



Lead Scoring



Using Machine Learning Predictive Models to
Assess Likelihood of Conversion



X Education



This problem and data set is created by UpGrad-IIIT-B for educational purposes

30%



Lead Conversion Rate

Leads Converted
Total # of Leads

Higher Conversion Rate
≠ More Sales

**Sales
Strategy:**

Identify different kinds of leads and adapt to their needs
(aka: convert more leads!)

**Marketing
Strategy:**

Determine who is more likely to convert and how we can reach them
(aka: recruit better leads!)

Lead Data

- **9240 leads**
- **38.5% conversion**
- **20 features (139 after encoding nominal features)**

Visits - # unique website visits

Visit time - Total time spent on website

Visit pages - Avg. number pages per visit

No email - lead requested no emails

Search - lead saw an ad from search

Recommendations - lead was recommended

MTI copy - lead wants free copy of “Mastering the Interview”

Modified - Last notable activity was “Modified”

Activity Index - Low/Med/High score based on Activity

Profile Index - Low/Med/High score based on Profile

Origin (5) - How lead was identified

Source (20) - Source of the lead

Last activity (17) - Most recent interaction from customer

Country (38)

Specialization (18) - Industry domain lead has worked in

Hear about (9) - How did they hear about X Education?

Occupation (6) - Currently working/studying/unemployed

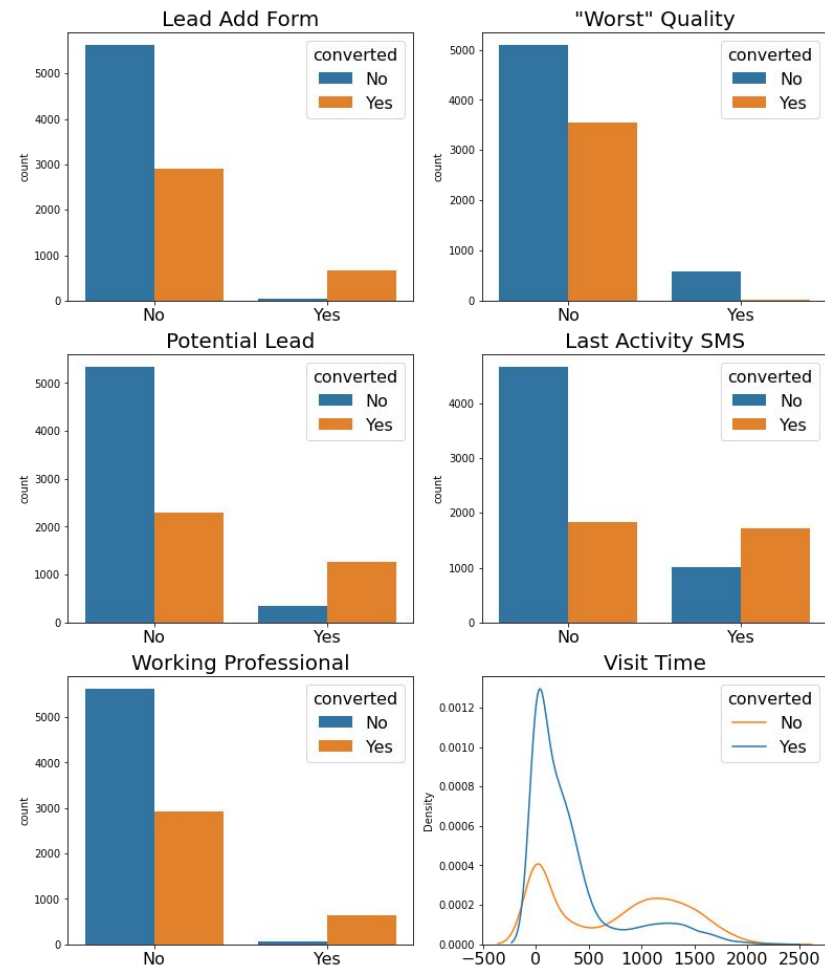
Quality (5) - Lead quality based on intuition of employee assigned

