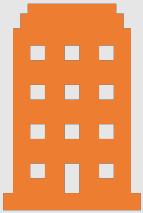# Safe Insurance

BEST CAR INSURANCE

Tan Zhi Yang (E0384308)
Lee Suan Ching (E0384993)
Norvin Chandra (E0384966)
Ci Rui Xiang (E0384938)

About Safe Insurance PLC

Safe Insurance PLC is a general insurance company in USA

Customers: personal and corporate customers

Sales channel:

Owned channel: Branch, Call Centre, Web
Intermediary channel: Agent

Background

Auto insurance in USA is facing stiff competition due to customers attrition, competitive premium as well as high accident and claim rate.

Safe Insurance has 4 auto insurance products which will be offered to existing policyholders on renewal.

External Data Analytics consultants have been enlisted to assist in understanding the portfolio segment

*targeting the right customer mix
*improving renewal success rate

Business Problem

Evaluate effectiveness of current sales channels for Auto Insurance

Identify Auto Insurance policyholders that bring value to Safe Insurance and retain these customers. Also, to classify newly joined customers to the different clusters

Improve renewal take up rate for Auto Insurance

Data Available

Data made available to Analytics team is for Auto Insurance expiring in Jan and Feb 2011 together with the response to the renewal offer

The data contains information from over 9000 customers from 5 states in USA

This mini project will attempt a prototype to showcase what can potentially be done better with bigger dataset and get management to commit resources into a full scale project

We do not have product information and the difference between respective renewal offers, therefore detailed analysis into this aspect is currently not possible

Data Exploration

Data Exploration – Overview

Data collected from 9134 customers and include 24 variables. No missing or duplicate data

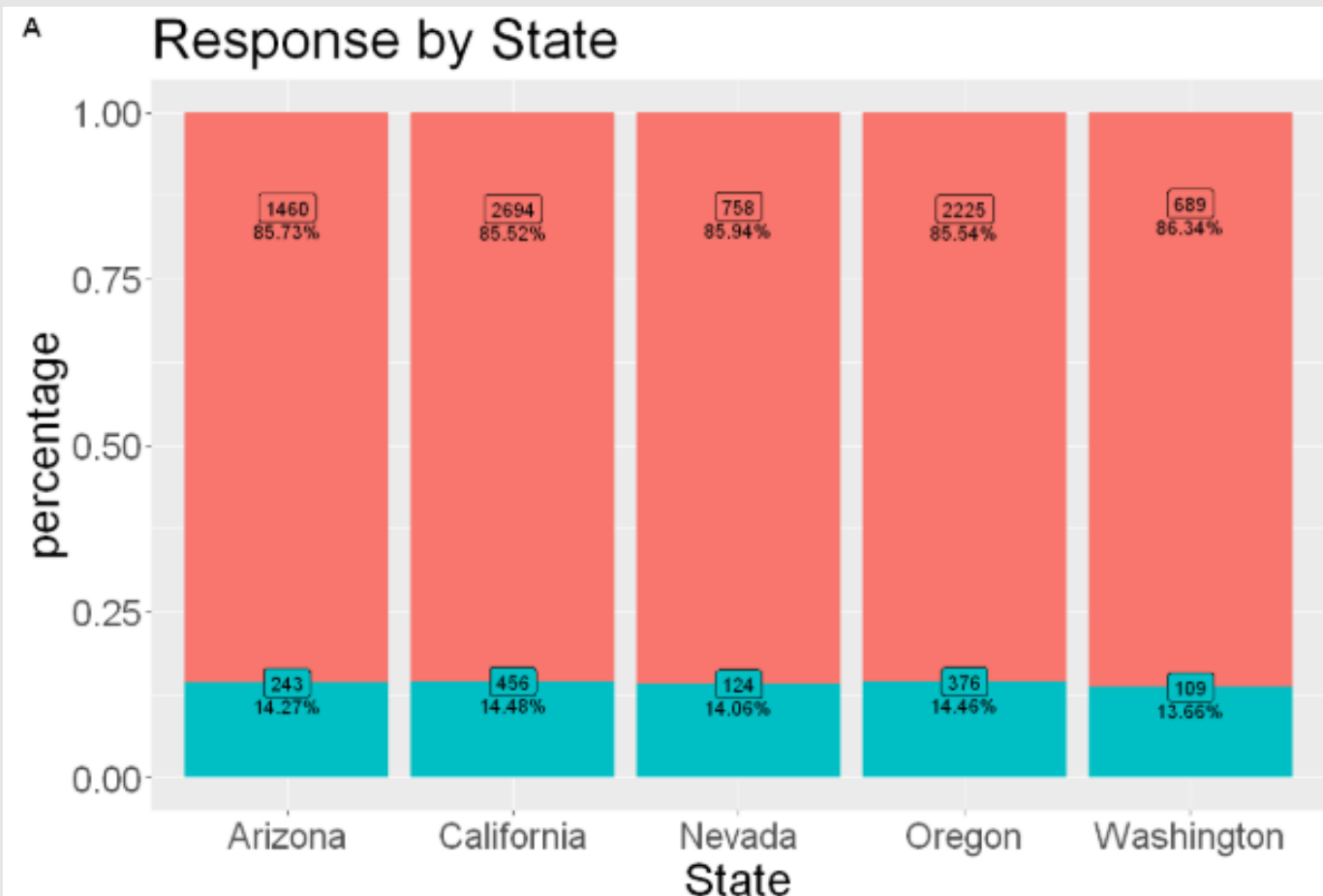Two major dependent variables – CLTV (pre-calculated) and Response to renewal

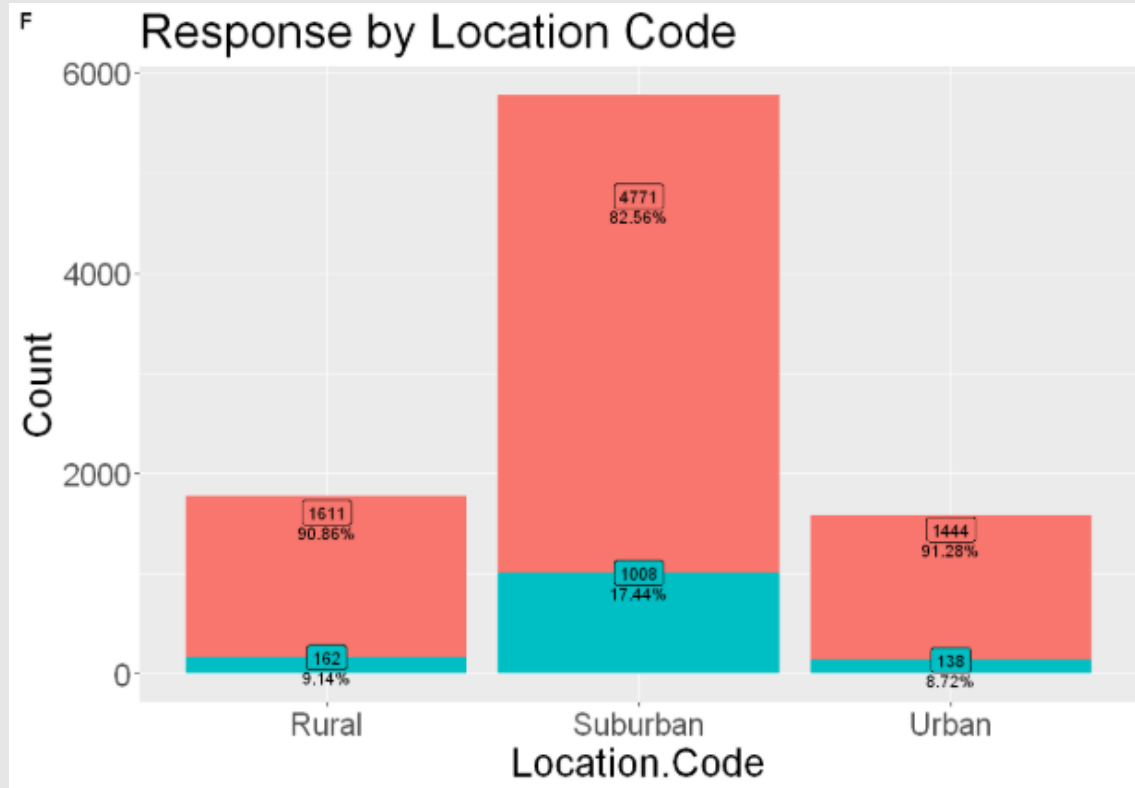Customer demographic information: Gender, Education, Employment, etc

