Lead scoring case study

Presented by
AshokRathod,
AlokSingh &
Baishali Acooli

Problem Statement

- An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.

- Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%.

- X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers.

Business Goal

☐ The company requires you to build a model wherein you need to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance.

☐ The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

Strategy

- Import dataset
- Clean and prepaid the dataset
- Exploratory data analysis
- Scaling features
- Prepare the data for model building
- Logistic regression model
- Assign a lead score for each leads
- Train and test model
- Evaluate the model
- Test the model in test set
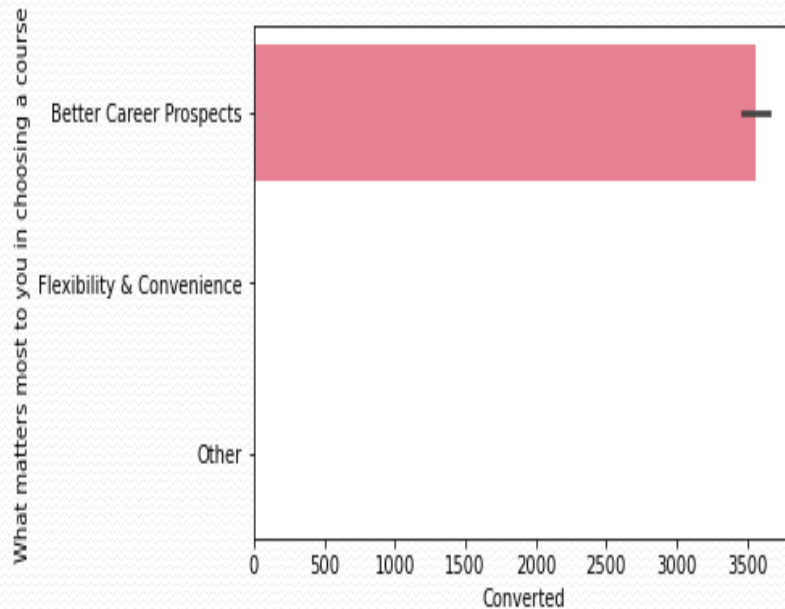- Measure the accuracy of the model

