samkemboi12 / Insurance-Cross-Sell-Prediction

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

Building a model to predict whether the policy holders clients from past year will also be interested in Vehicle Insurance provided by the company.

0 stars 0 forks 0 watching Branches Activity Tags

Public repository

master had recent pushes 18 minutes ago

Compare & pull request

2 Branches 0 Tags Go to file t Go to file + Add file Code

This branch is 14 commits ahead of, 1 commit behind `main` .

Contribute

samkemboi12 Update README.md bccb3de · 18 minutes ago

.ipynb_checkpoints	Final touches	1 hour ago
Insurance-Cross-Sell-Prediction	Initial commit with EDA	4 days ago
__pycache__	Updated Objective 3	yesterday
EDA_utils.py	Added EDA utilities file	2 days ago
Presentation.pdf	Final touches	1 hour ago
README.md	Update README.md	18 minutes ago
TruSecure Insurance Company...	Final touches	1 hour ago
TruSecure Insurance Company...	Final touches	1 hour ago
test.csv	Initial commit with EDA	4 days ago
train.csv	Initial commit with EDA	4 days ago

README

TruSecure Insurance Cross-Sell Prediction

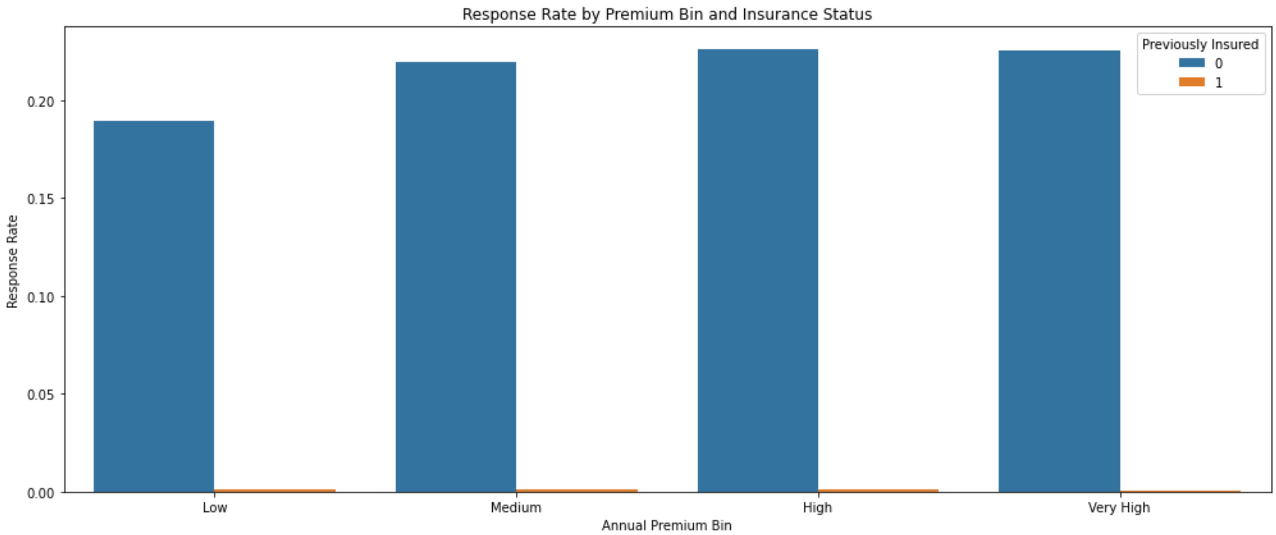
This project aims to help TruSecure Insurance Company identify which health insurance customers are likely to also purchase vehicle insurance. The solution involves exploratory data analysis (EDA), feature engineering, and predictive modeling using machine learning.

Business Problem

TruSecure is launching a new vehicle insurance product and wants to: Identify key customer characteristics that influence purchase decisions. Segment customers for targeted marketing. Predict customer conversion using a classification model.

