Lead Score Case Study

Lead Score Case Study for X Education

roblem Statement :

 Education sells online courses to industry professionals. The company markets its courses on several websites and search engines lik oogle.

nce these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When thes eople fill up a form providing their email address or phone number, they are classified to be a lead. Moreover, the company also gets eads through past referrals.

nce these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the eads get converted while most do not. The typical lead conversion rate at X education is around 30%.

usiness Goal:

 Education needs help in selecting the most promising leads, i.e. the leads that are most likely to convert into paying customers.

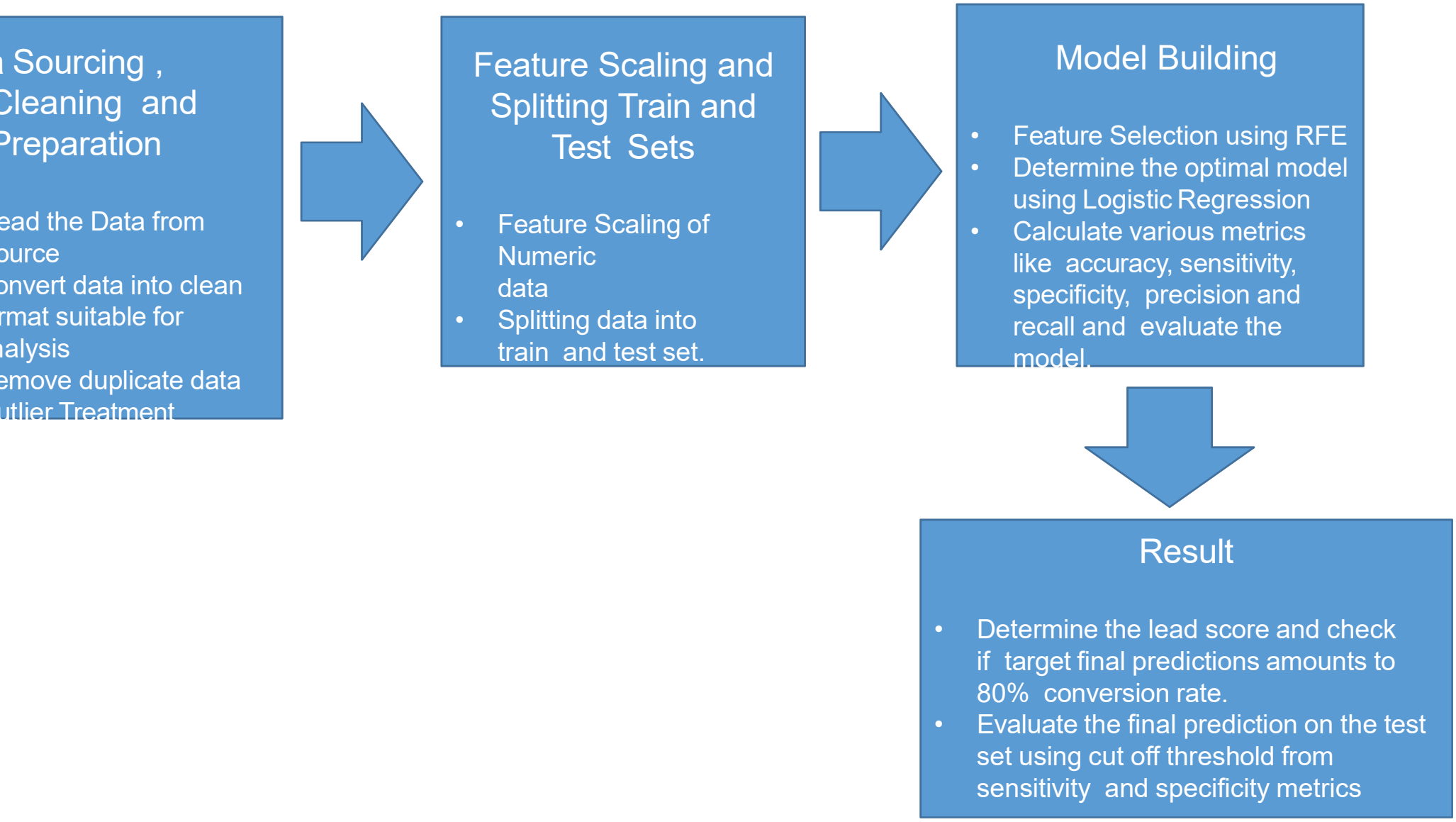
he company needs a model wherein you a lead score is assigned to each of the leads such that the customers with higher lead score ha higher conversion chance and the customers with lower lead score have a lower conversion chance.

