## SUBJECTIVE QUESTIONS

---

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

---

**ANSWER:**
1. lead_source_Reference
2. lead_source_Other Social Sites
3. do_not_email

**Explanation**

→ This question focuses on overall model feature importance.

**1. Analyze Feature Ranking:**

- **RFE:** which rank features based on their elimination order. The top-ranked features (with support_ = True) are considered more important.

```
[('do_not_email', True, 1),
 ('total_visits', True, 1),
 ('time_on_website', True, 1),
 ('page_views_per_visit', True, 1),
 ('mastering_interview', False, 3),
 ('lead_origin_Landing Page Submission', True, 1),
 ('lead_source_Google', True, 1),
 ('lead_source_Olark Chat', True, 1),
 ('lead_source_Organic Search', True, 1),
 ('lead_source_Other Social Sites', True, 1),
 ('lead_source_Reference', True, 1),
 ('specialization_Management Specializations', False, 2),
 ('occupation_Unemployed', True, 1),
 ('city_Non-Maharashtra Cities', True, 1),
 ('city_Non-Mumbai Maharashtra Cities', True, 1)]
```

- **GLM Coefficients:** Analyze the absolute value of the regression coefficients in the GLM results. Larger coefficients (positive or negative) indicate a stronger association with the target variable (conversion). Consider features with the highest absolute coefficients (excluding the intercept) as potentially important.

| | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| const | -0.3534 | 0.147 | -2.399 | 0.016 | -0.642 | -0.065 |
| #3 do_not_email | -1.2038 | 0.144 | -8.333 | 0.000 | -1.487 | -0.921 |
| total_visits | 0.1415 | 0.042 | 3.373 | 0.001 | 0.059 | 0.224 |
| time_on_website | 1.0413 | 0.036 | 29.257 | 0.000 | 0.972 | 1.111 |
| page_views_per_visit | -0.1825 | 0.048 | -3.775 | 0.000 | -0.277 | -0.088 |
| mastering_interview | 0.0007 | 0.094 | 0.007 | 0.994 | -0.183 | 0.185 |
| lead_origin_Landing Page Submission | 9.495e-05 | 0.092 | 0.001 | 0.999 | -0.181 | 0.181 |
| lead_source_Google | 0.3617 | 0.100 | 3.602 | 0.000 | 0.165 | 0.558 |
| lead_source_Olark Chat | 0.6850 | 0.137 | 5.016 | 0.000 | 0.417 | 0.953 |
| lead_source_Organic Search | 0.2099 | 0.116 | 1.811 | 0.070 | -0.017 | 0.437 |
| #2 lead_source_Other Social Sites | 1.6308 | 0.175 | 9.308 | 0.000 | 1.287 | 1.974 |
| #1 lead_source_Reference | 3.9581 | 0.221 | 17.921 | 0.000 | 3.525 | 4.391 |
| specialization_Management Specializations | 0.0273 | 0.069 | 0.394 | 0.693 | -0.108 | 0.163 |
| occupation_Unemployed | -0.8496 | 0.086 | -9.917 | 0.000 | -1.018 | -0.682 |
| city_Non-Maharashtra Cities | 0.0758 | 0.078 | 0.966 | 0.334 | -0.078 | 0.230 |
| city_Non-Mumbai Maharashtra Cities | 0.0939 | 0.076 | 1.234 | 0.217 | -0.055 | 0.243 |

**2. Combine Both Methods:**

- **Cross-Validate Importance:** Compare the feature rankings from RFE and GLM coefficients. Features consistently appearing at the top in both methods are likely the most influential.

**3. Domain Knowledge and Interpretation:**

- **Consider the context:** While both ranking methods are valuable, use the understanding of the data and domain knowledge to interpret the results.

