Machine Learning for Lead Management

A case of an online education company

Yusi Chen
July 10th, 2020

Project at a Glance

Problem Statement

An education company on sells online courses to industry professionals. The company markets its courses on several websites and search engines like Google. Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not. To make this process more efficient, the company wishes to identify the most potential leads by prioritized leads score.

Project Purpose

Using machine learning algorithms to predict the leads conversion including Yes/No outcome and conversion probability. Based on leads scoring pipeline, recommending the most effective marketing & sales strategies to boost leads conversion.

Project challenge

-Data cleaning: Handling a messy and unstructured dataset.

-Data preparation: Converting categorical columns into numerical ones. Analyzing large number of variables.

-Modeling: Improving the model by cross checking multiple techniques.

Project Introduction

How to boost the conversion rate with Predictive Lead Scoring
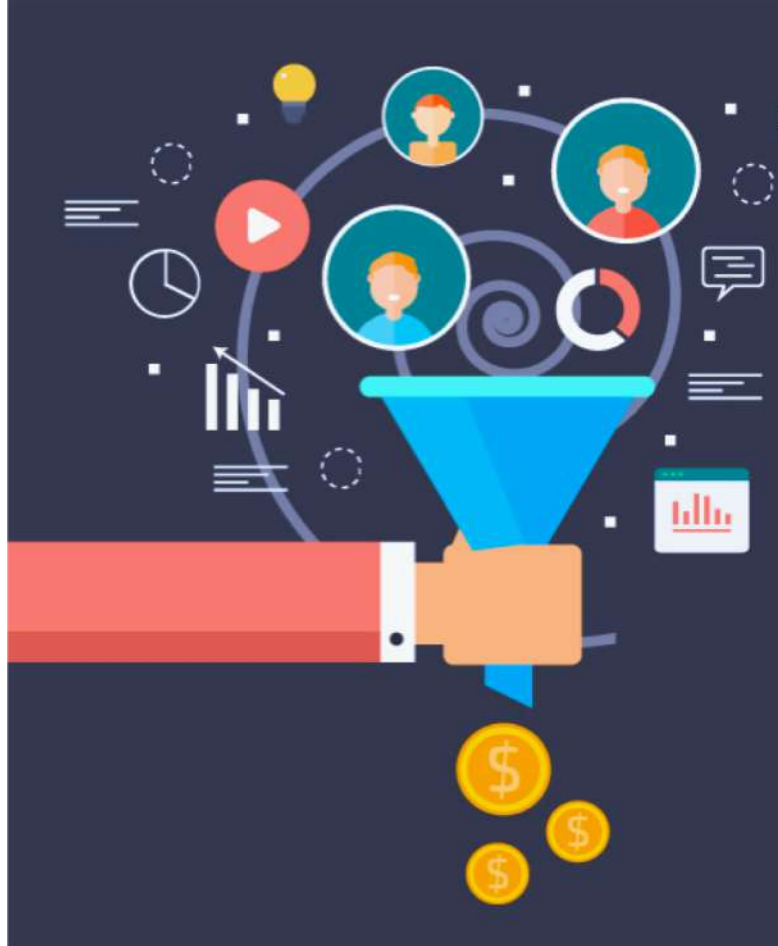
Industry:

Online professional education

Dataset:

10,000 entries, 37 columns

Goal:

Leads conversion 30%→80%

Method:

Using machine learning algorithms to predict conversion outcome
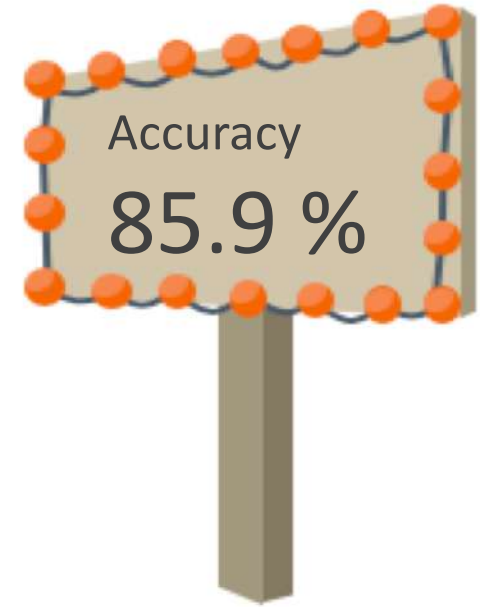
Algorithms ?

