Health Insurance Cross Sell Prediction

Team 03
(Bruce) Chang-Hung Hou | Chenli Qiu | Lequn Yu | Qiqi Tang | Yuchen Feng | Shamika Kalwe

Phase II

TO DO

- State the problem
- Tell us who cares about this problem and Why
- Describe you data – where it came from, what it contains
- **Present some interesting descriptive analyses (plots/tables) that inform the question your are answering**
- **Present your main results**
- **Which methods worked best for your particular problem?**
- **What were the challenges you faced? Tell us about the biggest challenge you faced and how you overcame it (or, not – that's fine too – not every problem has a solution.)**
- **Conclude – what have you learnt that can be put to practice?**
- **To submit: Slides and R markdown pdf**

Phase I

Problem Statement

To build a model to predict whether the existing health insurance customers will also be interested in Vehicle Insurance provided by the same company.

Cross Sell

Data Source

link

The kaggle link mentions **Analytics Vidhya** as its source for this dataset and problem. It also mentions relevant license for public sharing.

Structure of the Dataset (1/2)

Rows	**381,109**	**Columns**	**12**

No.	**Variable**	**Definition**
1	id	Unique ID for the customer
2	Gender	Gender of the customer
3	Age	Age of the customer
4	Driving_License	0 : No, 1 : Yes
5	Region_Code	Unique code for the region of the customer
6	Previously_Insured	0 : No, 1 : Yes
7	Vehicle_Age	Age of the Vehicle
8	Vehicle_Damage	0 : No, 1 : Yes (damaged in the past)
9	Annual_Premium	Health Insurance Premium per year
10	Policy*Sales*Channel	Anonymized Code for the channel of outreaching to the customer ie. Different Agents, Over Mail, Over Phone, In Person, etc.
11	Vintage	Number of Days, Customer has been associated with the company
12	**Response**	**0 : Not Interested, 1 : Interested**

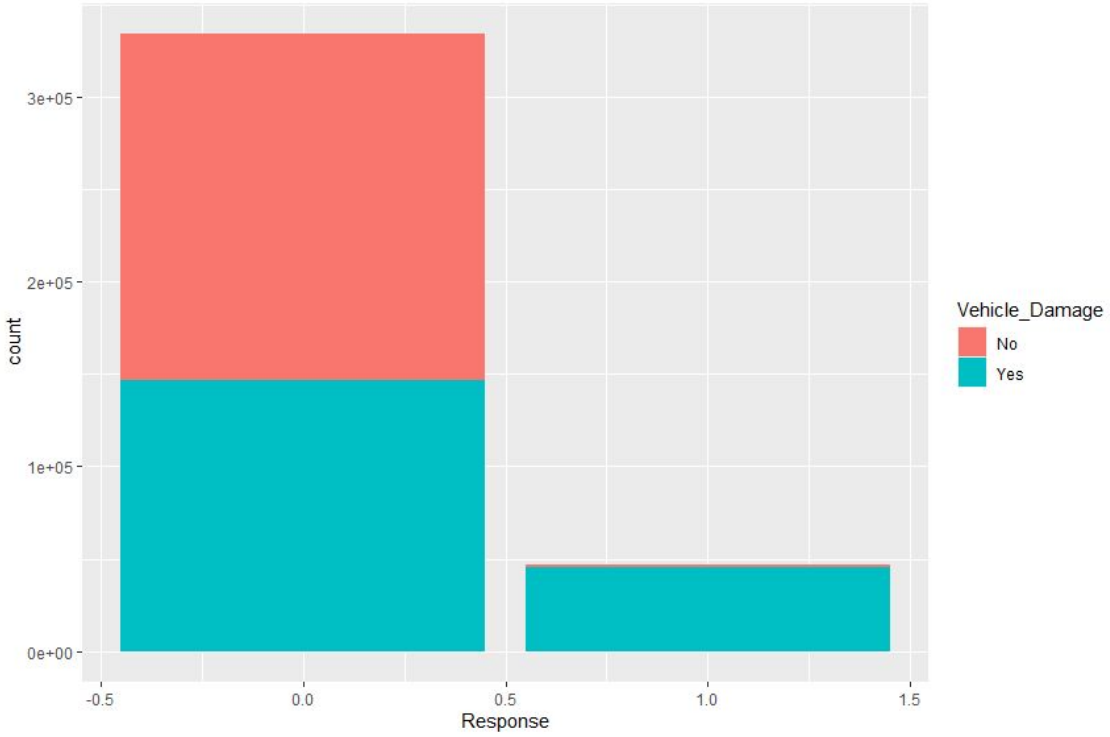
Target Variable

Structure of the Dataset (2/2)

```
Classes ‘data.table’ and 'data.frame':  381109 obs. of  12 variables:
 $ id                  : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Gender              : chr  "Male" "Male" "Male" "Male" ...
 $ Age                 : int  44 76 47 21 29 24 23 56 24 32 ...
 $ Driving_License     : int  1 1 1 1 1 1 1 1 1 1 ...
 $ Region_Code         : num  28 3 28 11 41 33 11 28 3 6 ...
 $ Previously_Insured  : int  0 0 0 1 1 0 0 0 1 1 ...
 $ Vehicle_Age         : chr  "> 2 Years" "1-2 Year" "> 2 Years" "< 1 Year" ...
 $ Vehicle_Damage      : chr  "Yes" "No" "Yes" "No" ...
 $ Annual_Premium      : num  40454 33536 38294 28619 27496 ...
 $ Policy_Sales_Channel: num  26 26 26 152 152 160 152 26 152 152 ...
 $ Vintage             : int  217 183 27 203 39 176 249 72 28 80 ...
 $ Response            : int  1 0 1 0 0 0 0 1 0 0 ...
 - attr(*, ".internal.selfref")=<externalptr>
```