Exploratory data analysis
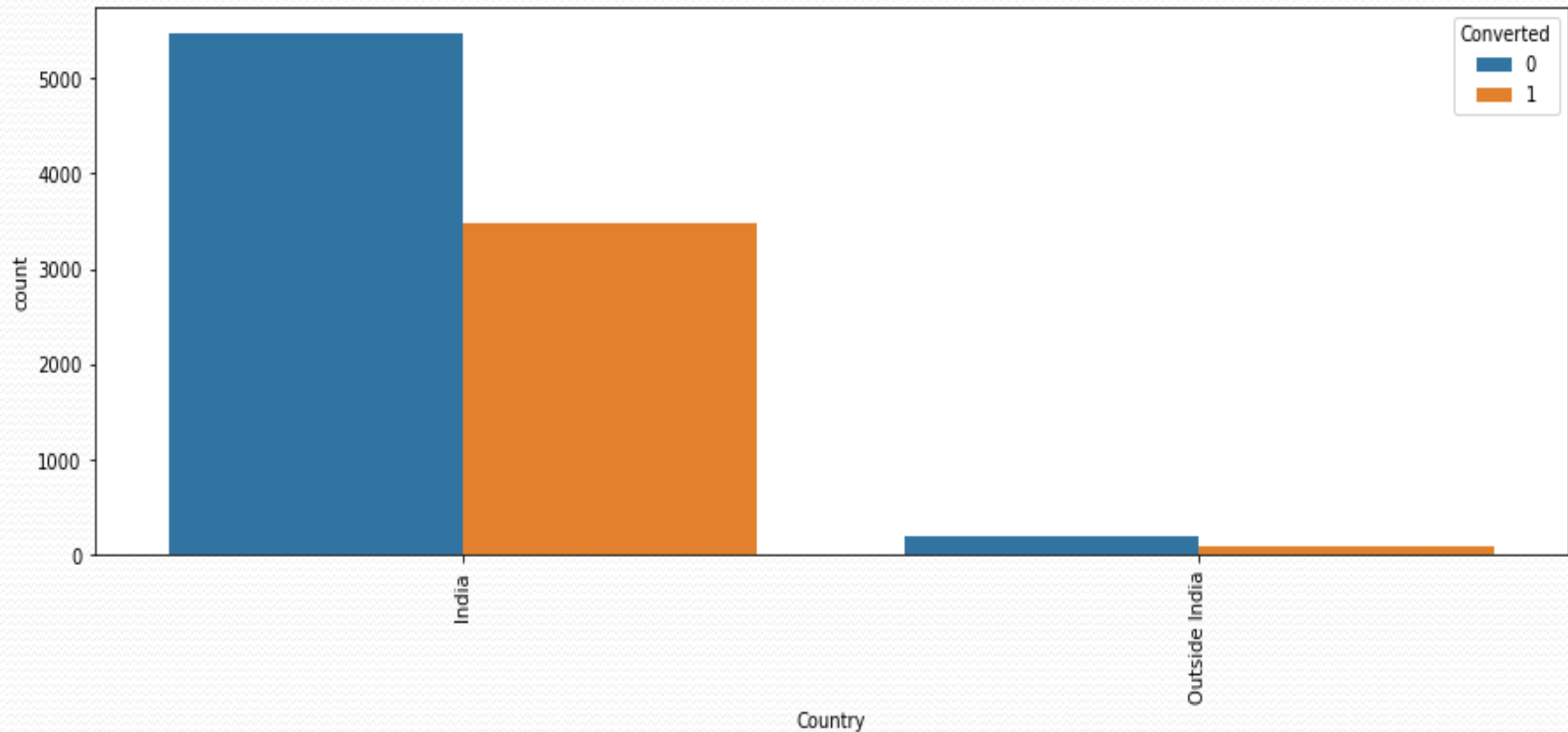
As we can see the Number of Values for India are quite high (nearly 97% of the Data), this column can be dropped.