Machine learning algorithms

- Logistic Regression (final selection)
- K-Nearest Neighbours
- Decision Tree
- Support Vector Machine (SVM)

Accuracy

85.9 %

Model improvement techniques

- Feature scaling

- Highly correlated feature filtering

- Train-test set overfitting checking

- Model evaluation with Confusion Matrix and ROC AUC

- Feature ranking with Recursive Feature Elimination(RFE)

- Principle Components Analysis (PCA)

Workflow

Data acquiring → data cleaning → data preparation →
data visualization (mainly in Tableau) → data reporting
Dataframe shape (9240 x 37) → (9074 x 79)

Algorithms

Build models to predict conversion outcome of future leads

Index	Lead Number	Conversion Real Value	Conversion Predict
9	660616	No	No
14	660553	No	No
15	660547	Yes	Yes
18	660522	Yes	Yes
20	660479	No	No
22	660471	Yes	Yes
23	660461	No	No
33	660339	No	No
38	660255	No	No
39	660255	Yes	Yes

Yes / No

Algorithms

Build models to predict conversion probability of future leads

Index	Lead Number	Conversion Real Value	Conversion Predict	Leading Scoring
9	660616	No	No	0.055877
14	660553	No	No	0.067691
15	660547	Yes	Yes	0.710651
18	660522	Yes	Yes	0.725876
20	660479	No	No	0.043348
22	660471	Yes	Yes	0.771458
23	660461	No	No	0.157629
33	660339	No	No	0.385444
38	660255	No	No	0.077897
39	660255	Yes	Yes	0.792177

How likely

Algorithms

Using predictive score to build leads pipeline

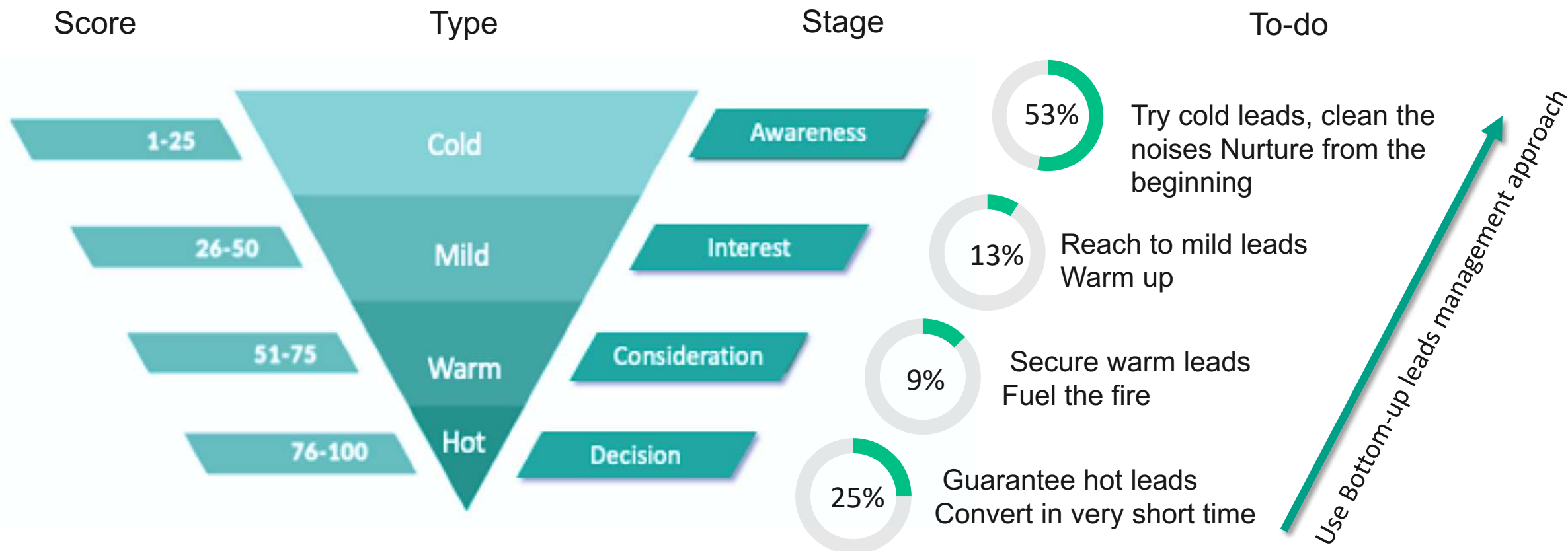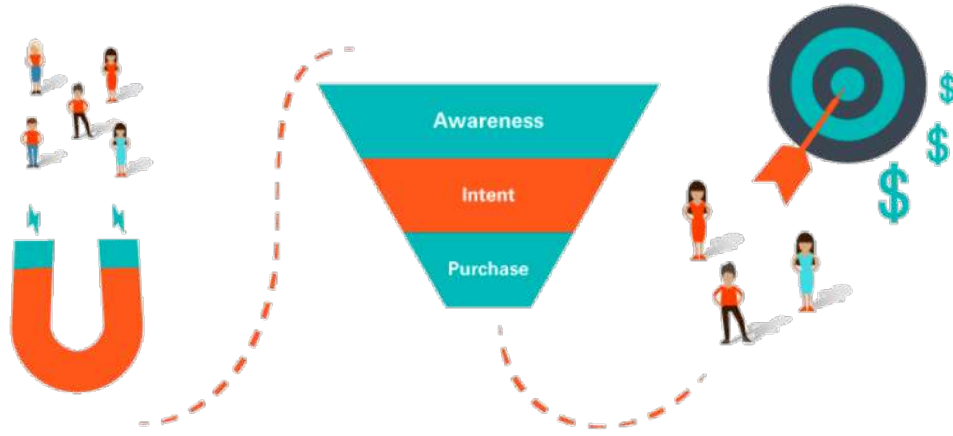| Index	Lead Number	Conversion Real Value	Conversion Predict	Leading Scoring	Lead Type	
9	660616	No	No	0.055877	Cold Lead	
14	660553	No	No	0.067691	Cold Lead	
15	660547	Yes	Yes	0.710651	Warm Lead	
18	660522	Yes	Yes	0.725876	Warm Lead	
20	660479	No	No	0.043348	Cold Lead	
22	660471	Yes	Yes	0.771458	Hot Lead	
23	660461	No	No	0.157629	Cold Lead	
33	660339	No	No	0.385444	Mild Lead	
38	660255	No	No	0.077897	Cold Lead	
39	660255	Yes	Yes	0.792177	Hot Lead	