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

Strategy

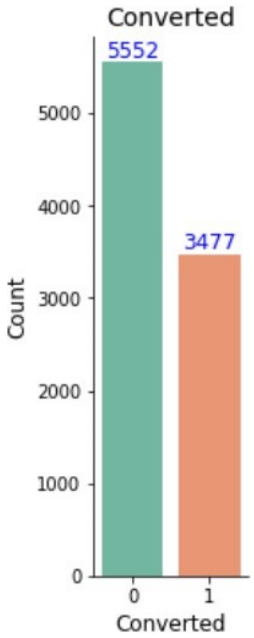
ource the data for analysis
lean and prepare the data
xploratory Data Analysis.
eature Scaling
plitting the data into Test and Train dataset.
uilding a logistic Regression model and calculate Lead Score.
valuating the model by using different metrics - Specificity and Sensitivity or Precision and Recall.
pplying the best model in Test data based on the Sensitivity and Specificity Metrics.

Problem solving methodology

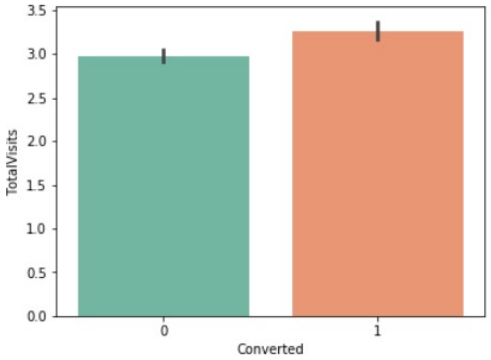


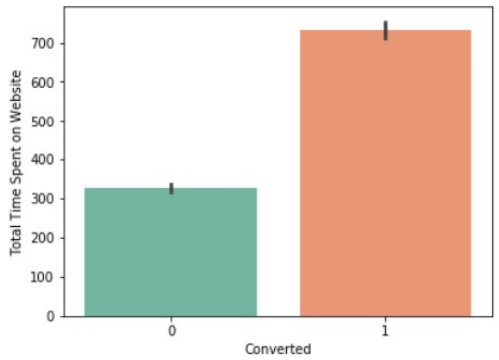
Exploratory Data Analysis

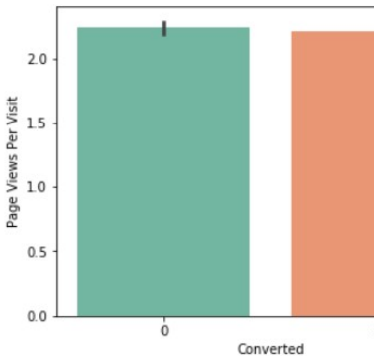
round 38.54% Conversion rate in



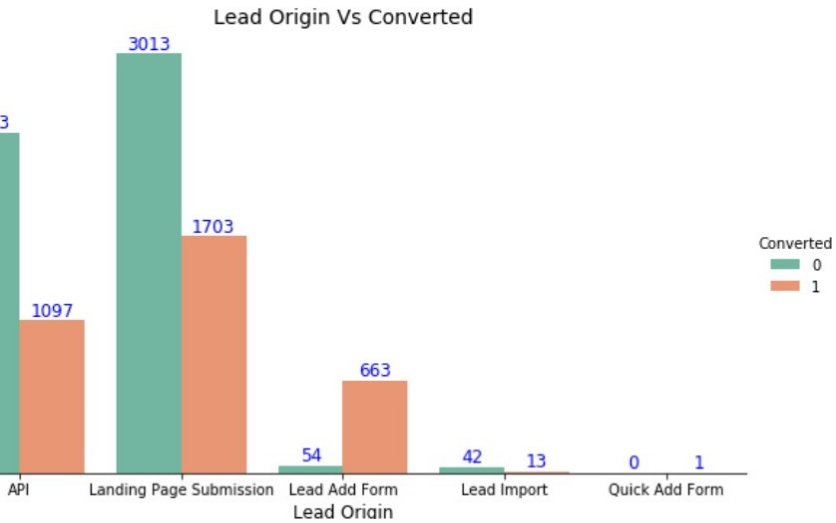
The conversion rates were high for Total Visits, Total Time Spent on Website and Page Per Visit



