Project Problem Statement - Potential Customers Prediction

**Context**

The EdTech industry has been surging in the past decade immensely, and according to a forecast, the Online Education market would be worth $286.62bn by 2023, with a compound annual growth rate (CAGR) of 10.26% from 2018 to 2023. The modern era of online education has enforced a lot in its growth and expansion beyond any limit. Due to having many dominant features like ease of information sharing, personalized learning experience, transparency of assessment, etc., it is now preferable to traditional education.

The online education sector has witnessed rapid growth and is attracting a lot of new customers. Due to this rapid growth, many new companies have emerged in this industry. With the availability and ease of use of digital marketing resources, companies can reach out to a wider audience with their offerings. The customers who show interest in these offerings are termed as leads. There are various sources of obtaining leads for Edtech companies, like:

* The customer interacts with the marketing front on social media or other online platforms.
* The customer browses the website/app and downloads the brochure.
* The customer connects through emails for more information.

The company then nurtures these leads and tries to convert them to paid customers. For this, the representative from the organization connects with the lead on call or through email to share further details.

**Objective**

ExtraaLearn is an initial stage startup that offers programs on cutting-edge technologies to students and professionals to help them upskill/reskill. With a large number of leads being generated on a regular basis, one of the issues faced by ExtraaLearn is to identify which of the leads are more likely to convert so that they can allocate the resources accordingly. You, as a data scientist at ExtraaLearn, have been provided the leads data to:

* Analyze and build an ML model to help identify which leads are more likely to convert to paid customers.
* Find the factors driving the lead conversion process.
* Create a profile of the leads who are likely to convert.

EDA (solutioning):

Leads will have different expectations from the outcome of the course and the current occupation may play a key role in getting them to participate in the program. Find out how current occupation affects lead status.

Given Input : current\_occupation

Soln:

Using pandas : Based on bi variate analysis

pd.crosstab(

Other analysis

1.

Find how age variable with the input has the conversion rate.

Select age, current\_occupation, status, count(\*) group by <>.

2.

time\_spent\_on\_website

Select current\_occupation, time\_spent\_on\_website

, status, count(\*) group by <>.

3.

page\_views\_per\_visit

Select current\_occupation, page\_views\_per\_visit

, status, count(\*) group by <>.

4. profile completed

<to do>

* age
* current\_occupation
* first\_interaction
* profile\_completed
* website\_visits
* time\_spent\_on\_website
* page\_views\_per\_visit
* last\_activity
* status

The company's first impression on the customer must have an impact. Do the first channels of interaction have an impact on the lead status?

Input: first\_interaction: Values include ‘Website’ and ‘Mobile App’

pd.groupby([‘first\_interaction’])