Pipeline

Leads Management Strategy

Differentiating marketing campaign to trigger leads upgrading

Score	Type	Stage	To-do

Score
- 1-25
- 26-50
- 51-75
- 76-100

Type
- Cold
- Mild
- Warm
- Hot

Stage
- Awareness
- Interest
- Consideration
- Decision

To-do

53% Try cold leads, clean the noises Nurture from the beginning

13% Reach to mild leads Warm up

9% Secure warm leads Fuel the fire

25% Guarantee hot leads Convert in very short time

Use Bottom-up leads management approach

Marketing & Sales

Did you do the right job ?

Customer segments

Marketing channels

Sales efforts

Data quality

Predictive Model Feature Importance Ranking

Top 20 indicators overview

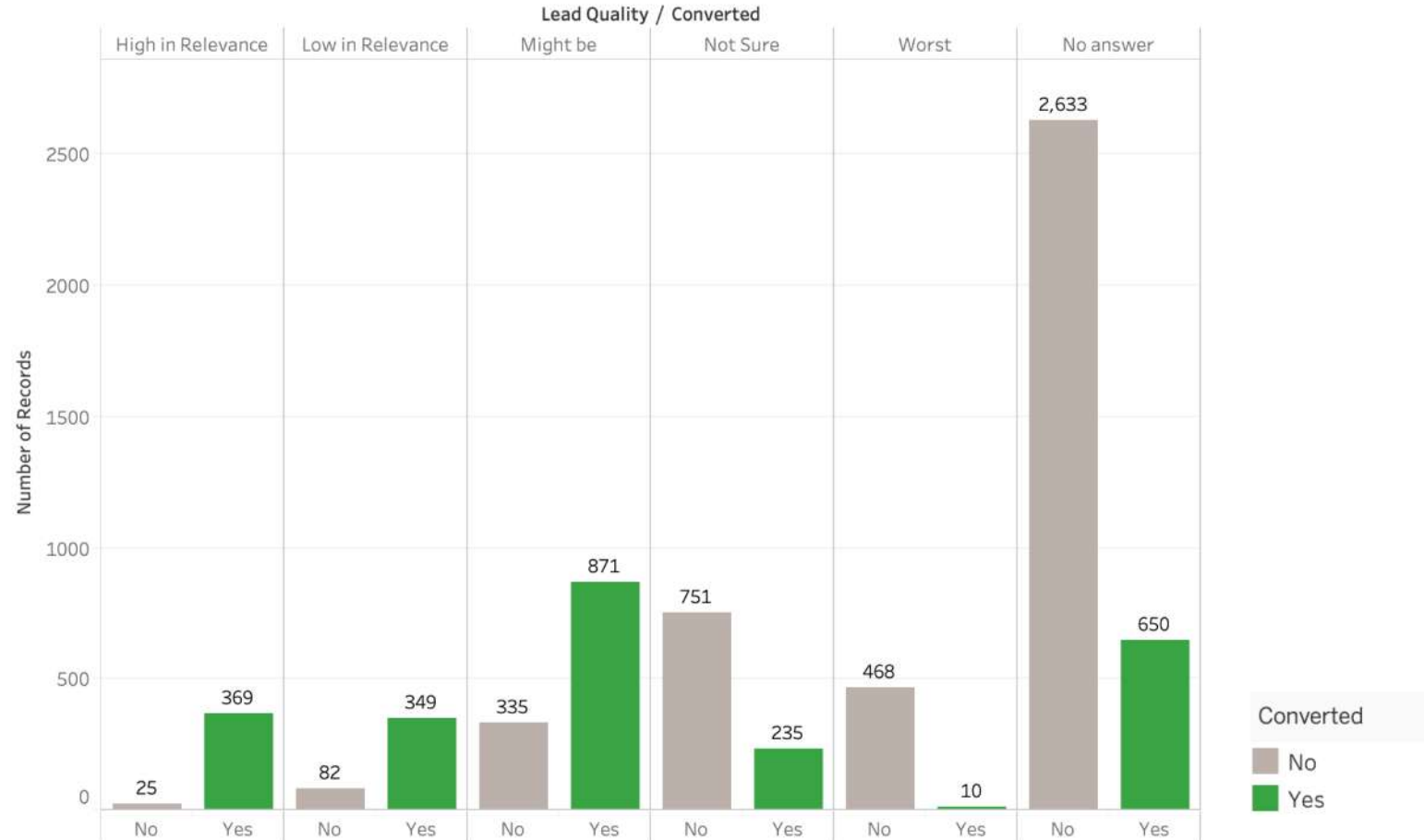|				
• Worst • No answer • Not Sure • Might be • Low in Relevance		• Welingkak Website • Direct Traffic • Google • Organic Search • Reference • Referral Sites	• SMS Sent • Olark Chat Conversation • Had a Phone Conversation • Converted to Lead	• Working Professional • Unemployed
Lead quality	**Total Times Spent on Website**	**Lead source**	**Last Activity**	**Current Occupation**