Customer transactional data: Premium, Policy type, Claim amount, etc
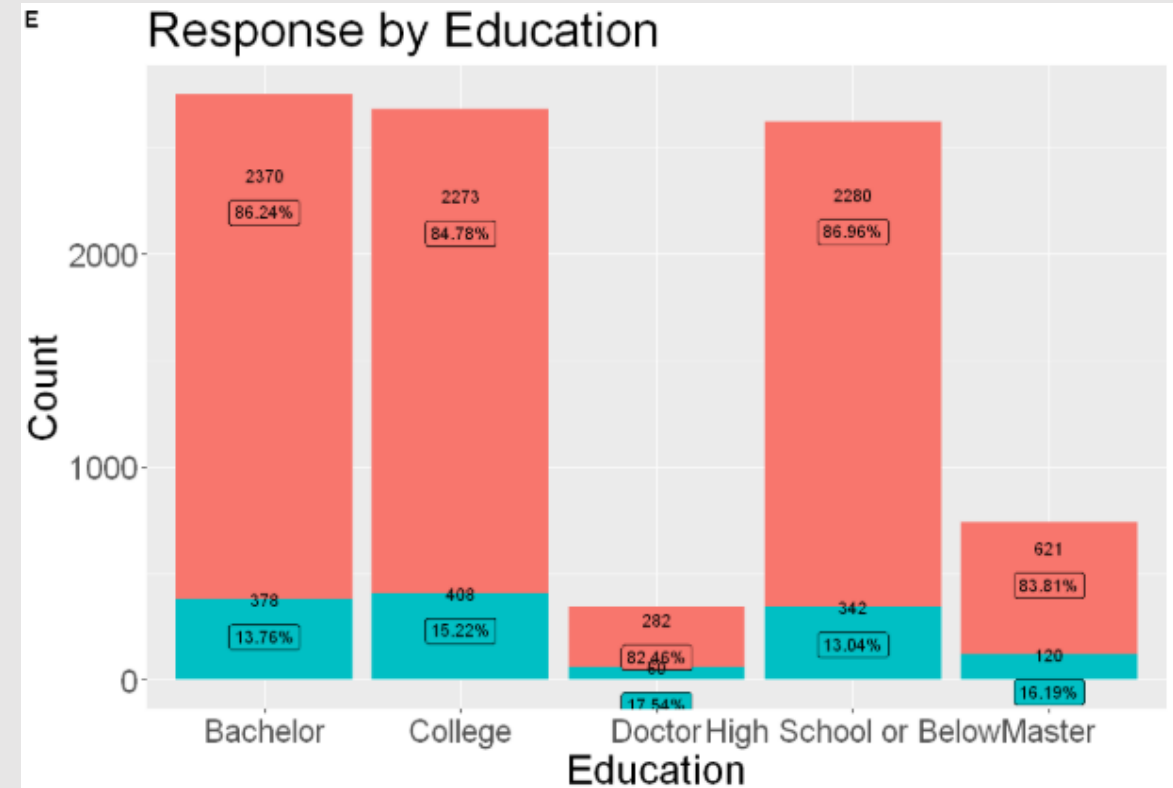
Data Exploration – Response (Part 1)

There is no difference in proportion of positive vs negative renewal response rate across 5 states
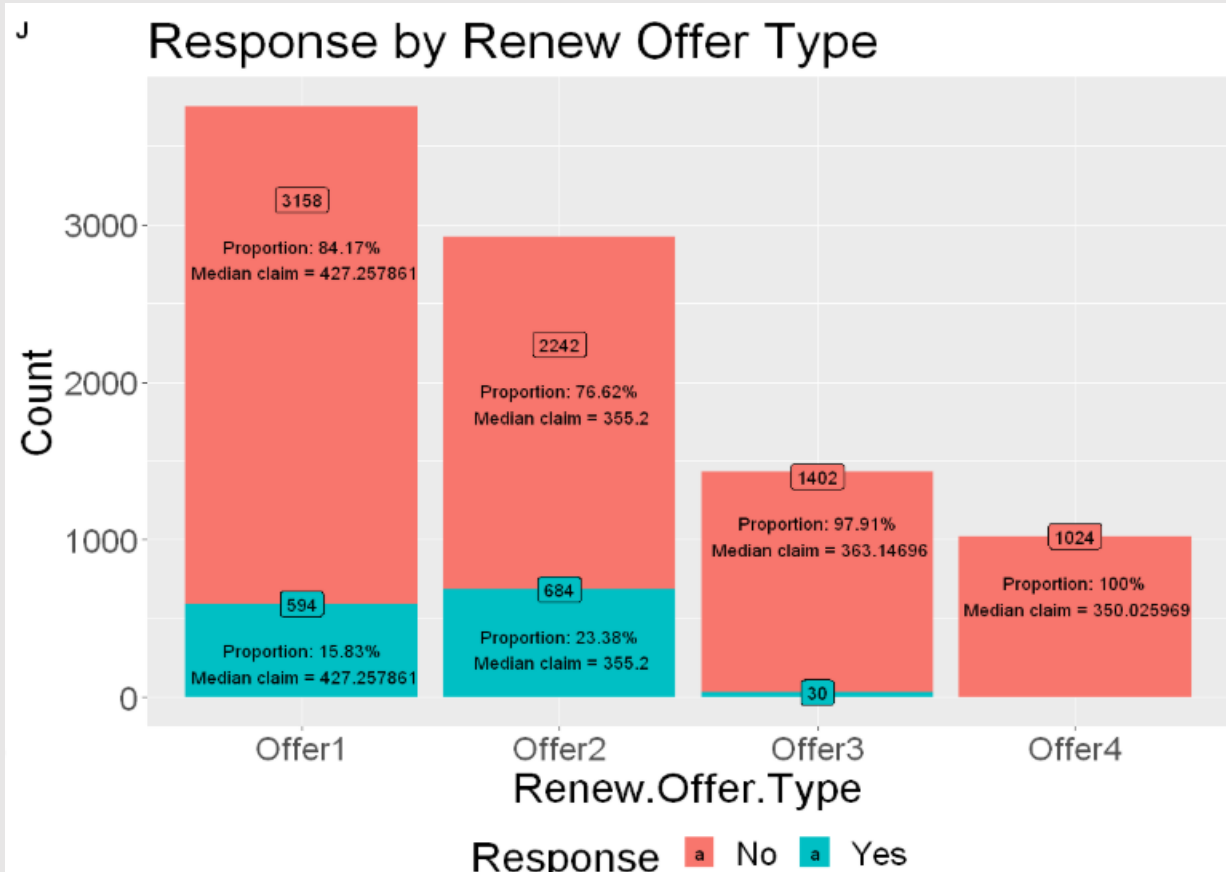
Data Exploration – Response (Part II)