---

*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:**
Categorical/dummy + High Positive contribution are

1. lead_source_Reference
2. lead_source_Other Social Sites
3. lead_source_Olark Chat

→ Should be focused the most

This question specifically asks for the top 3 categorical/dummy variables, implying focus on discrete features. → This targets **actionable variables** where we can directly take steps to influence conversion

| | | | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|---|---|
| | | const | -0.3534 | 0.147 | -2.399 | 0.016 | -0.642 | -0.065 |
| | | do_not_email | -1.2038 | 0.144 | -8.333 | 0.000 | -1.487 | -0.921 |
| | | total_visits | 0.1415 | 0.042 | 3.373 | 0.001 | 0.059 | 0.224 |
| | | time_on_website | 1.0413 | 0.036 | 29.257 | 0.000 | 0.972 | 1.111 |
| | | page_views_per_visit | -0.1825 | 0.048 | -3.775 | 0.000 | -0.277 | -0.088 |
| | | mastering_interview | 0.0007 | 0.094 | 0.007 | 0.994 | -0.183 | 0.185 |
| | | lead_origin_Landing Page Submission | 9.495e-05 | 0.092 | 0.001 | 0.999 | -0.181 | 0.181 |
| | | lead_source_Google | 0.3617 | 0.100 | 3.602 | 0.000 | 0.165 | 0.558 |
| Categorical/dummy | #3 | lead_source_Olark Chat | 0.6850 | 0.137 | 5.016 | 0.000 | 0.417 | 0.953 |
| | | lead_source_Organic Search | 0.2099 | 0.116 | 1.811 | 0.070 | -0.017 | 0.437 |
| Categorical/dummy | #2 | lead_source_Other Social Sites | 1.6308 | 0.175 | 9.308 | 0.000 | 1.287 | 1.974 |
| Categorical/dummy | #1 | lead_source_Reference | 3.9581 | 0.221 | 17.921 | 0.000 | 3.525 | 4.391 |
| | | specialization_Management Specializations | 0.0273 | 0.069 | 0.394 | 0.693 | -0.108 | 0.163 |
| | | occupation_Unemployed | -0.8496 | 0.086 | -9.917 | 0.000 | -1.018 | -0.682 |
| | | city_Non-Maharashtra Cities | 0.0758 | 0.078 | 0.966 | 0.334 | -0.078 | 0.230 |
| | | city_Non-Mumbai Maharashtra Cities | 0.0939 | 0.076 | 1.234 | 0.217 | -0.055 | 0.243 |

---

*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:**

1. **Tier leads by predicted probability:** Instead of calling all "1" predictions, prioritize leads closest to 1 (highest conversion likelihood). Then, segment by "lead_source" ("Reference", "Other Social Sites", and "Olark Chat"). This prioritizes the most promising leads within segments with positive conversion influence.
2. **Leverage interns for initial outreach:** Utilize interns for email/short call outreach while reserving experienced salespeople for higher-priority or challenging leads. Train interns on effective scripts and personalized messaging based on segments.

3. **Track individual performance:** Monitor both lead conversion and intern success to identify individuals excelling in specific segments and provide targeted coaching.
4. **Multi-channel approach:** Don't solely rely on phone calls. Use emails, social media messages, or SMS personalized based on lead profiles and segment preferences.
5. **Limited-time incentives:** Offer exclusive discounts or promotions during the internship period to encourage immediate conversion.
6. **Address "do_not_email" preferences:** If a lead opted out of email, respect their choice and prioritize alternative channels like phone calls or SMS with clear value propositions.

---

*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:**

1. **Recalibrate lead priority:** Instead of solely relying on predicted conversion probability, consider incorporating additional factors like:
    - **Engagement:** Prioritize leads actively interacting with emails, website content, or social media to gauge genuine interest.
    - **Lead source:** Focus on sources like "Reference" or "Olark Chat" with known positive conversion influence.
    - **Time since last contact:** Engage with leads who haven't been contacted recently to avoid oversaturation.
    - **Utilize scoring system:** Develop a scoring system that combines multiple factors (predicted probability, engagement, etc.) to rank leads and prioritize those most likely to benefit from contact without requiring aggressive phone calls.