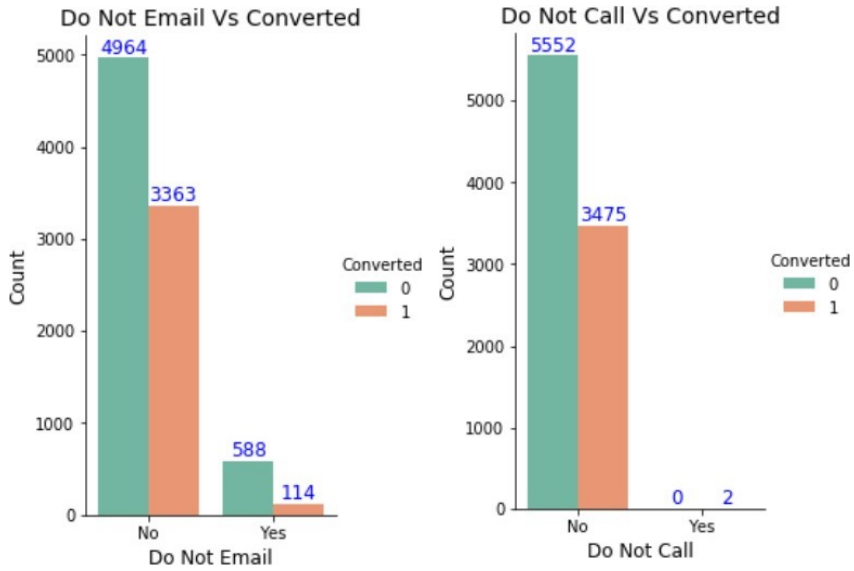




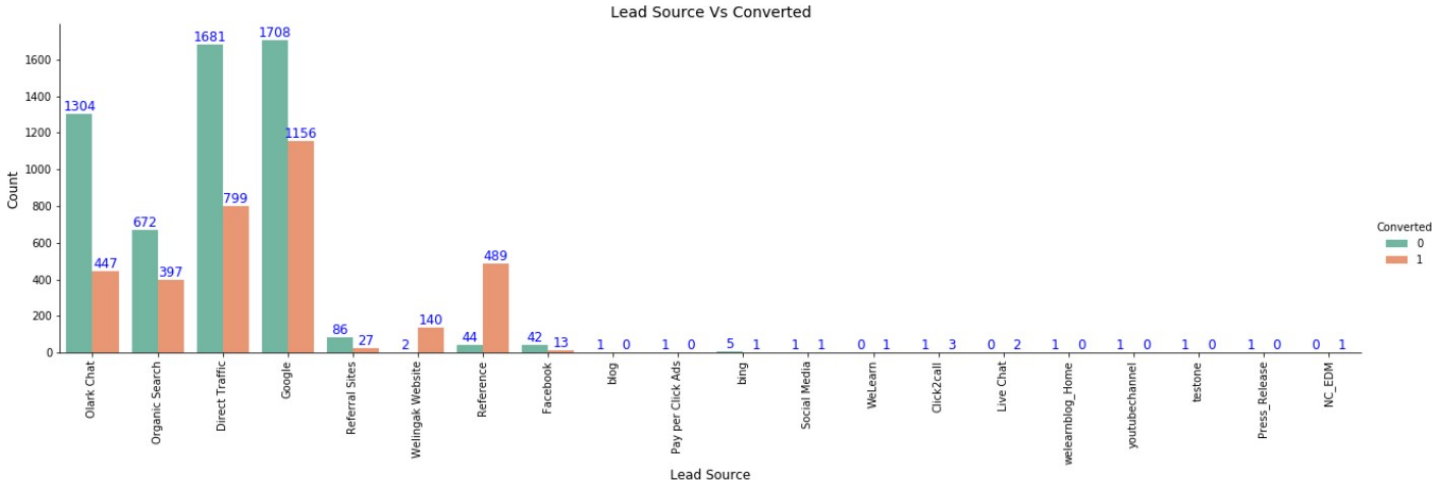
Origin, maximum conversion happened from Landing Page
ion



