FINAL PROJECT

PREDICTING VEHICLE INSURANCE

MUHAMMAD MUWAHIDDIN MASSAID

TABLE OF CONTENTS

DATA UNDERSTANDING

DATA UNDERSTANDING

1. Background

An insurance company want to offer vehicle insurance to their customers

2. Problems

They want to cross selling, but don't know who is interested in vehicle insurance

3. Goal

Predicting customers interested in vehicle insurance

4. Objective

Analysis features for gain more information & optimize metric ML to predicting data

DATASET INFORMATION

381.109 ROWS

12 FEATURES

ID
GENDER
AGE
DRIVING_LICENSE
REGION_CODE
PREVIOUSLY_INSURED
VEHICLE_AGE
VEHICLE_DAMAGE
ANNUAL_PREMIUM
POLICY_SALES_CHANNEL
VINTAGE

1 TARGET

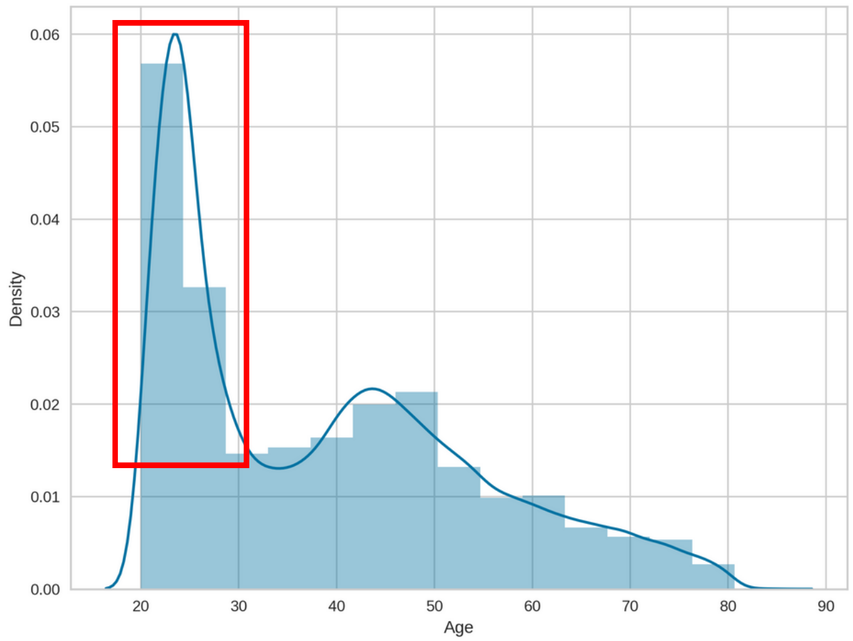
RESPONSE
(1: Interested,
0: Not Interested)

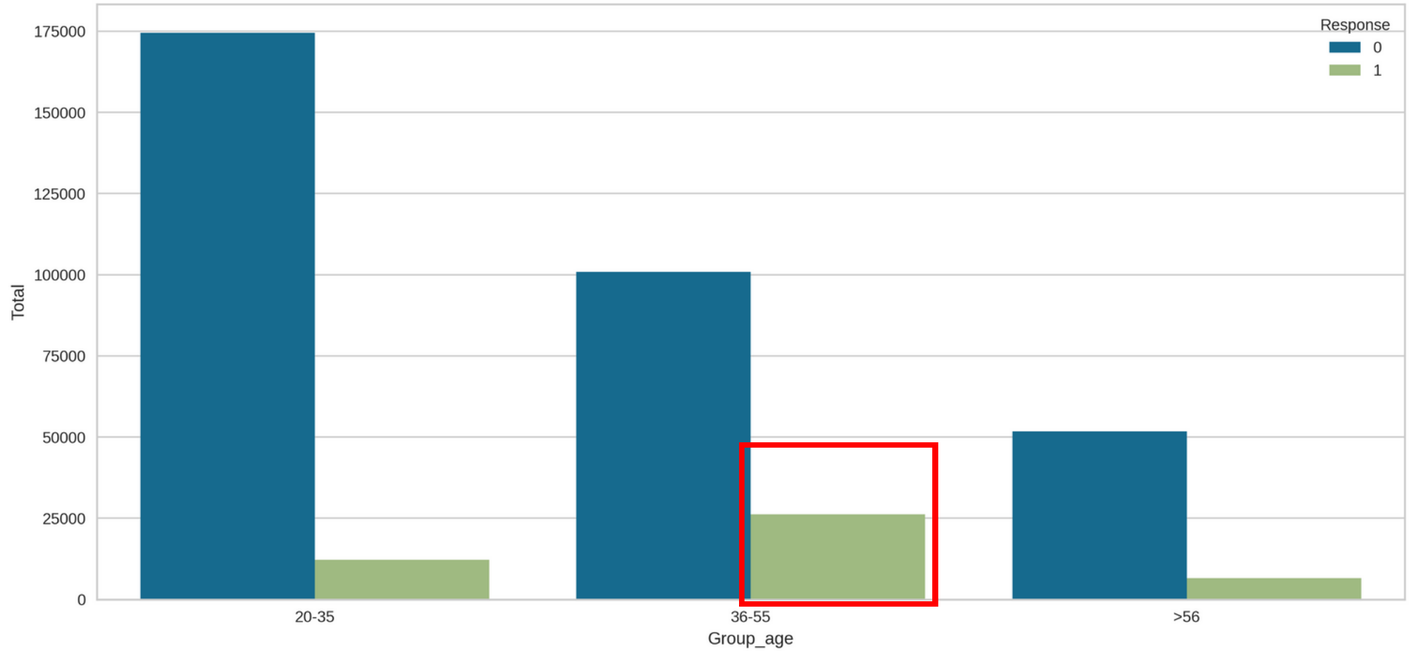
DATA PREPROCESSING

- Dropped 'id' columns
- No missing values & duplicated data
- Check data distribution & outliers
- Convert float data to integer
- Encoding data (Label & One Hot)
- Mulitolinearity check