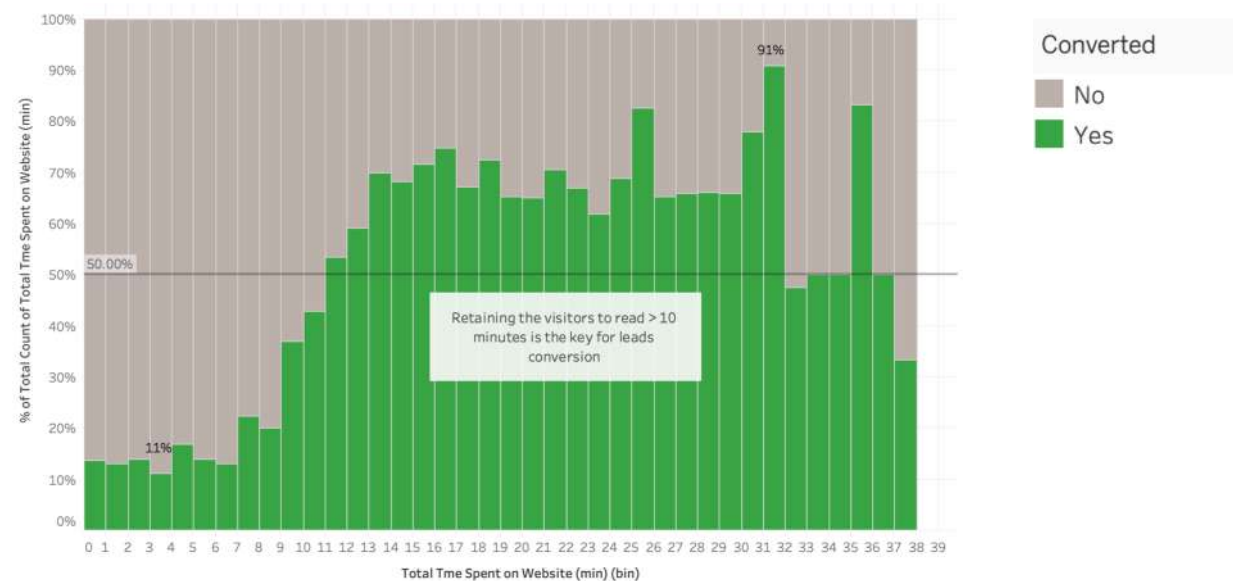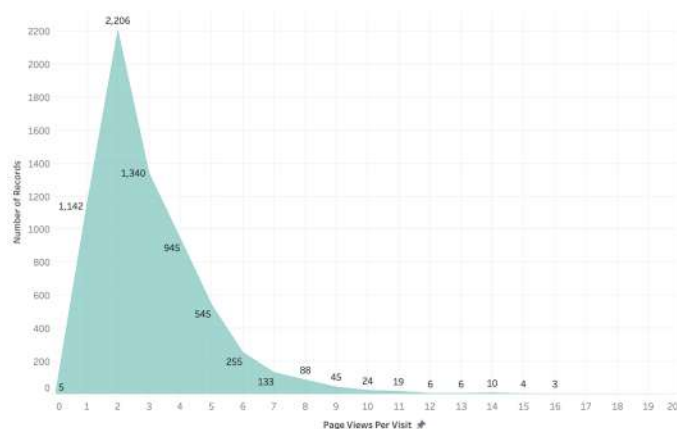| Importance	Importance	importance	Importance	Importance
90%	80%	70%	60%	50 %

Top Features

Lead Quality

Lead Quality / Converted

- The intuition of sales staff on the quality of leads.

- Conversion rate is highly correlated with the leads quality.