Response by Location Code

Suburban policyholders tend to respond better to renewal offers

Response by Education

Higher educated policyholders tend to respond better to renewal offer

Data Exploration – Response (Part III)

J Response by Renew Offer Type

Offer1: 3158 — Proportion: 84.17%, Median claim = 427.257861; 594 — Proportion: 15.83%, Median claim = 427.257861

Offer2: 2242 — Proportion: 76.62%, Median claim = 355.2; 684 — Proportion: 23.38%, Median claim = 355.2

Offer3: 1402 — Proportion: 97.91%, Median claim = 363.14696; 30

Offer4: 1024 — Proportion: 100%, Median claim = 350.025969

Response: No / Yes

M Response by Sales Channel

Agent: 2811 / 80.85; 666 / 19.15
Branch: 2273 / 88.55; 294 / 11.45
Call Center: 1573 / 89.12; 192 / 10.88
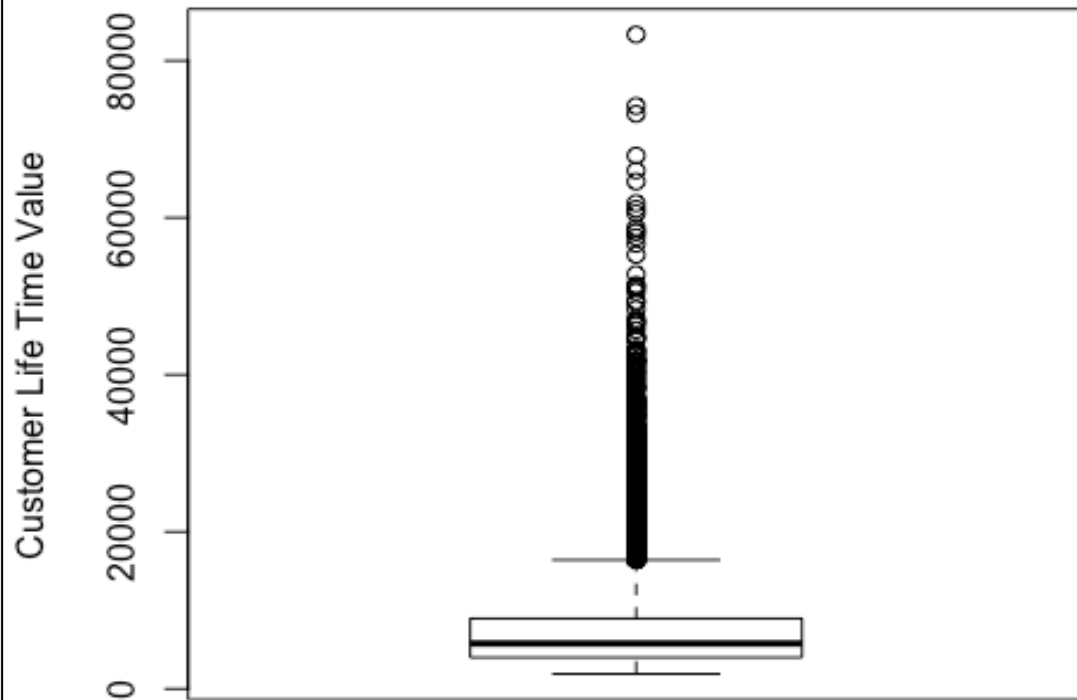Web: 1169 / 88.23; 156 / 11.77

Response: No / Yes

Offer 3 has very low take up rate and the take up rate for offer 4 is 0%
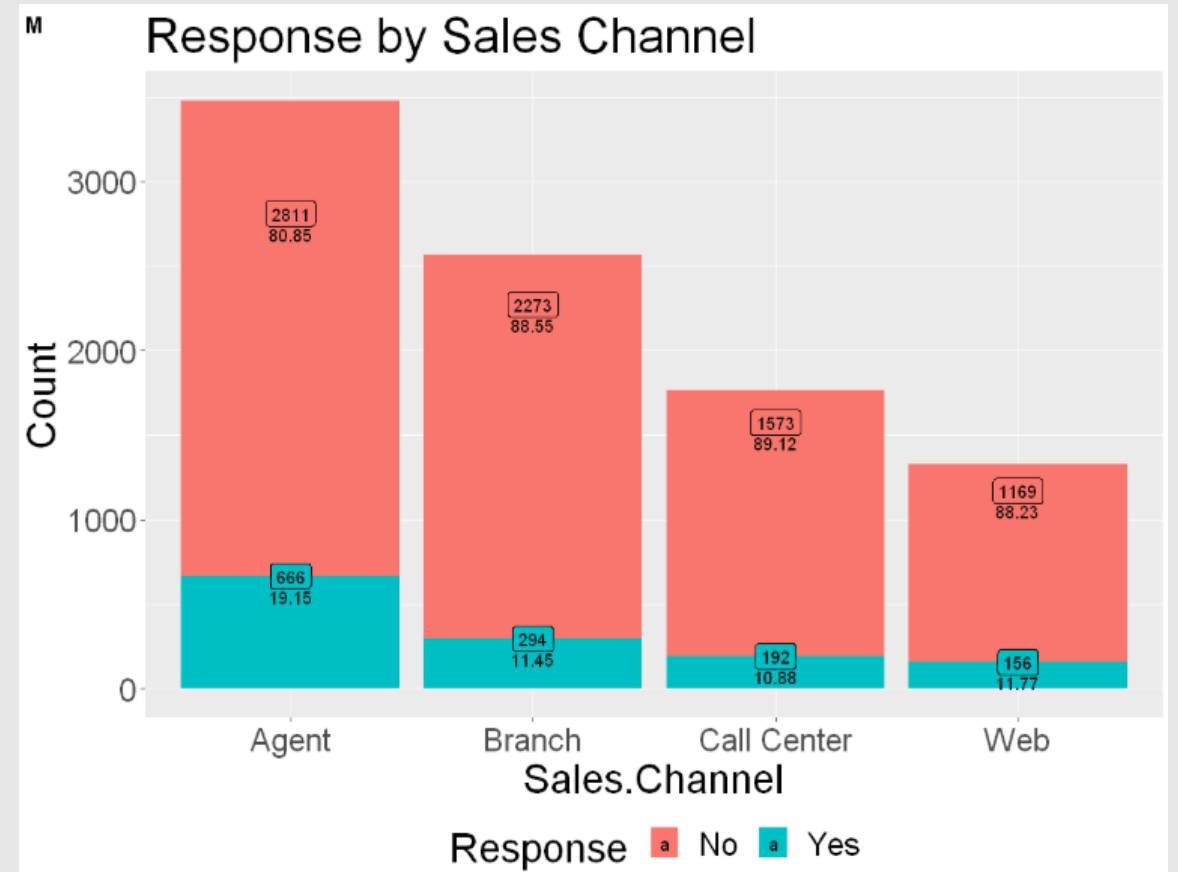
For sales channel, Agent has the highest percentage of response while web has the lowest percentage

