Overview

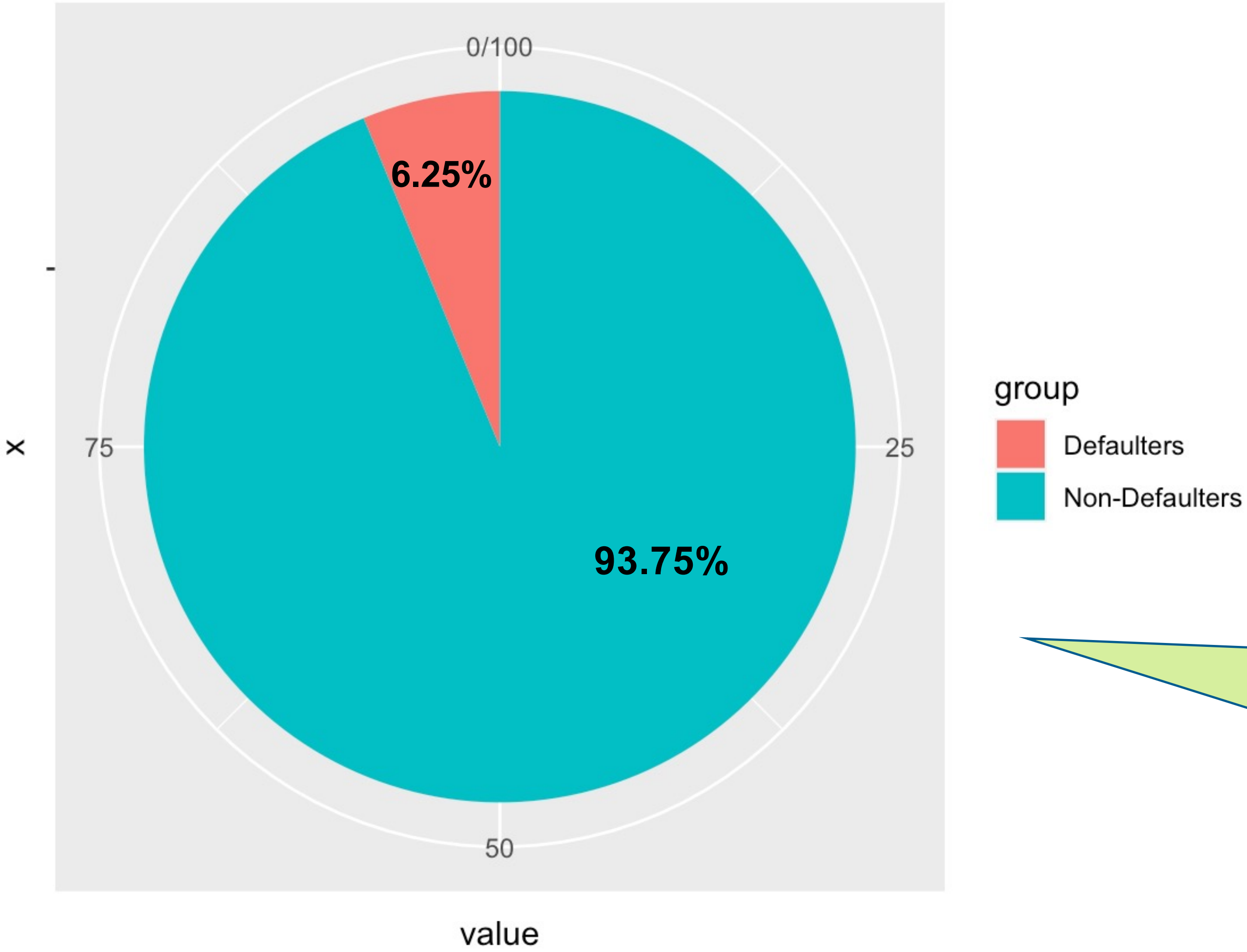
Premium paid by the customer is the major revenue source for insurance companies. Default in premium payments results in significant revenue losses and hence insurance companies would like to know upfront which type of customers would default premium payments.

Explore the data to identify customers with the propensity to default on the Premiums to be paid to the Insurance company.