Profile (5) - lead level assigned to customer based on profile

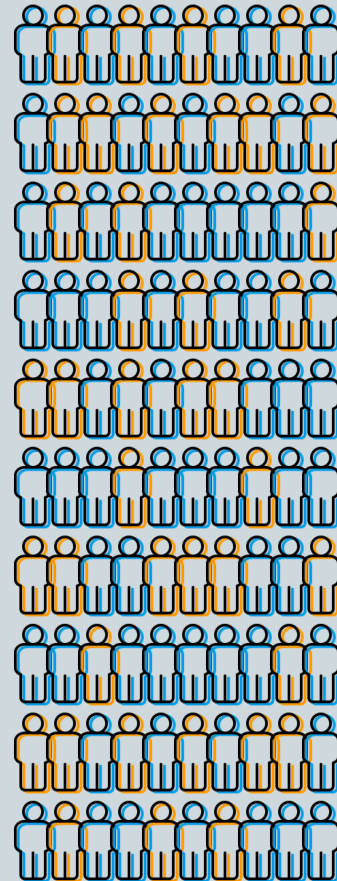
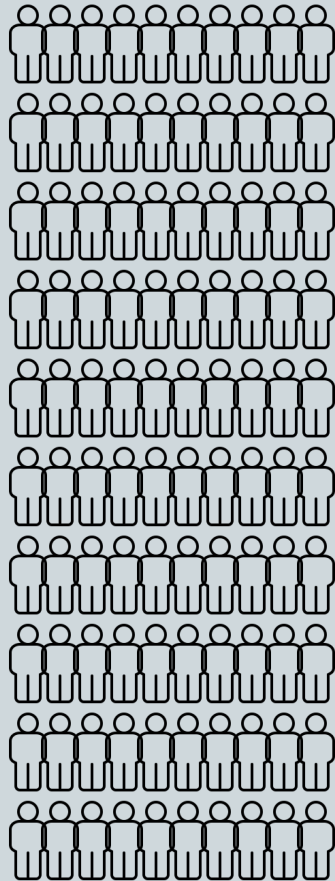
City (6) - type of city where lead currently lives

