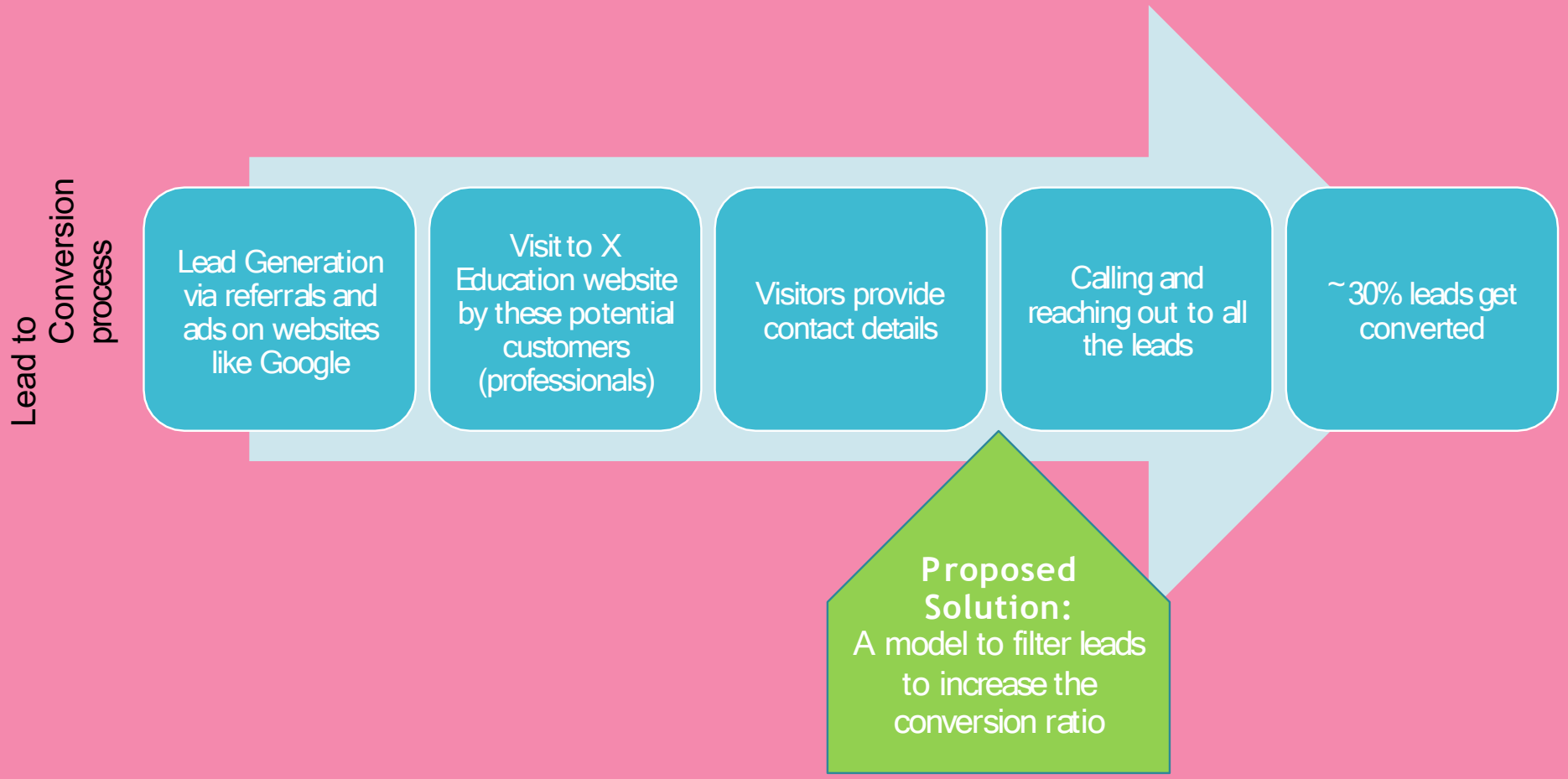
Problem Statement: To Identify Hot Leads to Increase Conversion Rate

Lead - Conversion Process



Proposed Solution

Selection of Hot Leads → Communicating with Hot Leads → Conversion of Hot Leads

Leads Clustering

We cluster the leads into certain categories based on their tendency or probability to convert, thus, reducing the sample size but getting more efficiency.

