



# Credit Consumption Prediction Challenge.

Predicting the average spend by the customers.

Shubham Abhange

# The problem

## Company

**The Common Man Bank Ltd (CMB)**

## Context

The Common Man Bank Ltd (CMB) wants to understand these patterns thoroughly and get insights on the customer persona and the spending patterns

## Problem statement

**Predict the average spend for a different set of customers for the coming 3 months.**

# Challenges deep-dive

## Challenge 1

### Knowing the current spend pattern

- So for the present data of the customer getting the current spend pattern.

## Challenge 2

### Predict the average spend.

Predicting the average spend of the customer for coming 3-months.

# Solution

1. Building the strategy partnership with the vendors for discounts and other rewards to retain the customers.

2. And based on the spending they can target customer with the loans and financial products.

---

# Overview of DATA

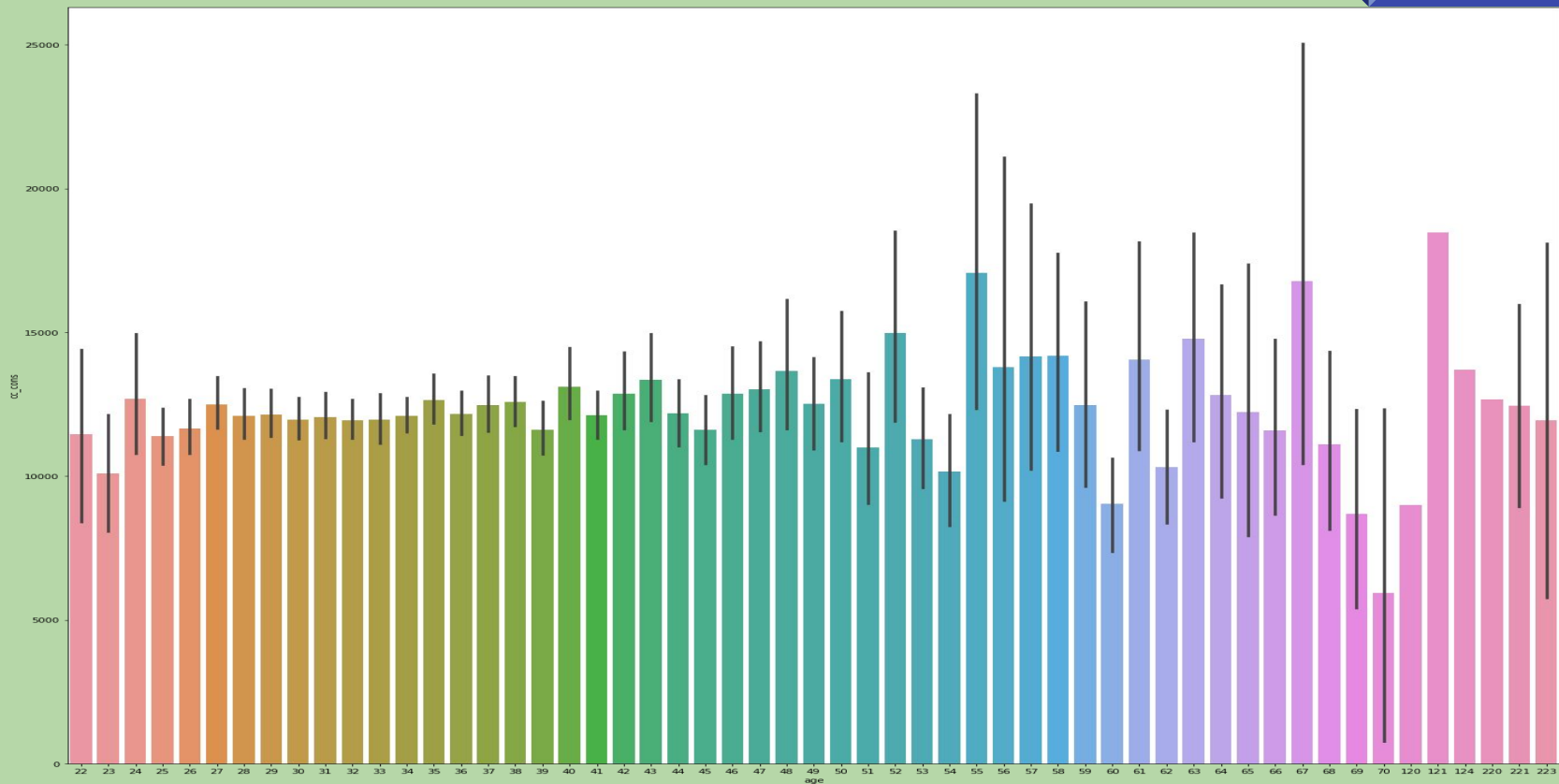
Variable	Description
id	Unique ID for every Customer
account_type	Account Type – current or saving
gender	Gender of customer
age	Age of customer
region_code	Code assigned to region of residence (has order)
cc_cons_apr	Credit card spend in April
dc_cons_apr	Debit card spend in April
cc_cons_may	Credit card spend in May
dc_cons_may	Debit card spend in May
cc_cons_jun	Credit card spend in June
dc_cons_jun	Debit card spend in June
cc_count_apr	Number of credit card transactions in April
cc_count_may	Number of credit card transactions in May
cc_count_jun	Number of credit card transactions in June
dc_count_apr	Number of debit card transactions in April
dc_count_may	Number of debit card transactions in May
dc_count_jun	Number of debit card transactions in June
card_lim	Maximum Credit Card Limit allocated
personal_loan_active	Active personal loan with other bank
vehicle_loan_active	Active Vehicle loan with other bank
personal_loan_closed	Closed personal loan in last 12 months
vehicle_loan_closed	Closed vehicle loan in last 12 months
investment_1	DEMAT investment in June
investment_2	fixed deposit investment in June
investment_3	Life Insurance investment in June
investment_4	General Insurance Investment in June
debit_amount_apr	Total amount debited for April
credit_amount_apr	Total amount credited for April
debit_count_apr	Total number of times amount debited in April
credit_count_apr	Total number of times amount credited in April
max_credit_amount_apr	Maximum amount credited in April
debit_amount_may	Total amount debited for May
credit_amount_may	Total amount credited for May
debit_count_may	Total number of times amount debited in May
credit_count_may	Total number of times amount credited in May
max_credit_amount_may	Maximum amount credited in May
debit_amount_jun	Total amount debited for June
credit_amount_jun	Total amount credited for June
debit_count_jun	Total number of times amount debited in June
credit_count_jun	Total number of times amount credited in June
max_credit_amount_jun	Maximum amount credited in June
loan_enq	Loan enquiry in last 3 months
emi_active	Monthly EMI paid to other bank for active loans
cc_cons	(Target) Average Credit Card Spend in next three months