Using the Data

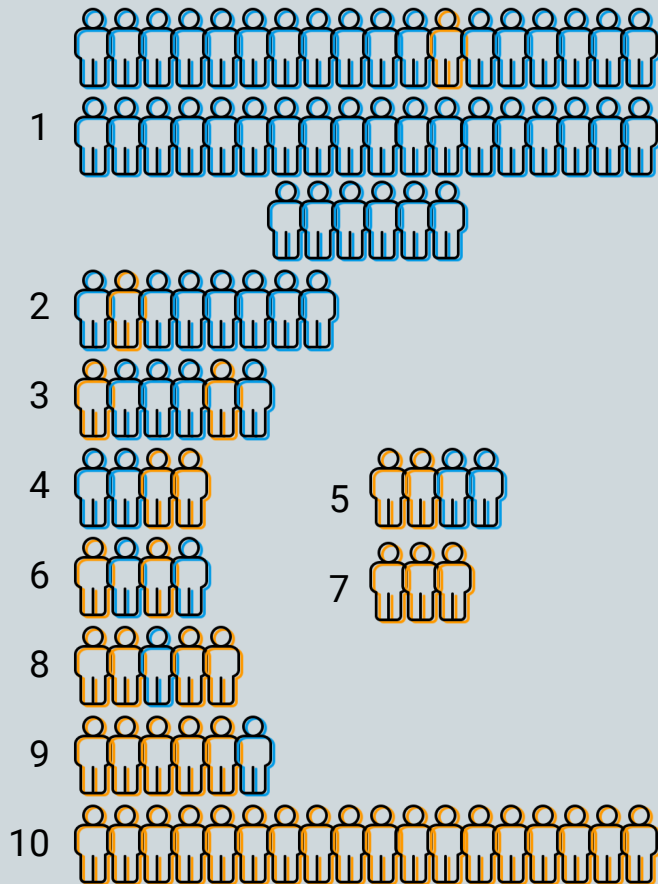
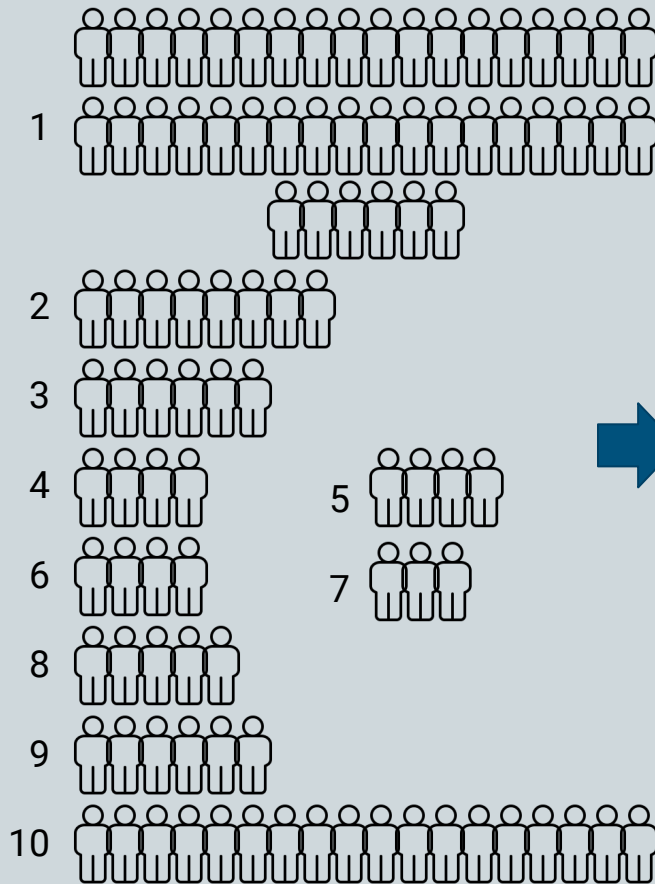
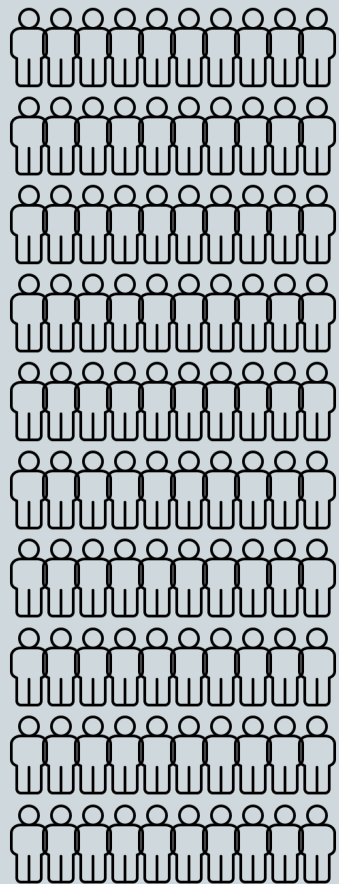
- XGBoost decision tree model outperformed LogReg and Random Forest



