Auto Insurance Cross-Sell

A Presentation by Samuel Middleton

Business Case

Can you use analytics data derived from customers who have a health insurance policy in order to target those most likely interested in an auto insurance policy?

The auto insurance industry is expected to make $308.8b this year.

Our Data

This dataset was sourced from Kaggle. The data is very clean and provides some very interesting insights into customer analytics data. Ultimately, we want to predict if the customer is 'Interested' as opposed to 'Not Interested' in a cross-sell.
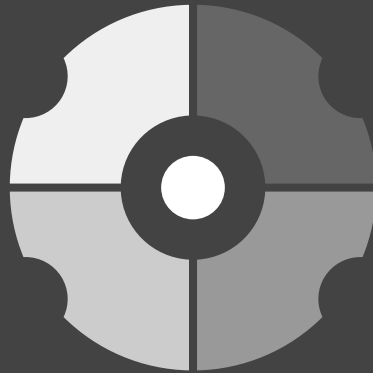
Data Definitions
In order to better understand the data.

Variable	Definition
id	Unique ID for the customer
Gender	Gender of the customer
Age	Age of the customer
Driving_License	0 : Customer does not have DL, 1 : Customer already has DL
Region_Code	Unique code for the region of the customer
Previously_Insured	1 : Customer already has Vehicle Insurance, 0 : Customer doesn't have Vehicle Insurance
Vehicle_Age	Age of the Vehicle
Vehicle_Damage	1 : Customer got his/her vehicle damaged in the past. 0 : Customer didn't get his/her vehicle damaged in the past.
Annual_Premium	The amount customer needs to pay as premium in the year
PolicySalesChannel	Anonymized Code for the channel of outreaching to the customer ie. Different Agents, Over Mail, Over Phone, In Person, etc.
Vintage	Number of Days, Customer has been associated with the company
Response	1 : Customer is interested, 0 : Customer is not interested

Exploring Data

We explored several elements of the data, but the most important in the end was the Response balance, and the Previously Insured by Age and Vehicle Damage by Age.
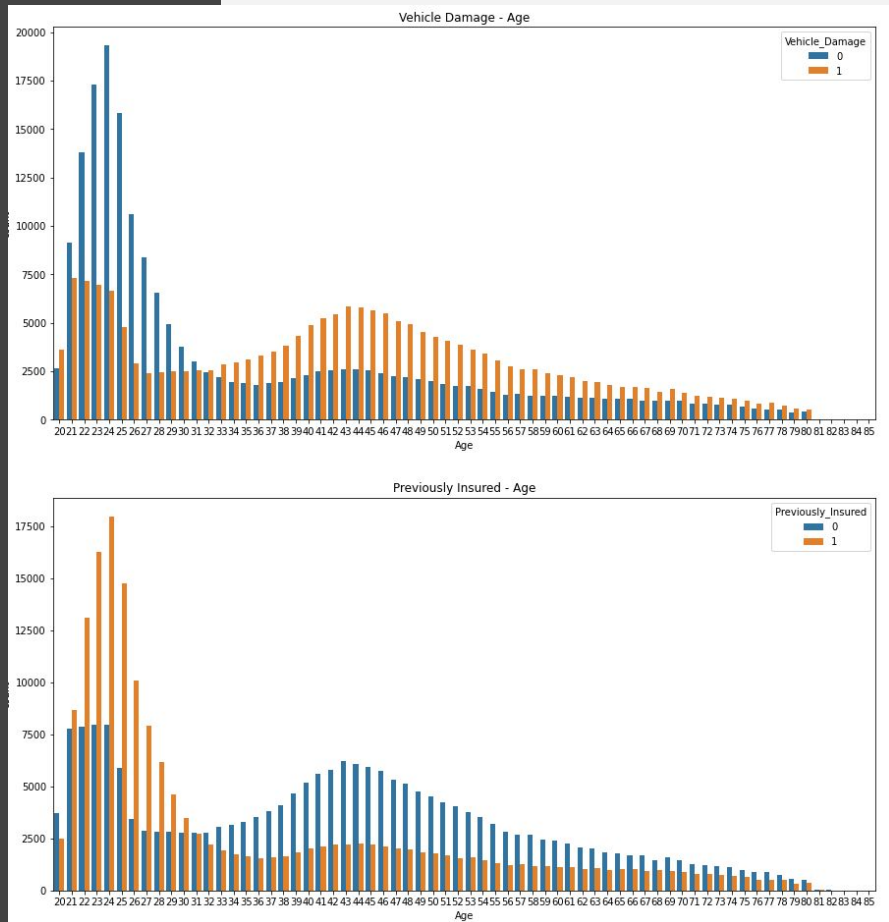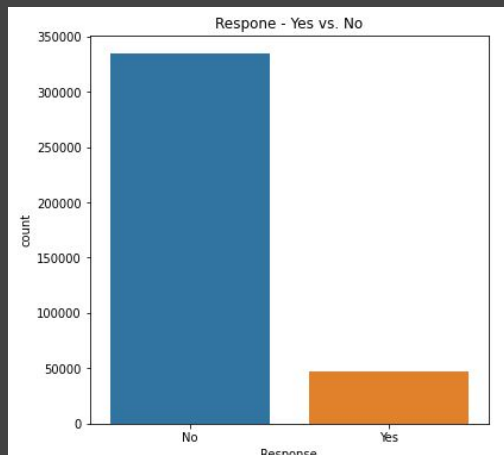
How do we do it?

