Lead Score

X Education

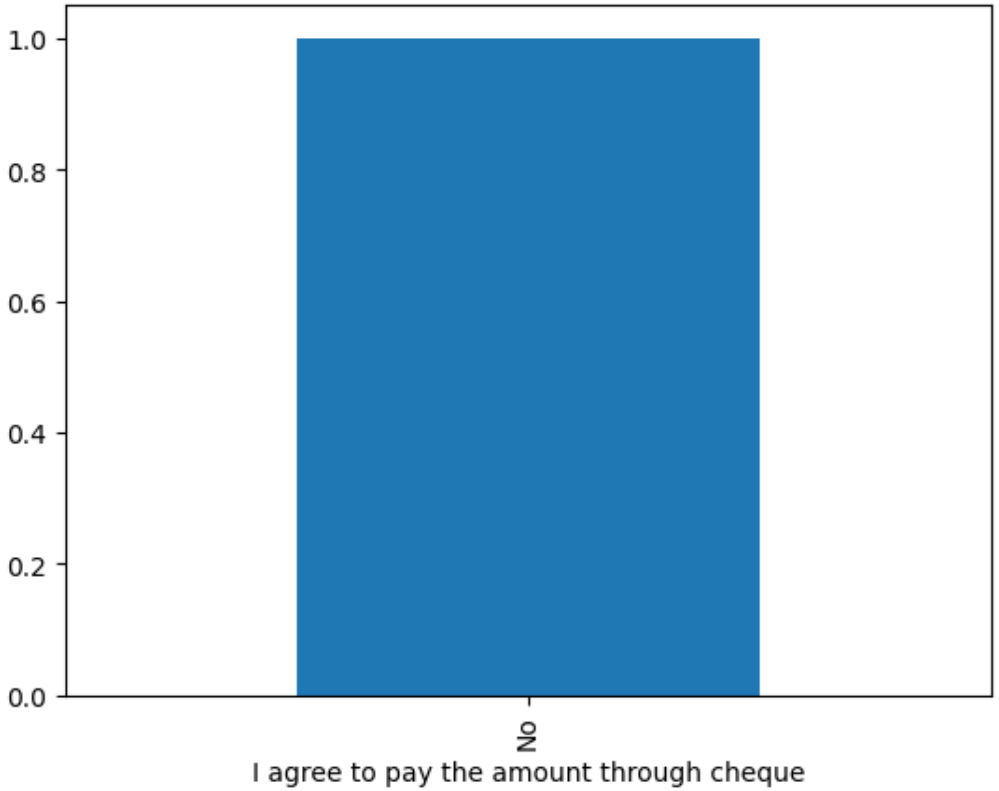
Problem Statement

- X Education has a low conversion rate(30%)
- X Education wants to increase the conversion rate to approx. 80%
- Data is collected and we have about 9000 data to deduce a pattern in conversion rates.
- Two strategies are required i.e. an aggressive sales strategy and a conservative strategy.

Analysis approach

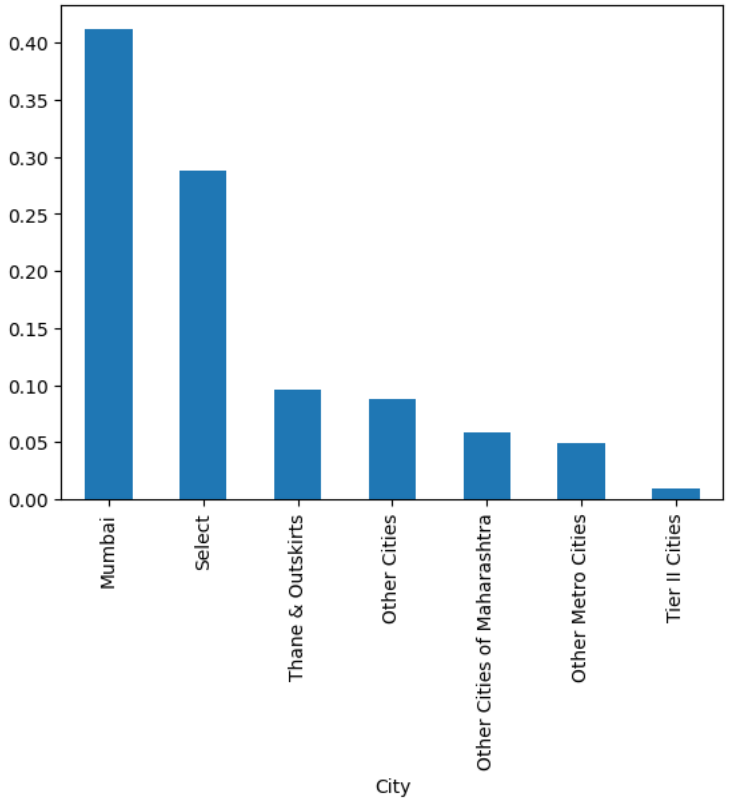
- Clear out any duplicates
- We'll drop any uniquely identifying columns, they add no value and increase the computation requirement, so we'll drop them as early as possible.
- When doing single variant analysis, we get 4 columns with "Select" as an option, these are the columns:
  - Specialization
  - How did you hear about X Education
  - Lead Profile
  - City