- Train sales staff to fill in the information in a more responsible manner.
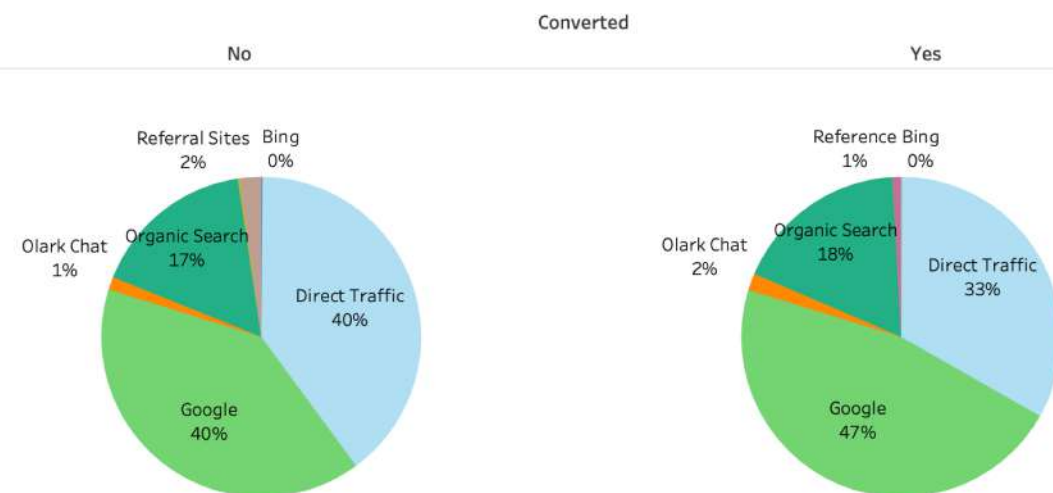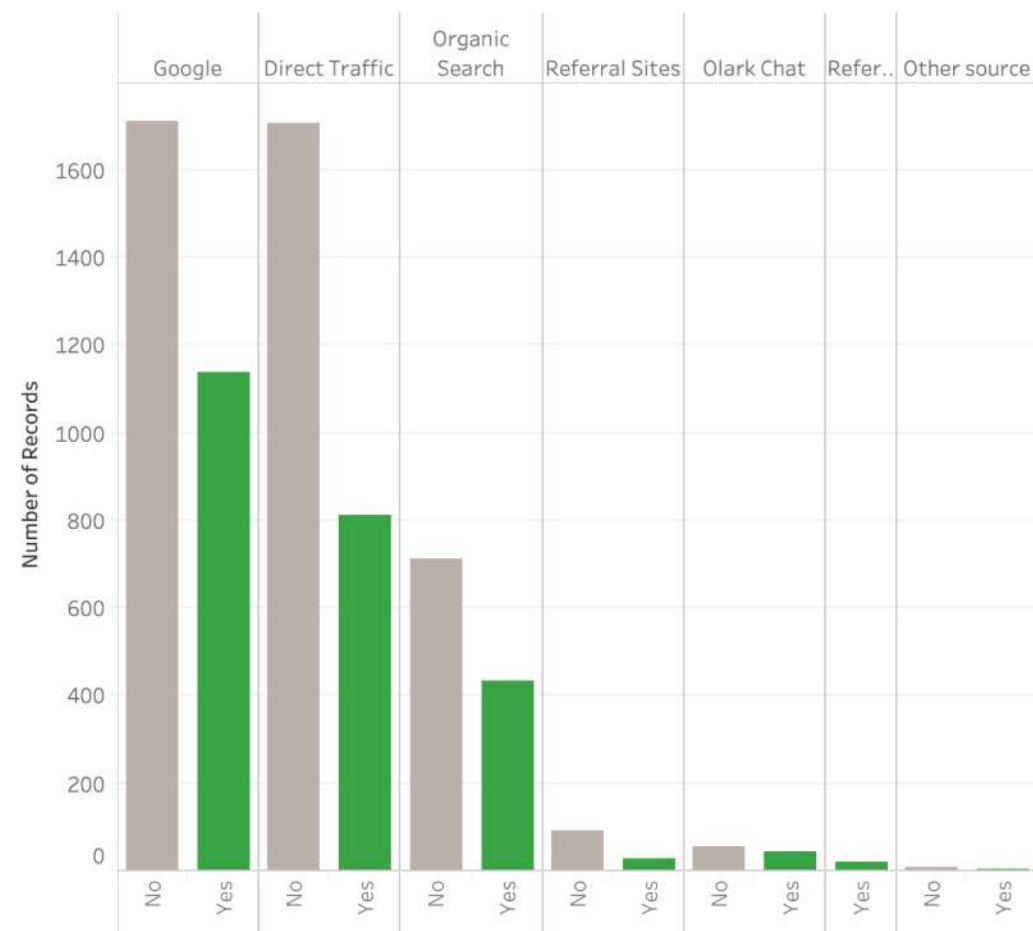
Top Features

Total Time Spent on Website

- Retaining web visitors for at least 10 minutes in 2 pages is the key to convert leads.