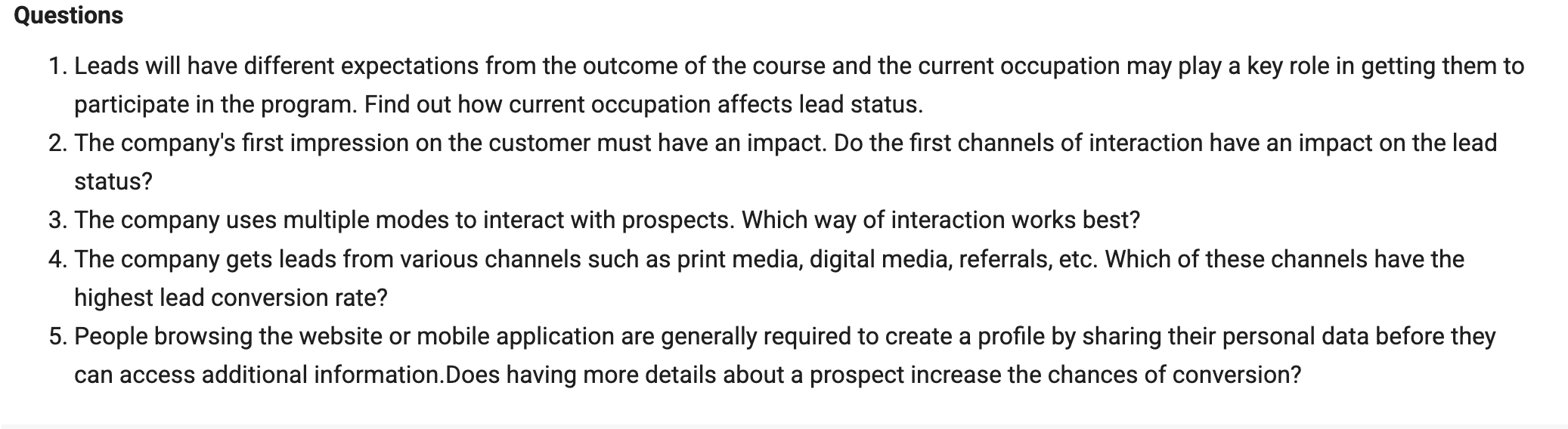
current\_occupation aggregated

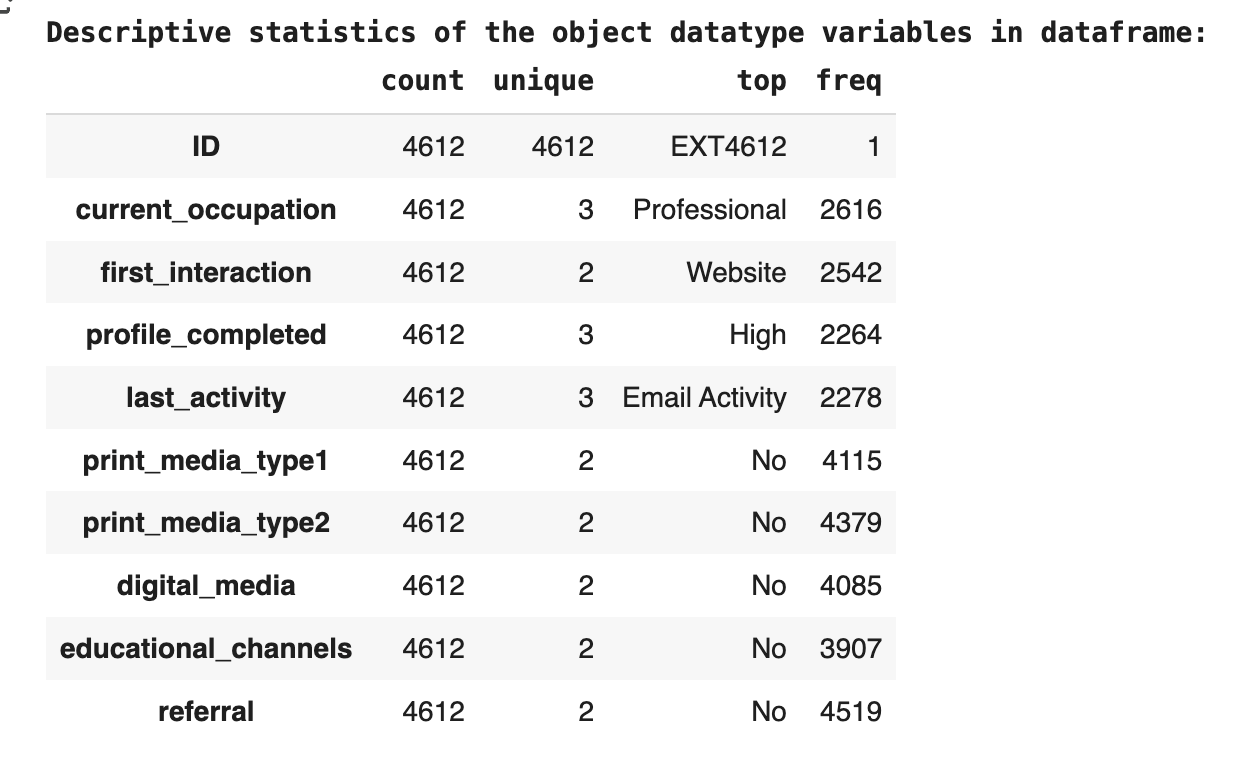
The company gets leads from various channels such as print media, digital media, referrals, etc. Which of these channels have the highest lead conversion rate?

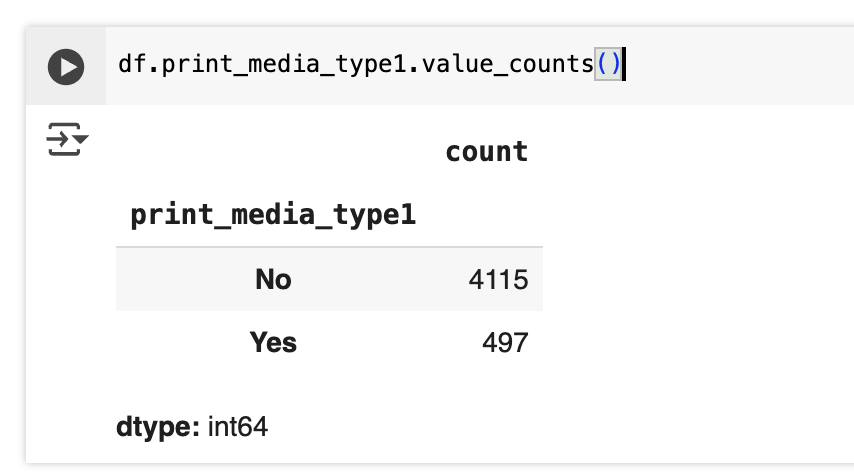
??