- “Select” is equivalent to having NaNs, i.e. we have no information on that.
- In case of Specialization and City, “Select”+NaNs are in an acceptable range, so we keep them and then drop them when making dummy columns.
- In case of Specialization, we could even say that a person that’s serious about buying the product would provide basic details like this one, so “Select” holds value.
- In case of How did you hear about X Education and Lead Profile, these columns could hold a lot of business value, example, a friend’s referral holds a strong value over any other, similarly in case of Lead Profile. But both these have a lot of “Select”+NaNs which will not provide any value.
- Fixing of data was required, which could lead to redundant features in when creating dummy variable, like Google=google.
- Fixing outlier in continues data, as after a point these values are the same. Example 15 units of viewing the website and 10 units of viewing the website are the same to and are enough to tell that the person is interested.

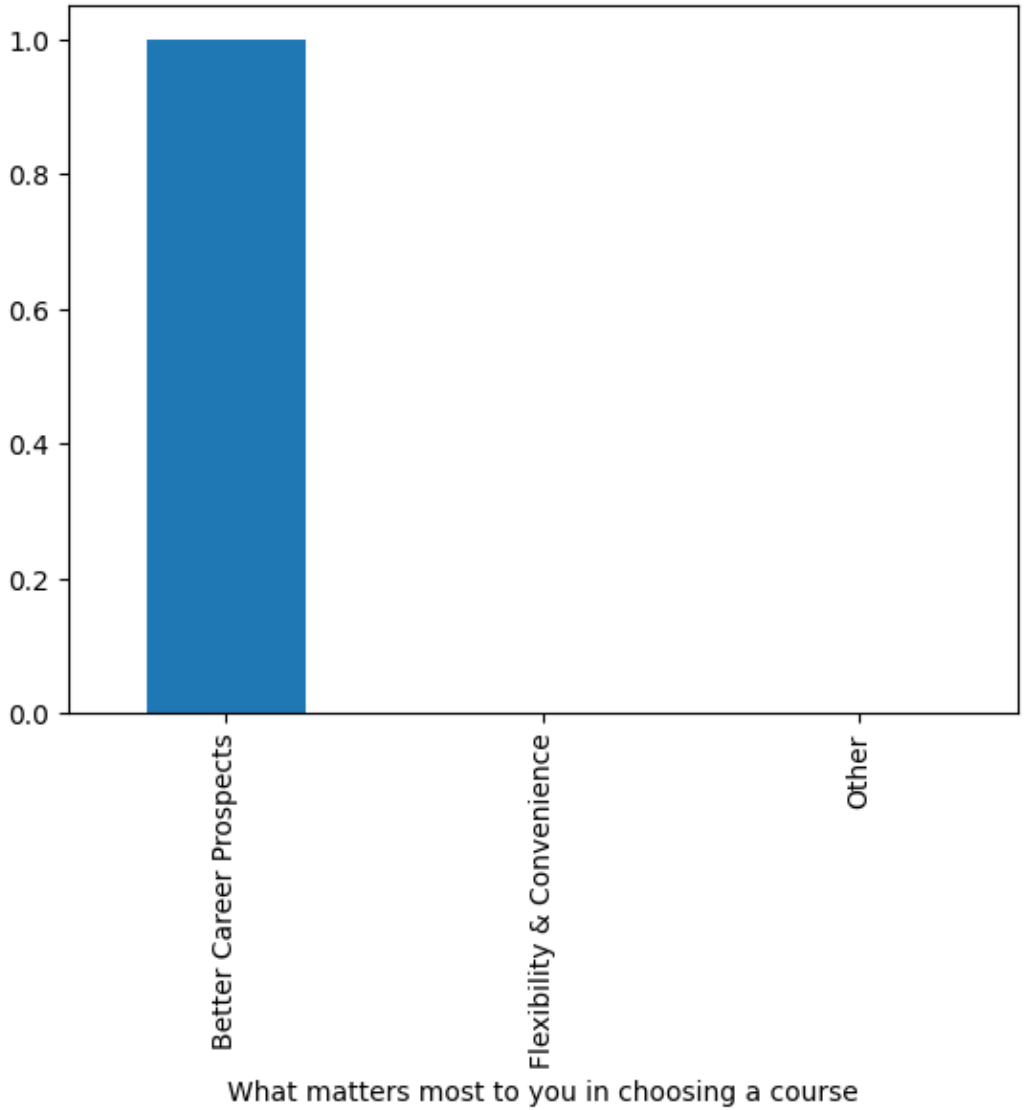


Columns like these don’t add value to our analysis, so we drop these, they have an extreme bias, it just means No one pays with cheques and so there is not trend to follow.