EXPLORATORY DATA ANALYSIS

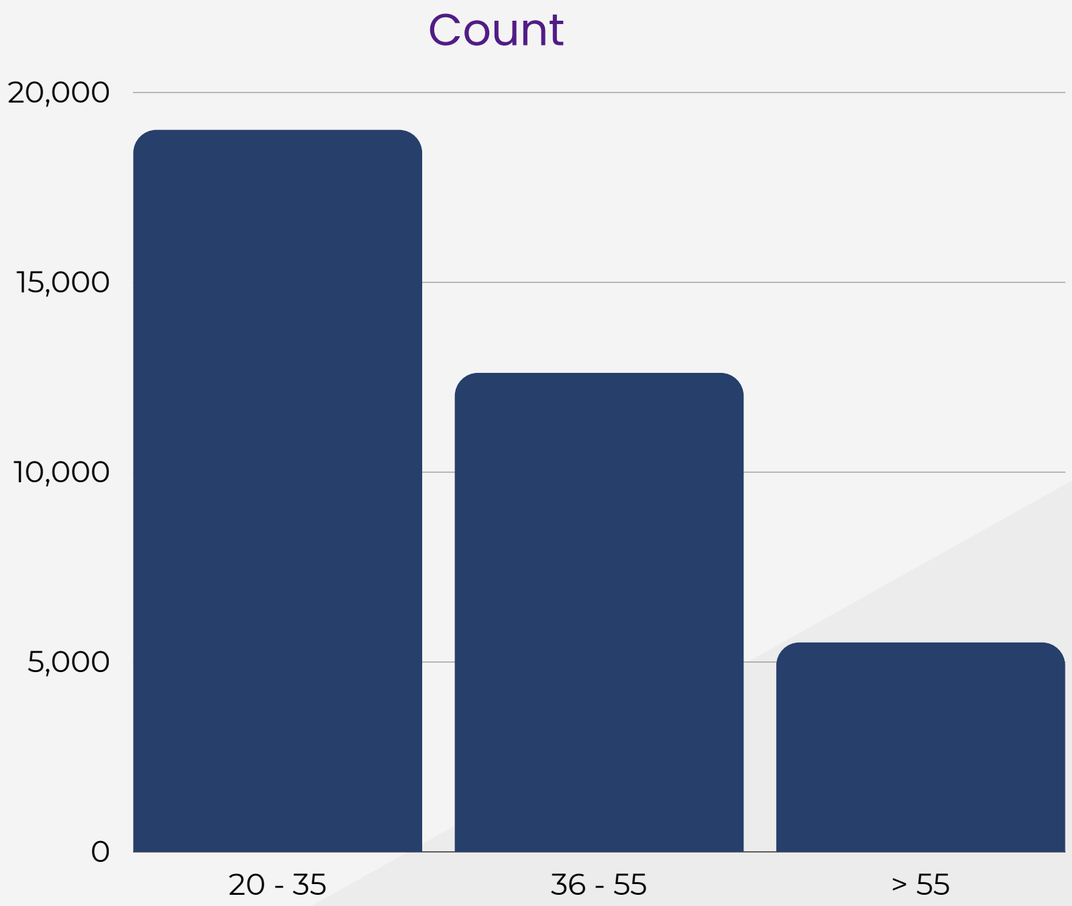
AGE & RESPONSE

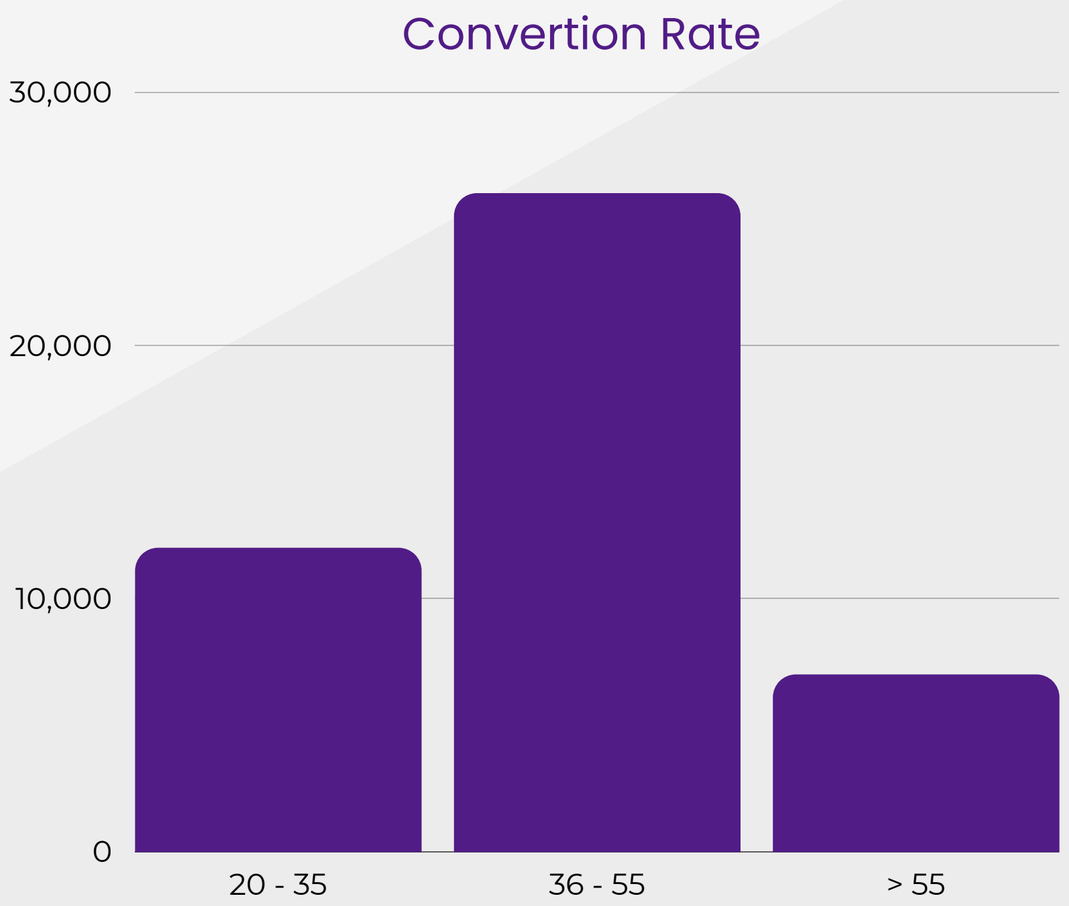




From the table, grouping `36 - 55` are the most interested age ranging in vehicle insurance, instead