Question 1 : In you own data set, is the task supervised / or unsupervised learning? Is it classification or regression? Provide the target variables and corresponding loss functions

Ans : The Data set task is supervised and the Hotel Booking Data is Classification.

Also I think below mentioned are the Target Variables used to predict future bookings.

Id

Age

Gender

Signup_method

Date_account_created

Timestamp_first_active

Date_first_booking

Signup_app

Country

Language

Question 2: What is your strategy to split your total train data when you are training models? In your setup, how many models you will have eventually? How can you use these trained models to predict for new input?

Ans : In my training model the Dataset is divided into 4 models

Training set : It is the subsection of a dataset from which the Machine learning algorithm uncovers or learns relationships between features and target variables.

Validation set : It is another subset of input data to which we apply the machine learning algorithm to see how accurately it identifies relationships.

Holdout set : It provides a final estimate of the machine learning model's performance after it has been trained and validated.

Cross validation : Which is resampling procedure used to evaluate models on a limites data sample.

Question 3: imbalanced data set in your own data set, if it is the classification task, please analyze the ratio between two classes (1 and 0)? If the ratio is very imbalanced (imbalanced data means the amount of 1 or 0 is extremely small comparing the other class), what is the possible methodologies to train models?

Answer: Below are the techniques to handle imbalanced Data

- Changing Evalauation metric
- Resampling Dataset
- Using K-Fold Cross validation
- Ensemble Different resampled data
- Resample with different ratios
- Cluster the abundant class