

## Answers for Subjective Questions

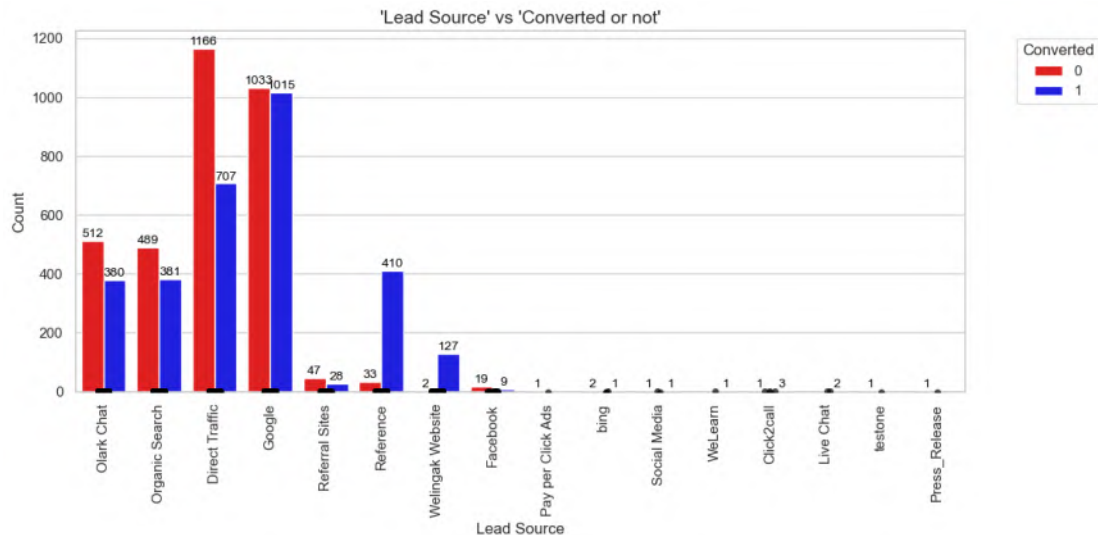
1. Which are the top three variables in your model which contribute most towards the probability of a lead getting converted?

Answer: Top 3 variables affecting conversion probability during our analysis:

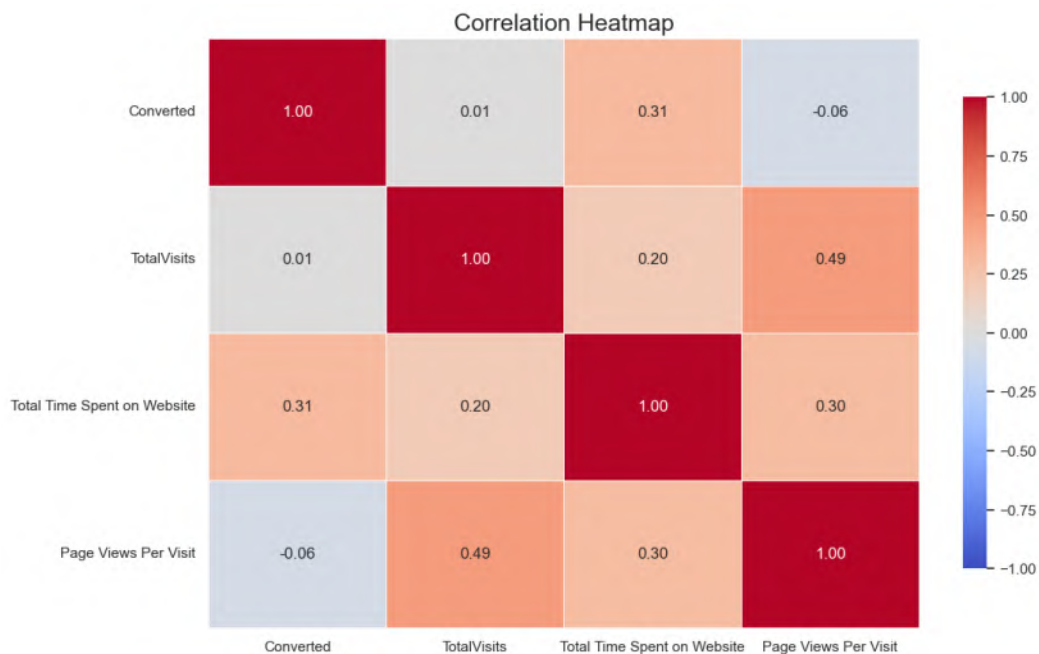
List of Features with the Descending Order of Coefficients:

	Feature	Coefficient	P-value
5	Lead Source_Welingak Website	5.455	5.205e-14
4	Lead Source_Reference	3.457	1.404e-51
12	Last Notable Activity_Unreachable	2.552	1.732e-03
2	Total Time Spent on Website	1.908	2.889e-124
1	Do Not Email	-1.537	9.869e-16
10	What is your current occupation_Unemployed	-1.506	9.906e-03
9	What is your current occupation_Student	-1.428	2.124e-02
3	Lead Source_Olark Chat	1.424	9.176e-34
7	Last Activity_Olark Chat Conversation	-1.336	3.436e-13
11	What is your current occupation_Working Professional	1.275	3.824e-02
6	Last Activity_Converted to Lead	-1.246	1.265e-07
8	Last Activity_SMS Sent	1.046	4.751e-36
0	const	0.074	8.995e-01

- **Lead Source:** The emphasis should be placed on cultivating high-quality lead sources, specifically the Welingak Website and referrals. These two channels have been identified as critical factors positively influencing conversion rates within the effectively evaluated logistic regression model. Furthermore, bivariate analysis related to the target variable "Converted" reveals a distinct difference between the "Converted" and "Non-converted" groups, underscoring the importance of these lead sources.



- **Last Notable Activity:** Attention should be paid to the last notable activities of the lead, especially when the lead is unreachable (Unreachable). This can help identify leads that need further assistance or follow-up.
- **Total Time Spent on Website:** The logistic regression model indicates a significant coefficient of 1.908 for this variable, highlighting that increased user engagement with website content significantly predicts conversion. Furthermore, among the three continuous numeric features analyzed, this variable exhibits the highest correlation with the target variable "Converted." This correlation, with a coefficient of 0.31, suggests a moderate positive association, reinforcing the finding that leads who dedicate more time to exploring the website are more likely to convert.



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?

Answer:

- **Lead Source\_Welingak Website:** This variable has a regression coefficient of 5.455, indicating that leads coming from the Welingak website have a significantly highest conversion probability compared to other sources.
- **Lead Source\_Reference:** Leads referred from referral sources have a regression coefficient of 3.457, indicating that this source brings in high-quality leads with good conversion potential.
- **Last Notable Activity\_Unreachable:** This variable demonstrates strong predictive power for conversion. With a substantial regression coefficient of 2.552, this category ranks third among high-quality lead sources associated with strong conversion potential. This unexpected finding suggests that leads categorized as "Unreachable" within this specific activity context may represent a valuable segment warranting further investigation and tailored outreach strategies.

**Top 3 Features Positively Impacting Conversion:**

	Feature	Coefficient	P-value
5	Lead Source_Welingak Website	5.455	5.205e-14
4	Lead Source_Reference	3.457	1.404e-51
12	Last Notable Activity_Unreachable	2.552	1.732e-03

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.

