

# 📁 Dataset Location Update - December 30, 2025

## ✓ What Changed

The raw dataset has been **moved inside the project directory** for better organization and portability.

## 📁 Old Structure

```
MLOpsExperimentalLearning_Assignment_1_Group_81/
└── heart-disease-mlops/          # Project directory
    ├── src/
    ├── data/
    └── ...
└── raw_dataSet/                  # Dataset was OUTSIDE project
    └── heart+disease/
```

## 📁 New Structure (Current)

```
heart-disease-mlops/          # Project directory
├── src/
├── data/
└── raw_dataSet/              # Dataset is NOW INSIDE project ✓
    └── heart+disease/
        ├── processed.cleveland.data
        ├── cleveland.data
        ├── heart-disease.names
        └── ...
    ...
```

## 🔄 Updated Files

The following files have been updated with the new path:

1. [src/download\\_data.py](#) ✓
  - Changed: `BASE_DIR.parent / "raw_dataSet"` → `BASE_DIR / "raw_dataSet"`
2. [README.md](#) ✓
  - Updated project structure diagram
  - Updated setup instructions
3. [LOCAL\\_DEPLOYMENT\\_GUIDE.md](#) ✓

- Updated all references to dataset location

#### 4. EXECUTION\_GUIDE.md

- Updated troubleshooting section

#### 5. docs/EXECUTION\_GUIDE.md

- Updated expected output paths

#### 6. docs/FINAL\_REPORT\_TEMPLATE.md

- Updated data acquisition section

#### 7. .gitignore

- Added raw\_dataSet/ to prevent committing large dataset files
- 

## Benefits of This Change

1.  **Self-contained Project:** Everything is in one directory
  2.  **Easier Distribution:** Just zip/clone the heart-disease-mlops folder
  3.  **Better Git Management:** Dataset excluded via .gitignore
  4.  **Simpler Paths:** No more ../ navigation needed
  5.  **Clearer Structure:** All project assets in one place
- 

## Verification

To verify the change works:

```
# Navigate to project
cd heart-disease-mlops

# Check dataset exists
ls -la raw_dataSet/heart+disease/

# Test the download script
python src/download_data.py
```

## Expected Output:

```
✓ Found local dataset at: raw_dataSet/heart+disease
✓ Data downloaded successfully!
✓ Dataset saved to data/processed/heart_disease.csv
```

## Important Notes

- The `raw_dataSet/` directory is added to `.gitignore` to avoid committing large files
  - All relative paths in code now reference `raw_dataSet/` within the project
  - No changes needed to model training or API functionality
  - The dataset files remain the same, only their location changed
- 

## 🔍 Path Reference Quick Guide

Old Paths (Don't use these anymore ✗)

```
source_path = BASE_DIR.parent / "raw_dataSet" / "heart+disease" # ✗ Old
```

New Paths (Use these ✓ )

```
source_path = BASE_DIR / "raw_dataSet" / "heart+disease" # ✓ New
```

In Documentation

- ✗ Old: `../raw_dataSet/heart+disease/`
  - ✓ New: `raw_dataSet/heart+disease/`
- 

## 🎓 For New Users

When setting up the project:

1. Clone/download the `heart-disease-mlops` directory
  2. The dataset should already be in `raw_dataSet/heart+disease/`
  3. Run `python src/download_data.py` to process the data
  4. Continue with the setup as described in `LOCAL_DEPLOYMENT_GUIDE.md`
- 

**Date:** December 30, 2025

**Status:** ✓ Complete - All references updated

**Verified:** ✓ Python syntax valid, dataset accessible