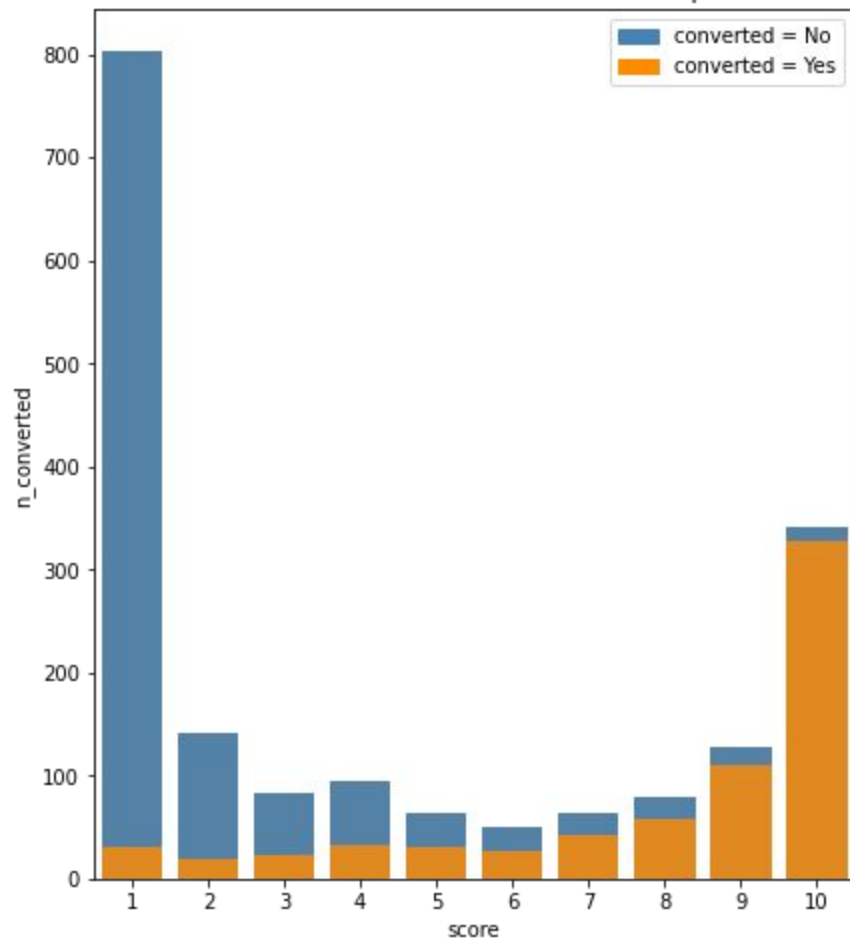
No Lead Scoring



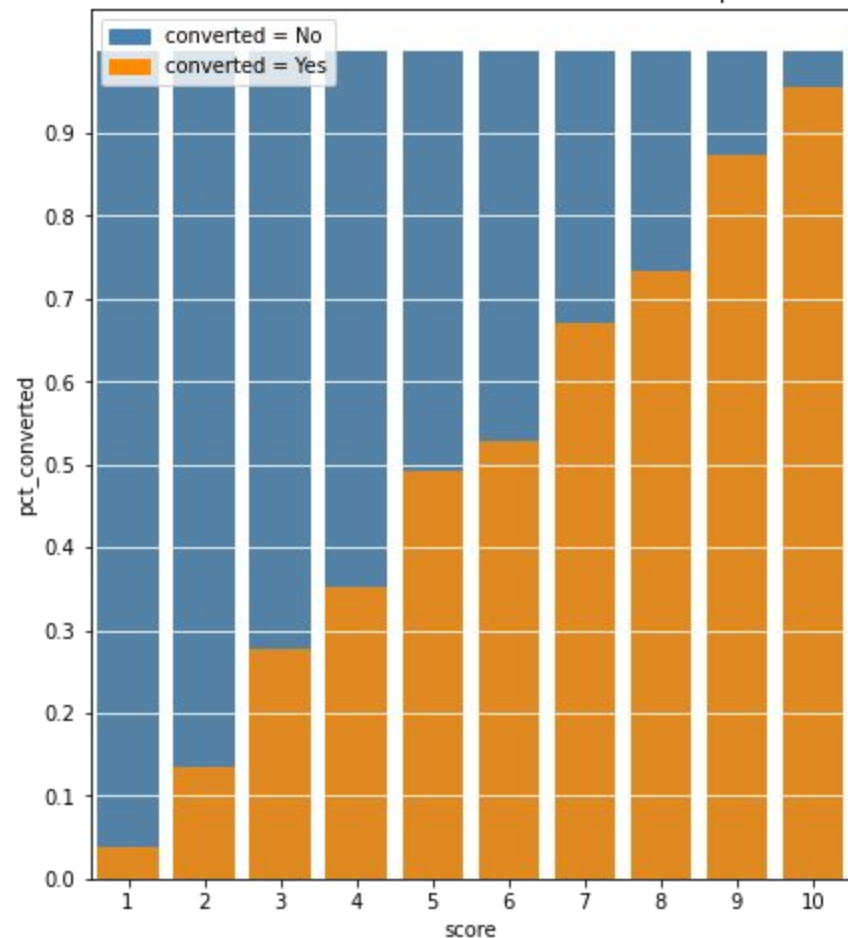
Lead Scoring



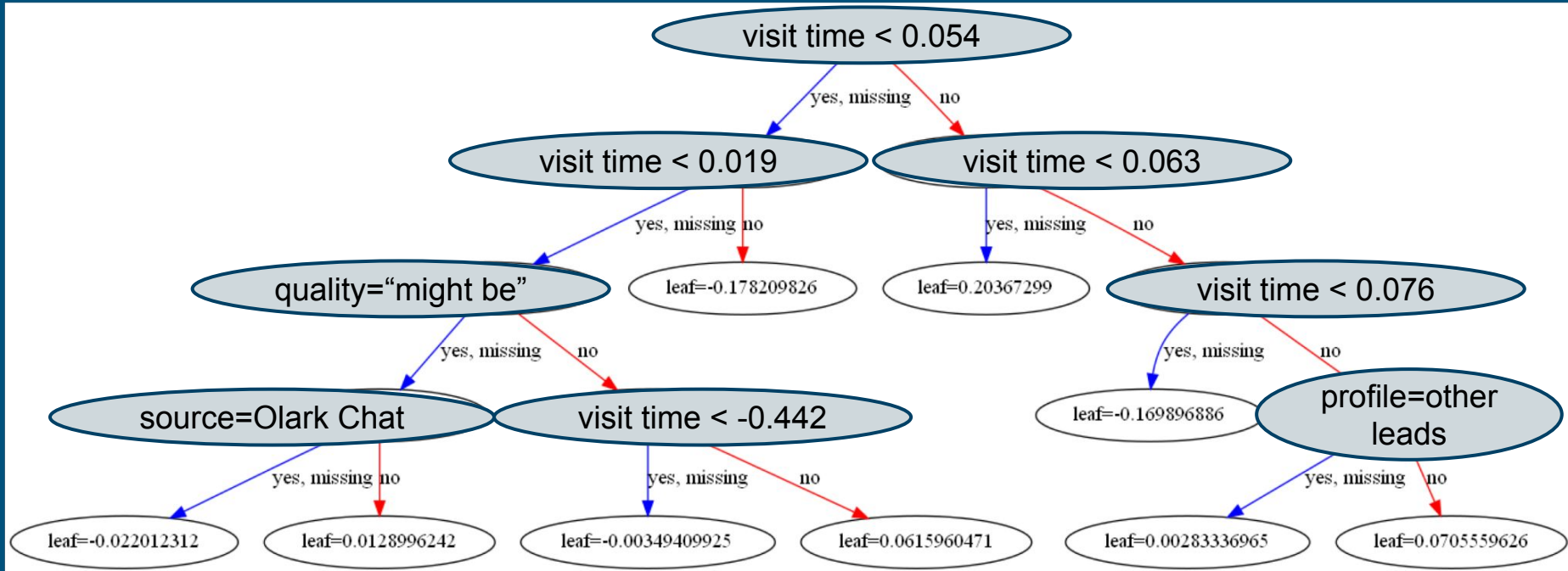