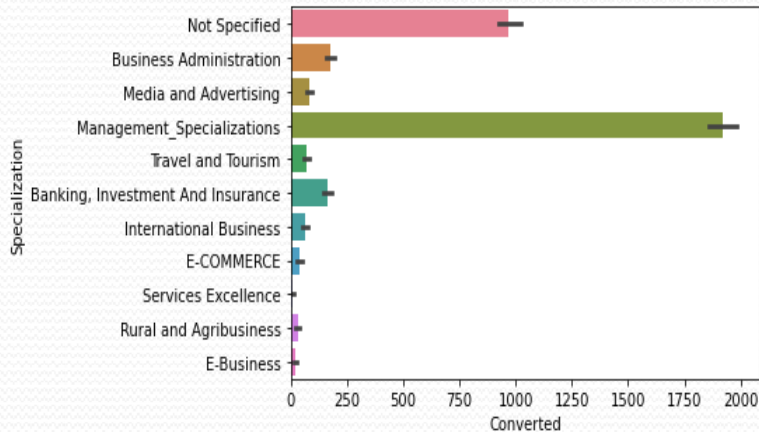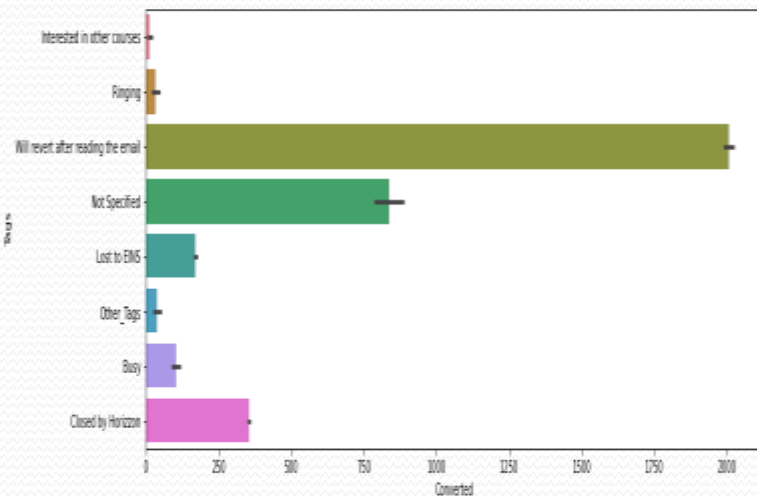
Most of the leads don't have any information about specialization. Management has high converted rates.
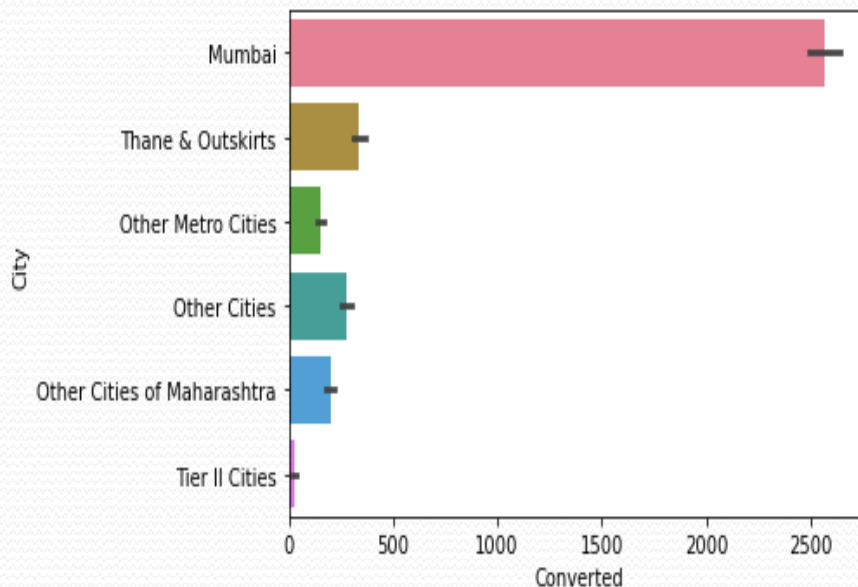