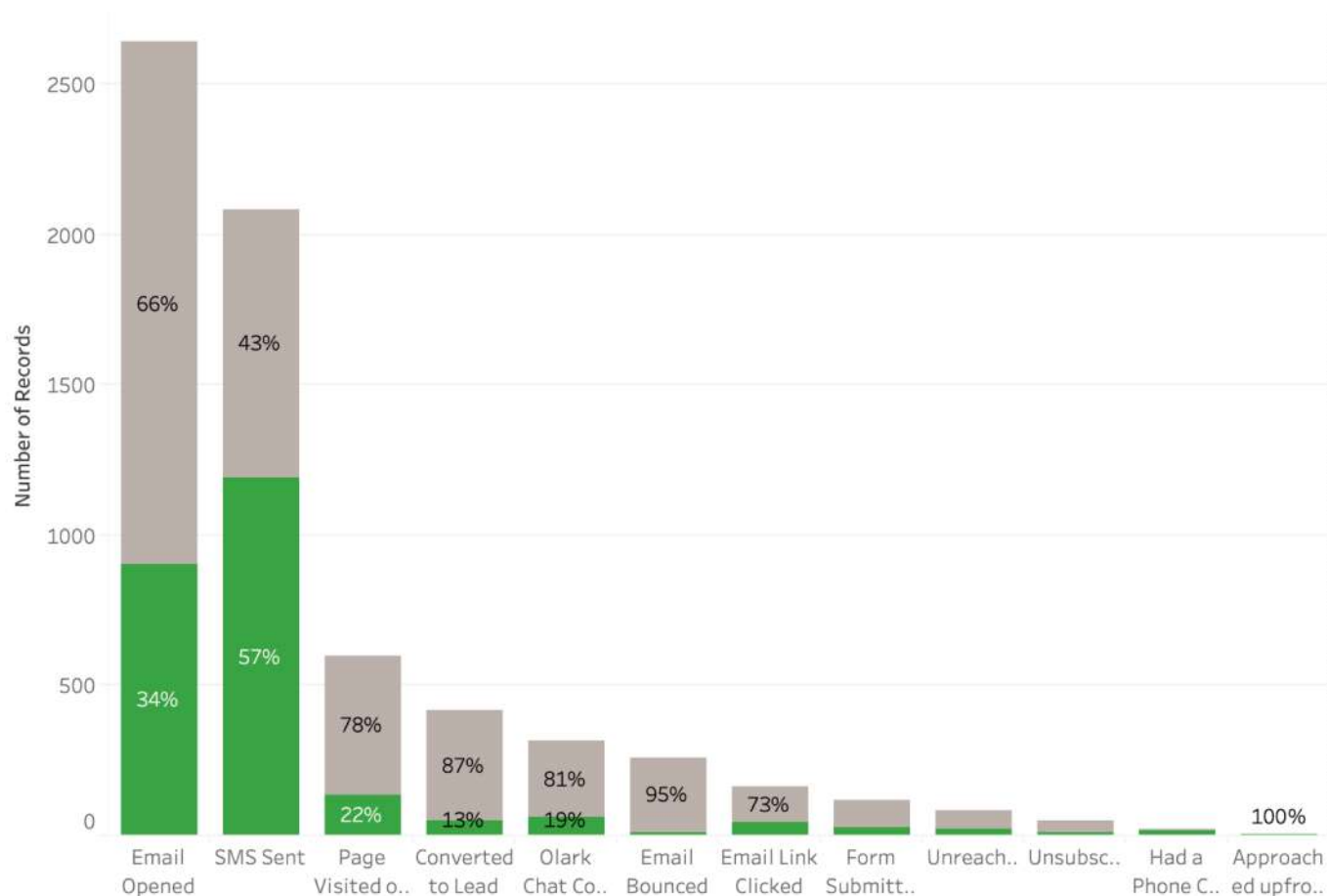- Working harder on Content Marketing. Building a compelling website.

Top Features

Lead Source

Converted

No / **Yes**

- Searching Engine is the major source for generating leads and the conversion rate is ¼ .