# EDA

1. Checking the null values in both the train and test set.
2. Also checking for the outliers in the data.
3. As this is a banking problem, there might be chances of data mislead, so using the null values will effect in the correlation.
4. Using the median to fill nan values in the data set.
5. To get the consumption of the three months, we create a new columns which is the sum of all count, and rest others.

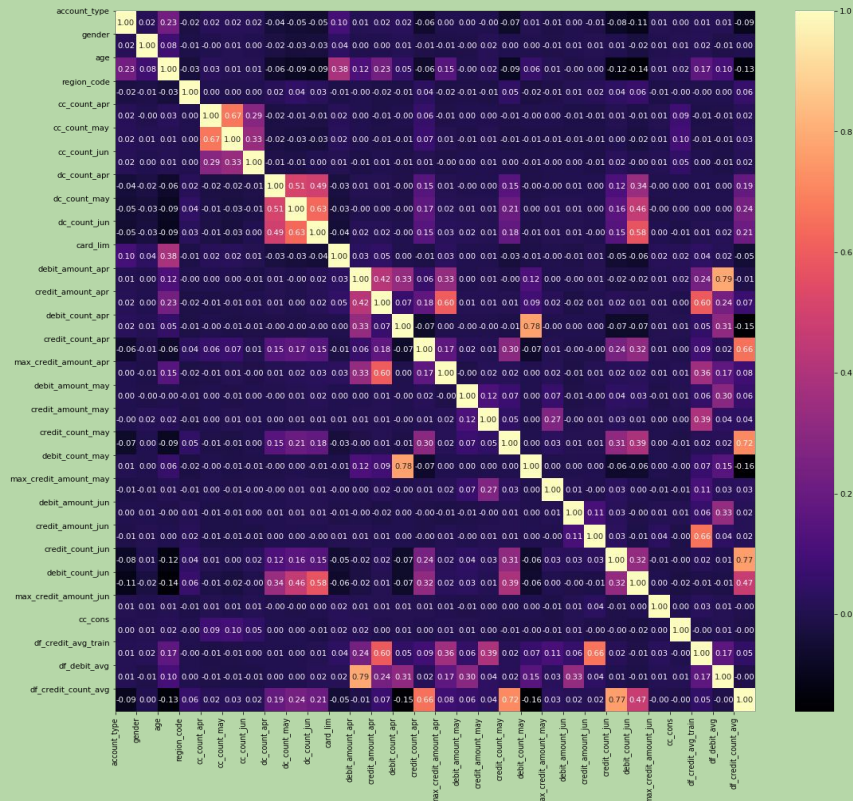
# EDA

Age Vs cc\_cons



# Correlation of Target and other variables

- We have plotted the correlation matrix to check the correlation between all the columns.
- Accordingly we can drop the columns which are not necessary.
- Apparently we can understand distribution of data between the strongest correlated columns.









# Model & Approaches

Three vanilla models were assessed without performing hyperparameter tuning.

The models were

- Lasso
- Ridge
- Random Forest Regressor

None of the Four models were able to give an RMSLE score Below 1.09.

Thus called for performing hyperparameter tuning using Gradient Boosting Regressor.