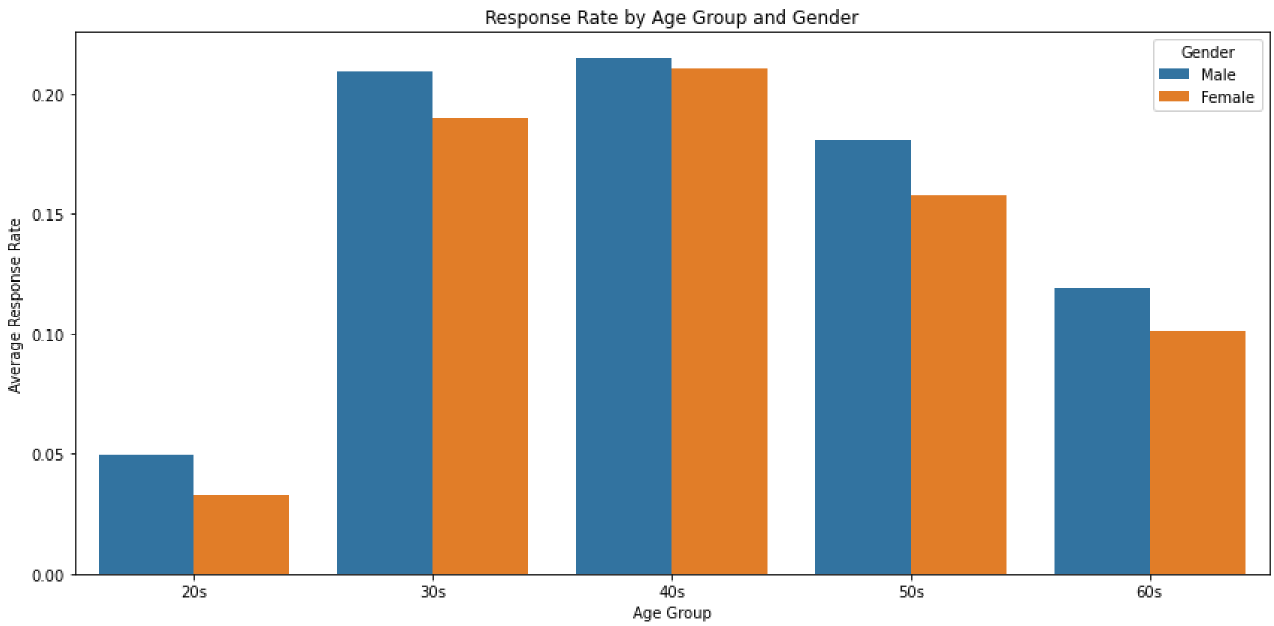
Exploratory Data Analysis (EDA)

Annual_premium level and Insurance status vs customer response rate.:



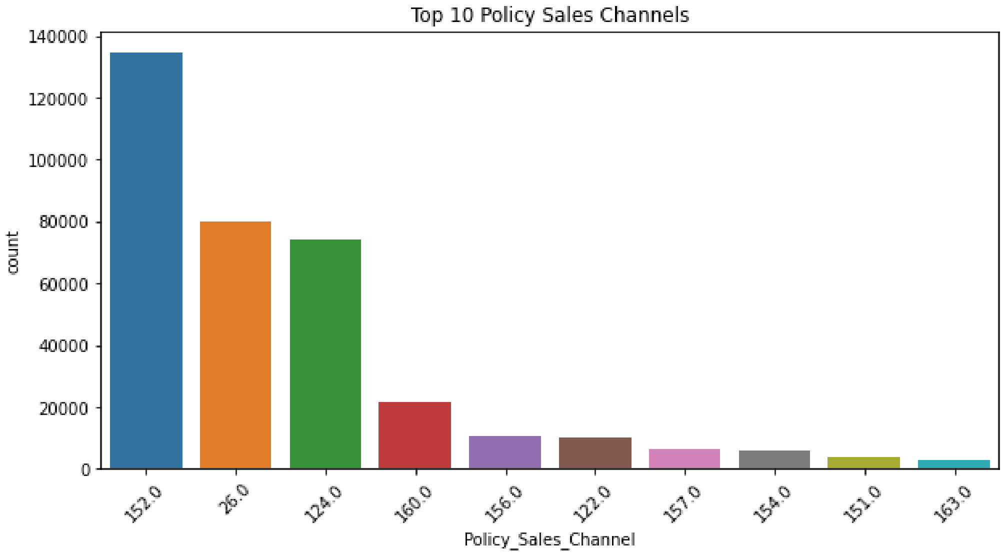
Response Rate By Age group and Gender

Most customers are middle-aged, with a higher interest in vehicle insurance seen in younger groups.



Top 10 Policy Channels Vs Response Rate

Certain policy sales channels show higher conversion rates.



Data Preparation

Cleaned and standardized categorical fields

Encoded features using One-Hot and Ordinal Encoding

Scaled numerical variables like age and premium

Split data into training and validation sets

Models Trained

Logistic Regression
Random Forest
LightGBM

XGBoost

LightGBM had the best ability to detect interested customers (high recall), which is critical in marketing use cases.

Model Performance Comparison

Model	Accuracy	Class 1 Recall	Class 1 Precision	F1 Score (Class 1)	ROC AUC
Logistic Regression	0.88	0.00	0.40	0.00	0.8342
XGBoost	0.72	0.90	0.29	0.44	0.8558
Random Forest	0.69	0.94	0.28	0.43	0.8548
LightGBM	0.70	0.93	0.28	0.43	0.8578

Recommendations

Prioritize outreach to customers with a history of vehicle damage.

Focus marketing on high-performing sales channels.

Use the model's probabilities to rank and target top prospects.

Consider using SHAP for model explainability in future phases.

Tech Stack

Monica

Instant

Repo Summary

Supports the most advanced models to help you quickly understand the contents of the repo

Summarize this repo

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package