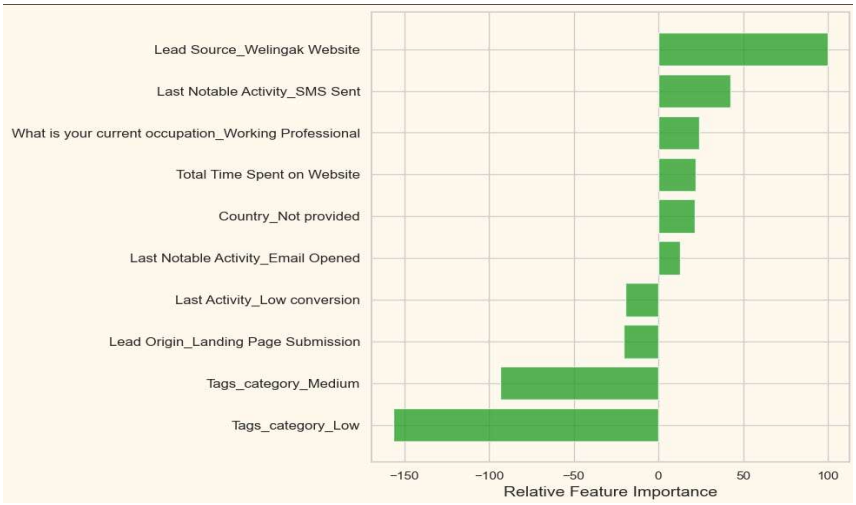
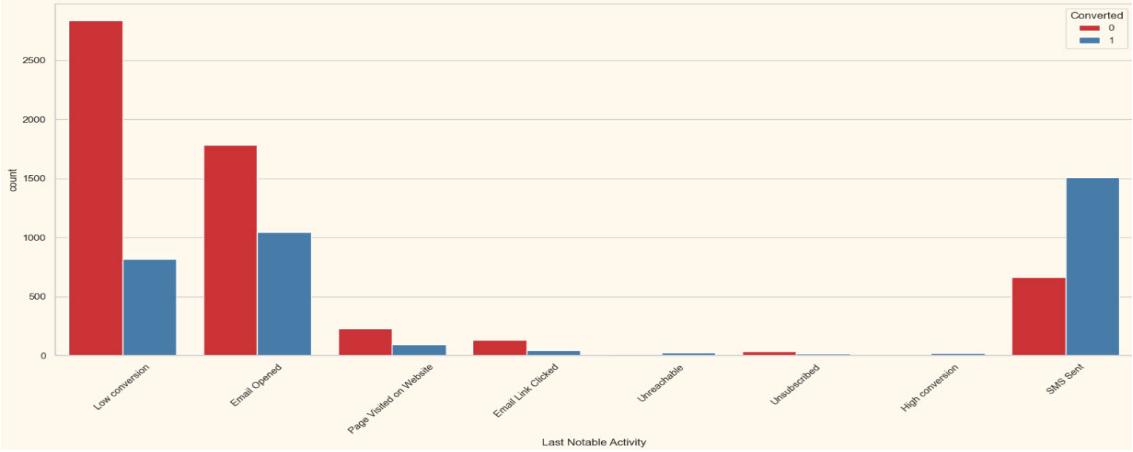
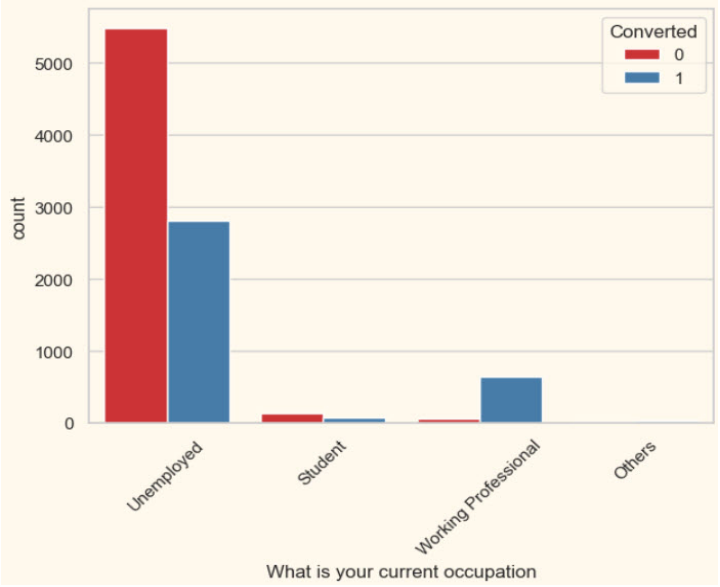
1. Which are the top three variables in your model which contribute most towards the probability of a lead getting converted?