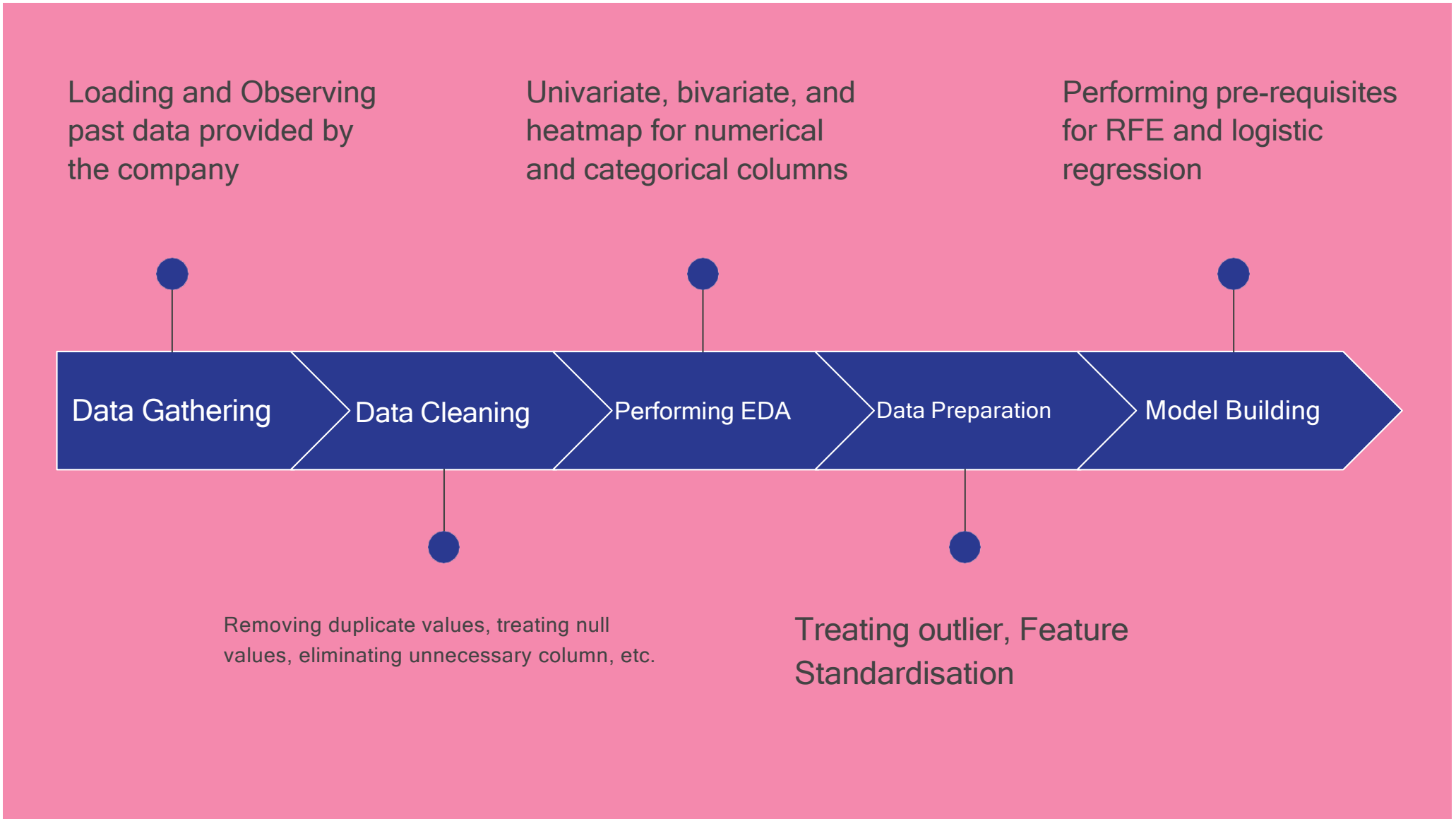
Focus Communication

Since we would know more about this smaller set of leads, we can have an impactful conversation.

Increase Conversion

Since we focussed on hot leads, which were more probable to convert, we would have a better conversion rate.

