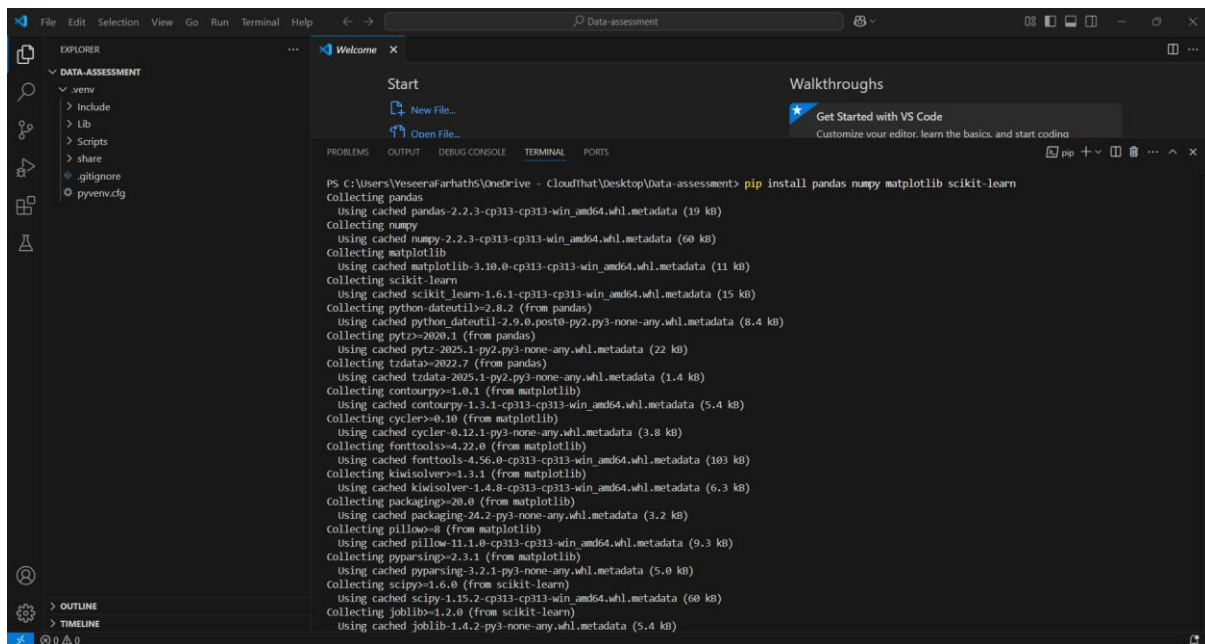
- Use VSCode Mandatory
- Create project structure (data, notebooks, scripts, output folders)
- Set up Jupyter Notebooks

**Solution:****I) Team Setup:****1. Set up virtual environments for Python dependencies**

- Created a new folder
- Created a python virtual environment.