Leads focus on carrier that's why they want better carrier prospects
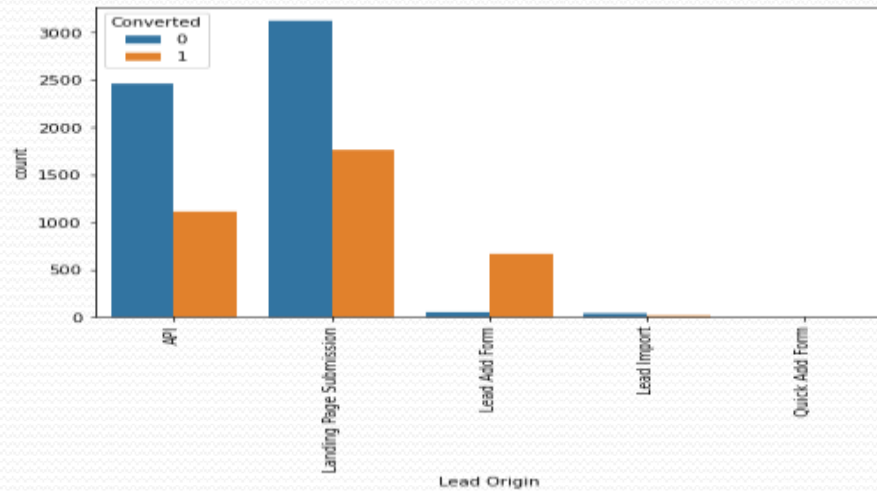
Converted vs tags and city