Machine Learning!

Machine Learning is the process of deriving relationships between data using mathematics!

The Models used:

There are numerous Machine Learning models that can be used, we applied:
- Decision Trees
- Random Forest
- K-Nearest Neighbors
- XGBoost
- CatBoost

Model Details

It sounds complicated but there are only 2 things you really need to know:

Trees and Forests

Decision trees ask simple 'Yes' or 'No' questions in order to follow a path to an answer.

Forests are just a group of trees where the best tree gets chosen.

Nearest Neighbors

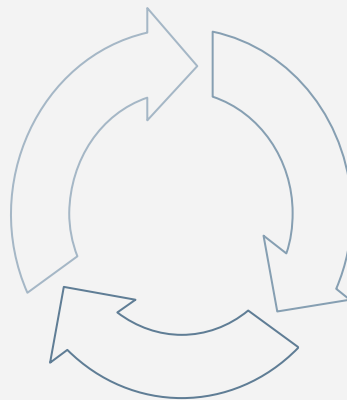Nearest neighbors classifies data based on the majority of points around it.

The Modeling Process:

An iterative Journey

Training

Tuning

Prediction

Recall Metric

We used the recall metric to score our modeling. This lets some false positives fall through the crack, but captures the most true positives.

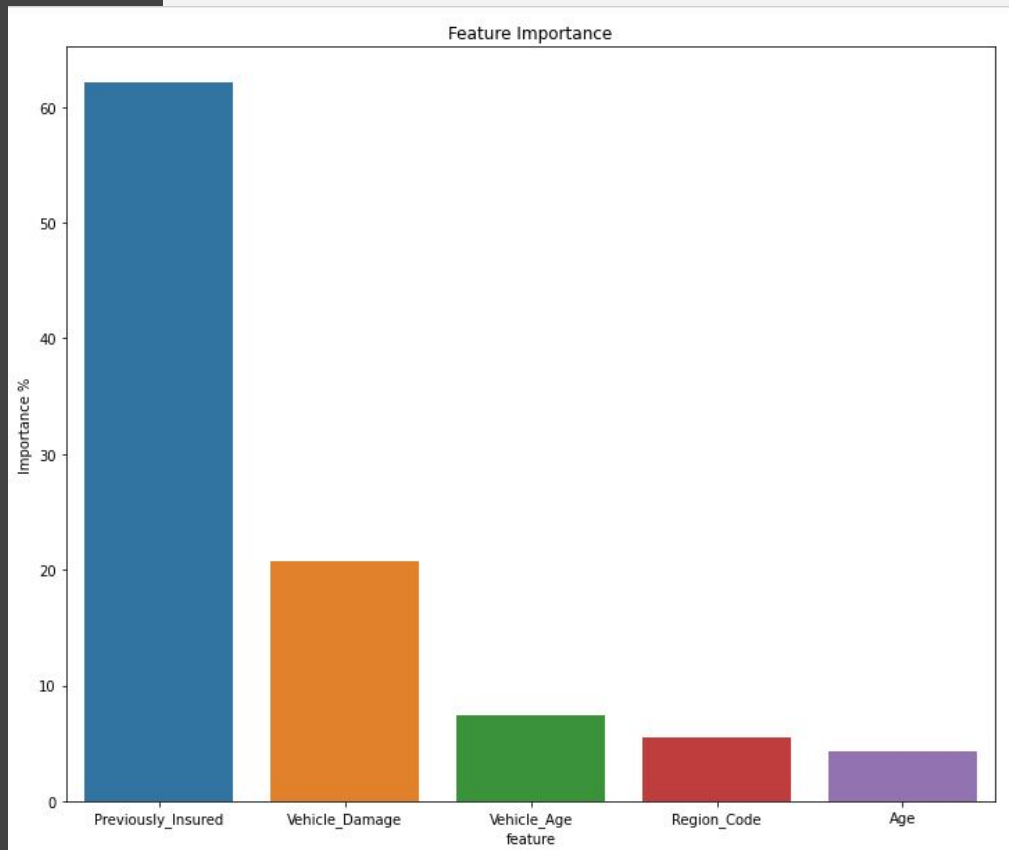**Recall —** Also called Sensitivity, Probability of Detection, True Positive Rate
The ratio of correct positive predictions to the total positives examples.

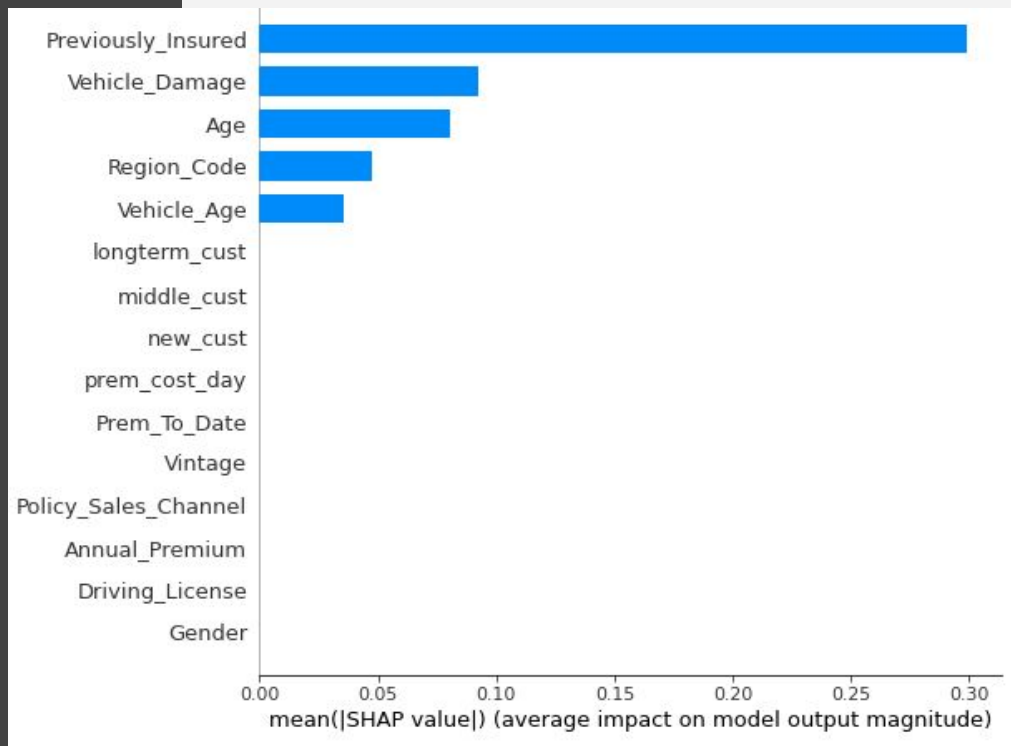$$R = \frac{TP}{TP + FN}$$

CatBoost Wins!