Conversions in Each Lead Score Group



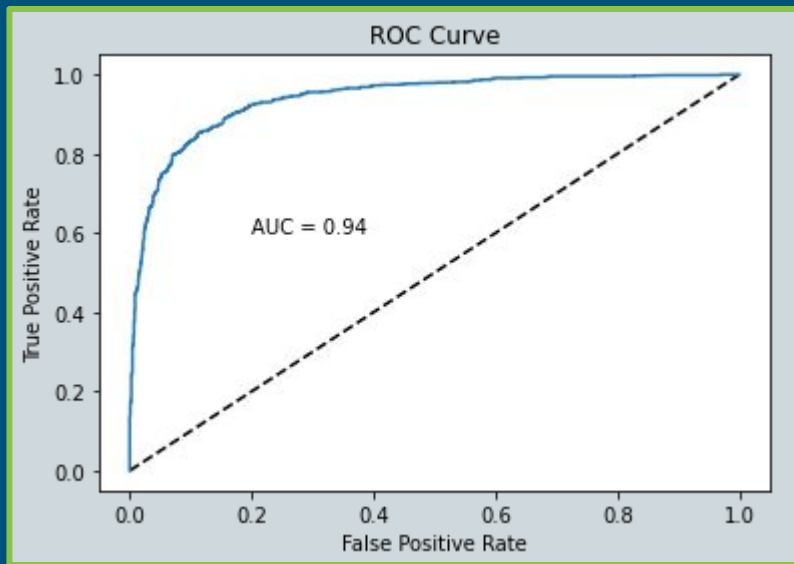
Conversion Rates in Each Lead Score Group