the biggest population are grouping age `20 -35`. It may happen by reason of better financial conditions to buy more comfortable living facilities.

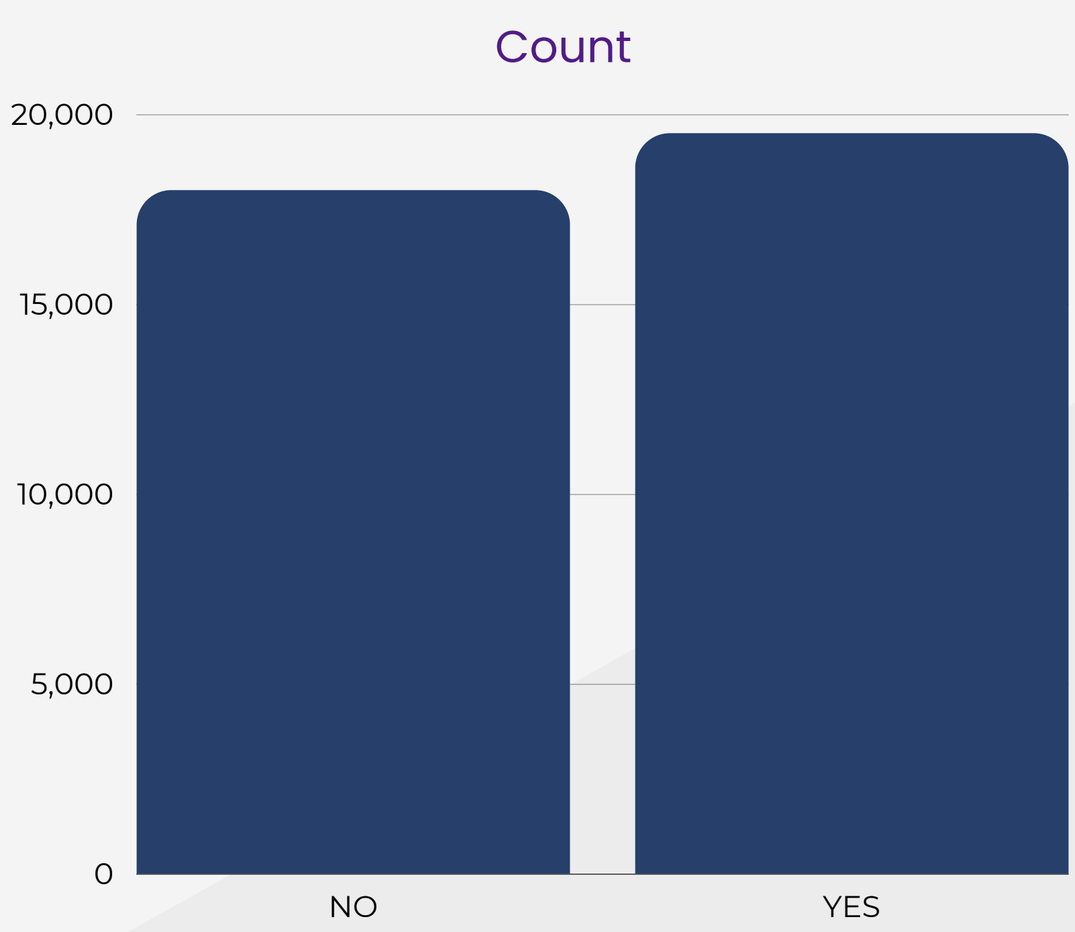
AGE & RESPONSE (CONVERTION RATE)