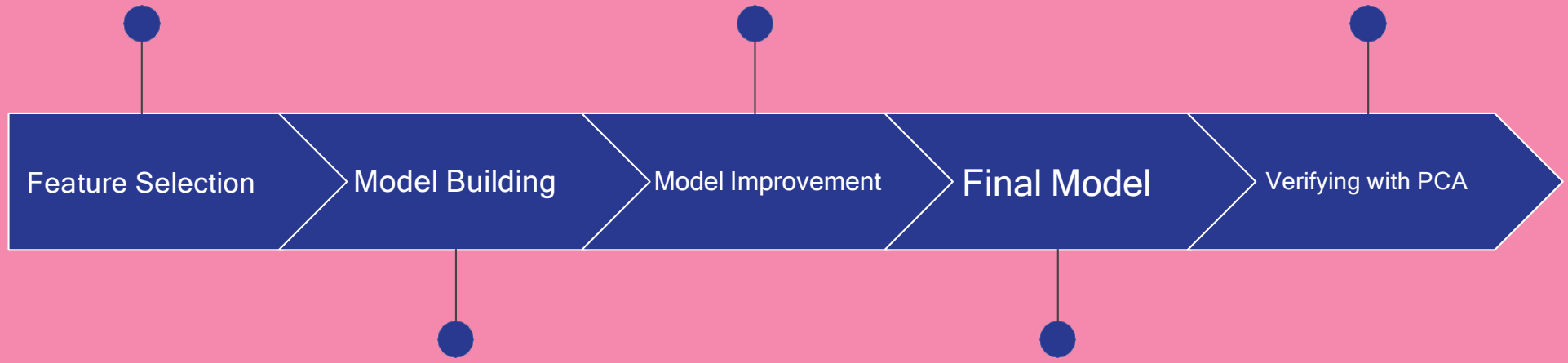
Implementation


Loading and Observing past data provided by the company
Univariate, bivariate, and heatmap for numerical and categorical columns
Performing pre-requisites for RFE and logistic regression
Data Gathering
Data Cleaning
Performing EDA
Data Preparation
Model Building
Removing duplicate values, treating null values, eliminating unnecessary column, etc.
Treating outlier, Feature Standardisation

Selection of top features using RFE

Reduction of columns and model re-building

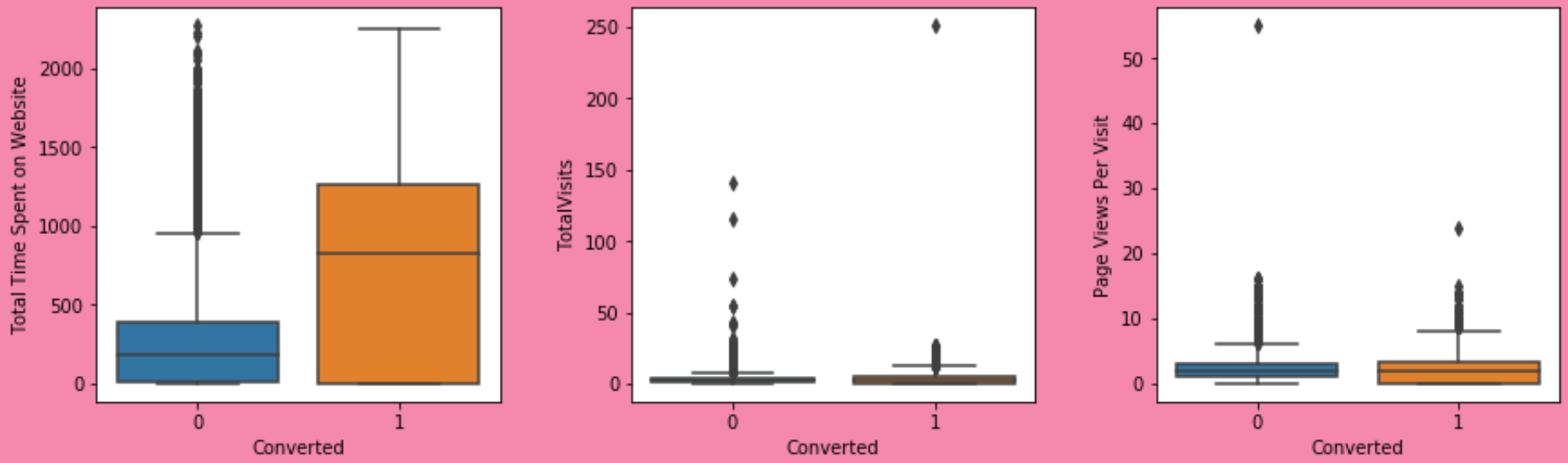
Verifying our final model accuracy with PCA



