**Metadata for Each File**

**1. Geo\_scores.csv**

* **Description**: Contains geographical scores associated with different IDs.
* **Key Columns**:
  + id (Unique identifier)
  + geo\_score (Numerical score related to geography)
* **Missing Values**: Present, handled using median imputation.

**2. Lambda\_wts.csv**

* **Description**: Contains lambda weight values for different groups.
* **Key Columns**:
  + Group (Categorical variable)
  + lambda\_weight (Numerical weight associated with each group)
* **Missing Values**: Checked but not explicitly mentioned as an issue.

**3. Qset\_tats.csv**

* **Description**: Contains Qset Turnaround Time (TAT) scores for different IDs.
* **Key Columns**:
  + id (Unique identifier)
  + qsets\_normalized\_tat (Numerical turnaround time score)
* **Missing Values**: Present, handled using median imputation.

**4. instance\_scores.csv**

* **Description**: Contains instance scores for different IDs.
* **Key Columns**:
  + id (Unique identifier)
  + instance\_score (Numerical score for instances)
* **Missing Values**: Checked and addressed.

**5. test\_share.csv**

* **Description**: Test dataset used for model predictions.
* **Key Columns**:
  + id (Unique identifier)
  + Group (Categorical variable)
  + Other relevant features for prediction
* **Missing Values**: Checked and handled accordingly.
* **Target Variable**: Not present (Predictions generated by the model).

**6. train.csv**

* **Description**: Training dataset containing features and the target variable.
* **Key Columns**:
  + id (Unique identifier)
  + Group (Categorical variable)
  + Target (Binary or categorical target variable)
  + Other numerical and categorical features
* **Missing Values**: Checked and handled.
* **Target Variable**: Present (Used for model training).