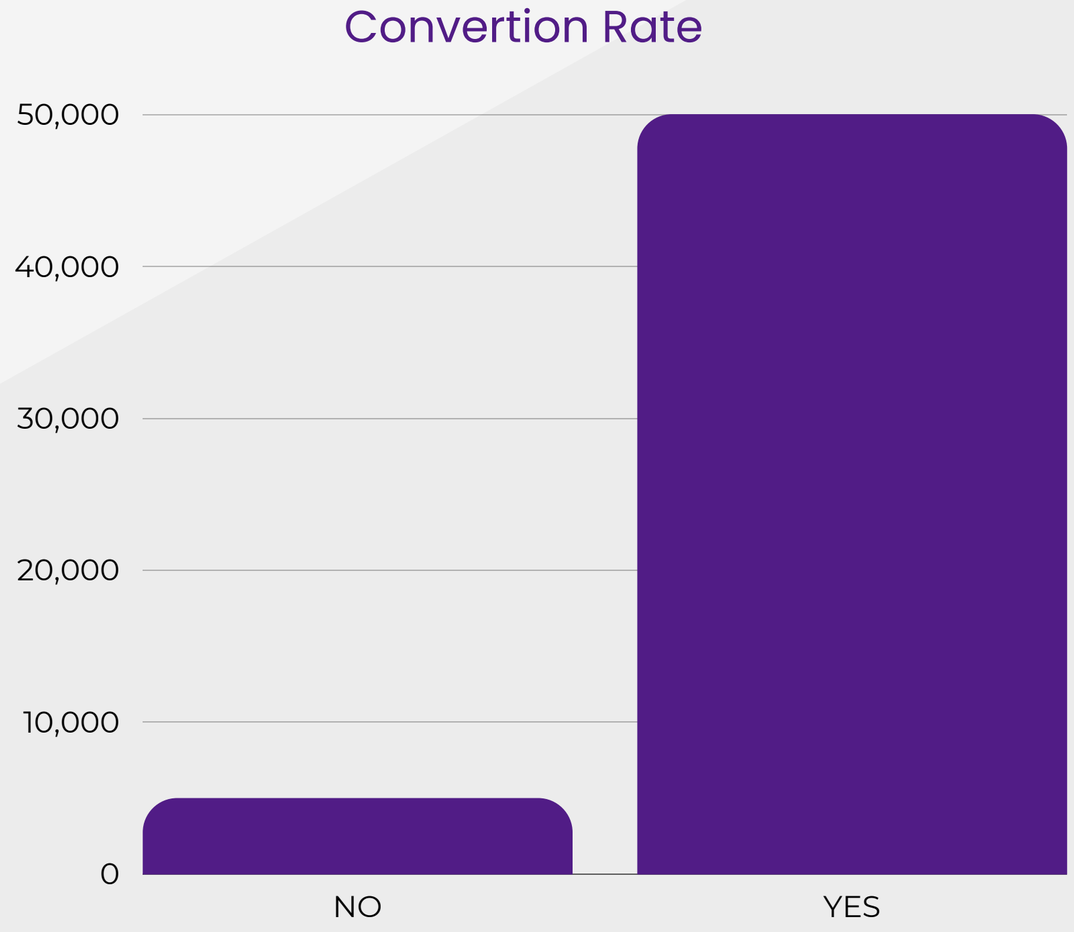




``**Recomendation:`** You need some strategy to nurture them until they are interested in vehicle insurance. The fact that customers have interest comes from the group_age 36-55 which is the time they have maximum ability to generate income.

VEHICLE_DAMAGE & RESPONSE (CONVERTION RATE)

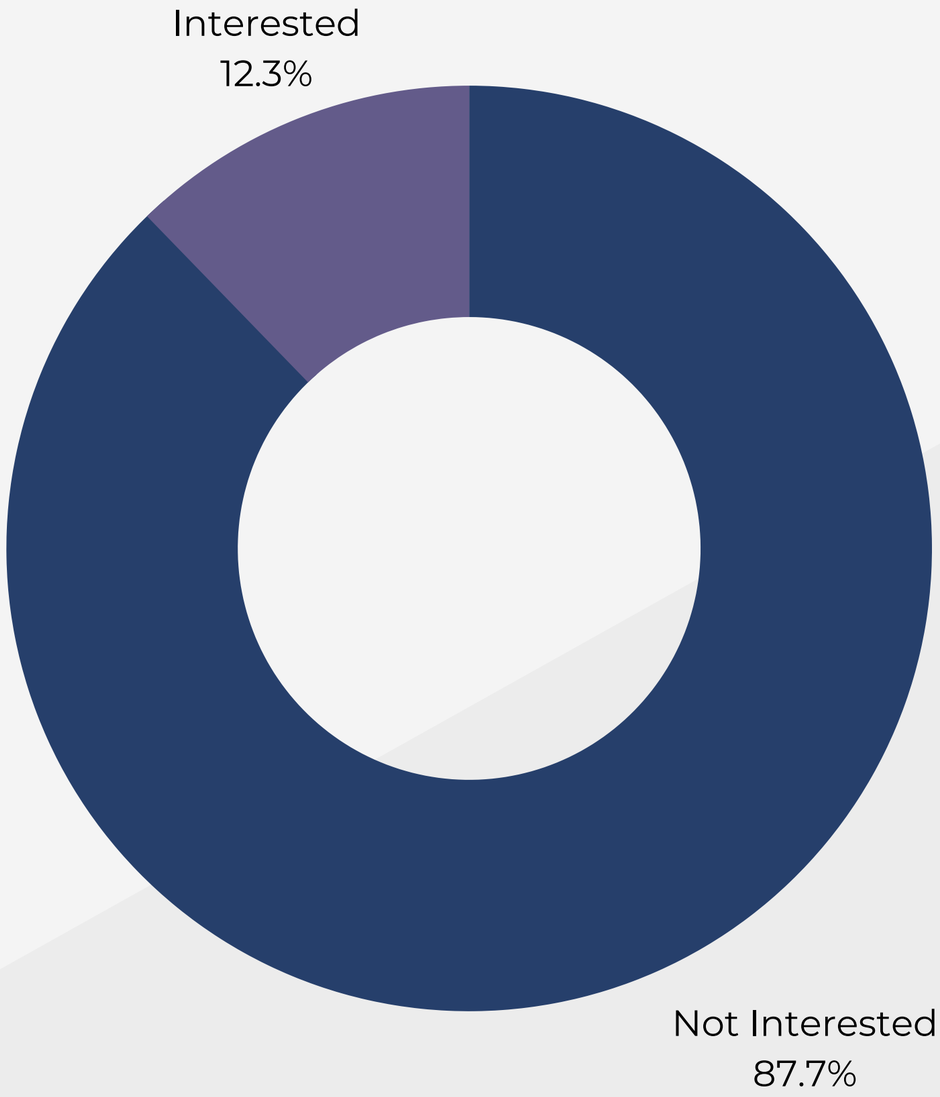




It is a good insight that majority of customers who had experienced vehicle damage are very interested in vehicle insurance, so company can offering them vehicle insurance.