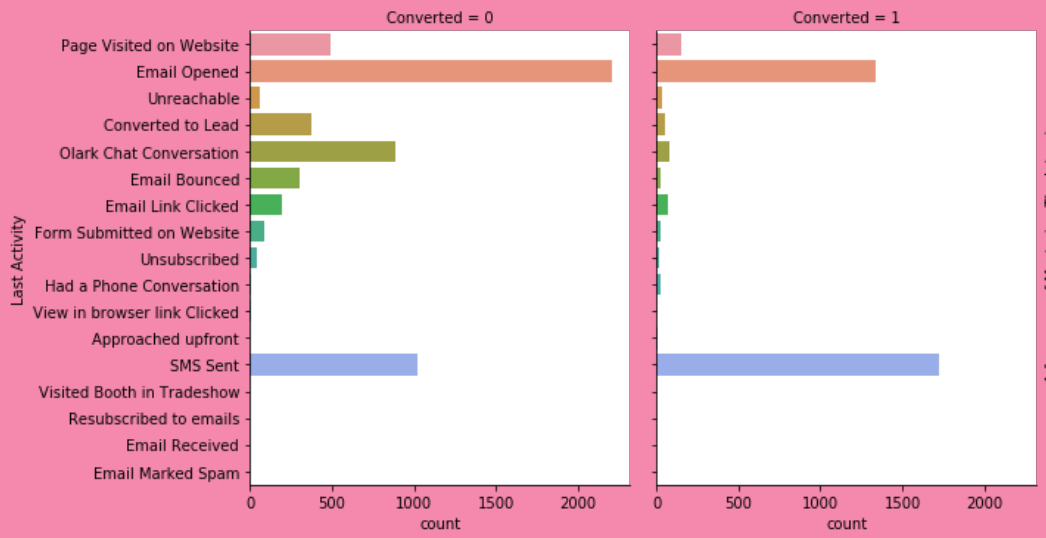
Model building using RFE for selected columns

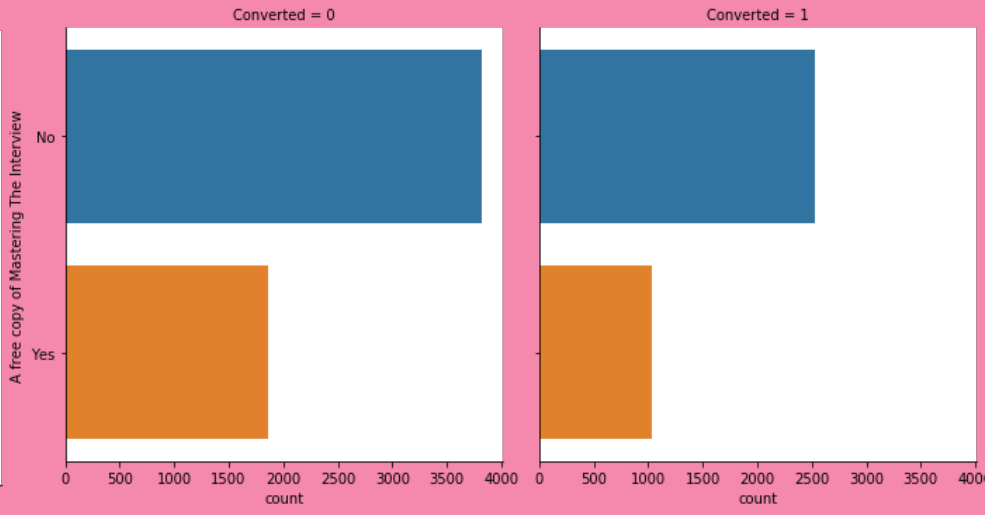
Final model analysis and performance on test data

Plots (Visualisation)