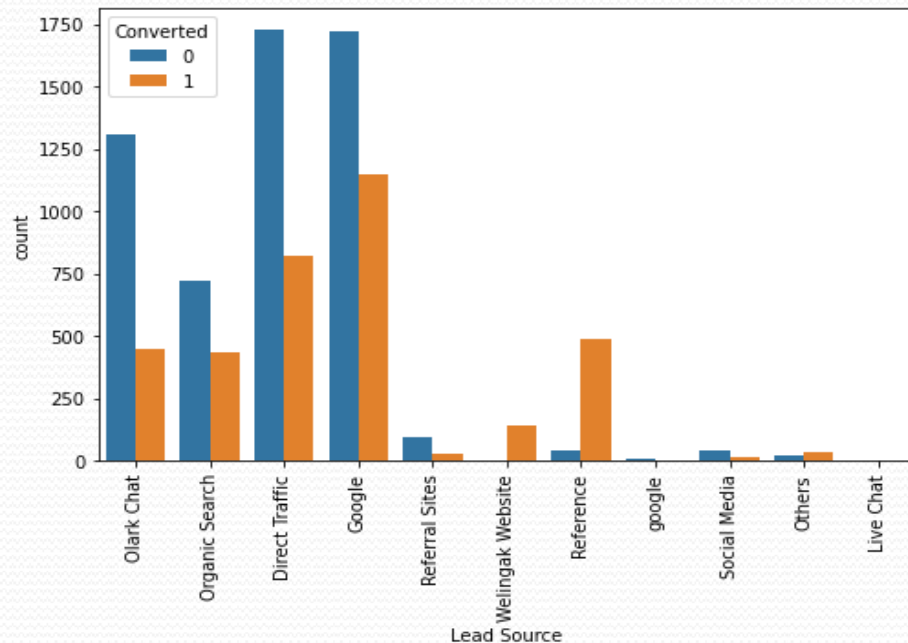Most of the leads will revert after reading the mail.

Most of the leads from Mumbai are converted.
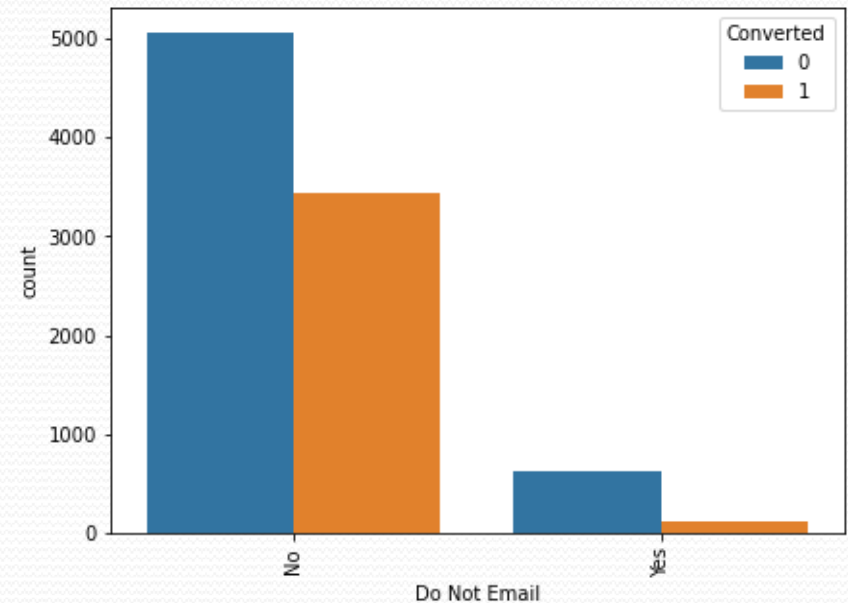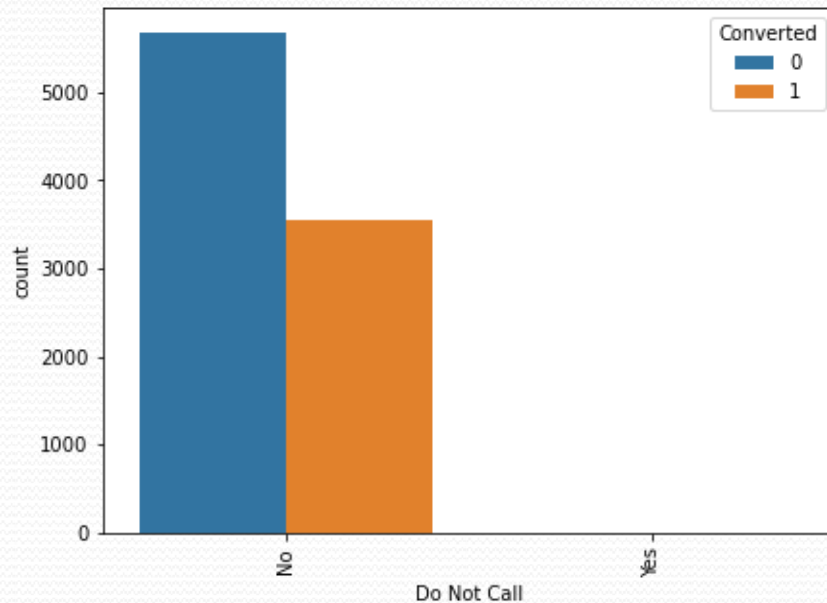
Converted vs. lead origin and lead source