Our CatBoost model was the most accurate with ~87% to predict a 'Yes' response.

It also gave us insight on important data features.
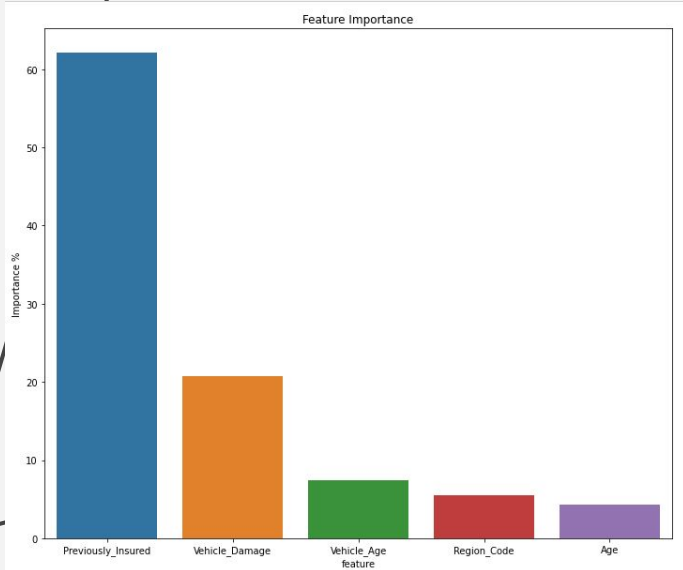

Feature Importance

Feature Influence

We use a mathematical method known as Shapely Values that allow us to determine what impact these features have on our model's ability.
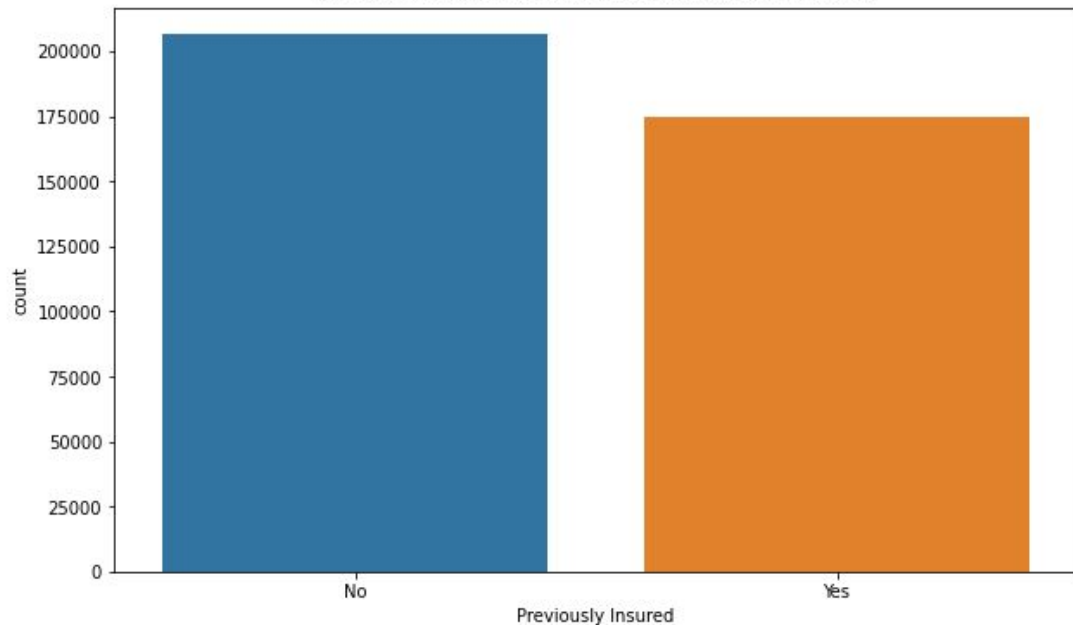
What Do They Mean