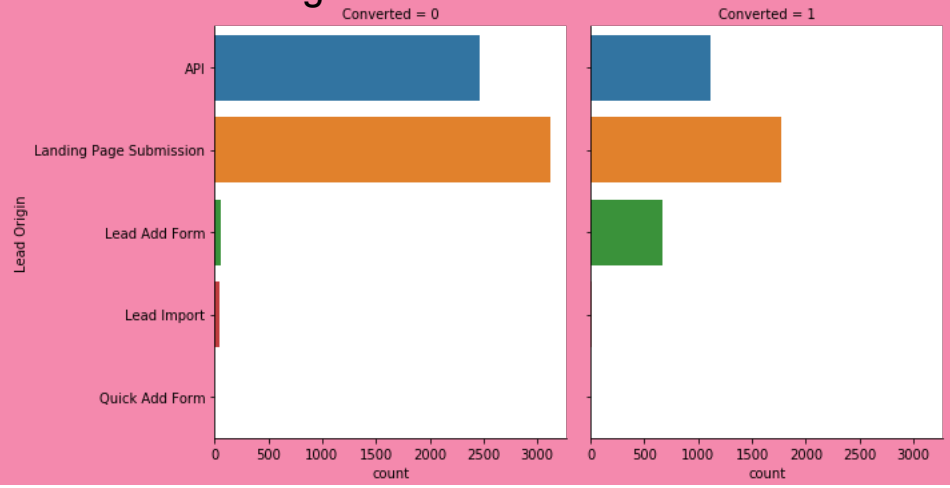


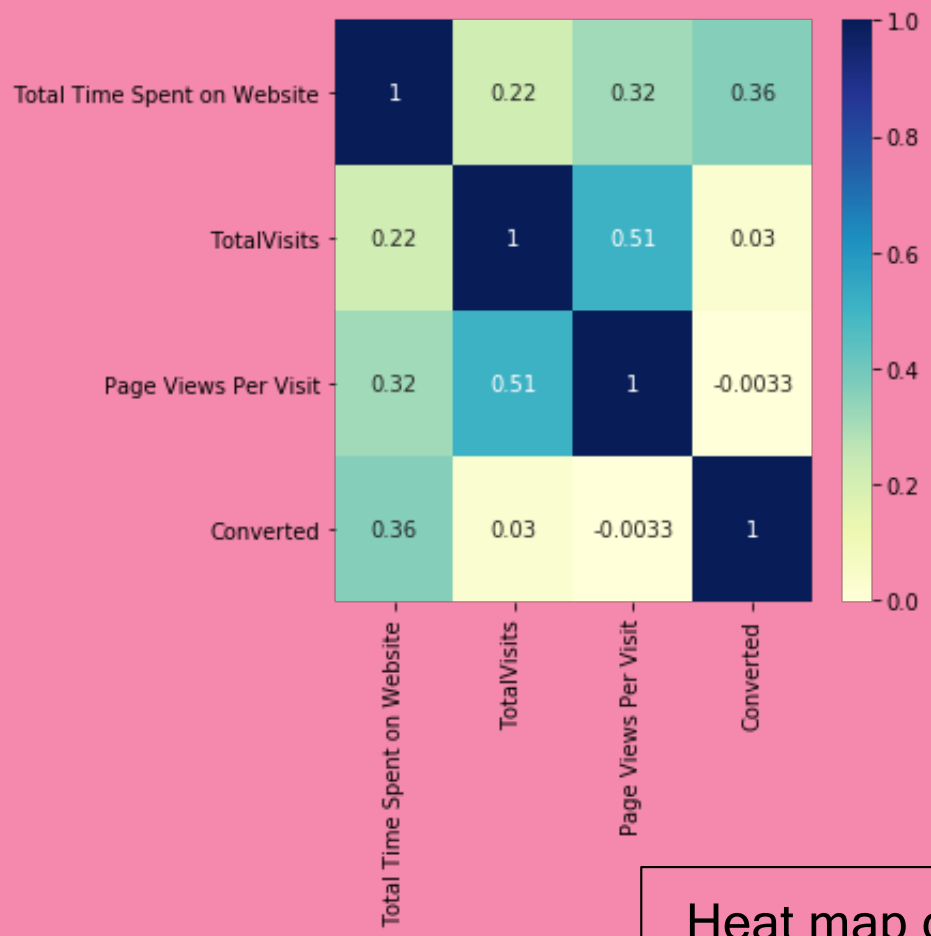
Plots depicting variation in numerical columns with converted



