Flight Price Prediction

- Importing the important libraries for reading the dataset.
- Because data is present in two different sets.
- Joining the both dataset in single variable.
- Checking for null values.
- Checking the datatypes of dataset, in which form data is present.
- All columns are present as object datatype except one.

- Now, we import encoder for encoding the required columns.
- Importing label encoder and ordinal encoder.
- Because target variable is present as object datatype, so we have to encode it also.

- Now we import libraries for visualization the dataset.
- First, we plot distribution plot.
- Second, we plot boxplot.
- By distribution plot, data is shown which is not in normal shape.
- By boxplot, some outliers are shown in dataset.

- Now, we take out quantiles and inter quantile range, which help us to remove outliers.
- After, removing outliers, again we check for data shape by distribution plot.
- Now, after removing outliers, data starts looking in some normal shape.
- By boxplot, it seems that a lot of outliers has been removed from dataset.

- Now, we plot heatmap, for checking the correlation.
- No such relation is shown by heatmap.

- Now, its time to split the dataset into two variables.
- Standardize the data and train test split.
- Because target variable is present as continuous data so we apply Regression models on this dataset.

- Importing metrics for checking the model error.

- Now, first model, Linear Regression.
- By, Linear Regression, a very low score is given on train and test data both.

- Second model, Decision Tree.
- By second model, good accuracy is given on train data but on test data it became very low.

- Third model, Random Forest.
- By third model, good accuracy on train dataset is given, but low on test data.

- Fourth model, Support Vector Machine.
- By fourth model, a very-very low accuracy on both train and test data is given.

- Fifth model, Ada Boost.
- By fifth model, very low accuracy on both train and test dataset.

- Now, we do hyperparameter tuning on Random Forest and Ada Boost.
- After doing hyperparameter tuning, by Random Forest, accuracy on train dataset decreases and on test dataset increases.
- After doing hyperparameter tuning, by Ada Boost, accuracy remains same on both train and test dataset.

- Random Forest performed well, so we can save it by using pickle.