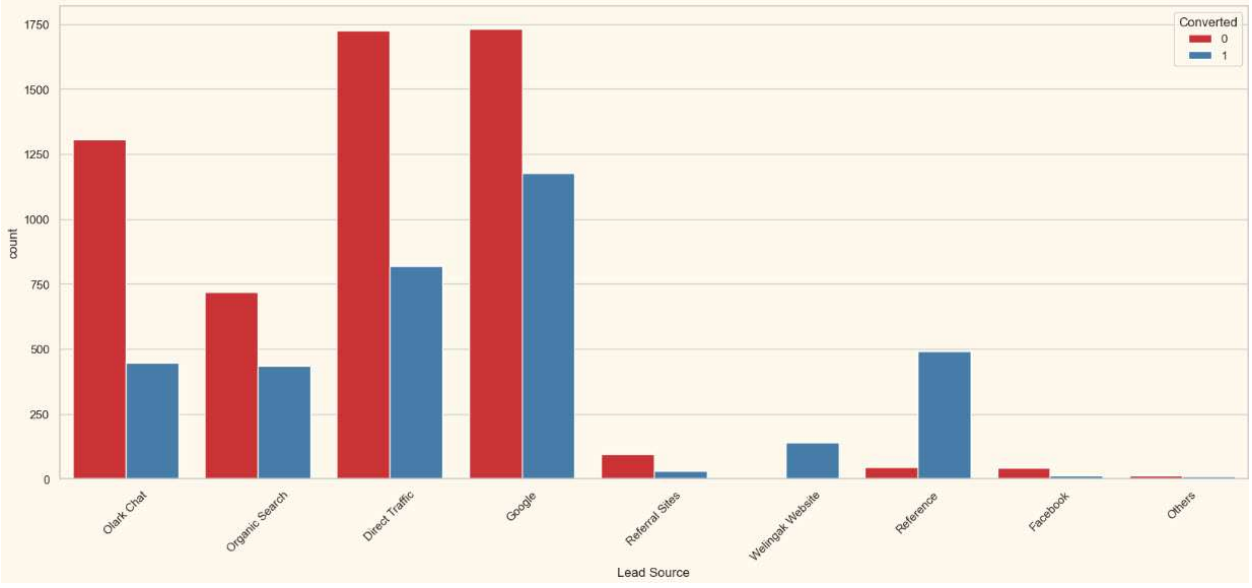
   Based on our final model, the top three variables which contribute most to the lead conversions are –
   - Lead_Source Welingak Website
   - Last Notable Activity_SMS sent
   - What is your current occupation_working professional



2. What are the top 3 categorical/dummy variables in the model which should be focused the most on in order to increase the probability of lead conversion?
   The top three variables which should be focused most to increase the Conversion rates are:
   - Lead Source – Welingak Website and References as they have very high conversion rates
   - Last Notable activity – SMS_Sent
   - What is your current occupation – Working Professionals


Converted
0
1
count
2500
2000
1500
1000
500
0
Low conversion
Email Opened
Page Visited on Website
Email Link Clicked
Unreachable
Unsubscribed
High conversion
SMS Sent
Last Notable Activity


Converted
0
1
count
5000
4000
3000
2000
1000
0
Unemployed
Student
Working Professional
Others
What is your current occupation


Converted
0
1
count
1750
1500
1250
1000
750
500
250
0
Olark Chat
Organic Search
Direct Traffic
Google
Referral Sites
Welingak Website
Reference
Facebook
Others
Lead Source

All the three features have very high conversion rate.