Landing page submission has had high lead conversions.

Goggle source has had high conversions compared to other modes ,while references has had high conversions rate.
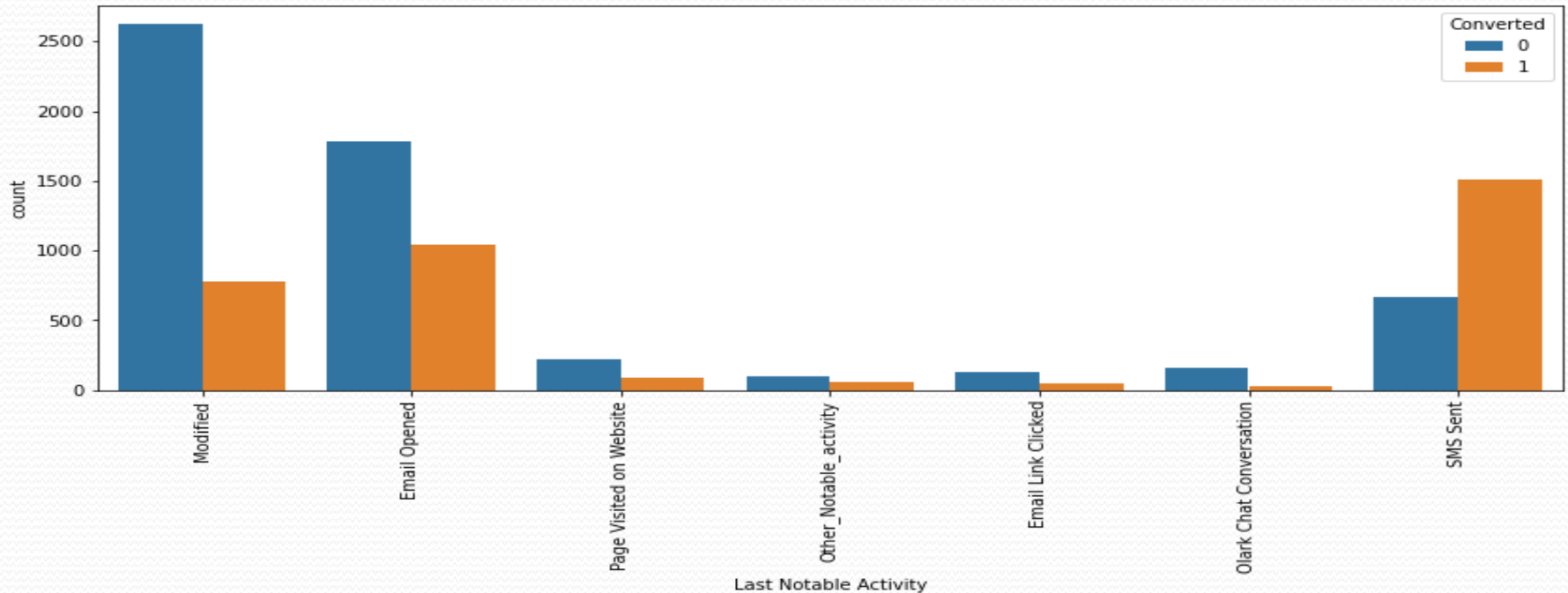
Converted vs. do not call and email

• Most leads prefer not to inform through phone.

