

# Insights from Customer Data

---

## Problem Statement

Given the dataset, we have to predict if a customer would buy a product when he was approached by the store.

## Insights from Given Data

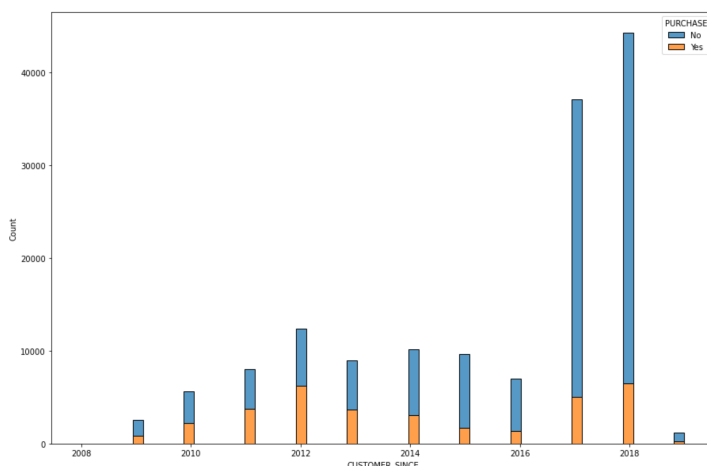
The dataset given had quite over 60,000 missing values in total. I handled the missing values by using the mode for STATE and INCOME\_GROUP columns and made use of a new missing category for the variables OCCUPATION and LOYALTY\_PROGRAM. There were a few outliers in the PAST\_PURCHASE column. These were handles by removing those rows which were more than 3 standard deviations away. After this I performed EDA on this dataset to gain some information about the given data. Some of them have been listed below in no specific order of importance.

### 1. The company has had exponential growth in the number of customers from 2017

When we plot the CUSTOMER\_SINCE column's counts, we see that the number of customers have had a very high growth from the year 2016 to 2017. It has almost grown from less than 10,000 customers to close to 40,000 customers in a single year.

Assuming that the data is accurate, this means that the company has had great growth in the number of customers.

The issue with this is that the percentage of those who would buy the product when reached by the store has reduced.

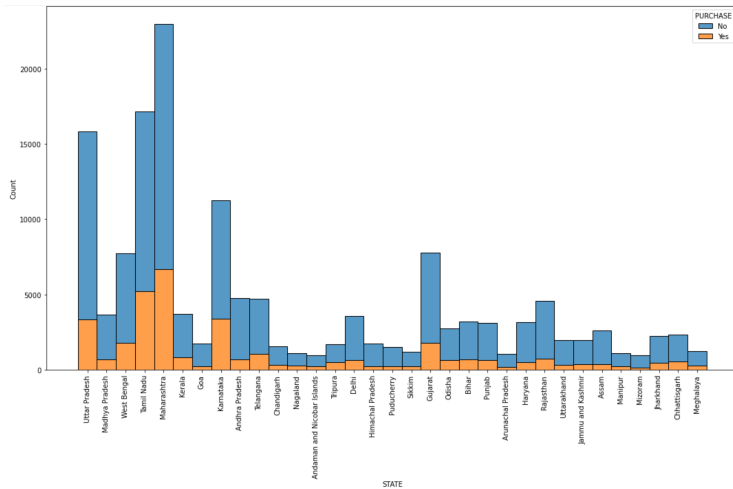


### 2. Targetting Karnataka is better than UP.

When I drew a plot of counts of the number of customers in each state, I could see that the most number of orders are from Maharashtra, Tamil Nadu, Uttar Pradesh and Karnataka in a decreasing order.

An important point to note here is that while UP has a higher number of customers, the percentage of those purchasing a product when approached by the store is higher in Karnataka. This way, by

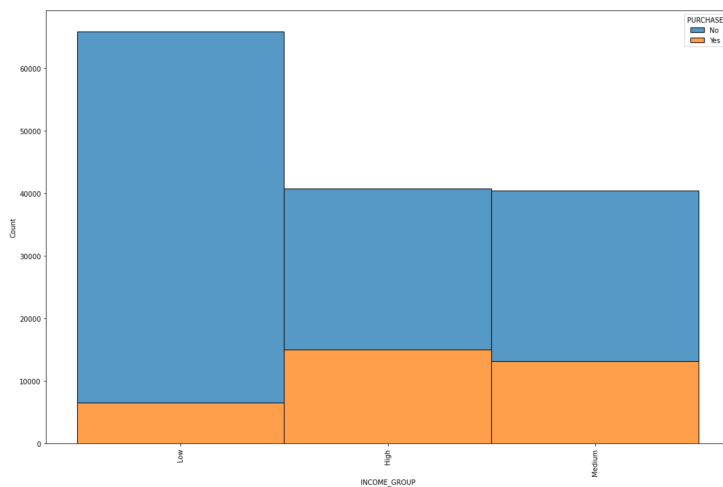
targeting Karnataka over UP for the same cost, the company can make more sales in Karnataka.



### 3. People from high income groups more likely to buy the product.

When I drew a plot of INCOME\_GROUP stacked with PURCHASE, it came to my attention that people belonging to the high and medium income group are much more likely to buy the product when reached out by the store. The percentage of those buying in high and medium income groups is significantly higher than those from the low income group.

What this means is that it would make much more sense for us to pitch our products to people from high and medium income group than to those from low income group keeping the number of pitches fixed.



### 4. Self Employed and Business Men/Women are more likely to accept when approached by the store.

When plots for income class, occupation type, and purchase is drawn, we see that the self employed and business men are in the higher and middle income groups. What this means is that they are more likely to buy the product than those from the lower income group.