MACHINE LEARNING & INTEPRETATION

TARGET VARIABLE



Target variable is imbalanced. So we preferred AUC Score to evaluate models than Accuracy.

BASELINE MODEL

MODELS	ACCURACY	AUC	RECALL	PRECISION	F1 SCORE	TT (SEC)
Light Gradient Boosting Machine	0.8780	**0.8427**	0.0002	0.2143	0.0004	2.26
Gradient Boosting Classifier	0.8780	0.8413	0.0000	0.0000	0.0004	40.49
Ada Boost Classifier	0.8780	0.8389	0.0000	0.0000	0.0000	10.06
Random Forest Classifier	0.8644	0.8142	0.0997	0.3205	0.1521	39.29

Accuracy is percentage of prediction were correct.
AUC (Area Under the Curve) means model has a good measure of separability if it has AUC near to 1.
Recall is actual positive rate.

Precision is predicted positive rate
F1-Score combines Recall and Precision to one performance matrix.
TT(SEC) is training time models in second

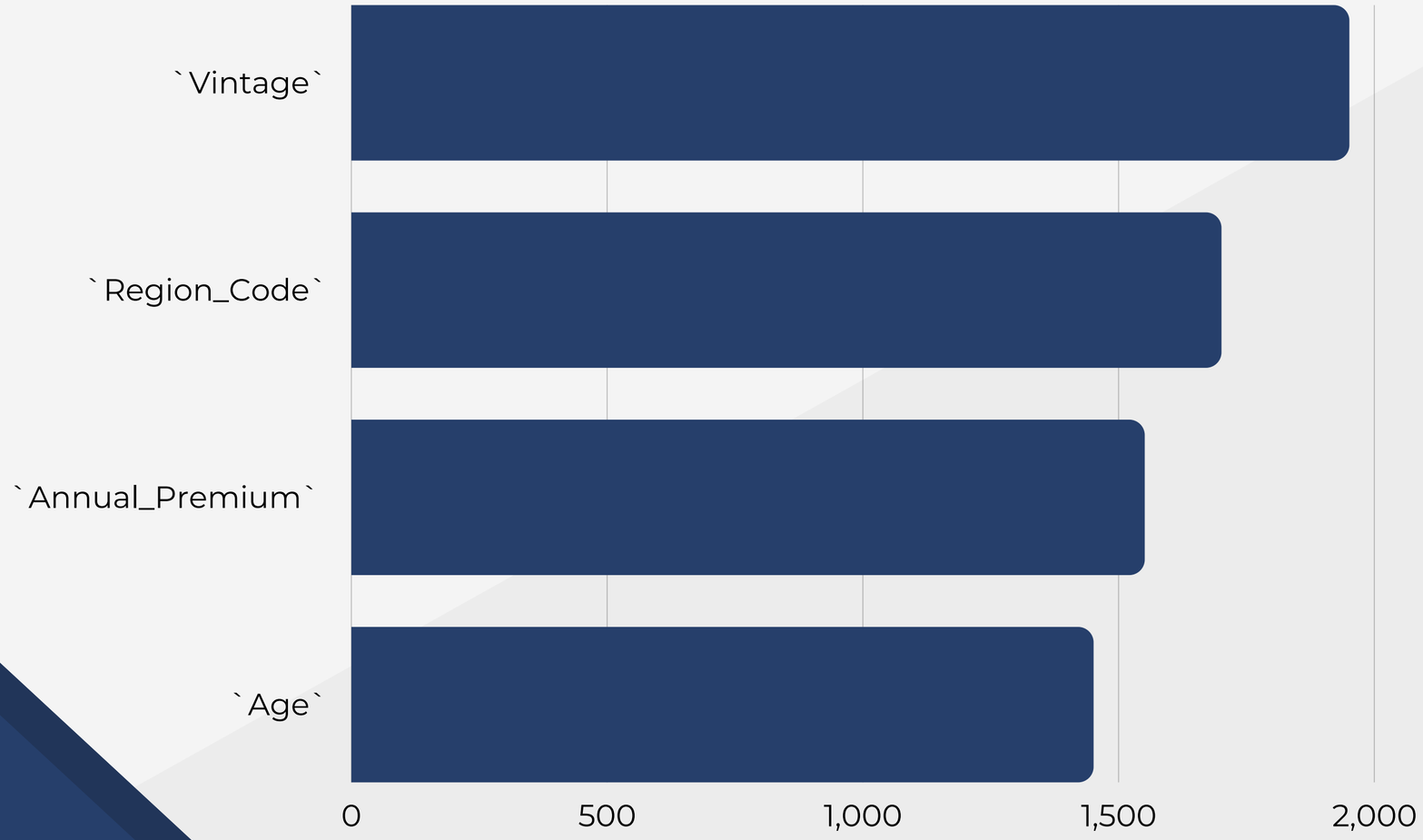
HYPERPARAMETER TUNING & EVALUATION MODEL

MODELS	ACCURACY	AUC	RECALL	PRECISION	F1 SCORE
Baseline LGBM	0.8780	0.8427	0.0002	0.2143	0.0004
Tuning LGBM	0.8775	0.8611	0.0211	0.5391	0.0405
Tuning LGBM in Test Data	0.8778	0.8574	0.0173	0.468	0.0333

MACHINE LEARNING IMPACT

- Model can predict 86% of interested customers. It means if there are 1.000 data and 120 data has interested in vehicle insurance, then 103 (86% of 120) probably are really interested in your offer.
- Assume the annual premium for vehicle insurance is $ 500, then you have potential additional income from this cross-selling is $ 51,500 from existing customers without spent any budget advestising. It is effective & efficient too cause you only approach them that interested

MOST FEATURED IMPORTANCE



The most important featured

- `Vintage`
- `Region_Code`
- `Annual_Premium`
- `Age`.

