Logistics Regression Assignment

Blending Business Insight with Technical Execution

Agenda

- Introduction
- Problem Statement
- Technical Solution
- Most Critical Features Influencing the Lead Conversion
- How the Features impact
- Predictions on the basis on probability
- Insights from the model
- Conclusion and Next Steps

Introduction

- From Business Perspective:
  - Need to understand the leads which has a chance for conversion
  - Decision making from the outcome of the model using a scoring mechanism, to make decision faster and efficient

- From Technical Perspective:
  - Use Data science techniques to clean, analyze and get scoring for each leads
  - Automated process to provide information on the scoring process

Problem Statement

- X Education aims to improve its lead conversion rate from 30% to a target of 80% by focusing on high-potential leads.
- The current process generates numerous leads, but only a small fraction converts into paying customers.
- The goal is to build a predictive model that assigns a lead score to prioritize leads most likely to convert.
- A higher lead score will help the Marketing team focus efforts on nurturing promising leads, increasing overall conversion efficiency.

Technical Solution

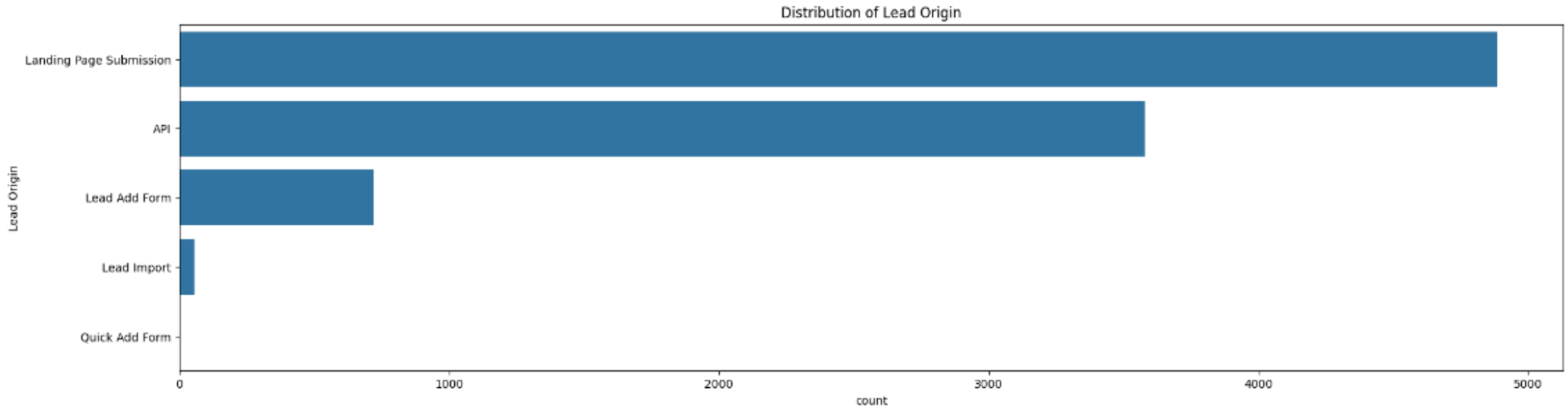
- Since the Target Variable is a categorical type, we will be using Logistic Regression Model to assign lead scores to each leads.
- Based on the above scores the company can differentiate between the Hot and cold leads.
- From analyzing the parameters for each variable, we can decide upon the future strategies of communication by sales team.
- This will also strategize the company to decide on the future investment areas.

Model Selected Features for Lead Conversion

Lead Origin_Lead Add Form	3.083239
const	2.283911
What is your current occupation_Working Professional	1.525493
Total Time Spent on Website	1.028789
Lead Source_Olark Chat	0.831544
Lead Origin_Landing Page Submission	-0.270319
Lead Source_Referral Sites	-0.706979
What is your current occupation_Unknown	-0.742412
Lead Source_Facebook	-0.942725
Do Not Email	-0.950541
Specialization_Hospitality Management	-1.059615
Lead Quality_Low in Relevance	-1.320248
Lead Quality_Might be	-1.627085
Lead Quality_Unknown	-3.397079
Lead Quality_Not Sure	-3.525185
Lead Quality_Worst	-6.135582

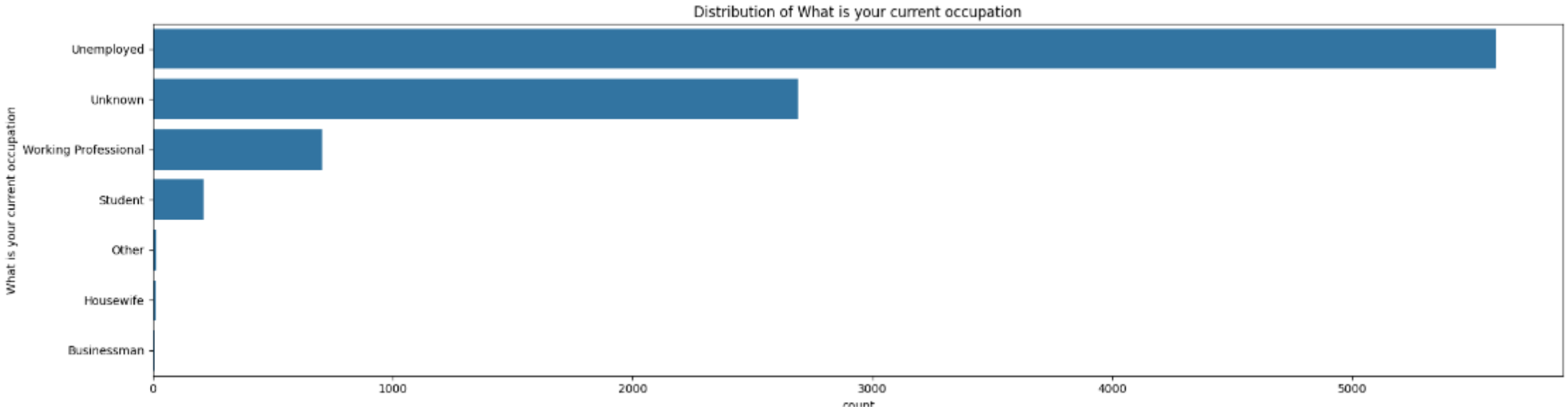
Lead origin

- Sales Team should focus more on the Lead Add form which have high chance of conversion.
- Sales team needs to improve the Landing page submission and API as it has large distribution but its not showing the in the lead conversion.



