

Repeat Customer Prediction

Problem Statement

Predict the Customers who are more likely to shop again and also recommend products to them.

Approach:

Though I think this problem could be modelled as a supervised learning problem, the problem would be:

- We cannot know for certain after how many visits a Customer can be labelled as a **Loyal Customer**(a customer with a high probability to return).
- Also predicting over a new Customer to know if he will revisit would give inaccurate predictions.

In our data only a handful percentage of Customers are contributing to majority of our revenue.

Since we don't want to leave out on these Loyal Customers, it would be very important to know if the these customers would revisit again.

- If we find out that if we are almost losing him, we should be able to make good recommendations and make him shop again.
- So effectively what we want is a boundary between several clusters of Customers and a recommender system.

So, I modelled this problem as a unsupervised one and made clusters of the Customer base.

Feature Engineering

From the given data we can use:

- Age
- Income (instead of Income category)
- SUB_CAT of products the customer shopped.

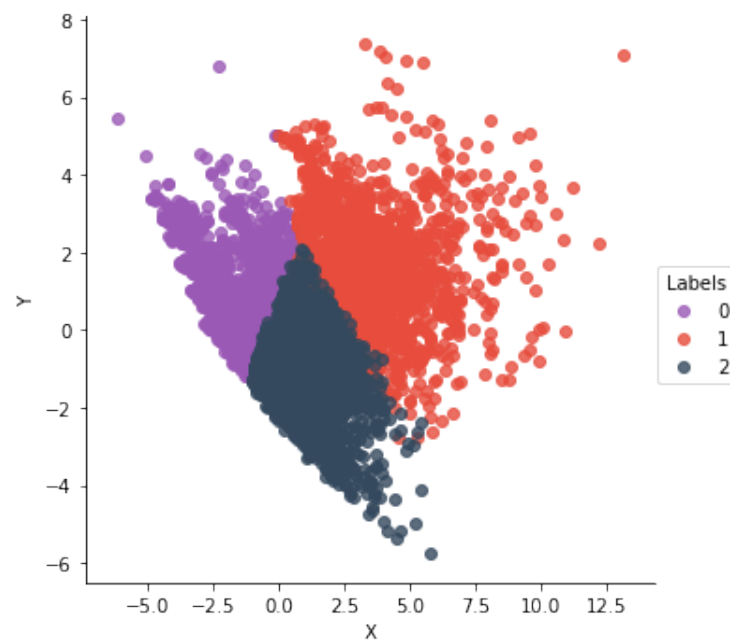
In addition to these features, the features extracted are:

- Total money the Customer spent.
- Total money the Customer spent in the last one month.
- Total money the Customer spent in the last two months.
- Total money the Customer spent in the last three months.
- Total money the Customer spent in the last six months.

- Total number of times the Customer shopped.
- Total number of times the Customer shopped in the last one month.
- Total number of times the Customer shopped in the last two months.
- Total number of times the Customer shopped in the last three months.
- Total number of times the Customer shopped in the last six months.

- Numbers of months since the Customer last shopped.

Clustering



CLUSTER COLOR	CUSTOMERS SHARE	SPENDS PER CUSTOMER	RECENCY IN MONTHS	FREQUENCY PER PERSON
	45.75%	675	1.65	30
	15.07%	1718	1.15	39
	39.18%	935	1.01	40

- Clearly Customers in one and two spends more money, visits more frequently and also they are take less months to visit again.
- So these Customers are our interest.

Recommending Products

Since the data is large a subset of data is considered.

- A user-product matrix is built with Customers on the vertical axis and Products as columns.
- If a customer had bought the product in the past 1 is placed on the corresponding Customer and Product
- Using cosine similarity the item-item matrix is filled on how similar the products are.
- Finally using cosine and item-item matrix the Costumer-Item matrix is filled to recommend 10 products to each user.