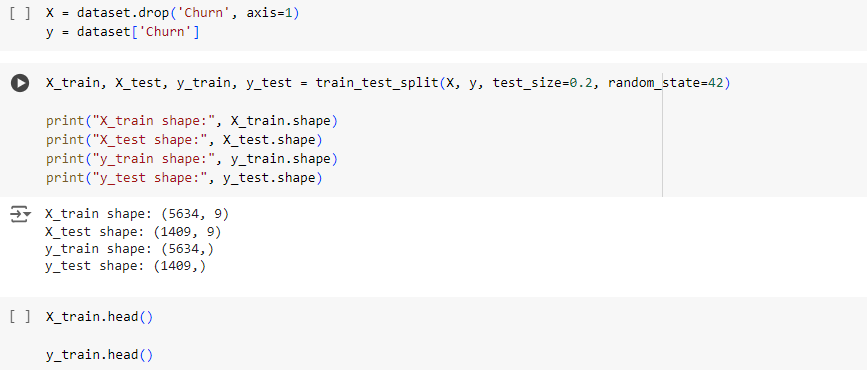
1. **Training Set**:
   * **Purpose**: The training set is used to build and train the model. It contains the majority of the data, allowing the model to learn patterns and relationships.
   * **Size**: The training set consists of approximately 70-80% of the total dataset, ensuring enough data for robust model training.
   * **Composition**: The training set is representative of the overall dataset, containing a balanced mix of features and labels to accurately reflect the data distribution.



A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

1. **Testing Set**:
   * **Purpose**: The testing set is used to evaluate the model's performance on unseen data. It helps in assessing how well the model generalizes to new data.
   * **Size**: The testing set comprises the remaining 20-30% of the total dataset.
   * **Composition**: Like the training set, the testing set is carefully selected to represent the overall dataset, ensuring that it covers the same data distribution.

**Documentation**:

* Each set is saved in its respective subfolder with clear naming conventions for easy identification.
* A document is included that provides detailed information on the size and composition of both the training and testing sets, including the criteria used for splitting the data.