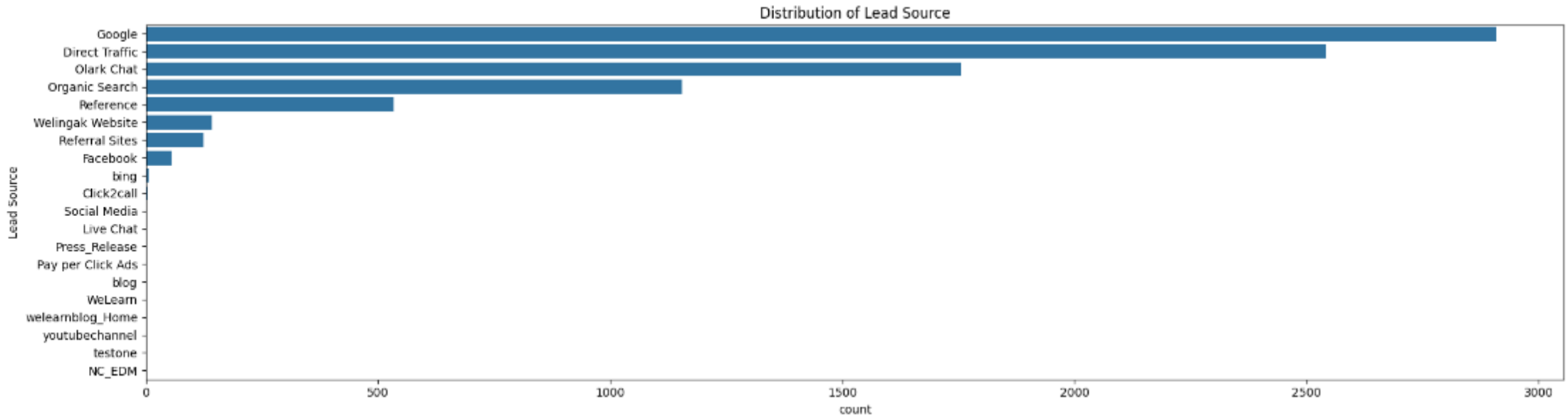
What is your current occupation

- Sales Team should focus more on the Working professional and student's category as they have high chances of lead conversion
- Sales team has a possibility of bringing more unemployed and unknown sector to the conversions by providing job assurance based on their merit, this may be treated as a strategy, as there is a huge potential of distribution.



Lead Source

- Sales Team should focus on leads from Olark Chat as they have high chances of lead conversion
- Sales team should focus in developing more chatbot type app’s rather than direct traffic and Google platforms. As we see although we have high distribution of Google Source and social media, the conversion rate is very less.



Train Data

Accuracy	Sensitivity	Specificity	Precision	Recall
82.37%	82.56	82.25%	74.36%	82.56%

Test Data

Accuracy	Sensitivity	Specificity	Precision	Recall
82.90%	84.02	82.18%	74.97%	84.02%

```
y_pred_final.loc[y_pred_final["Lead_Score"]>=80]
```

✓ 0.0s

	Converted	Converted_probability	Lead Number	predicted	Lead_Score
318	1	0.977002	318	1	98
1570	1	0.987289	1570	1	99
8086	1	0.996364	8086	1	100
1840	1	0.876902	1840	1	88
6157	1	0.964775	6157	1	96
...	...	...	...	...	...
548	1	0.950440	548	1	95
7288	1	0.890980	7288	1	89
616	1	0.884772	616	1	88
622	1	0.890980	622	1	89
7036	1	0.976509	7036	1	98

595 rows × 5 columns

Predictions on the basis on probability

- The model has predicted around 595 leads which has 80% chance of getting converted.

- Train Data and Test Data model are showing good stability with more than 80% Accuracy

Insights from the model

- **Sensitivity (Recall): 84.02%:**
  - Higher than the training set (82.56%), indicating the model is doing better at capturing positives in the test data.
  - This also indicates that more than 80% conversion for the leads are achieved
- **Balanced Model:**
  - The model appears well-balanced between sensitivity and specificity, meaning it is performing well in identifying both positive and negative classes.
- **Generalization:**
  - The model generalizes well since the test metrics are quite close to the training metrics, which is a good indicator of the model's robustness and its ability to avoid over fitting.
- **Accuracy and Recall:**
  - With accuracy around 82-83% and recall over 84% on the test data, the model shows strong performance in showing the positive cases, which is crucial in lead conversion predictions.

Conclusion

- 1. Marketing team should focus more on "working professionals" as they are more likely to get converted.
- 2. Marketing team should focus more on leads who spent time on website as they are more likely to get converted.
- 3. Marketing team should focus more on leads coming from Olark Chat as they are more likely to get converted.
- 4. Marketing team should skip calls to the leads who chose the option "Do not Email" as "yes" as they are not likely to get converted.
- 5. Marketing team should skip calls to the leads who has the Lead Origin as "Landing Page Submission" as they are not likely to get converted.