Independent Variables to be explored

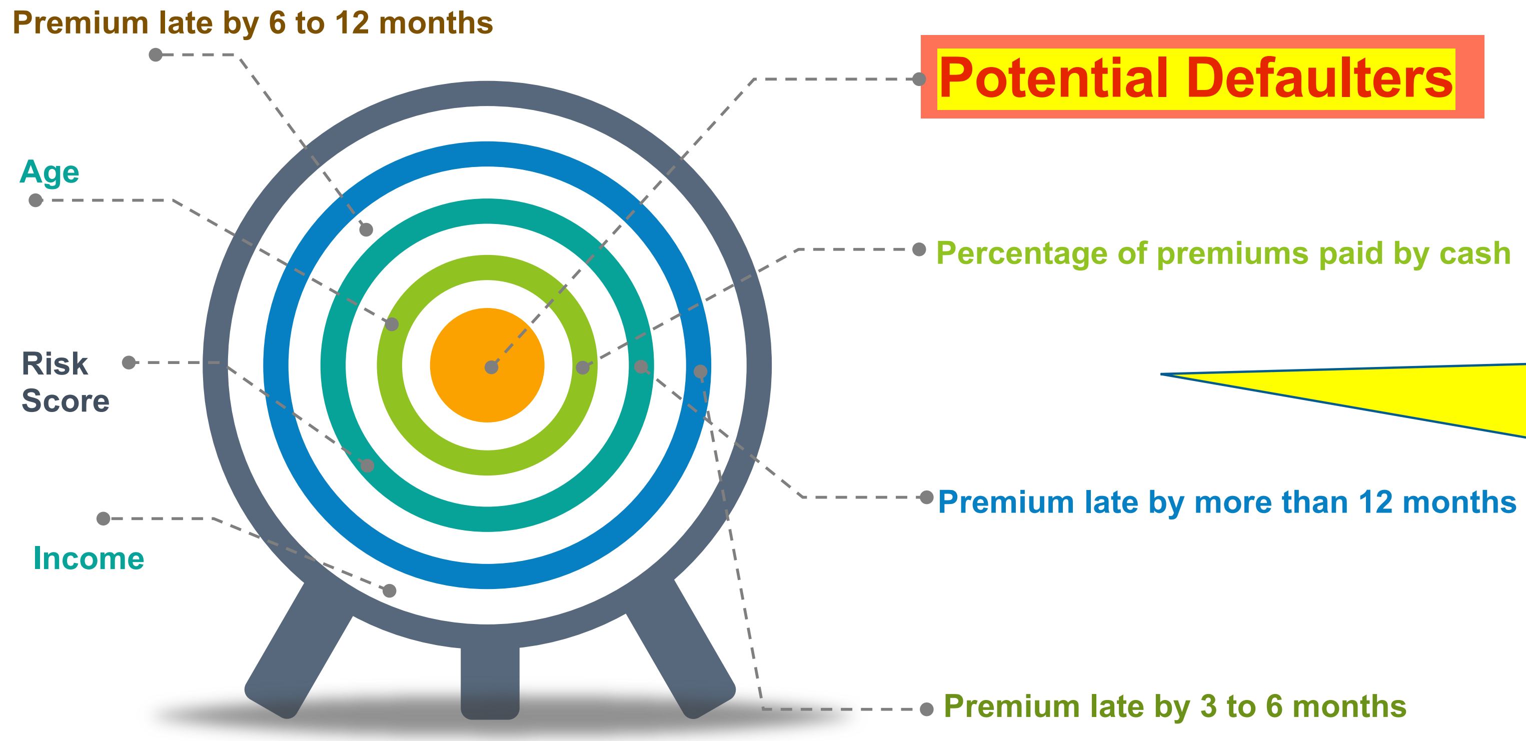
- Premium
- Marital Status
- Nos. of Dependents
- Vehicle owned
- Risk Score
- Income
- Premium paid Late by 6 to 12 months
- Sourcing Channel
- Num of Premium Paid
- Age
- Accommodation
- Premium Paid in Cash
- Premium paid Late by 3 to 6 months
- Premium paid Late by more than 12 months
- Residence Area

Exploring the unbalanced Data



An unbalanced data is a challenge to explore and filter the information/cohorts especially if we have 16 variable to dig into

Exploring various attributes to identify the potential defaulters among the cohorts



Exploring Categorical & Numerical Data to discover patterns, similarities, correlations, user behaviors, etc. to identify the potential defaulters

Solution Design

The journey of exploring the data for identifying the potential defaulters

1. Clean & Filter Data

Data Treatment

- *Missing Value*
- *Outlier Treatment*
- *Variable Transformation*
- *Adding/removing variables*
- *Column Names*

2. Uni Variant Analysis

EDA Uni Variant

The categorical and numerical variables explored to check on aspects like the mean, median, mode, outliers and other influencing aspect of each variable individually.