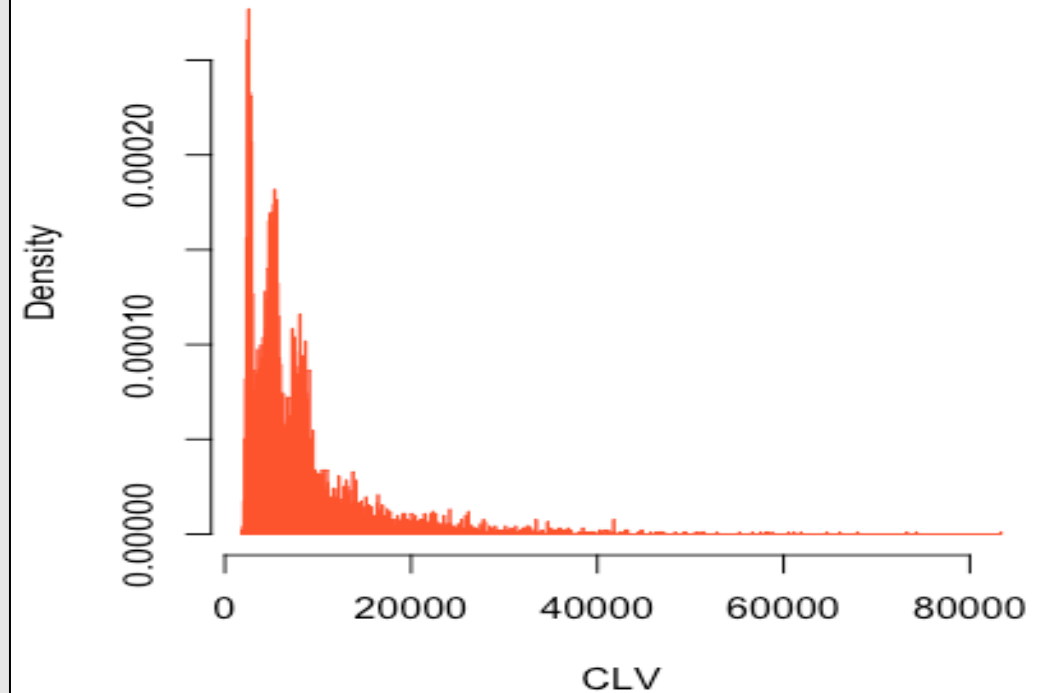Boxplot Distribution of Customer Life Time Value
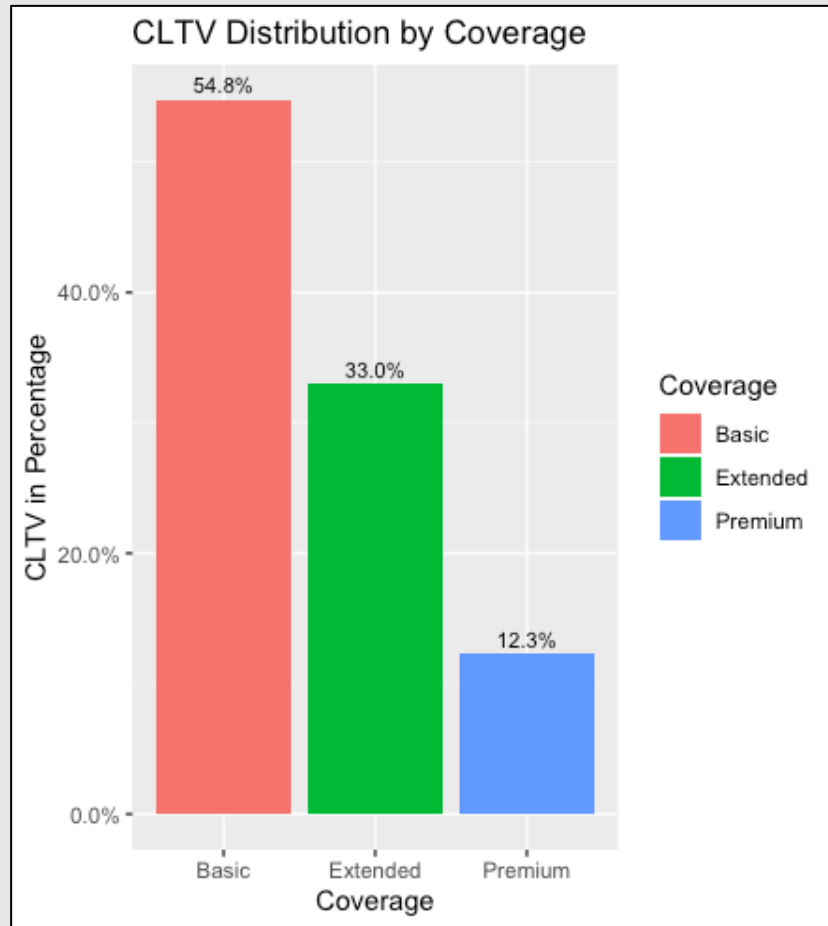

CLV Histogram

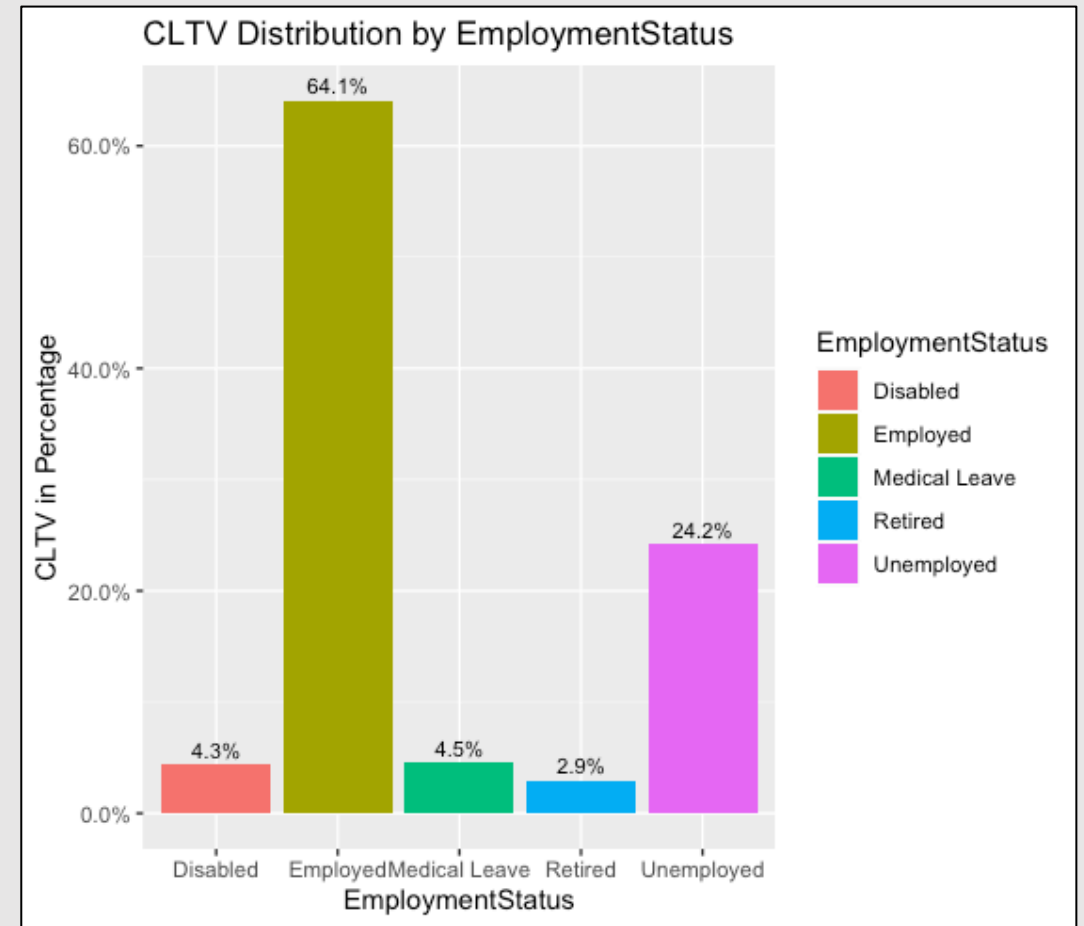Range: 1898   Mean: 8010   SD: 6870

CLTV is heavy-tailed, highly skewed to the left and not normally distributed

Data Exploration – CLTV (Part II)