People browsing the website or mobile application are generally required to create a profile by sharing their personal data before they can access additional information.Does having more details about a prospect increase the chances of conversion?

??

****

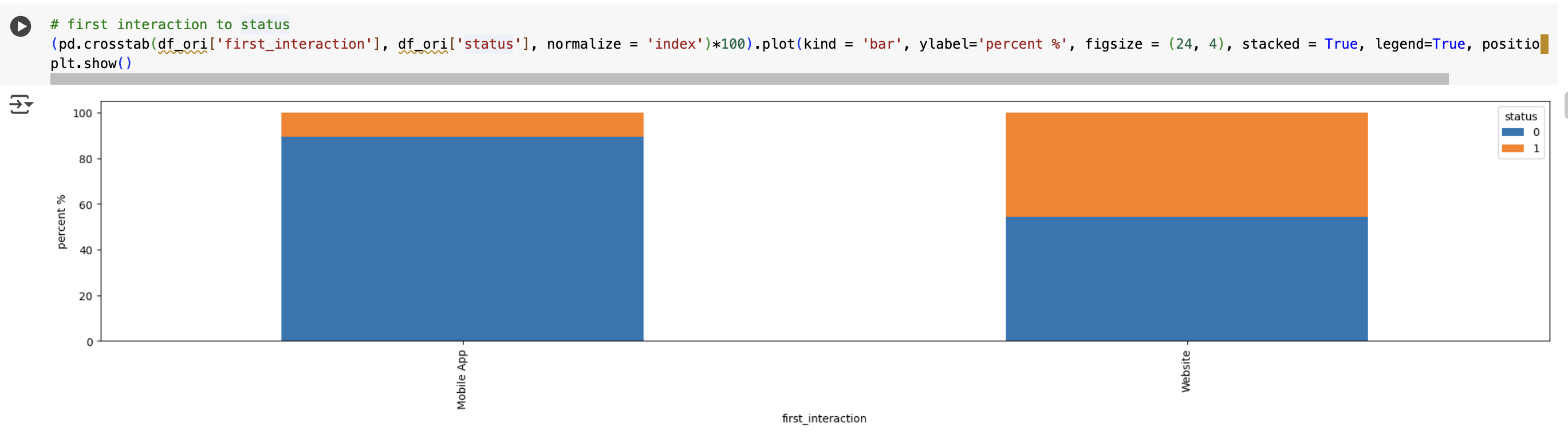
****

****

**Data Dictionary**

The data contains the different attributes of leads and their interaction details with ExtraaLearn. The detailed data dictionary is given below.

* ID: ID of the lead
* age: Age of the lead
* current\_occupation: Current occupation of the lead. Values include 'Professional', 'Unemployed', and 'Student'
* first\_interaction: How did the lead first interact with ExtraaLearn? Values include 'Website' and 'Mobile App'
* profile\_completed: What percentage of the profile has been filled by the lead on the website/mobile app? Values include Low - (0-50%), Medium - (50-75%), High (75-100%)
* website\_visits: The number of times a lead has visited the website
* time\_spent\_on\_website: Total time spent on the website
* page\_views\_per\_visit: Average number of pages on the website viewed during the visits
* last\_activity: Last interaction between the lead and ExtraaLearn
* Email Activity: Seeking details about the program through email, Representative shared information with a lead like a brochure of the program, etc.
* Phone Activity: Had a phone conversation with a representative, had a conversation over SMS with a representative, etc.
* Website Activity: Interacted on live chat with a representative, updated profile on the website, etc.
* print\_media\_type1: Flag indicating whether the lead had seen the ad of ExtraaLearn in the Newspaper
* print\_media\_type2: Flag indicating whether the lead had seen the ad of ExtraaLearn in the Magazine
* digital\_media: Flag indicating whether the lead had seen the ad of ExtraaLearn on the digital platforms
* educational\_channels: Flag indicating whether the lead had heard about ExtraaLearn in the education channels like online forums, discussion threads, educational websites, etc.
* referral: Flag indicating whether the lead had heard about ExtraaLearn through reference.
* status: Flag indicating whether the lead was converted to a paid customer or not.



# --<remove>The company's first impression on the customer must have an impact. Do the first channels of interaction have an impact on the lead status?

Mobile app usage is high (89%) compared to website than using it for conversion.

Website usage of 45% is the first interaction for lead conversion to paid customer.

* Highest mobile app usage by leads 89%.
* About 45% paid customers use website as first interaction compared to mobile app being 11%.