Insurance Analytics – Executive Summary

Katayoon Sadeghi mehr – May 30th 2020

Back ground:

The insurance company website provides the potential customers an application to be downloaded and quote on different type of insurance. If the applicant is interested in the quote would request to be contacted by an agent from sales team and if the applicant finds everything appealing would purchase the insurance. The time of each action is recorded. After the sale is done, additional information about the customers are collected and the final premium for each type of insurance is determined.

Objective:

We are interested to know: The conversion rate of sale, the effect of each recorded time on sale, the features that affect the amount of premium.
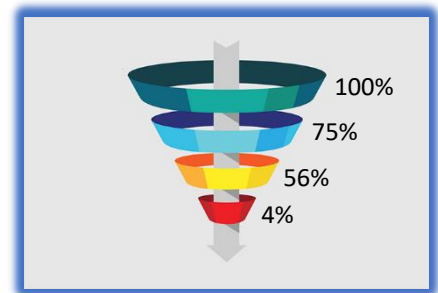
Findings:

74.88% got the quote after the app installation, 56.23% of total applicants/visitors were contacted by an agent, this is 75.10% of the applicants/visitors who got the quote were contacted by as agent. Similarly, 24.90% were not interested to be contacted. Sales Conversion rate is 4.34%.

The most installation, getting quote, being contacted and sale happens in summer time, specially in September. However, the conversion rate stays the same through the year. people use their free/ convenient time to look for insurance and normally if they start buying for insurance in Tuesday or Wednesday, they finalize it before weekend.

The most insurance is sold for Auto and the City with the highest number of customers is New York. But Los Angeles has the highest rate of the premium.

Auto and Home insurance have some outliers on upper side which belongs to luxury homes and cars. Life insurance have a few outliers on lower side. They can be those people who use some kind of subsidies or incentives from their work place.

Box plot of Primum based on Type of insurance and Gender

	Gender	F	M	Total
Accident	0	631	2015	2646
	1	429	100	529
	Total	1060	2115	3175
	%Accident	40.47	4.73	16.66
Ticket	0	652	1334	1986
	1	408	781	1189
	Total	1060	2115	3175
	%Ticket	38.49	36.93	37.45
Insurance Cancelation	0	982	1931	2913
	1	78	184	262
	Total	1060	2115	3175
	%ins_Cancelation	7.36	8.70	8.25
Premium_Mean		4036.52	3879.62	3932.00
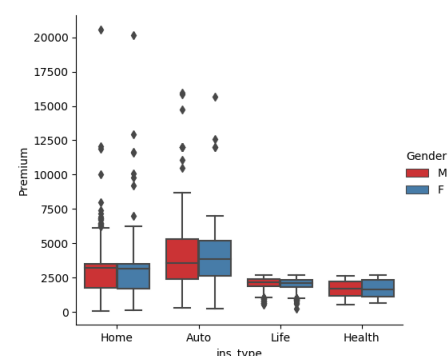
Although women have significantly more accident the rate of premium for Auto insurance is almost the same for both genders.

Premium for Auto is affected by the history of clients. If they have insurance cancelation, ticket or accident. But age, gender, city, time of installation, getting quote, being contacted by the agent or sale, does not affect the premium.

Using Backward Regression, only Inc_Cancelation, Ticket and Accident and constant stay in the model.

$$Y= 3305.85 + 350.99 * ins\_Cancelation + 1032.68 * Ticket + 182.97 * Accident.$$

R-square for train is 31.46% and R-square for test is 33.67%.

Applying Adaboost model shows that the model is the best at n=10. the accuracy is about 62%. Applying Random Forest model shows that the best model is a forest of n=1000 trees with accuracy of 75%. Applying SVR model shows that the best model is for {'C': 1, 'gamma': 1, 'kernel': 'poly'} with accuracy of 60%. Applying KNN model shows that the best model is for n=40. but the score is still very low (15%). And having too many classes in the model is not efficient. Applying SVR model shows that the best model is for {'C': 1, 'gamma': 1, 'kernel': 'poly'} with accuracy of 60%.

5 Models were fitted on the data, Linear Regression, Adaboost, Random Forest and SVR, KNN. Among these models, Random Forest with n=1000 is the best. The accuracy is about 75%.

Linear Regression	34%
Adaboost	64%
Random Forest	75%
SVR	60%
KNN	15%