The total CLTV distribution by coverage is highest for customers subscribed to basic coverage

The total CLTV distribution by employment status is highest for customers who are employed

Data Exploration – CLTV (Part III)

CLTV Distribution by SalesChannel
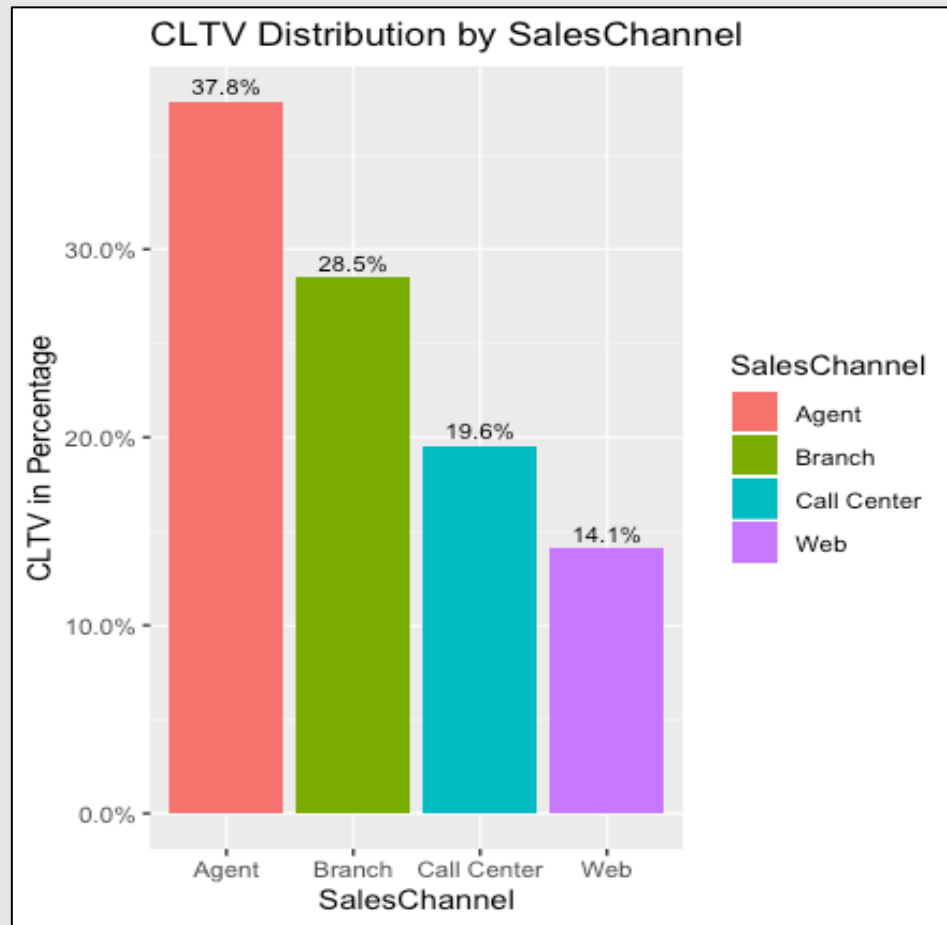
The total CLTV distribution by Sales Channel is highest for customers that respond to agent

CLTV Distribution by Gender
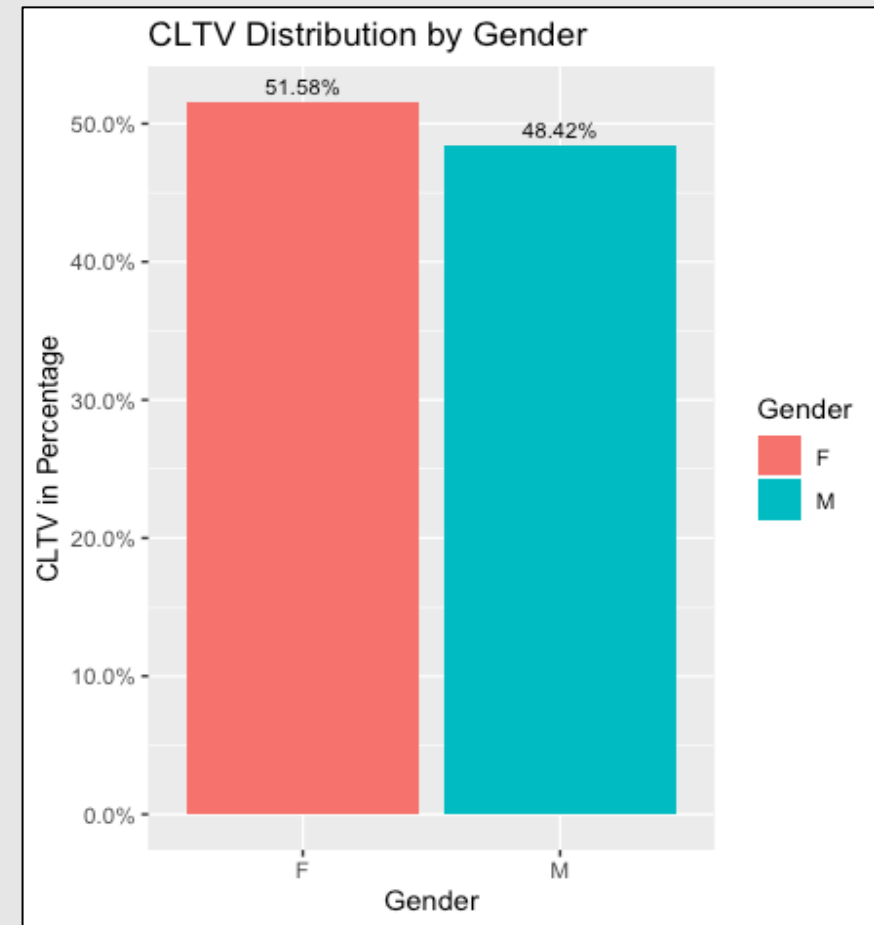
The total CLTV distribution is about the same for both genders, with males being marginally higher

Business Problem 1

Evaluate Effectiveness of Sales Channel

BUSINESS PROBLEMS

Business Problem 1 – Approach

BUSINESS PROBLEMS

Using existing data, plot bar chart of number of customers, CLTV value and response rate, with respect to Sales Channel

Analyse which sales channel results in the highest response rate as well as the highest CLTV

Sales channel with highest response rate/CLTV would be the most effective. Safe Insurance should invest more in that channel