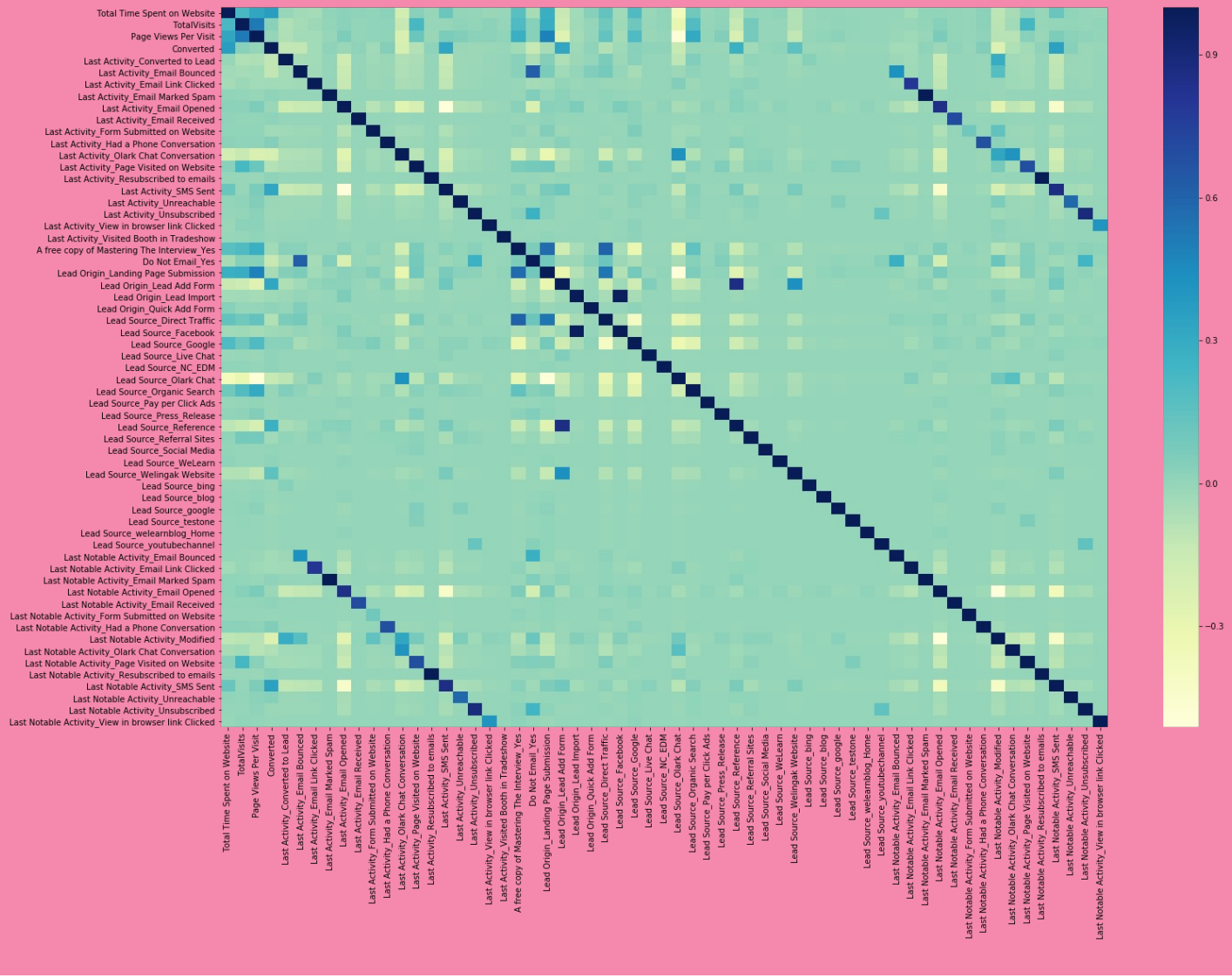


Plots depicting variation in Last Activity, Mastering the interview and Lead origin for those who Converted vs those who didn't