**2. Create a new GitHub repository for the project**

### 3. Install required Python libraries (NumPy, Pandas, Matplotlib, scikit-learn)

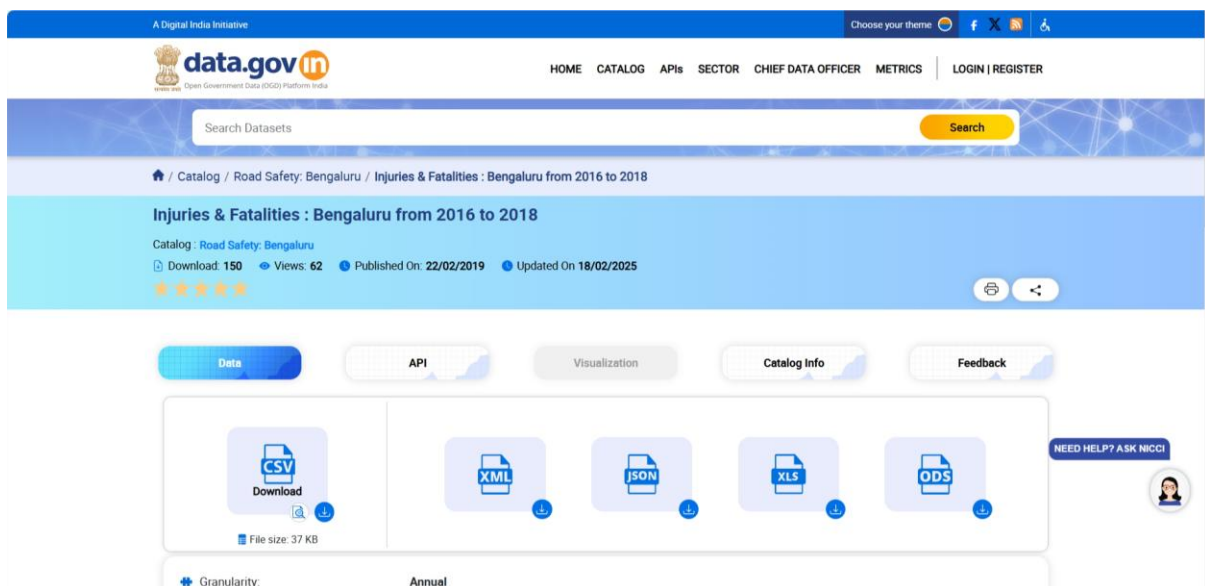


```

PS C:\Users\YeseeraFarhath\OneDrive - Cloudhat\Desktop\Data-assessment> pip install pandas numpy matplotlib scikit-learn
Collecting pandas
  Using cached pandas-2.2.3-cp313-cp313-win_amd64.whl.metadata (19 kB)
Collecting numpy
  Using cached numpy-2.2.3-cp313-cp313-win_amd64.whl.metadata (60 kB)
Collecting matplotlib
  Using cached matplotlib-3.10.0-cp313-cp313-win_amd64.whl.metadata (11 kB)
Collecting scikit-learn
  Using cached scikit-learn-1.6.1-cp313-cp313-win_amd64.whl.metadata (15 kB)
Collecting python-dateutil>=2.8.2 (from pandas)
  Using cached python-dateutil-2.9.0.post0-py2.py3-none-any.whl.metadata (8.4 kB)
Collecting pytz>=2020.1 (from pandas)
  Using cached pytz-2025.1-py2.py3-none-any.whl.metadata (1.4 kB)
Collecting contourpy>=1.0.1 (from matplotlib)
  Using cached contourpy-1.3.1-cp313-cp313-win_amd64.whl.metadata (5.4 kB)
Collecting cycler>=0.10 (from matplotlib)
  Using cached cycler-0.12.1-py3-none-any.whl.metadata (3.8 kB)
Collecting fonttools>=4.22.0 (from matplotlib)
  Using cached fonttools-4.56.0-cp313-cp313-win_amd64.whl.metadata (103 kB)
Collecting kiwisolver>=1.3.1 (from matplotlib)
  Using cached kiwisolver-1.4.8-cp313-cp313-win_amd64.whl.metadata (6.3 kB)
Collecting packaging>=20.0 (from matplotlib)
  Using cached packaging-24.2-py3-none-any.whl.metadata (3.2 kB)
Collecting pillow>=8 (from matplotlib)
  Using cached pillow-11.1.0-cp313-cp313-win_amd64.whl.metadata (9.3 kB)
Collecting pyparsing>=2.1.1 (from matplotlib)
  Using cached pyparsing-3.2.1-py3-none-any.whl.metadata (5.0 kB)
Collecting scipy>=1.6.0 (from scikit-learn)
  Using cached scipy-1.15.2-cp313-cp313-win_amd64.whl.metadata (60 kB)
Collecting joblib>=1.2.0 (from scikit-learn)
  Using cached joblib-1.4.2-py3-none-any.whl.metadata (5.4 kB)