CONCLUSION

CONCLUSION

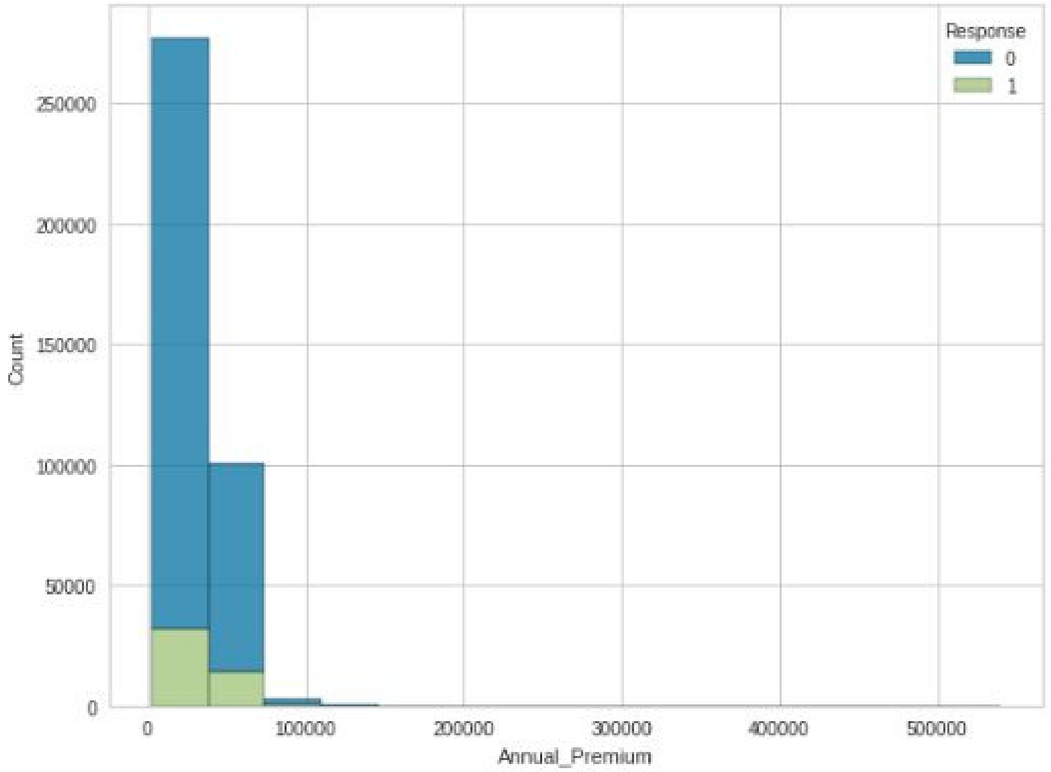
- Since most customers are under 30th, then you need some strategy to nurture them until they are interested in vehicle insurance. The fact that customers have interest comes from the group_age 36-55 which is the time they have maximum ability to generate income.
- The most important features based on our model are *`Vintage`, `Region_Code`, `Annual_Premium` & `Age`*. Focus on them. We need more information about these features to have more insight such as city/country of region code, and type of insurance (hirarcial) to match with `Annual_Premium`

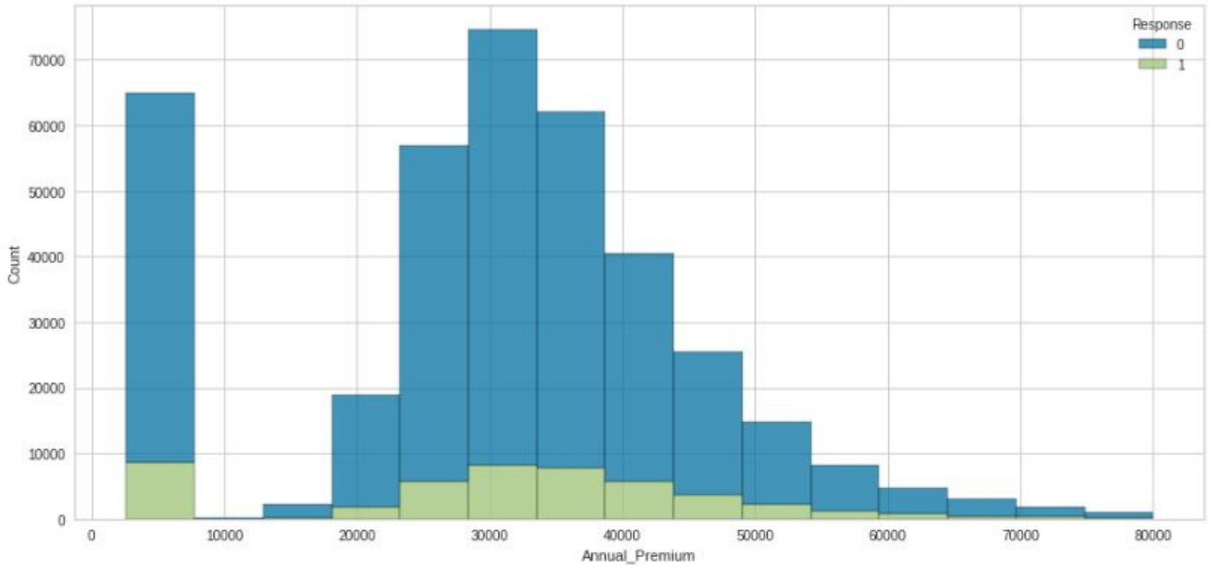
RECOMENDATION

- If possible, you can provide vehicle type and financial measuring such as job name, position in a company, or anything financial symbol that can be measured. It can be improving machine learning prediction.
- Start this cross-selling and evaluating. If you have a good result then you can do the same with your other insurance without spent any budget advertising

