**Summary Report**

**Exploratory Data Analysis Summary:**

The following points have to be noted:

* The following segments, the conversion rate has to be focused upon:
  + Lead Source:  'Olark Chat', 'Organic Search', 'Direct Traffic', 'Google'
  + Lead Origin: 'API', 'Landing Page Submission'
  + Last Activity: ‘Email Opened’
  + Occupation: ‘Unemployed’
  + Last Notable Activity: ‘Modified’, ‘Email Opened’
* In the following segments, the lead count has to be focused upon:
  + Lead Source: 'Reference', 'Welingak Website'
  + Lead Origin:  'Lead Add Form'
  + Last Activity: ‘SMS Sent’
  + Occupation: ‘Working Professional’
  + Tags: 'Will revert after reading the email', 'Lost to ENS', 'Closed by Horizon'
  + Last Notable Activity: ‘SMS Sent’
* Despite this, the Website has to be interactive in order to make the prospects spend more time on the website.
* A well-built mechanism can be established for assigning the ‘Lead Quality’ tags for the prospects, since this a significant feature and can greatly improve conversion rates.

**Conclusion:**

* The logistic regression model predicts the probability of the target variable having a certain value, rather than predicting the value of the target variable directly. Then a cutoff of the probability is used to obtain the predicted value of the target variable.
* Here, the logistic regression model is used to predict the probabilty of conversion of a customer.
* Optimum cut off is chosen to be 0.275 i.e. any lead with greater than 0.275 probability of converting is predicted as Hot Lead (customer will convert) and any lead with 0.275 or less probability of converting is predicted as Cold Lead (customer will not convert)
* Our final Logistic Regression Model is built with 14 features.
* Features used in final model are ['Do Not Email', 'Lead Origin\_Lead Add Form', 'Lead Source\_Welingak Website', 'Last Activity\_SMS Sent', 'What is your current occupation\_Working Professional', 'Tags\_Busy', 'Tags\_Closed by Horizzon', 'Tags\_Lost to EINS', 'Tags\_Ringing', 'Tags\_Will revert after reading the email', 'Tags\_switched off', 'Lead Quality\_Not Sure', 'Lead Quality\_Worst', 'Last Notable Activity\_Modified',]
* The top three categorical/dummy variables in the final model are ‘Tags\_Lost to EINS’, ‘Tags\_Closed by Horizzon’, ‘Tags\_Busy’ with respect to the absolute value of their coefficient factors.
* ‘Tags\_Lost to EINS’, ‘Tags\_Closed by Horizzon’, ‘Tags\_Busy’ are obtained by encoding original categorical variable ‘Tags’.
* Tags\_Lost to EINS (Coefficient factor = 9.606615)
* Tags\_Closed by Horizzon (Coefficient factor = 8.359193)
* Tags\_Busy (Coefficient factor = 3.925069)
* The final model has Sensitivity of 0.884, this means the model is able to predict 88% customers out of all the converted customers, (Positive conversion) correctly.
* The final model has Precision of 0.847, this means 84% of predicted hot leads are True Hot Leads.