- Previously Insured is the largest single influencing factor on if people are interested in auto insurance.
- Previous vehicle damage is the second most influential factor.
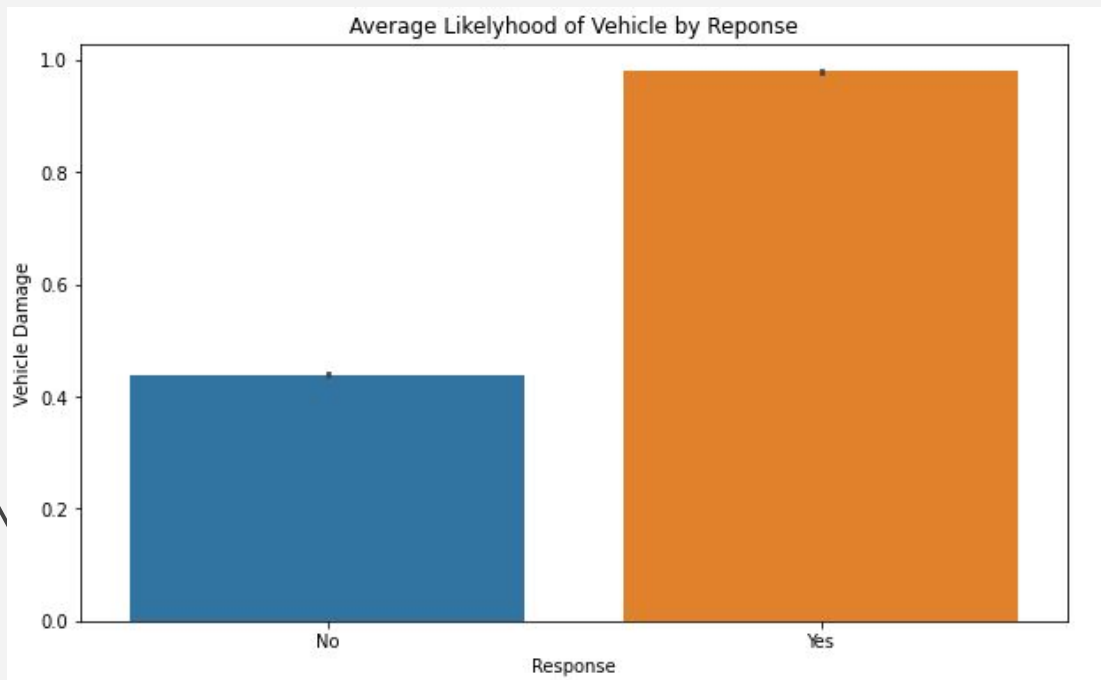- Age is third most important.
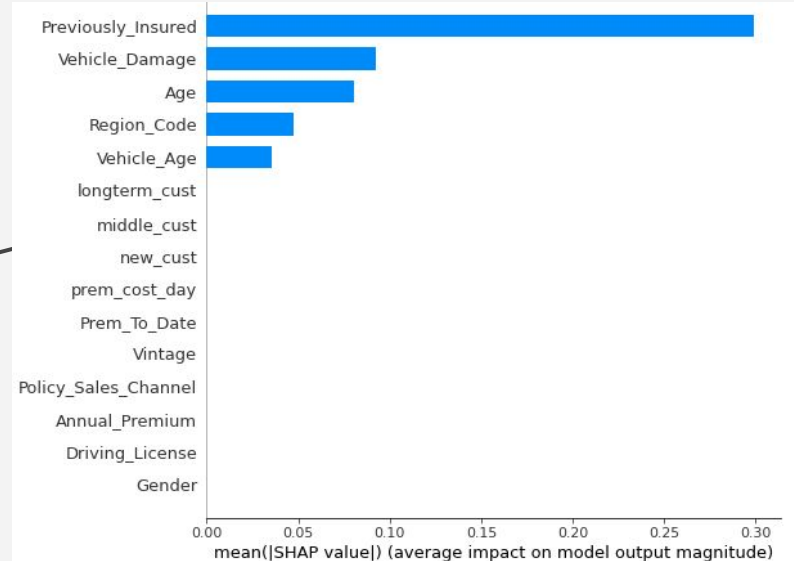
Previously Insured