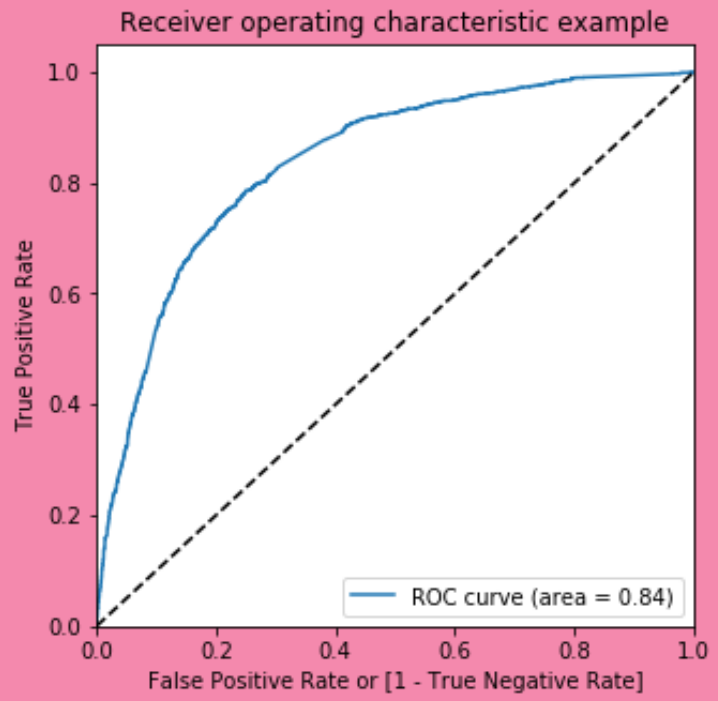


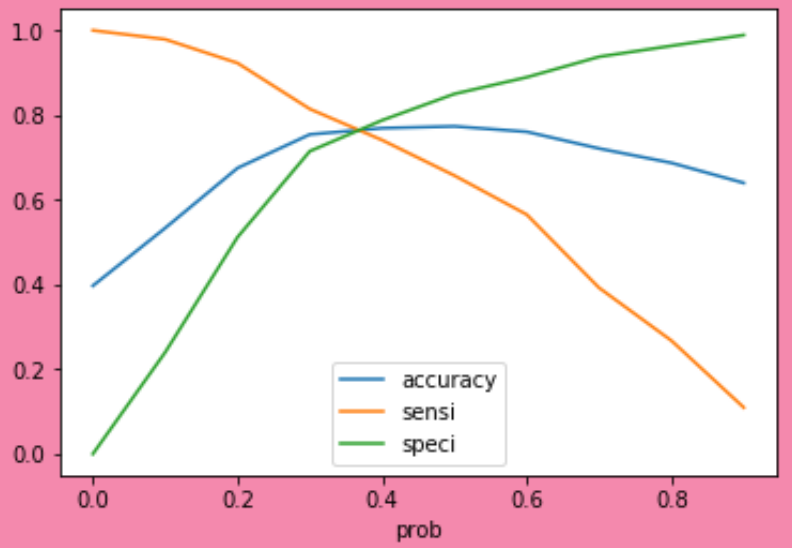


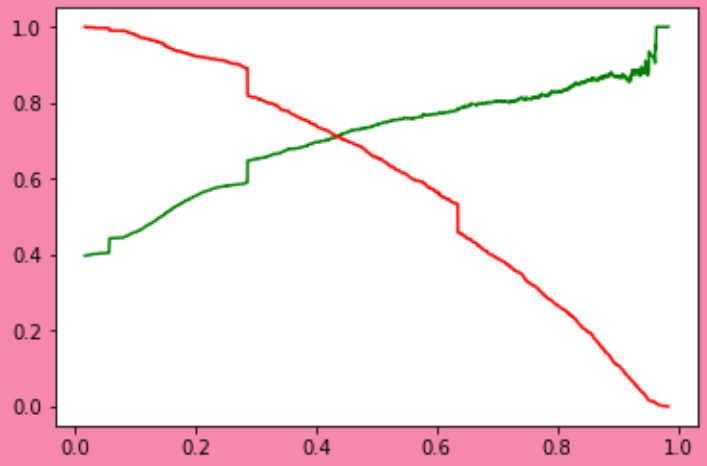
Heat map of all selected numerical columns