Business Problem 1 – Analysis and Results

BUSINESS PROBLEMS

Approach: Using existing data, plot bar chart of number of customers, CLTV value and response rate, with respect to Sales Channel
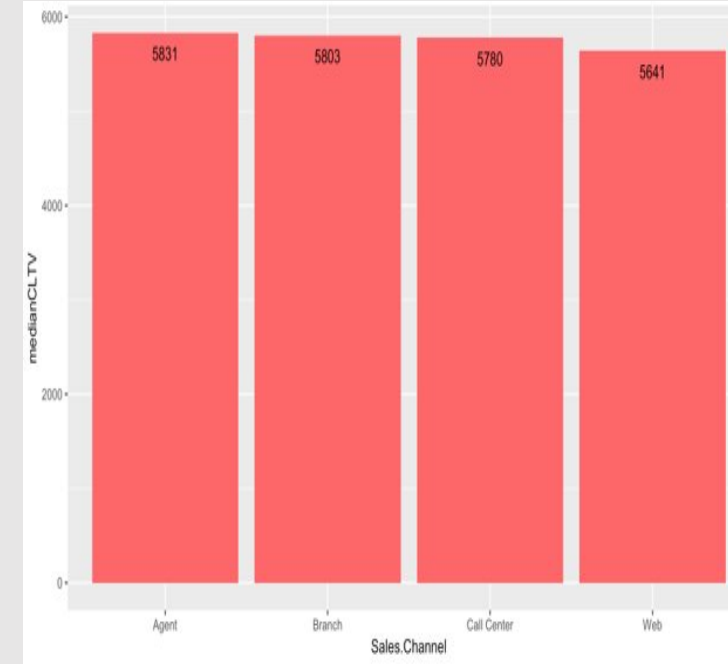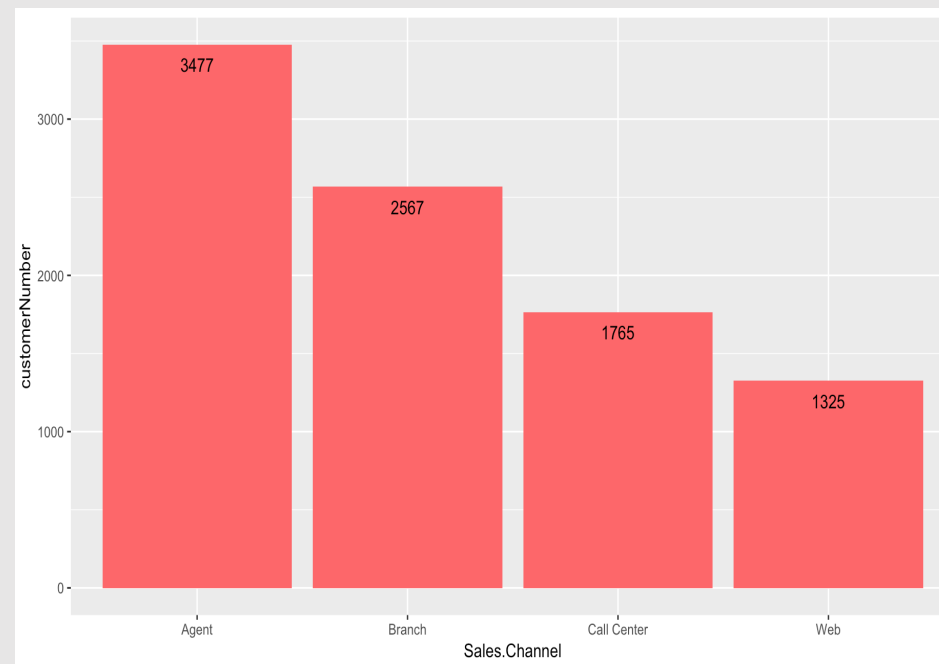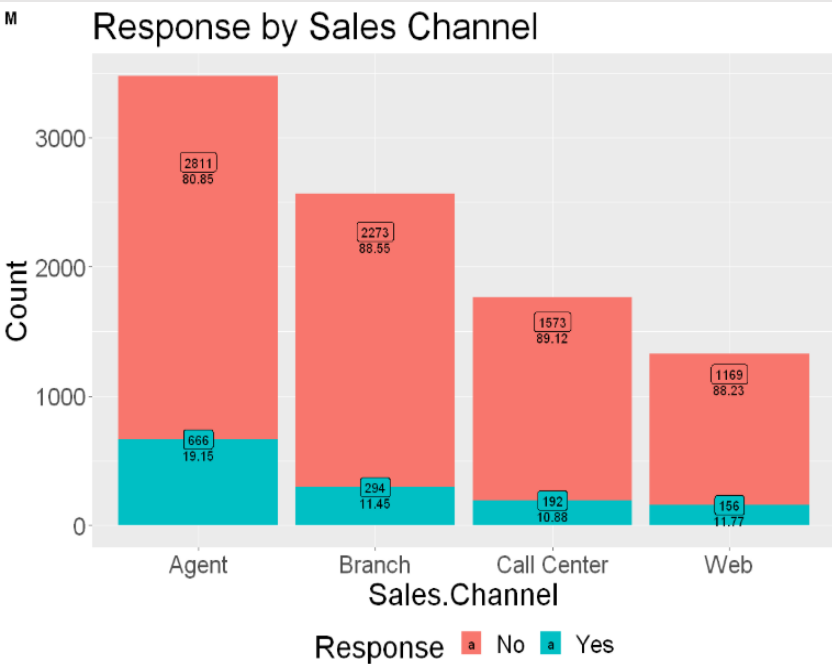
Response by Sales Channel

Analysis: 1) Agent is the most effective sales channel
2) Web is the least effective sales channel

BUSINESS PROBLEMS

- Safe Insurance can consider recruiting more sales agents to increase revenue since agent is the most effective channel

- As web is the least effect channel in terms of customer acquisition, it might be due to the unattractive design or user interface. Safe Insurance can look into re-designing their web page

- Call centre is also less effective than agent. Hence, if Safe Insurance were to retain the call centre, can re-access the skill set of call centre staff and look out for areas of improvement.
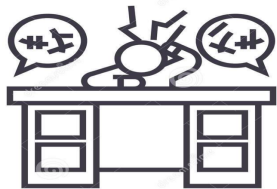
Business Problem 2

Identify valuable customers for customer retention

Business Problem 2 – Approach

To use clustering analysis for customer segmentation of existing customers.

o Determine the value of the customer based on two metrics
- Net CLTV(CLTV-Claims)
- Business Understanding of the Auto insurance Business

To use a decision tree model to classify new customers into the defined clusters. To utilize strategies to deepen customers relations