- Consider to invest more SEO and SEA (if budget allowed)
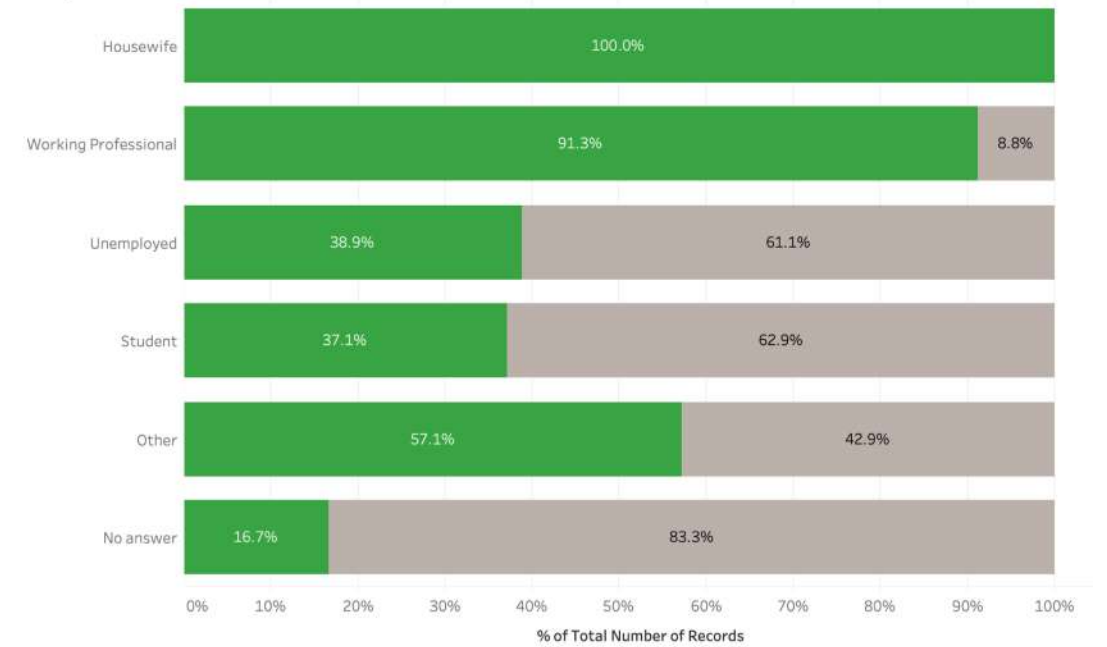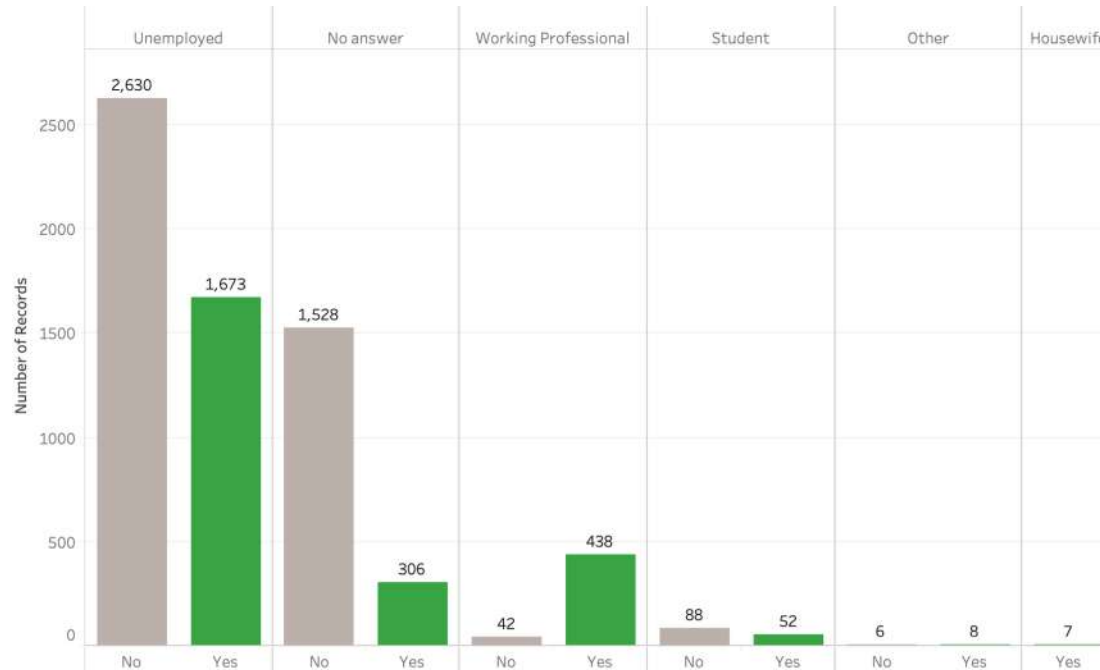
Top Features

Last Activity

- SMS and email marketing are the most effective tactics.

- Improving email opening rate
    - Perfect your sending Timing
    - Make your subject line stand out
    - Personalize the letter
    - Write amazing content, every time
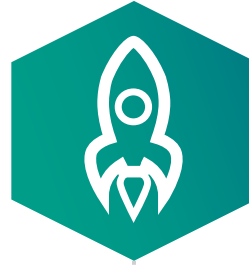
Top Features

Current Occupation

- Unemployed group of people contributed the highest number of leads
- Working professionals are more likely to buy the course, and the rate is extremely high, at 91%.
- Focus on generating more leads from working groups.

Conclusion

Strategy that you need

Lead Generating

Organic SEO
Targeting Working
professional

Lead Scoring

Differentiated
campaign
depends on the
stages in the
funnel

Lead Nurturing

Email Marketing
SMS Marketing

Content is the King

Why I am the best
Web Design
Landing page
Testimony
Blog
Video

Data Strategy

A company-wide
solution;
Forward-thinking
investment

Empowering businesses through data analytics

Keep in Contact

www.linkedin.com/in/yusichen

yusi0928@gmail.com