# **Summary of Hotel recommendations**

* Loaded 200000 rows from training csv file
* Original destination distance, check in date and check out date have been removed as there are many NA values and they didn’t add any value to predict the hotel cluster
* The date\_time also has been removed from the final dataset with the same reason
* Only rows where the value of is\_booking = 1 are taken into consideration because when there is no booking there is no hotel cluster to choose
* KNN is used to predict the hotel cluster using training and test data
* The training data is divided into 80% training and 20% testing for initial testing
* With KNN the scores are as shown below (with k=1)

Precision Score: 0.13451072954377877

Recall Score: 0.13179949744521458

Accuracy Score: 0.14127174565086983

F1 Score: 0.1289744025611782

* The accuracy score is the highest when k=1
* Not much correlation is found between different variables and hotel cluster
* Multiple regression models are also plotted with user\_location\_country, user\_location\_city, srch\_destination\_id, srch\_adults\_cnt, and src\_childrens\_count.
* The adjusted R2 value is found to be very low in all these models and not much correlation is found between the predictors and the hotel cluster
* Adjusted R2 value for the first one is 0.001525, for the second model Adjusted R-squared: 0.001284, and for third model the value is -0.0002304. None of the input variables have any correlation with the hotel cluster
* Compared to all models, KNN has better accuracy score with predicted hotel cluster values.