BUSINESS PROBLEMS
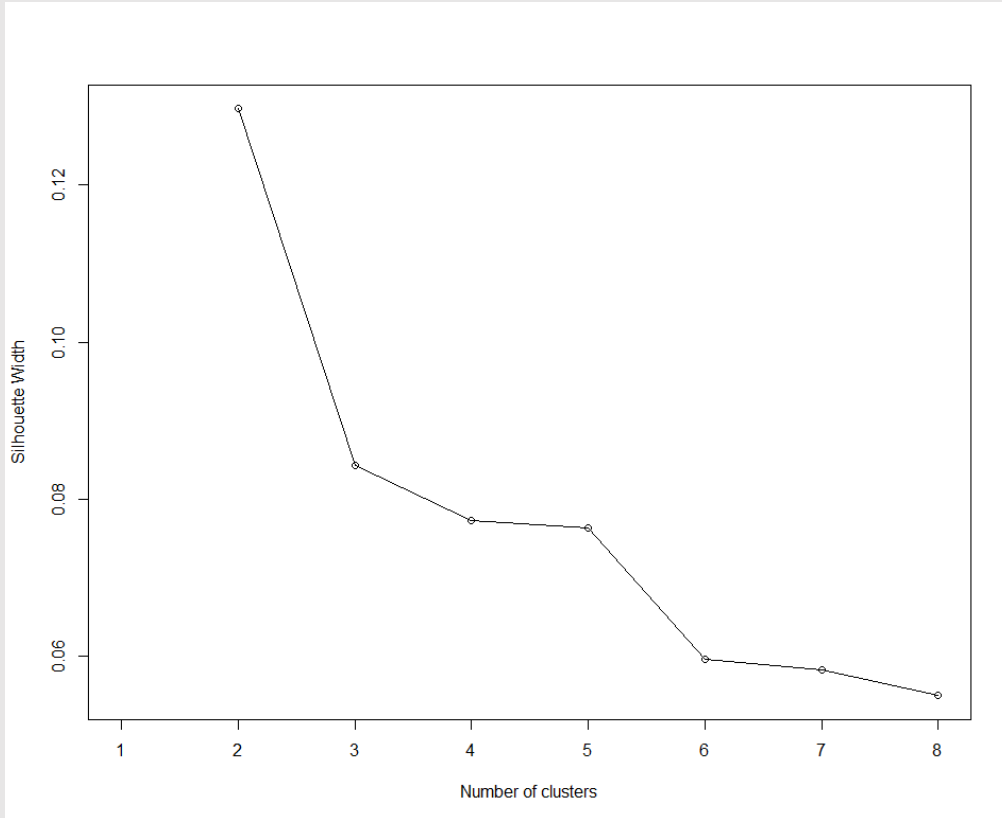
From Silhouette Width analysis:
- Number of clusters which provide largest difference = 2 clusters
- 3 clusters provided as it would provide more room for strategic decisions
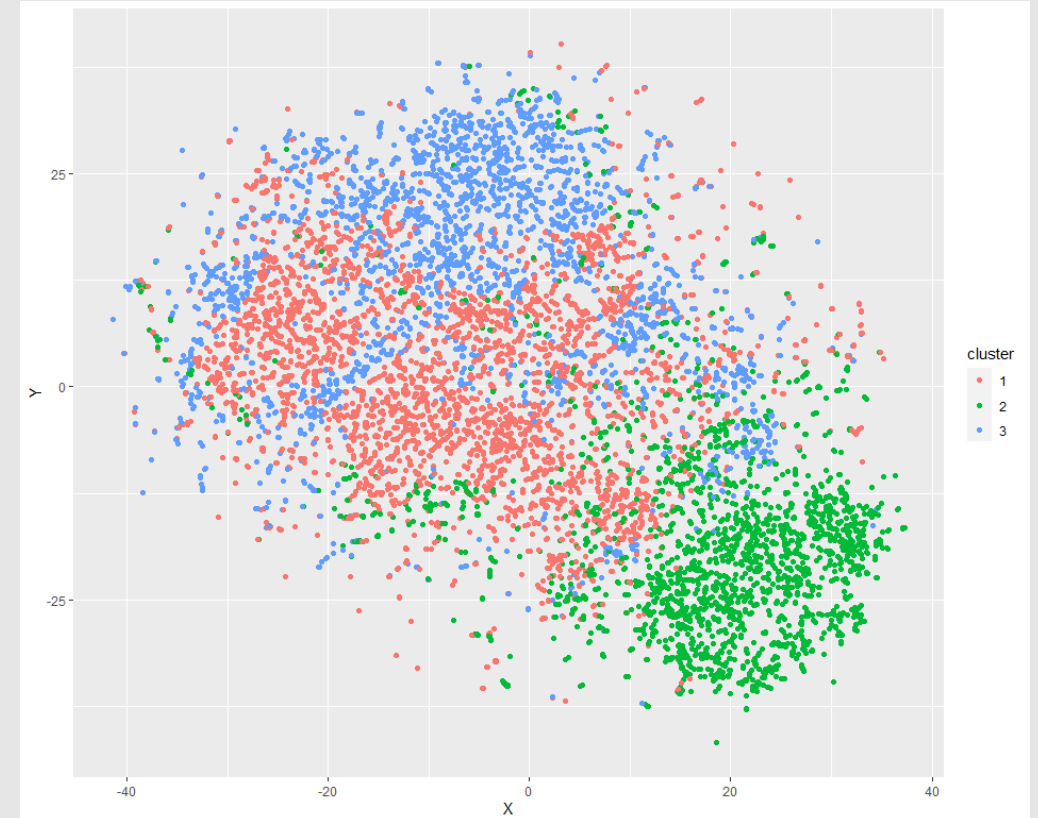
Results:
- Cluster 2 is very distinct from 1 and 3
- Clusters 1 and 3 are similar in terms of distance

Cluster 1	Cluster 2	Cluster 3
• Well educated	• Not highly educated	• Well Educated
• Employed	• Un-employed	• Employed
• Mostly Female	• Mostly Male	• Equal Male/Female
• Married	• Single	• Married
• High median income	• Lowest median income	• Highest median income
• Mostly from California	• Mostly from California	• Mostly from Oregon
• Mostly lives in suburban areas	• Mostly lives in suburban areas	• Living in Rural/Urban areas
• Most have Policy personal L3	• Most have Policy personal L3	• Most have Policy personal L2
• Most purchased from Branches or agents	• Most purchased from Branches or agents	• Most purchased from Branches or agents
	• Highest median claim amount	• Lowest median claim amount.

BUSINESS PROBLEMS

Cluster	Median CLTV	Median Total Claims	Median Net CLTV
1	5703	336	5362
2	5842	532	5217
3	5847	350	**5629**

- Based on the Median Net CLTV, it can be gathered that:
  - Cluster 3's customers are the most valuable to the company, followed by Cluster 1
  - Cluster 2 which has the lowest Median Net CLTV, which is perceive as less valuable to Safe Insurance
    - Most members of Cluster 2 are unemployed which may indicate an issue with paying premiums.
    - Cluster 2 has the highest premium claims among the 3 clusters

BUSINESS PROBLEMS

- For new customers, a decision tree classifier was created
  - To sort customers into their respective clusters at the point of sales based on the following Variables
    - ❖ Demographics of New Customer
    - ❖ Sales Channel
    - ❖ Type of Vehicle
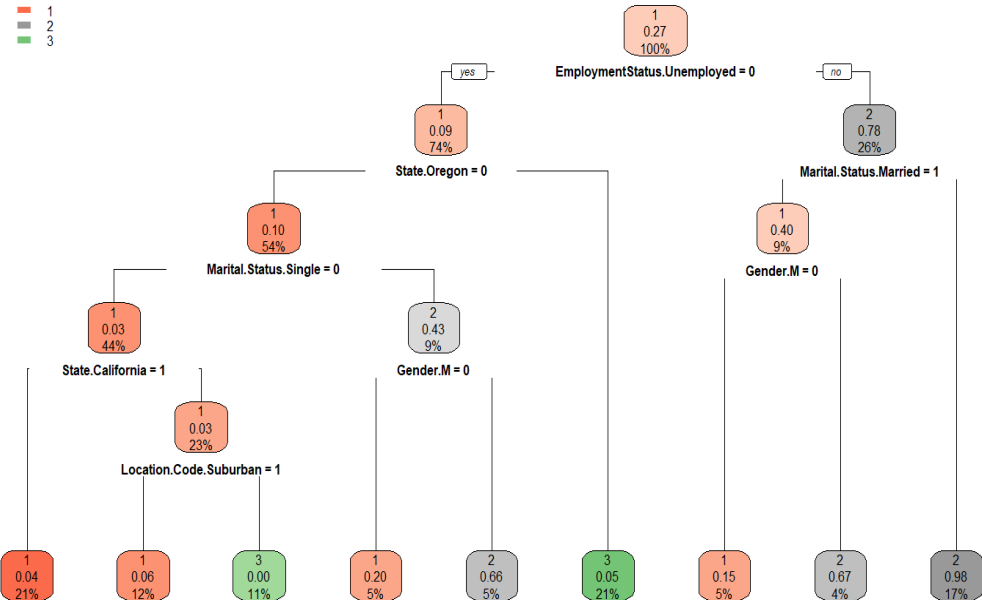
Cluster/Predicted Actual	1	2	3
1	863	56	238
2	80	615	31
3	207	30	620