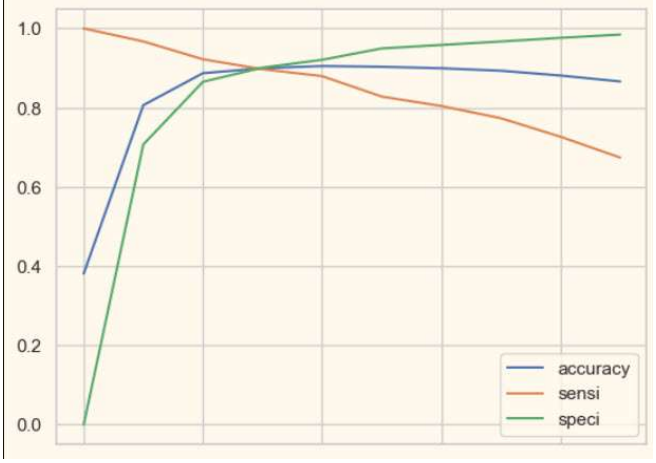
3. X Education has a period of 2 months every year during which they hire some interns. The sales team, in particular, has around 10 interns allotted to them. So during this phase, they wish to make the lead conversion more aggressive. So they want almost all of the potential leads (i.e. the customers who have been predicted as 1 by the model) to be converted and hence, want to make phone calls to as much of such people as possible. Suggest a good strategy they should employ at this stage.

Solution: Since X Education wants to **maximize lead conversion** and **make phone calls to as many potential leads as possible** they should take the following steps:

- **Prioritize recall (sensitivity)** over precision. This ensures they capture **most of the potential leads (customers predicted as 1)** and avoid missing out on possible conversions. By revising the threshold to 0.3 our recall **Improved** from **82.83%** to **90.78%** (More potential conversions are identified)
- **Specificity slightly decreased** from **94.95%** to **90.94%** (More false positives). But still it's high so model is still rejecting irrelevant customers well.
- Instead of calling all predicted leads at once, sort them by **conversion probability** and **call high-confidence leads first**. This will optimize intern's time
- **Implement multi touch follow up strategy:** Use a **CRM system** to track interactions and automate follow-ups.
- **Optimize call strategy** by dividing the interns into shifts to maximize the coverage over different time slots(morning, afternoon, evening)

	prob	accuracy	sensi	speci
0	0.0	0.381262	1.000000	0.000000
1	0.1	0.806277	0.967153	0.707146
2	0.2	0.886982	0.922141	0.865317
3	0.3	0.899505	0.896999	0.901049
4	0.4	0.905071	0.879562	0.920790
5	0.5	0.903216	0.828062	0.949525
6	0.6	0.899505	0.804136	0.958271
7	0.7	0.893166	0.773317	0.967016
8	0.8	0.881107	0.726683	0.976262
9	0.9	0.866110	0.673966	0.984508

```
cutoff_df.plot.line(x='prob',y=['accuracy','sensi','speci'])
plt.show()
```



4. Similarly, at times, the company reaches its target for a quarter before the deadline. During this time, the company wants the sales team to focus on some new work as well. So during this time, the company's aim is to not make phone calls unless it's extremely necessary, i.e. they want to minimize the rate of useless phone calls. Suggest a strategy they should employ at this stage.

   Solution

   Since the company has already met its targets for the quarter and now wants to **avoid unnecessary phone calls**, the focus should be on **precision (positive predictive value, PPV)** rather than recall.

   This could be achieved by raising the threshold to (0.7 or 0.8) will classify fewer customers as leads, but those who remain will have a **higher probability of conversion**.