• Goggle searches has had high conversions compare to other modes , whilst references has had high conversions rate.

Converted vs. last notable activity

Most leads are converted with messages.
Email also include leads.

Model building

- Splitting into train and test set
- Scale variables in train set
- Building the first model
- Use RFE
- Build the next model
- Check VIF
- Predict using train set
- Evaluate accuracy
- Predict using test set
- Precision and recall analysis on test predictions

Model evaluation(TRAIN)
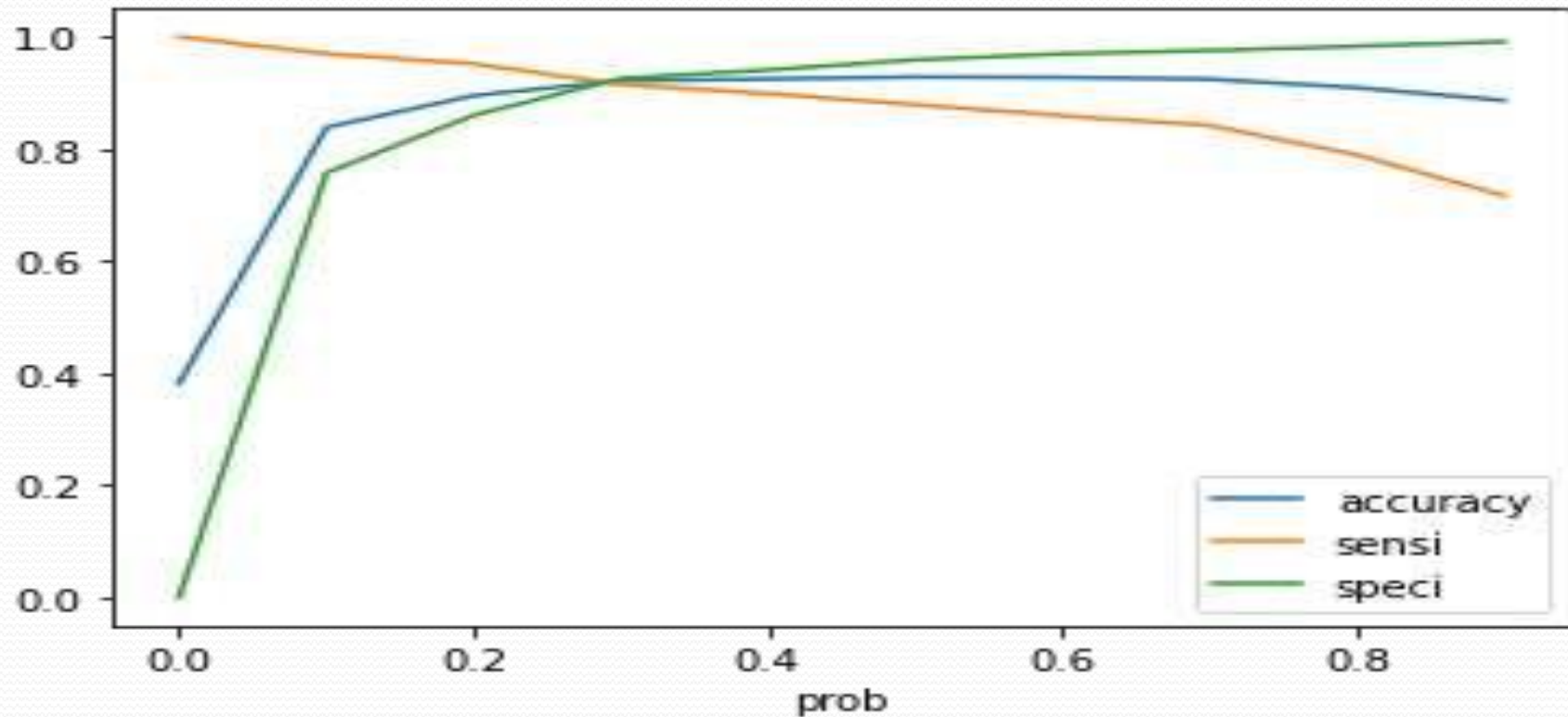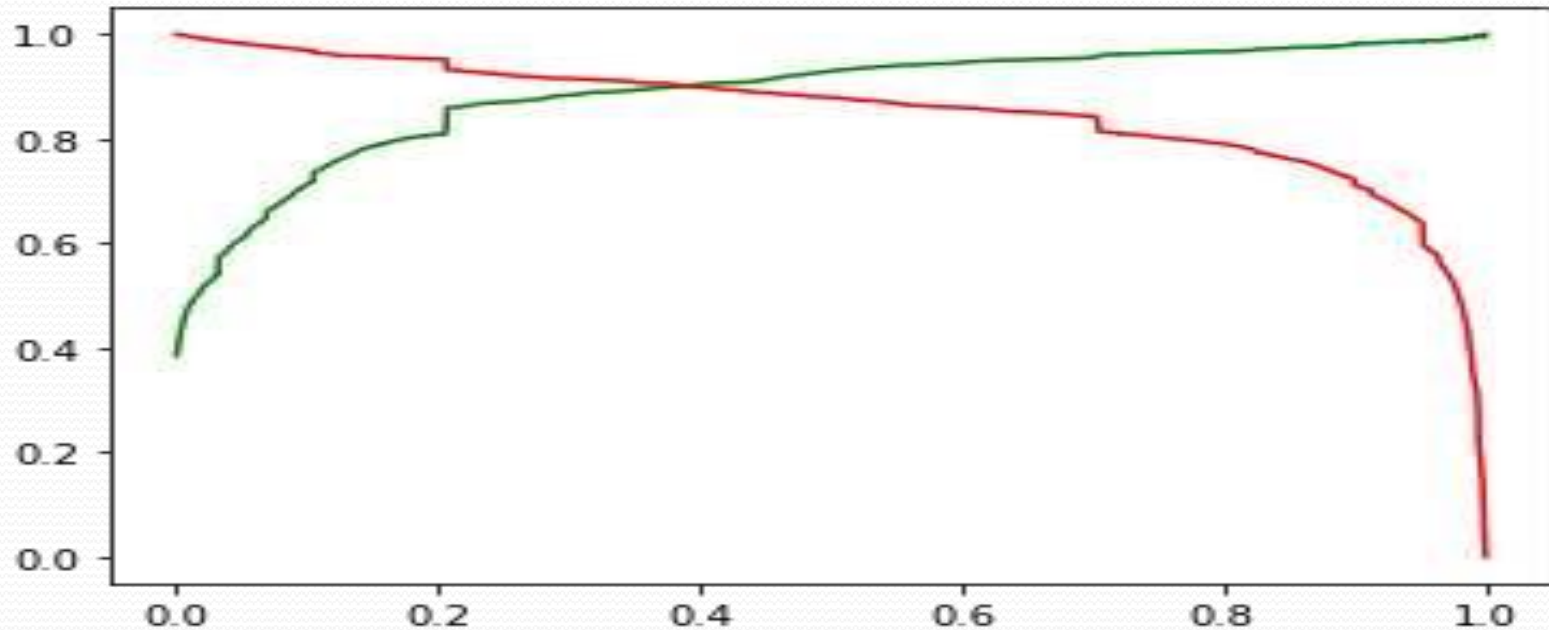

Accuracy sensitivity and specificity

Train Data:

Accuracy : 92.14%
Sensitivity : 91.49%
Specificity : 92.54%

Model evaluation(TEST)