Modeling



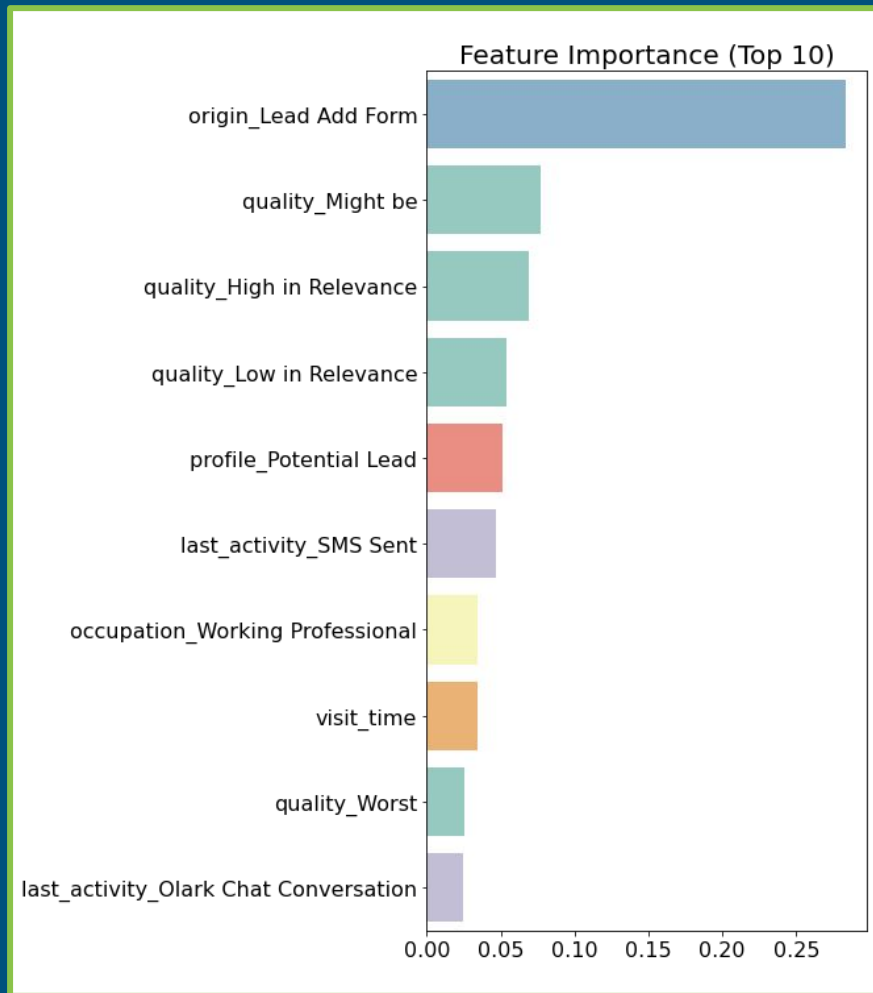
Model Performance



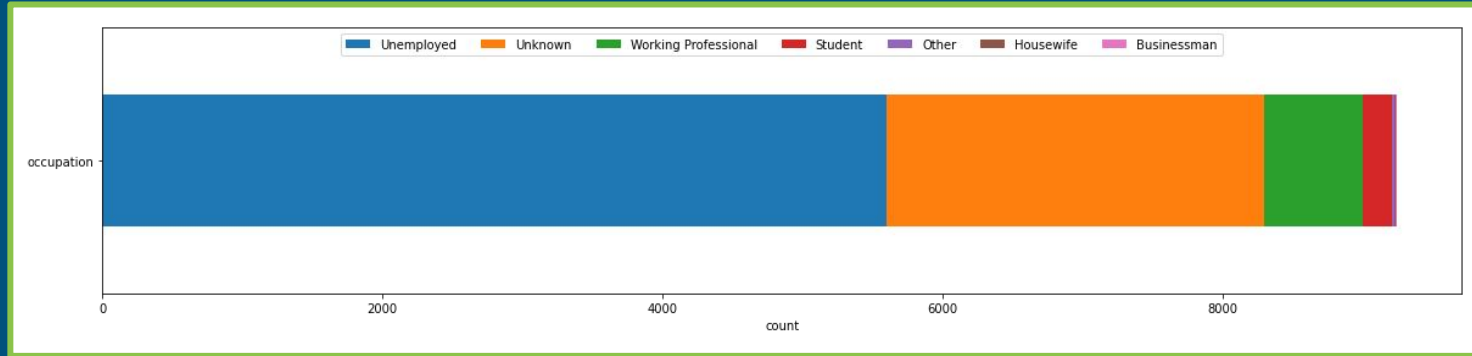
	Converting	Not Converting
High Score	True Positives: Good Leads	False Positives: Wasted Time
Low Score	False Negatives: Missed Opportunities	True Negatives: Bad Leads

Important Features

- Do we have any underlying knowledge of these features?
- Can these features be leveraged?



Example: Working Professionals



- 7.6% of this dataset (706 leads)
- 91.6% conversion (647 leads)

▶▶▶**TARGET THIS MARKET!**◀◀◀

Next Steps



1. Target important features
2. Further data collection
3. Develop new sales strategies
4. Review model periodically