

DS project

Dataset

- > NY-House-Dataset.csv

Data Exploration and Cleaning

- > Load the Dataset
- > Understand the Data
- > Data Cleaning

Data Visualization

- > Univariate Analysis
- > Bivariate Analysis

Feature Engineering

- > Feature Selection
- > Feature Transformation

Model Building

- > Split the Data
- > Choose a Model
- > Train the Model
- > Evaluate the Model

Model Tuning

- > Hyperparameter Tuning
- > Cross-Validation

01

DS project

Model Interpretation

- > Feature Importance
- > Partial Dependence Plots

Finalize the Model

- > Retrain on Full Dataset
- > Save the Model

Present the Results

- > Create a Report
- > Dashboard (Optional)

Project Structure

```
data_science_project/
|
├── data/
|   ├── raw/                # Original dataset
|   └── processed/          # Cleaned and processed data
|
├── notebooks/
|   ├── 01_data_exploration.ipynb
|   ├── 02_data_visualization.ipynb
|   ├── 03_feature_engineering.ipynb
|   ├── 04_model_building.ipynb
|   ├── 05_model_evaluation.ipynb
|   ├── 06_model_tuning.ipynb
|   ├── 07_model_interpretation.ipynb
|   └── 08_final_model.ipynb
|
├── models/                 # Saved models
|   └── final_model.pkl
|
├── scripts/               # Python scripts for various tasks
|
├── README.md              # Project overview and instructions
├── requirements.txt        # Required packages
└── .gitignore             # Ignore unnecessary files for version control
```

02