Answer: X Education can leverage the provided logistic regression model and its internship period to maximize lead conversion with this strategy:

**A. Prioritize High-Probability Leads:**

- Utilize the model to predict the conversion probability of each lead, with the optimised cutoff (0.431) from the Precision-Recall trade-off (*see Note 24 in my Jupyter notebook*)

The Effectively Evaluated Logistic Regression Model Equation is:  

$$\text{Logit}(p) = \text{Log}(p/1-p) = 5.455 * \text{Lead Source\_Welingak Website} + 3.457 * \text{Lead Source\_Reference} + 2.552 * \text{Last Notable Activity\_Unreachable} + 1.908 * \text{Total Time Spent on Website} + 1.424 * \text{Lead Source\_Olark Chat} + 1.275 * \text{What is your current occupation\_Working Profession} + 1.046 * \text{Last Activity\_SMS Sent} + 0.074 * \text{const} - 1.537 * \text{Do Not Email} - 1.506 * \text{What is your current occupation\_Unemployed} - 1.428 * \text{What is your current occupation\_Student} - 1.336 * \text{Last Activity\_Olark Chat Conversation} - 1.246 * \text{Last Activity\_Converted to Lead}$$
 (with p as the Probability of Conversion)

- Focus on leads with the highest probability (closer to 1). Pay close attention to the features with high positive coefficients, such as "Lead Source\_Welingak Website", "Lead Source\_Reference", "Last Notable Activity\_Unreachable", "Total Time Spent on Website", etc.

#### Top 3 Features Positively Impacting Conversion:

	Feature	Coefficient	P-value
5	Lead Source_Welingak Website	5.455	5.205e-14
4	Lead Source_Reference	3.457	1.404e-51
12	Last Notable Activity_Unreachable	2.552	1.732e-03

#### B. Optimize Communication:

- Assign interns to contact prioritized leads via phone calls.
- Develop effective call scripts tailored to different lead segments based on source, recent activity, occupation, etc.
- Train interns on product knowledge, communication skills, and objection handling.
- Monitor performance, gather feedback, and refine the communication strategy.

#### C. Leverage Multiple Channels:

- Integrate phone calls with other channels like email, SMS, and chat to increase reach and conversion.
- Personalize messaging based on individual lead information.

#### D. Acknowledge/ Consider Negative Factors:

- The model shows certain factors can negatively impact conversion, such as "Do Not Email", "Unemployed/Student" occupation, and "Last Activity\_Olark Chat Conversation/Converted to Lead".
- Exercise caution when approaching these leads and adapt communication accordingly.

### Top 3 Features Negatively Impacting Conversion:

	Feature	Coefficient	P-value
1	Do Not Email	-1.537	9.869e-16
10	What is your current occupation_Unemployed	-1.506	9.906e-03
9	What is your current occupation_Student	-1.428	2.124e-02

This strategy empowers X Education to utilize its interns effectively, driving lead conversion during the internship period.

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.

Answer: When X Education achieves its quarterly targets ahead of schedule, it's wise to shift the sales team's focus to new initiatives while minimizing unnecessary calls. Here's a strategy to achieve this:

- **Identify Low-Probability Leads:** Utilize the logistic regression model to identify leads with a low probability of conversion. These are the individuals where phone calls are least likely to yield positive results.
- **Set a Higher Threshold for Contact:** Instead of contacting all leads above a certain probability, raise the threshold significantly. For example, only contact leads with a predicted conversion probability above 80% or 90%. This ensures that only the most promising leads are pursued.
- **Prioritize Other Communication Channels:** For leads below the contact threshold, focus on less resource-intensive channels like email marketing, targeted content, or social media engagement. These channels can nurture leads and keep them engaged without requiring direct phone calls.
- **Automate Communication:** Implement automated email sequences or chatbots to provide information and answer basic questions, reducing the need for human intervention.
- **Empower Leads with Self-Service:** Enhance the company's website and online resources to provide comprehensive information, FAQs, and self-service options. This allows potential customers to find answers independently, reducing their need to contact the sales team.
- **Focus on Relationship Building:** Shift the sales team's focus towards building relationships with existing clients, exploring upselling and cross-selling opportunities, and generating referrals.

By implementing this strategy, X Education can effectively minimize unnecessary calls while maintaining customer engagement and pursuing new opportunities when quarterly targets are

met early. This approach ensures efficient resource allocation and allows the sales team to contribute to broader company objectives.