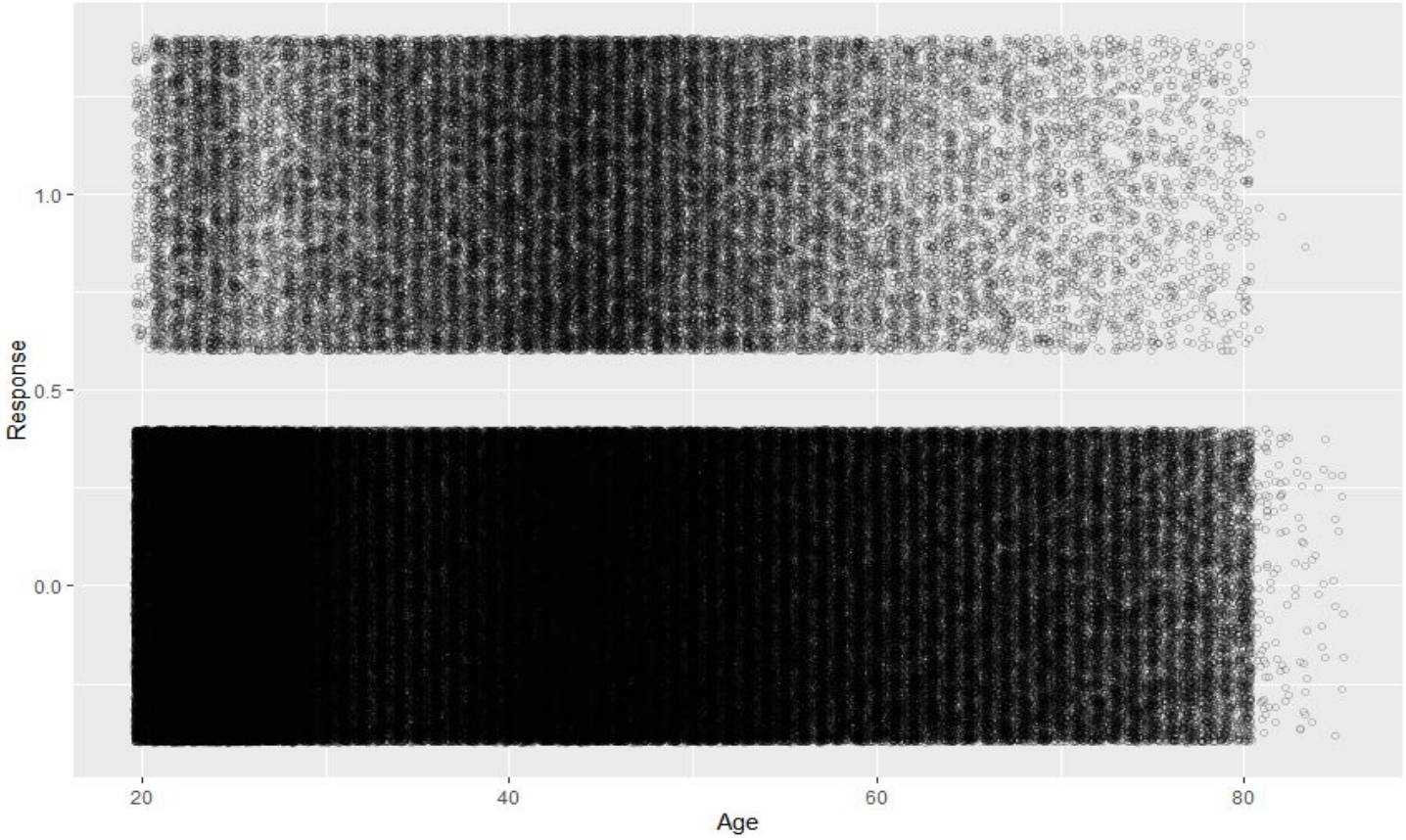
A couple of interesting figures (1/2)

Response v.s Vehicle_Damage



A couple of interesting figures (1/2)

Response v.s Age



Anticipated results

- Explore relationship between X variables and the target variable
- Prediction Model to predict possibility of Cross-sell. Thus predict whether a given Health Insurance customer will be interested in Vehicle Insurance or not

No.	Variable
1	id
2	Gender
3	Age
4	Driving_License
5	Region_Code
6	Previously_Insured
7	Vehicle_Age
8	Vehicle_Damage
9	Annual_Premium
10	Policy*Sales*Channel
11	Vintage
12	**Response**

Target Variable

Results and Implications

- Evaluate potential to cross-sell: selling vehicle insurance to healthcare insurance customers
- Estimate enrollment rate of vehicle insurance
- Define target audience: age, gender, etc.
- Learn about data-based decision-making in insurance sector

Thank you!

Any Questions?