Project Structure and File Traversal

1. Recommended Environment: Jupyter Notebook
Using Jupyter Notebook is highly recommended to navigate the files easily.
The notebooks are organized to group code, outputs, and explanations together to follow the workflow without confusion.

2. Files and Folders Included in the Zip File
The project folder contains the following:

- **9 Python code files** (.ipynb):
  - Each notebook corresponds to a specific step in the data processing, data exploration, model development, and model evaluation.
- **4 Saved Model files**:
  - 3 Tree-based models saved in .pkl format.
  - 1 ANN model saved in .h5 format.
- **8 Data files** (.csv):
  - **2 Raw data files**:
    - adult_train.csv
    - adult_test.csv
  - **2 Combined files** (training and testing combined after preprocessing):
    - adult_combined_train_80.csv
    - adult_combined_test_20.csv
  - **2 Imputed data files** (after missing data imputation):
    - adult_combined_train_80_imputed.csv
    - adult_combined_test_20_imputed.csv
  - **2 Feature-selected and imputed files** (final processed datasets):
    - adult_combined_train_80_imputed_feature_selected.csv
    - adult_combined_test_20_imputed_feature_selected.csv
- **1 Folder named adult**:
  - Contains the original raw data files downloaded from the UCI Machine Learning Repository.
  - Also includes a supporting notebook Understanding_Data.ipynb for basic exploration of the raw data.

3. How to Follow the Code Files
To follow the development process step-by-step, it is recommended to open and run the code files in the following order:

1. **Missing_Data_Exploration.ipynb**
   (Initial exploration and understanding of missing values.)
2. **Missing_Data_Imputation.ipynb**
   (Handling and imputing missing values.)
3. **EDA_Feature_Selection_and_Model_Selection.ipynb**
   (Exploratory Data Analysis, feature selection, and multicollinearity check.)

4. **GBM_model.ipynb**
   (Training and evaluating the Gradient Boosting Machine model.)
5. **Adaboost_model.ipynb**
   (Training and evaluating the AdaBoost model.)
6. **RF_model.ipynb**
   (Training and evaluating the Random Forest model.)
7. **ANN_model.ipynb**
   (Training and evaluating the Artificial Neural Network model.)
8. **Models_Evaluation.ipynb**
   (Evaluationg the performance of the models and select the best model)
9. **EDA_Finding_Insight_from_Data.ipynb**
   (Exploring the data and finding insights)

4. Important Note

- **Each code file** contains:
  - An **overview** at the beginning that explains the purpose of the notebook.
  - **Explanations** on findings after key steps.
  - **Brief summaries** about the code logic, particularly in developing models.

5. Project Workflow

- The following workflow diagram will help to quickly grasp the logical flow of tasks, see how each stage connects, and understand where specific code files and analysis steps fit into the overall work.