User Conversion Rate

Exploratory Data Analysis:
1. Combine the 4 tables into one by using a join over the Unique registration value i.e 'registration_id'
2. Check Data Distribution
3. Normalize data/formats/types
4. Check for missing values
5. Delete duplicates

Data Pre-Processing:
1. Impute the missing values
2. Try changing the distribution of numerical values (gaussian preferred)
3. Change the categorical data to numerical data
4. Rescale continuous variables
5. Calculate **ANOVA F** value between each feature and target vector.
6. Split the data into test and train set (80-20).

Model Building:
1. Build the model with **XGBoost** and learn the best parameters with gridsearch.
2. Perform k-fold cross-validation.
3. Plot AUC-ROC curves for every model.

Model Evaluation:
1. **AUC ROC** curves along with confusion metrics measures.
2. Calculate **Specificity** as it is important to identify the non-converting users.
3. Analyze feature importance using **SHAP** values to determine the important features impacting the conversion rate.