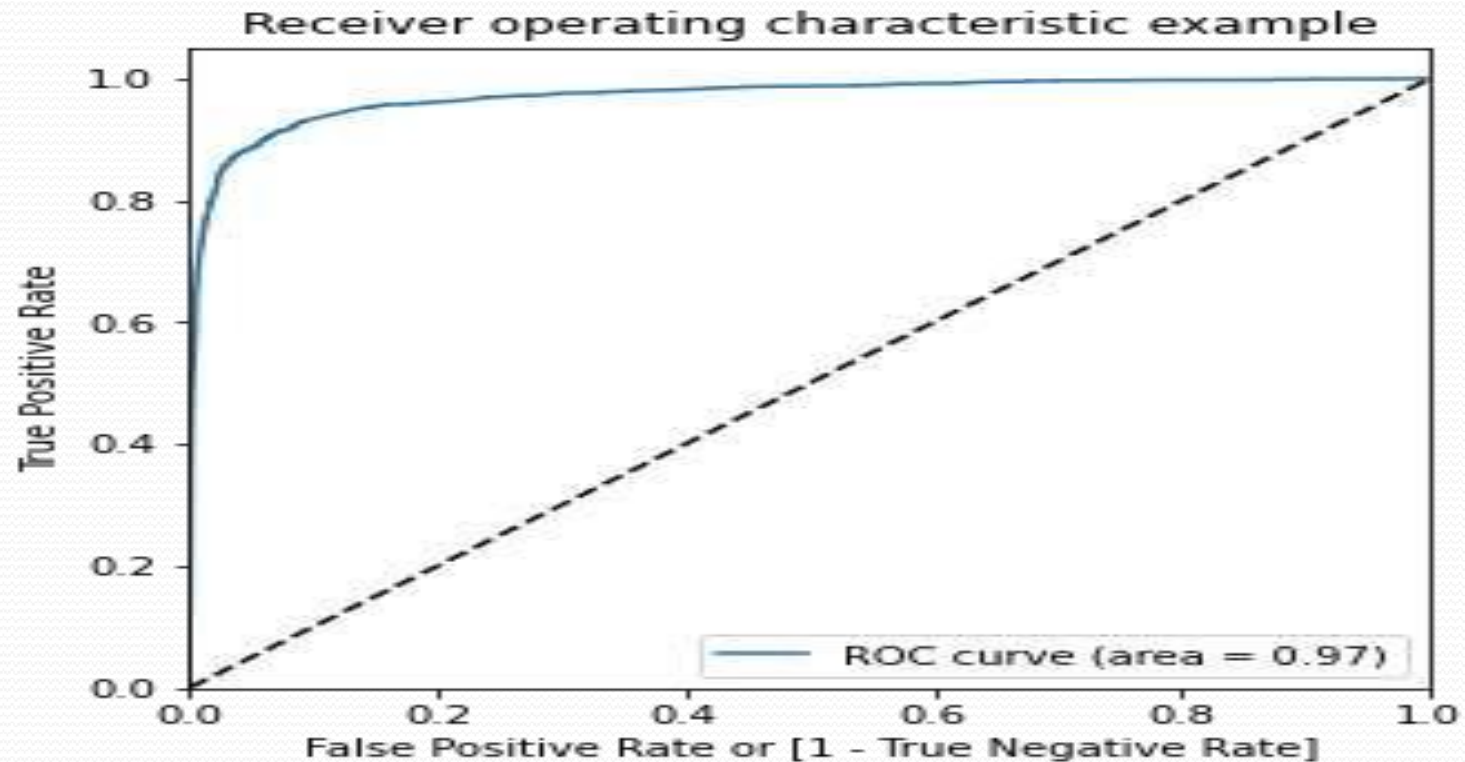Accuracy sensitivity and specificity
Test Data:
Accuracy : 92.57%
Sensitivity : 91.18%
Specificity : 93.46%

ROC curve

Optimal cutoff probability is that prob where we get balanced sensitivity and specificity.

Conclusion

- The model seems to be performing well. The ROC curve has a value of 0.97, which is very good.
- Train Data:
- Accuracy : 92.14%
- Sensitivity : 91.49%
- Specificity : 92.54%
- Test Data:
- Accuracy : 92.57%
- Sensitivity : 91.18%
- Specificity : 93.46%

- The model seems to be performing well. Can be recommend this model in making good calls based on this model.

Thank You!!