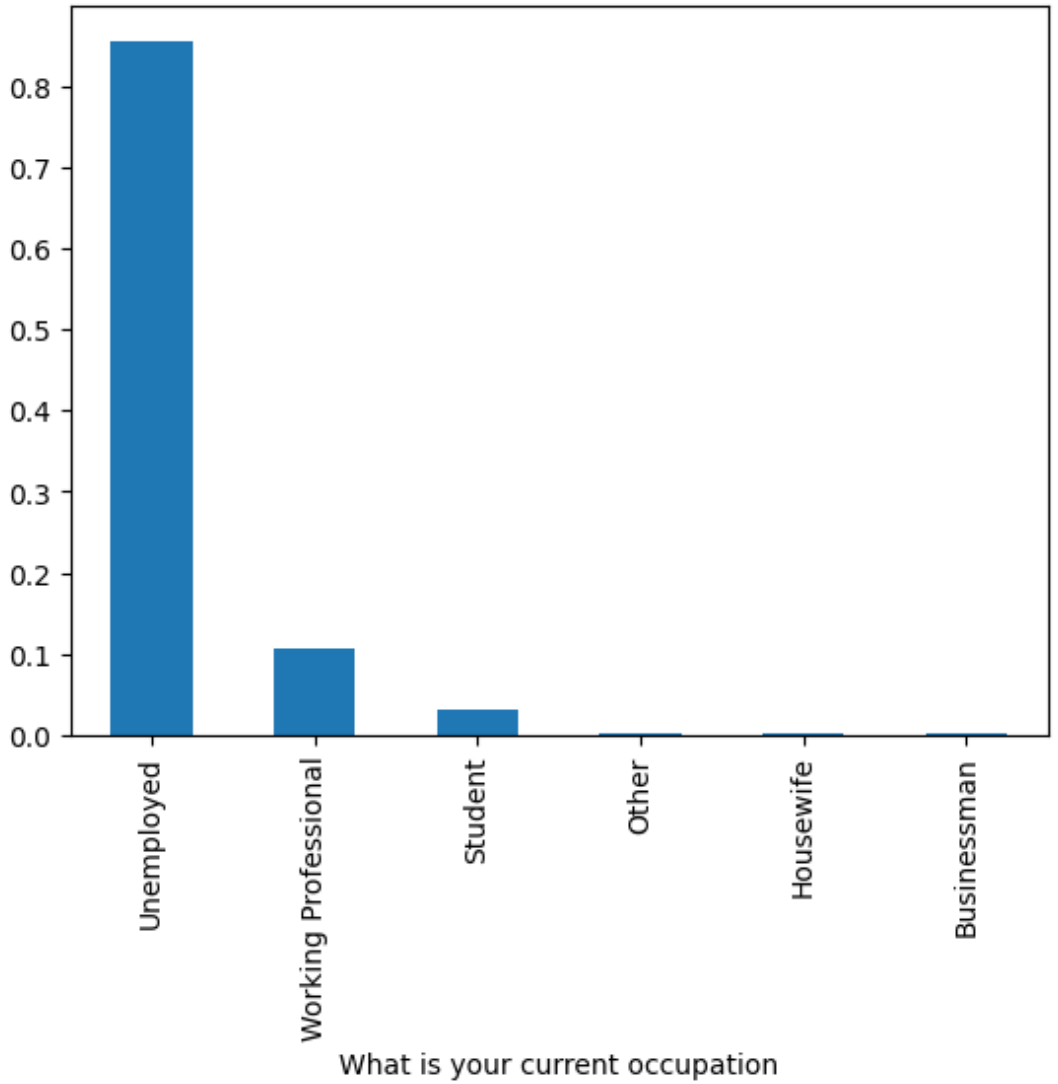
Observations



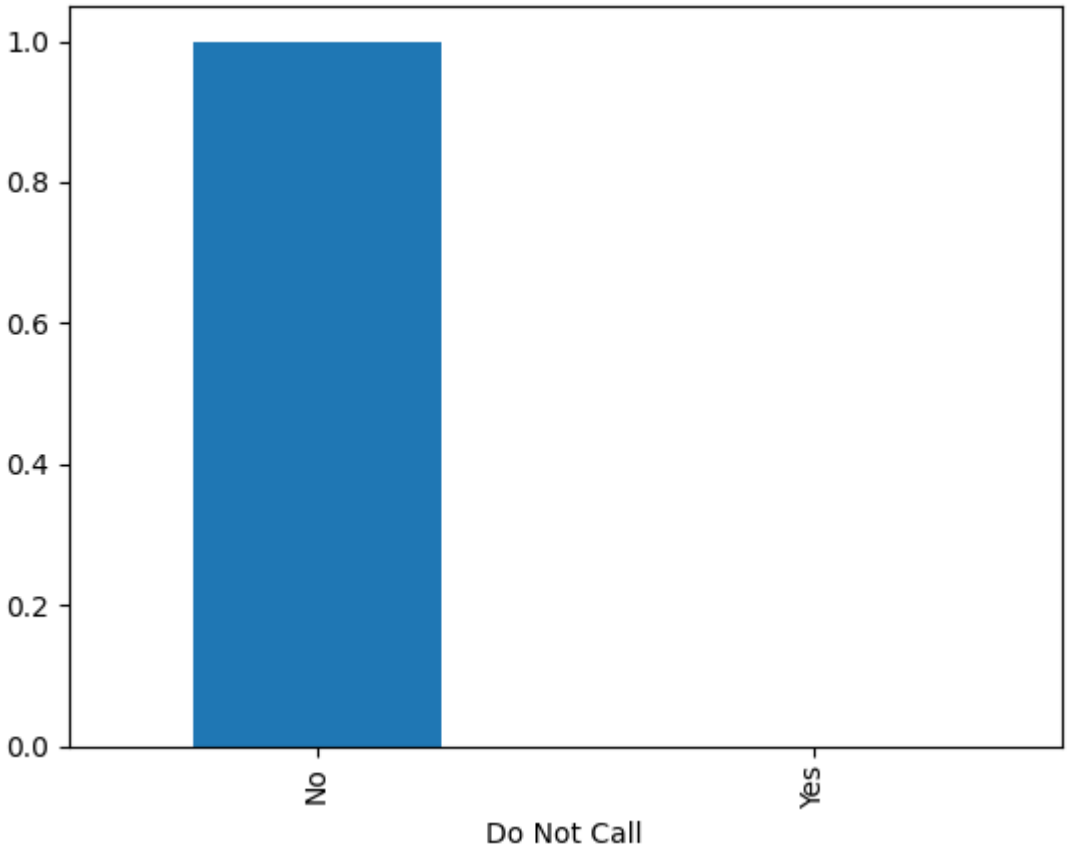
- Most of the potential customers are from a tier-1 city, like Mumbai.
- Select has a high percentage of the data too.



Almost all the people who are our potential customers are driven by a need to get better career opportunites.



- Most of the people that X Education gets in contact are Unemployed.
- Later on we see that even though we have high percentage of unemployed people that we get in contact with, they aren't the ones that are driving our sales.
- This shows where the company might be losing out on time and money.



- This is a biased column and is rejected.
- But this goes on to show how most people feel about getting called.
- This becomes more of a reason to avoid calling a person who has low chances of buying our product.

Results

- Our results are heavily driven by people who have visit our websites. It's a good idea to invest on the person visits our websites for long hours. Adding things like course schedule, topics covered can lead to a higher conversion rate.
- A person who fills out the forms provided to them have a higher chance of converting into a sale, filling a form takes time and effort, who's ready to put in that shows that they are interested in the product.
- Olark Chats decline the sales. People who have last used Olark chats are likely to not by the product, it is better to avoid them to save on costs.
- Working class people are the people who are good candidates to sell the product to.

Summary

- Avoiding calls saves the company it's resources and increases customer satisfaction, as we had seen that most people don't like being called.
- We should focus more on the people who already have a requirement of the product, these are people of working class, who have researched the course on their own and are ready to put the effort to fill the lead forms.
- To be in an aggressive sales strategy 30-100 of lead score produced from the model.
- For a more conservative approach 50-100 is a good lead score to look out for.

“

Thank you

”