3. Bi Variant Analysis

EDA Bi Variant

see how the dependent variable (Defaulter) is behaving amidst all the categorical and numerical variables.

4. Correlation Analysis

Correlation

understand the degree of correlation and if they would be any high correlations which we need to consider for our analysis.

5. Building Analytical Models

Building Basic Models, Bagging & Boosting

- *CART model*
- *Logistic Regression Model*
- *Naïve Bayes*
- *KNN*
- *Random Forest*
- *Gradient Boosting*
- *XG Boost*

6. Finalize/ Select Model

Best Model for Analysis

Model performing best of aspects like:

- *Confusion Matrix*
- *Accuracy*
- *Specificity*
- *Sensitivity*
- *Precision*
- *AUC*
- *Gini*
- *Concordance*

7. Identify Important variables

Important Variables

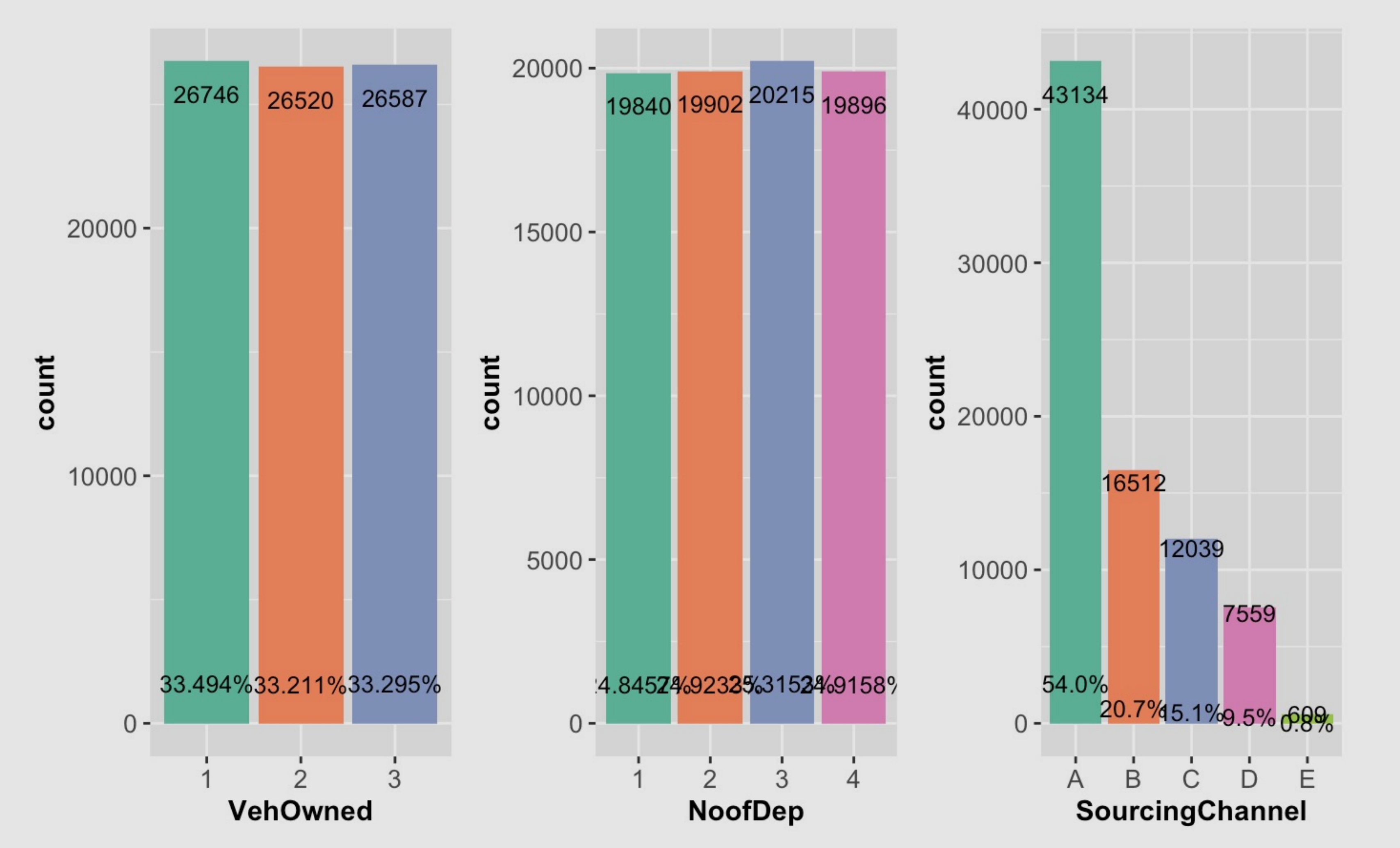
Discover the important influencing variables which indicate pattern, user behavior, tendency, characteristics and introspect to come out with solutions & recommendations