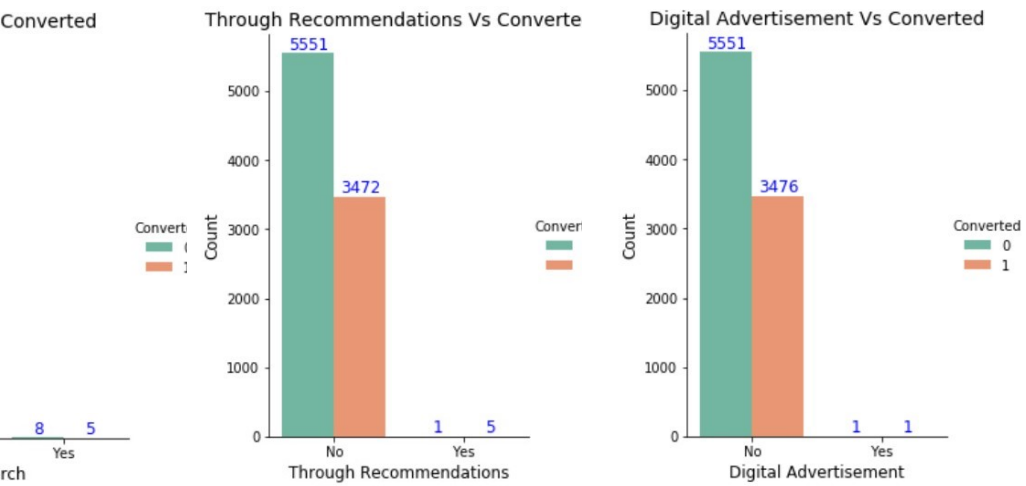
Major conversion has happened from Emails sent and Calls made



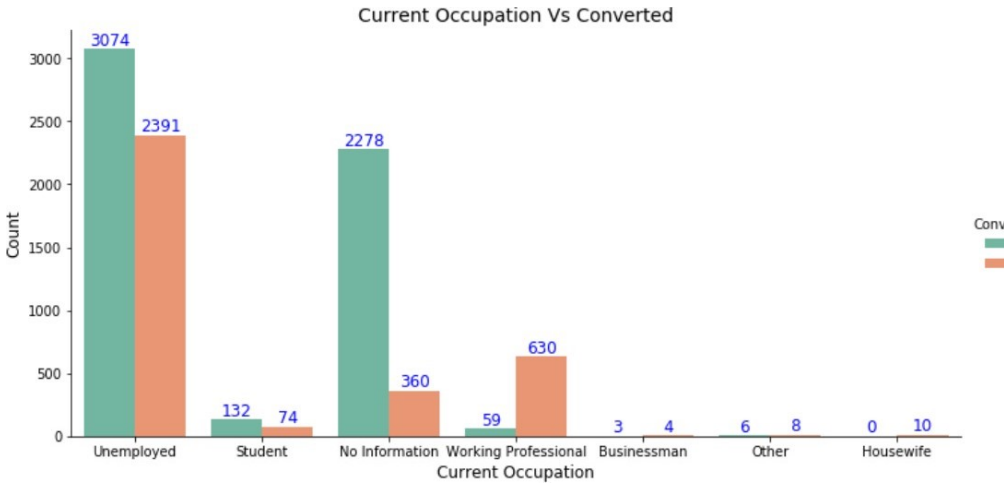
Major conversion in the lead source is from Google