ATTACHMENT

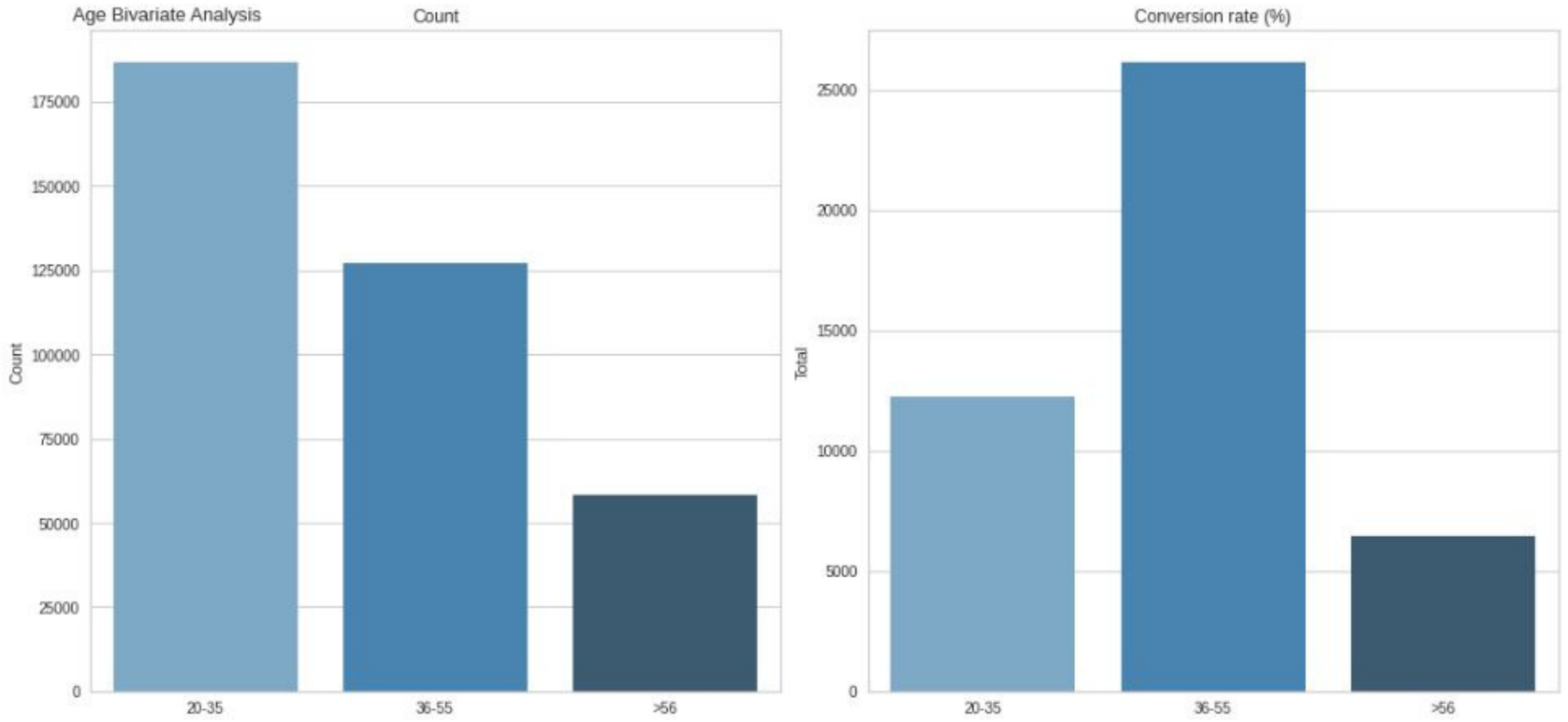
Annual_Premium & Response





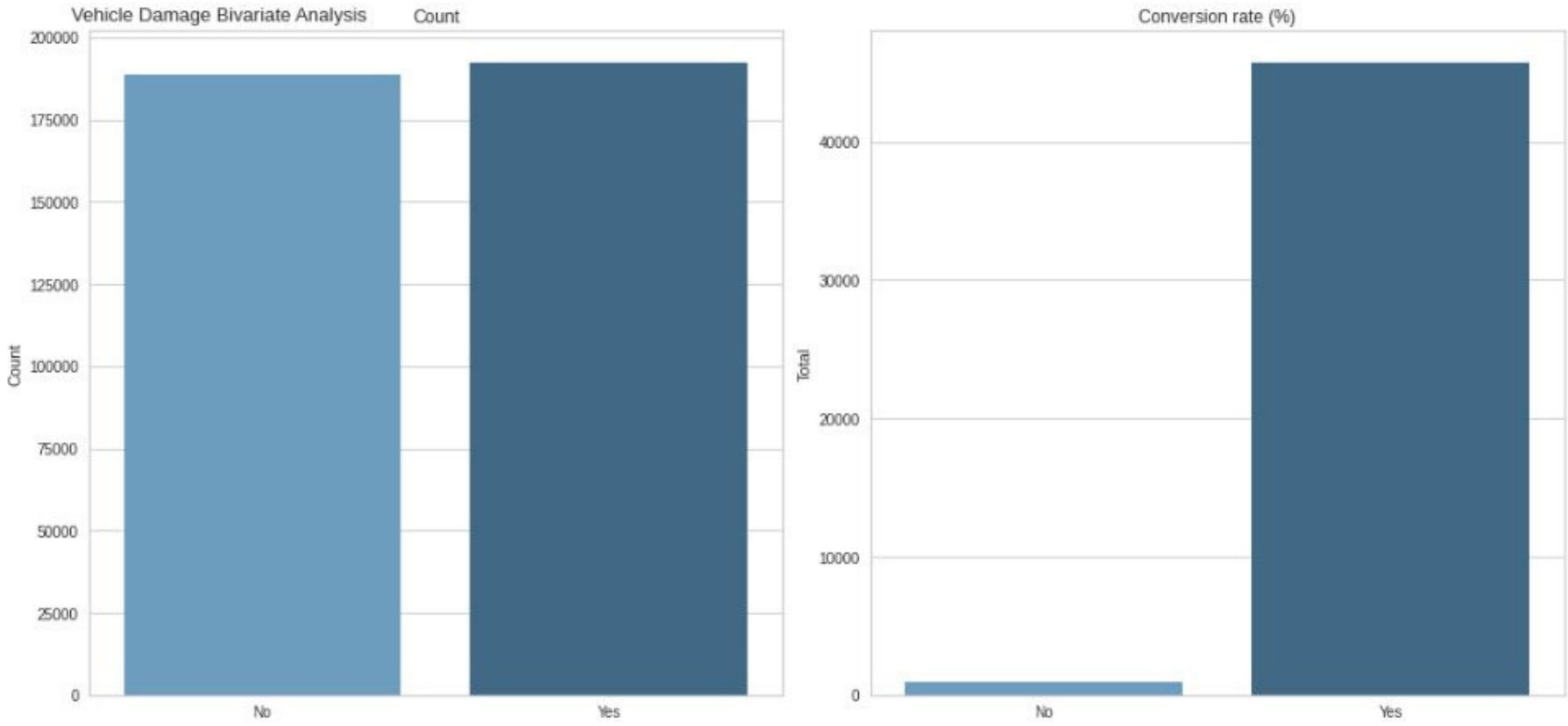
ATTACHMENT

Age & Response (Convertion Rate)



