# Model building and interpretation.

## Build various models (You can choose to build models for either or all of descriptive, predictive or prescriptive purposes)

For this process we have implemented the following models

* Linear Regression
* Naive Bayes Regression
* Random Forrest
* SVR

## Test your predictive model against the test set using various appropriate performance metrics

We took mse as our benchmark and the results of the above-mentioned models are provided below

* Linear Regression

mse 422806.529263

mae 511.988021

* Naive Bayes Regression

mse 759621.9292035399

mae 653.622418879056

* Random Forrest

mse 668379.9938745537

mae 640.2521429333992

* SVR

mse 1401289.7876900986

mae 907.1755789458981

We also ran Randomized CV search on random forest we got the following as the best params

* 'n\_estimators': 1600,
* 'min\_samples\_split': 5,
* 'min\_samples\_leaf': 1,
* 'max\_features': 'auto',
* 'max\_depth': 10,
* 'bootstrap': True

## Interpretation of the model(s)

Based on the several permutations we have identified that model provides if the following features are used

* 'CustTenure',
* 'Occupation',
* 'EducationField',
* 'Designation',
* 'MonthlyIncome',
* 'Complaint',
* 'ExistingPolicyTenure',
* 'SumAssured',
* 'LastMonthCalls',
* 'Age\_bin'

Even on conducting PCA we find that we need 11 components to explain 75% of data

The PCA mapping of the same is provided below

Graphical user interface, text, application

Description automatically generated

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Table, calendar

Description automatically generated with medium confidence

The following is the performance matrix of PCA linear regression is

mse 579345.3838874538

mae 605.4617366173047

# 2). Model Tuning and business implication

## Ensemble modelling, wherever applicable

We have used AdaBoostRegressor to see how the model behaves the following is the performance metrics

mse 668379.9938745537

mae 640.2521429333992

## Any other model tuning measures (if applicable)

Currently we are working on the model to select the best model using the best feature and best param. Due to computational bottleneck, this process is taking some time

## Interpretation of the most optimum model and its implication on the business

We find that the linear regression model providing ther best model as of now.