1. To remind after modelling, if accuracy is low then check data\_clean file and remove ffill part.
2. After that, in feature engineering file in airlines col change tol=0.12.
3. The code snippet provided defines a pipeline air\_transformer that performs a series of transformations on a column (presumably named "airline") in your dataset X\_train. Let's go through each step of the pipeline:
4. Imputer: Uses SimpleImputer to fill missing values with the most frequent value in the column.
5. RareLabelEncoder: Groups rare labels (those that appear less frequently than the specified tolerance tol=0.1) into a single category called "Other". It also ensures that at least n\_categories=2 are kept as separate categories.
6. OneHotEncoder: Encodes categorical values into one-hot (dummy) variables, ensuring that any unknown categories encountered during transformation are ignored.
7. The goal is to fit this pipeline to the airline column of X\_train and then transform it.
8. air\_transformer = Pipeline(steps=[ ("imputer", SimpleImputer(strategy="most\_frequent")), ("grouper", RareLabelEncoder(tol=0.1, replace\_with="Other", n\_categories=2)), ("encoder", OneHotEncoder(sparse\_output=False, handle\_unknown="ignore")) ]) air\_transformer.fit\_transform(X\_train.loc[:, ["airline"]])#.airline.value\_counts()