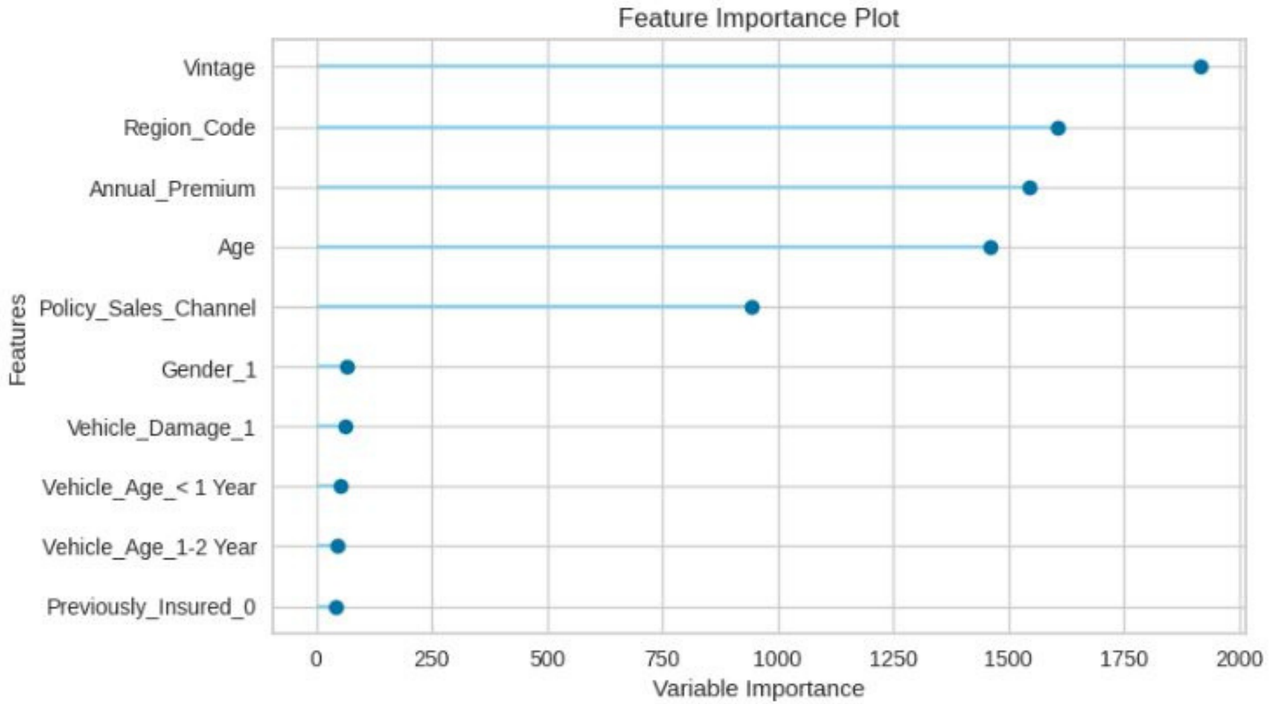
ATTACHMENT

Vehicle_Damage & Response (Convertion Rate)



ATTACHMENT

Featured Importance



Plot ROC-AUC