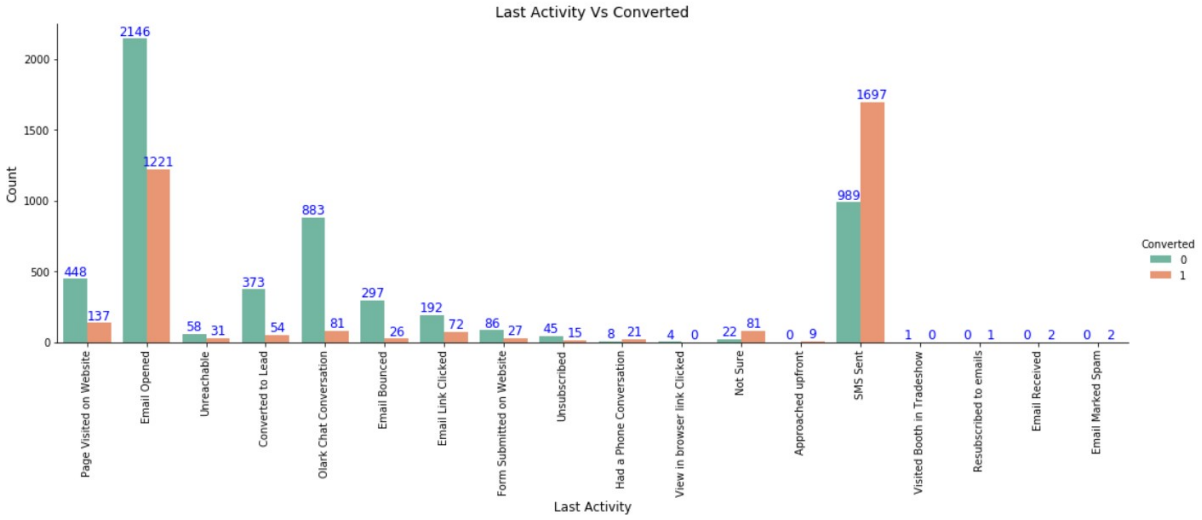
uch impact on conversion rates through Search, digital
isements and through recommendations



More conversion happened with people who are unemployed



Last Activity value of SMS Sent' had more conversion.

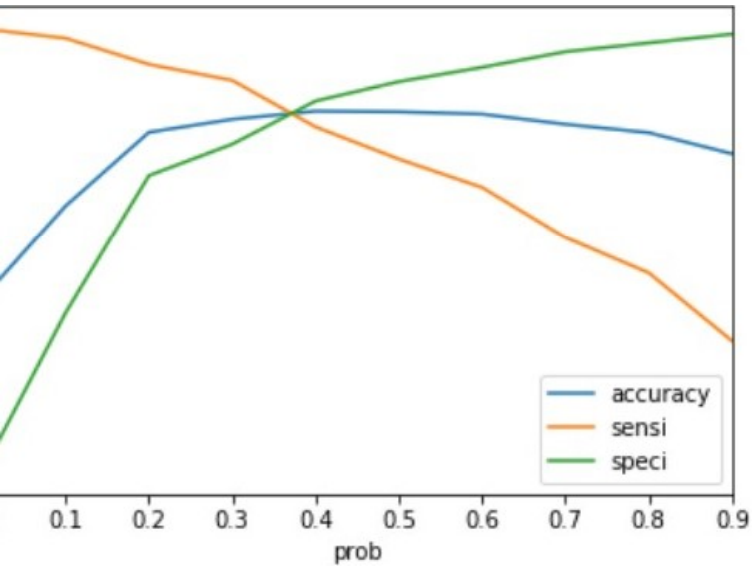


Variables Impacting the Conversion Rate

Not Email
al Visits
al Time Spent On Website
d Origin – Lead Page Submission
d Origin – Lead Add Form
d Source - Olark Chat
t Source – Welingak Website
t Activity – Email Bounced
t Activity – Not Sure
t Activity – Olark Chat Conversation
t Activity – SMS Sent
rent Occupation – No Information
rent Occupation – Working Professional
t Notable Activity – Had a Phone Conversation
t Notable Activity - Unreachable

Model Evaluation - Sensitivity and Specificity on Train Data Set

graph depicts an optimal cut off of 0.37 based on Accuracy, tivity and Specificity



Confusion Matrix

```
[[3514  488]
 [ 759 1707]]
```

- Accuracy - 79%
- Sensitivity - 79 %
- Specificity - 78 %

Model Evaluation – Sensitivity and Specificity on Test Dataset

Confusion Matrix

```
[1360,  317],
[ 240,  855]],
```

- Accuracy -79.76%
- Sensitivity - 78.26 %
- Specificity – 80.74 %

onclusion

e we have checked both Sensitivity-Specificity as well as Precision and Recall Metrics, we have considered the optimal cut off based
tivity and Specificity for calculating the final prediction. –

racy, Sensitivity and Specificity values of test set are around 79%, 79% and 78% which are approximately closer to the respective val
lated using trained set.

the lead score calculated shows the conversion rate on the final predicted model is around 79% (in train set) and 80% in test set

op 3 variables that contribute for lead getting converted in the model are

- Welingak Website, lead_source_Olark Chat and Reference from lead_source.
- SMS Sent and Other Activity from last_activity.
- time_on_website.

e overall this model seems to be good.