- After a 70/30 training test split, 76% classification accuracy was achieved.

- The classification decision tree can be used for new customers as it uses variables which are only applicable to new customers.

Customised Programme for each cluster

Cluster 1 programme	Cluster 2 programme	Cluster 3 programme
• Offer Family Travel insurance at a special rate • Offer House insurance at special rate	• No special program as not considered a focus customer segment.	• Offer complementary insurance for technology gadgets. With added cover • Offer Travel insurance at a special rate • Offer House insurance at special rate

Having Customised Programme helps to build relationship with customer; Less likely to churn
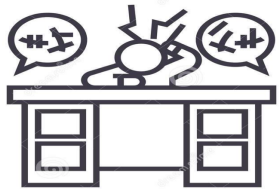
BUSINESS PROBLEMS

Business Problem 3

Low Renewal Retention Ratio

BUSINESS PROBLEMS

Renewed: **16%**

Total **Net.CLTV*** was lost due to non-renewal: **USD31,327,295**

Response	Value	Count	PortionValue	PortionCount
<chr>	<chr>	<chr>	<chr>	<chr>
0	31,372,925	4,142	0.84	0.85
1	5,908,896	756	0.16	0.15

- *Net.CLTV is derived from Customer.Lifetime.Value – Total.Claim.Amount
- Value of policyholder should also take into account the amount of claim incurred by them

BUSINESS PROBLEMS

Build a churn prediction model to identify policyholders likely to churn (not renewing their policy)

The model will generate probability of churning which we use to score policyholders.

The score combined with the list of policyholders and their information will be used to strategically target the policyholders with high likelihood to churn and attempt to salvage them via renewal campaign

BUSINESS PROBLEMS

Random Forest model tuned with grid search (rf3) works best in identifying policyholders who are not going to renew.

	Reference 0	**Reference 1**
Predicted 0	3676	504
Predicted 1	8	48

- **Implementation:**
  - Combine scored policyholder data with other profiles & Net. CLTV to filter and scope the target policyholders
    - > 60% probability of non renewal
    - Net.CLTV > = 75th percentile Net.CLTV of the renewal population

```
Confusion Matrix and Statistics

                Reference
Prediction    0     1
         0 3676   504
         1    8    48

               Accuracy : 0.8791
                 95% CI : (0.8689, 0.8888)
    No Information Rate : 0.8697
    P-Value [Acc > NIR] : 0.03469

                  Kappa : 0.1372

 Mcnemar's Test P-Value : < 0.000000000000002

            Sensitivity : 0.08696
            Specificity : 0.99783
         Pos Pred Value : 0.85714
         Neg Pred Value : 0.87943
             Prevalence : 0.13031
         Detection Rate : 0.01133
   Detection Prevalence : 0.01322
      Balanced Accuracy : 0.54239

       'Positive' Class : 1
```

BUSINESS PROBLEMS

Ideally we would like to have 11 months of observation period, 1 month of scoring period and 2 months of performance period as auto insurance policy is usually annual contract.

Feb-10	Mar-10	Apr-10	May-10	Jun-10	Jul-10	Aug-10	Sep-10	Oct-10	Nov-10	Dec-10	Jan-11	Feb-11	Mar-11
Observation / Data Collection											Scoring on 1 Jan 2011	Performance	

BUSINESS PROBLEMS

- For Feb 2011, there are more than 3,600 policies with Net.CLTV worth of more than USD28m not renewed

- To allow effective renewal follow up, we need to focus our effort on high Net.CLTV policyholders with high likelihood to lapse

- For a start, we select 60% as the score threshold and Net.CLTV value in 75th percentile of population Net.CLTV as threshold.

- Using threshold above, we narrow down the list to about 1000 policyholders – 28% of policyholder count in exchange for 59% of Net.CLTV worth USD16.7m

Net.CLTV at stake	Net.CLTV retainable	Net.CLTV ratio	No. of Policyholder at stake	No. of Policyholder retainable	Policy count ratio
<chr>	<chr>	<chr>	<chr>	<chr>	<chr>
28,088,537	16,693,916	0.594332	3,684	1,037	0.2814875

Conclusion

Recommendation

Function	Recommended Action	Time Frame	Metrics
Overall	• Improve data collection to train better Churn Prediction Model • Offer relationship strategies to deepen ties with valuable customers based on cluster. • Set up effective CRM System	• Immediate • Immediate • Medium – Long Term	Measure against current overall renewal ratio of 16%
Agent	• Utilize scored renewal list to run renewal campaign (measure success by renewal ratio) • Further strengthen the Agent performance through regular sales and product training	• Immediate • Medium Term	Measure against current renewal ratio of 19%
Call Center	• Utilize scored renewal list to run renewal campaign (measure success by renewal ratio) • Investigate reason for poor performance • Provide product and sales training • Establish renewal retention KPI and incentive to spur call center personnel	• Immediate • Immediate • Immediate • Medium Term	Measure against current renewal ratio of 11%

Recommendation

Function	Recommended Action	Time Frame	Metric
Web	• Utilize scored renewal list to run renewal campaign • UX and UI redesign • Search Engine Optimisation • Establish simple and fuss free policy purchase and renewal flow via website	• Immediate • Medium Term • Medium Term • Medium Term	Measure against current renewal ratio of 12%
Branch	• Utilize scored renewal list to run renewal campaign • Look into branch design and branding at branches • Train branch personnel on sales and product • Establish renewal retention KPI and incentive to spur branch personnel	• Immediate • Medium Term • Immediate • Medium Term	Measure against current renewal ratio of 11%

Recommendation

Channel	Renewal Notice (1 month before expiry)	Follow Up (2 weeks before expiry)	Reminder (1 week before expiry)	Early Renewal Promo on Target Policyholders
Agent	Mail	Call	SMS	Petrol Voucher
Branch	Mail	Call	SMS	Dashboard camera
Web	Email	Email	SMS	Dashboard camera
Call Center	Mail	Call	SMS	Dashboard camera

Limitations/Gaps

1. Dataset only spans across two months (Jan and Feb), model could be more accurate if a few years' worth of data is available (there could be seasonality and insurance policy is usually an annual contract)

2. Product information and Renewal offer details are not available, which makes it difficult to study the underlying reason for not renewing and hamper the effort to improve renewal strategy

3. No information on costing making it difficult to conduct costing analysis and ROI analysis for the different plans and sales channels.

Q&A