Heat map for all selected columns (numerical columns





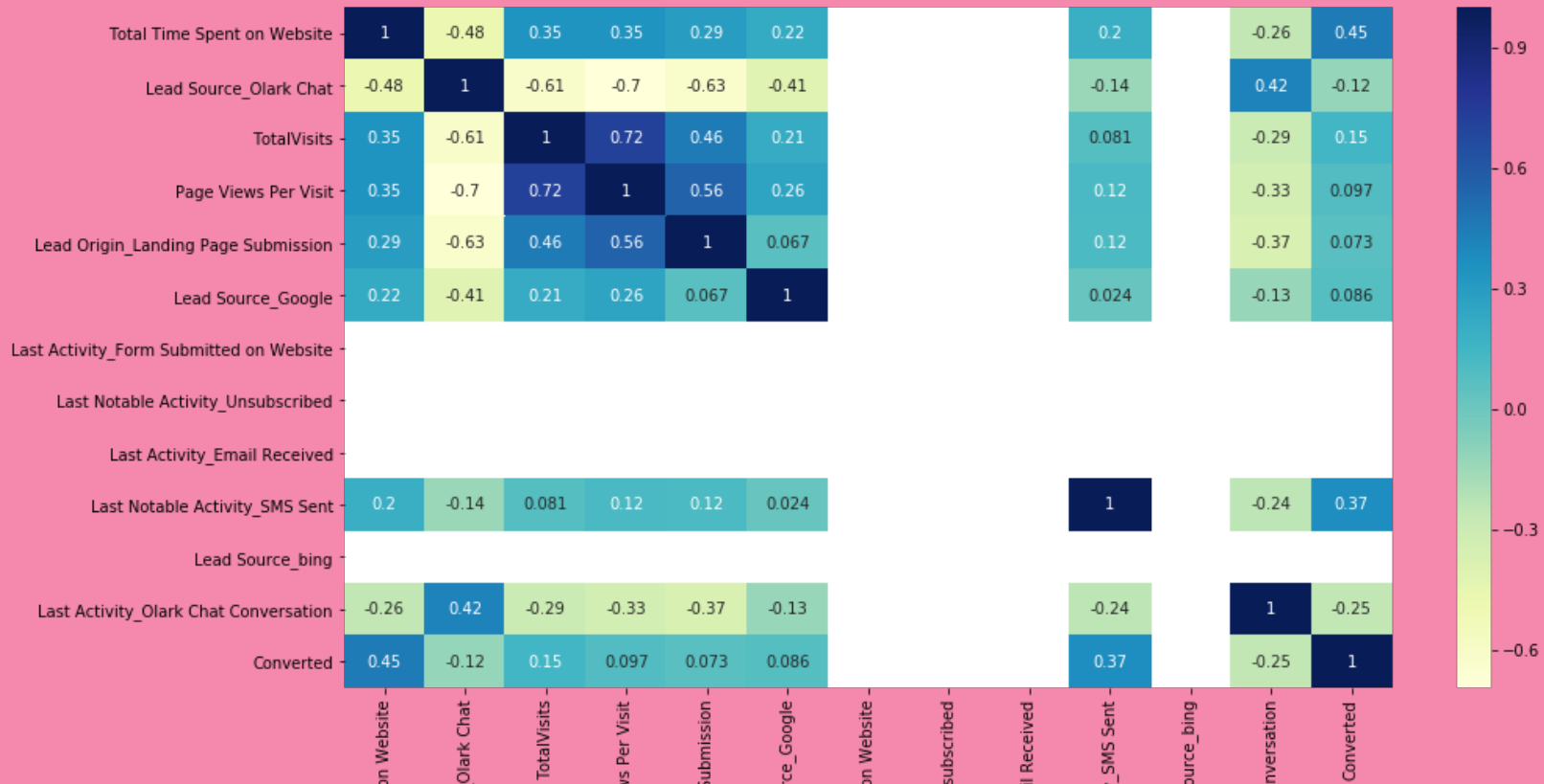


Linear Regression Final Model Parameters

Area under ROC = 0.84

Intermediate cut-off = 0.35

Final cut-off = 0.42



Heat Map of all selected columns in our final model

Inference / Conclusion

Model Analysis

Performance of our Final Model

Overall accuracy on Test set: 0.81

Sensitivity of our logistic regression model: 0.82

Specificity of our logistic regression model: 0.82

Inferences from Model

Business Insights Derived from our Model

Top 3 variables in our model that contribute towards lead conversion are:

- Total Time Spent on Website
- TotalVisits
- Lead Origin

Inferences from Model

Business Insights Derived from our Model

Top 3 variables in my model, that should be focused are:

- Last Activity_SMS Sent (positive impact)
- Last Activity_Olark Chat Conversation (negative impact)
- Lead Source_Olark Chat (negative impacting)