8. Insights & Recommendations

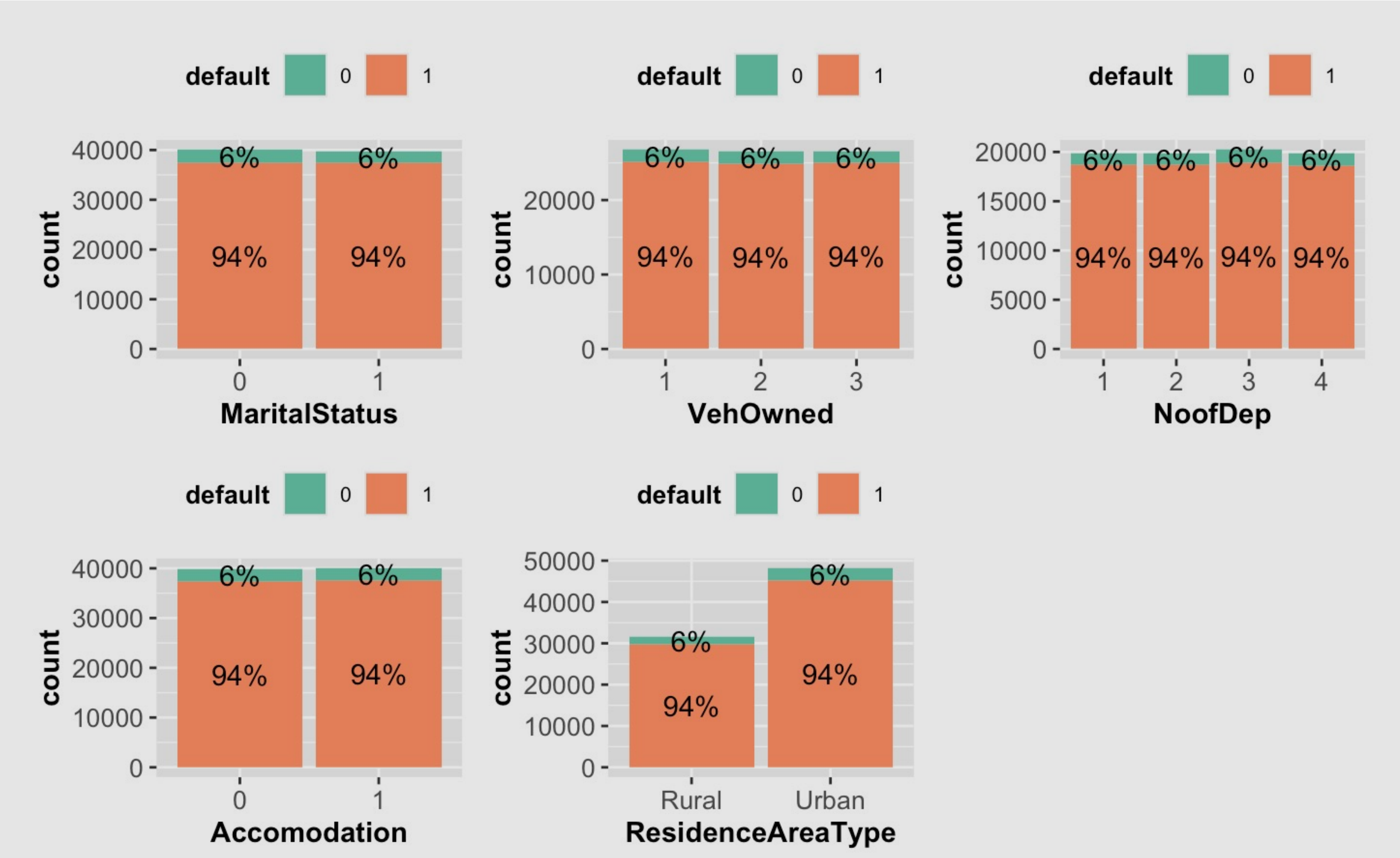
Findings & Solutions

- *Long Term – Short term*
- *One Fit or tailored solution model*
- *Conclusion*
- *Recommendations with some real-world example*