Number of Customers Without vs With Auto Insurance

Previously Insured individuals are the minority, therefore it is a great demographic to target since it is also the most important feature of our model.

Vehicle Damage

Average Likelyhood of Vehicle by Reponse

Vehicle Damage we can see with this chart influences the interest response in individuals we approach with the cross-sell.

Age

Average Likelyhood of Cross-Sale by Age
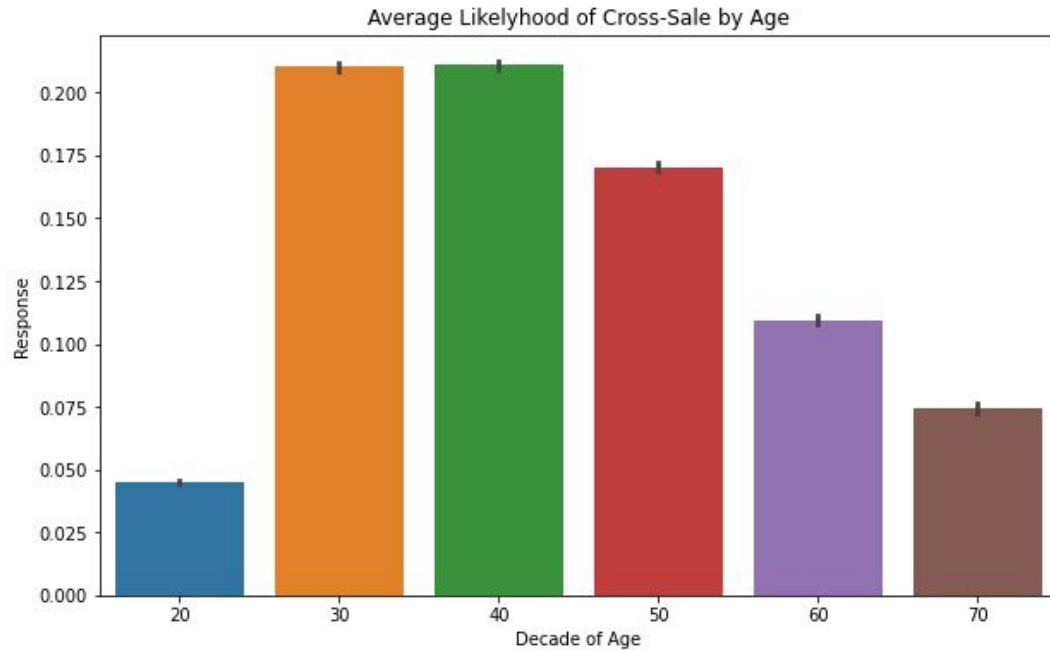
Age is an influencing factor on if an individual is interested in the cross sell or not. The very young and the elderly are generally not interested in the cross-sell.

Recommendations

- Leverage those individuals that lack auto insurance to drive cross-sells.

- Take a risk, insure those with previous vehicle damage.

- Target the middle aged, as the young and the elderly are less interested in a cross-sell.

Future Work:

Work with further models and tuning to boost the performance.

Derive more powerful features.

Try new methods of Machine Learning to get a market segmentation.

Thank You

For further questions contact me at:

doomedc4@gmail.com

Github:

https://github.com/emperorner0