QUALIFICATION OF THE PROBLEM

GOAL: TO PREDICT WHO WOULD BE INTERESTED IN BUYING A CARAVAN INSURANCE POLICY AND EXPLAIN WHY?

WHY IS THE ISSUE RELEVANT:

- It's important for every insurance company to understand their customers and be able to build a customer-centric marketing model

- If Insurance company could predict the behavior of their customers, they can optimize marketing strategies for maximal cost-efficiency

DATA:

- The raw data was downloaded from The Insurance Company Benchmark – KDD Cup

- 86 predictor variables describing 5822 customers in the training and 4000 customers in the testing set

METHOD:

- Classification problem – machine learning algorithms can be used to predict the probability of customer having a caravan insurance policy
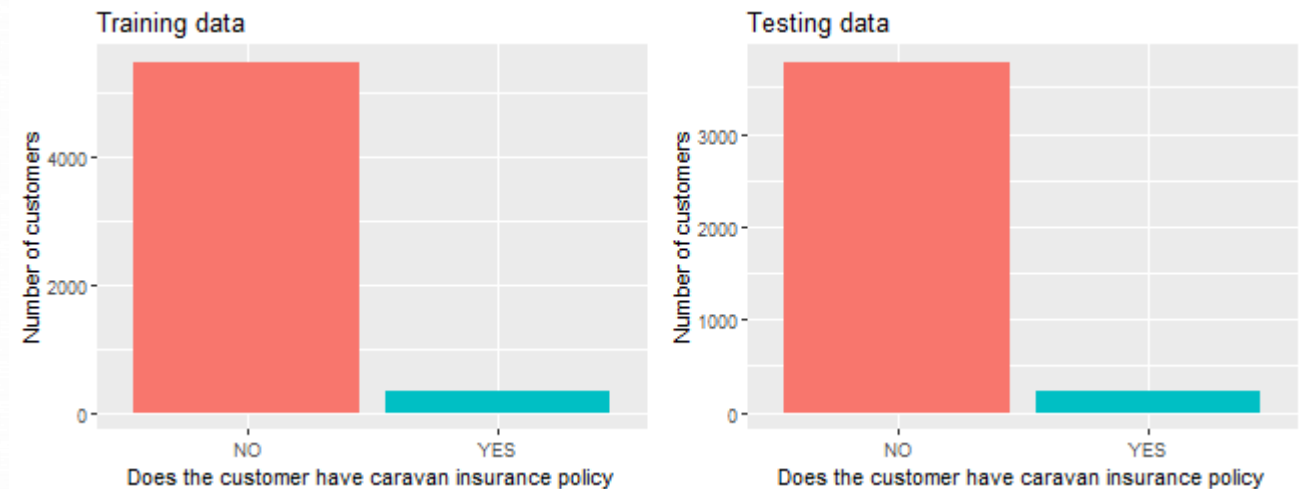
TRAINING DATA

Caravan insurance policy: **5474 NO   348 YES**

TESTING DATA

Caravan insurance policy: **3762 NO   238 YES**

ABOUT 6% OF CUSTOMERS HAVE CARAVAN INSURANCE POLICY

- a very valuable customer segment!

CONCLUSION:

- The proposed analysis is of utmost importance for any insurance company and company can gain major competitive advantage

- The data available should be sufficient to build a prediction model using classification machine learning algorithm

- The exploratory data analysis and data preprocessing are the next steps in building the model