```

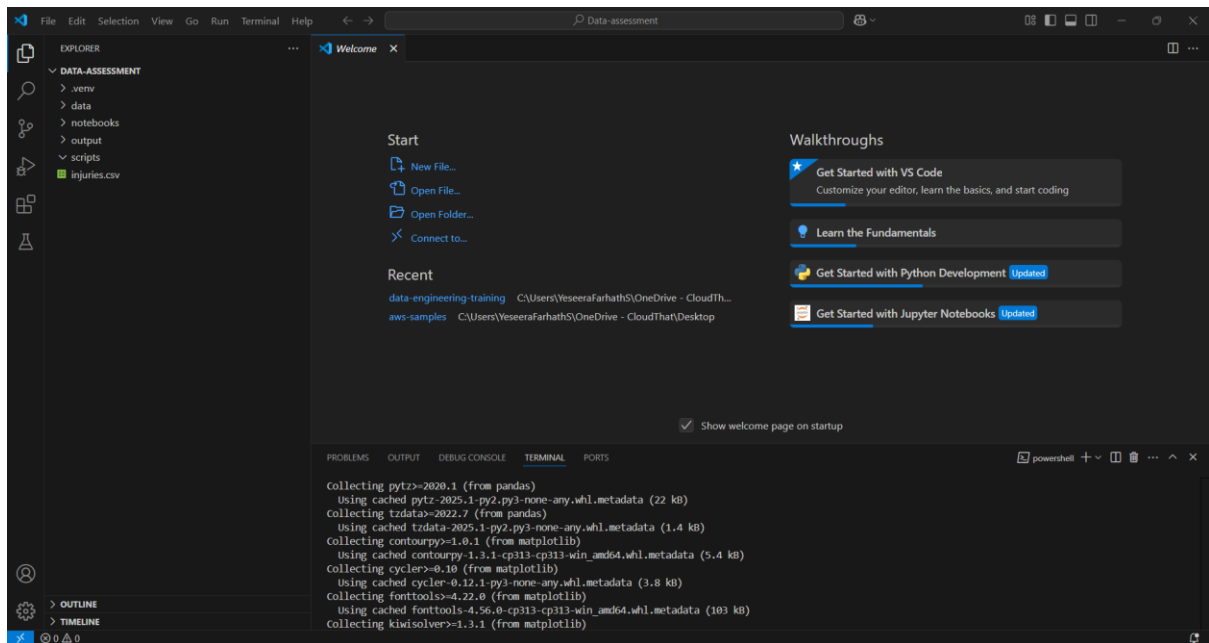
### II) Dataset Selection: Choose any from one dataset from data.gov.in



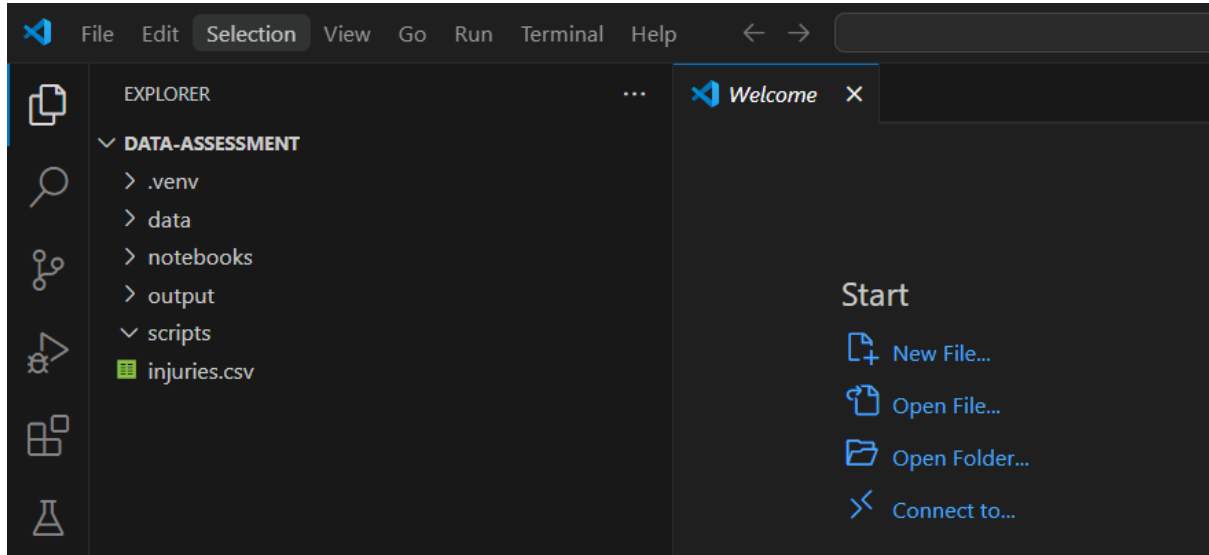
- Selected Injuries and fatalities dataset

### III) Environment Setup

#### 1. Use VSCode Mandatory



#### 2. Create project structure (data, notebooks, scripts, output folders)



### 3. Set up Jupyter Notebooks