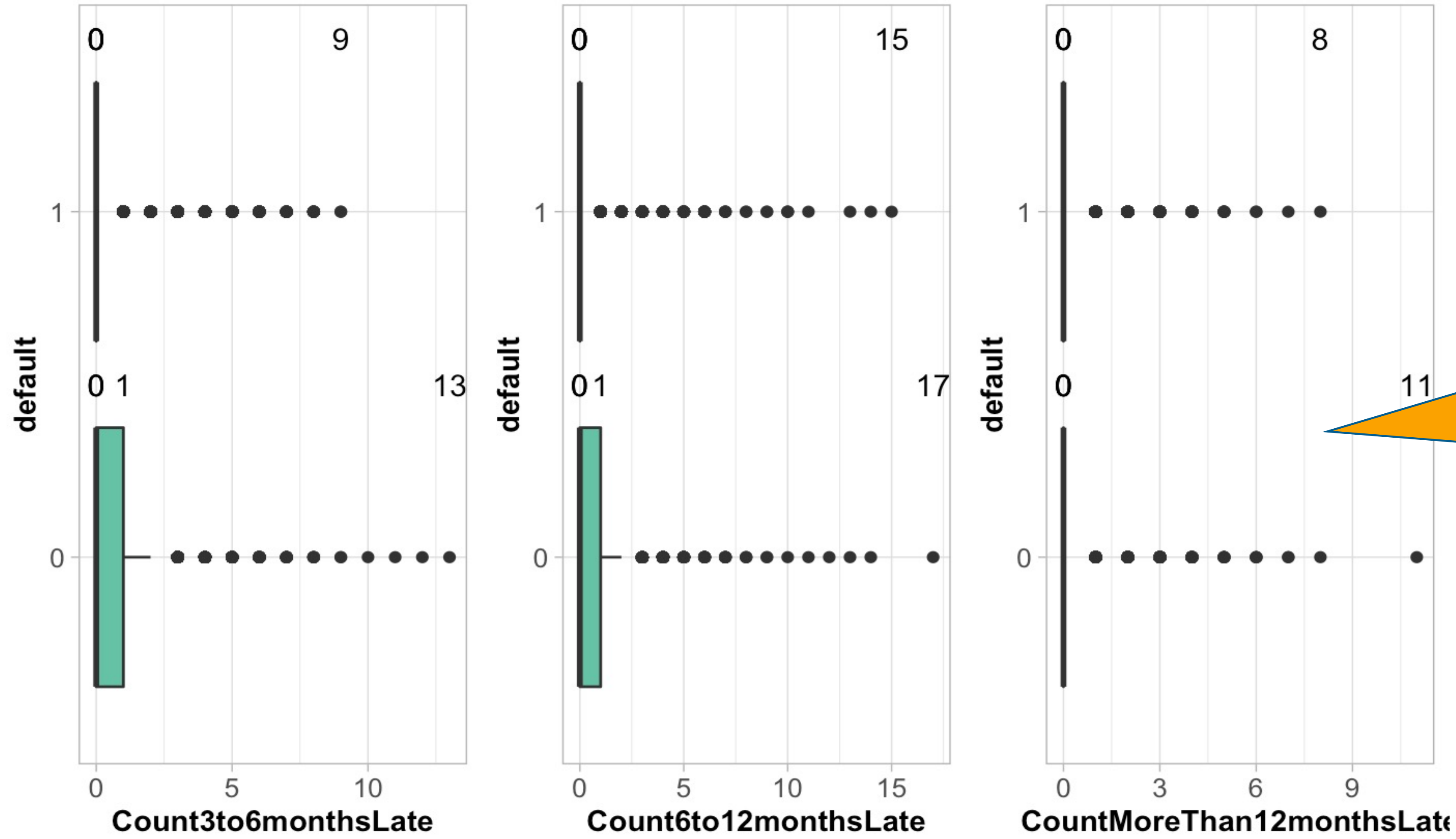
Uni Variant Analysis of Categorical variables



Bi Variant Analysis of Categorical variables



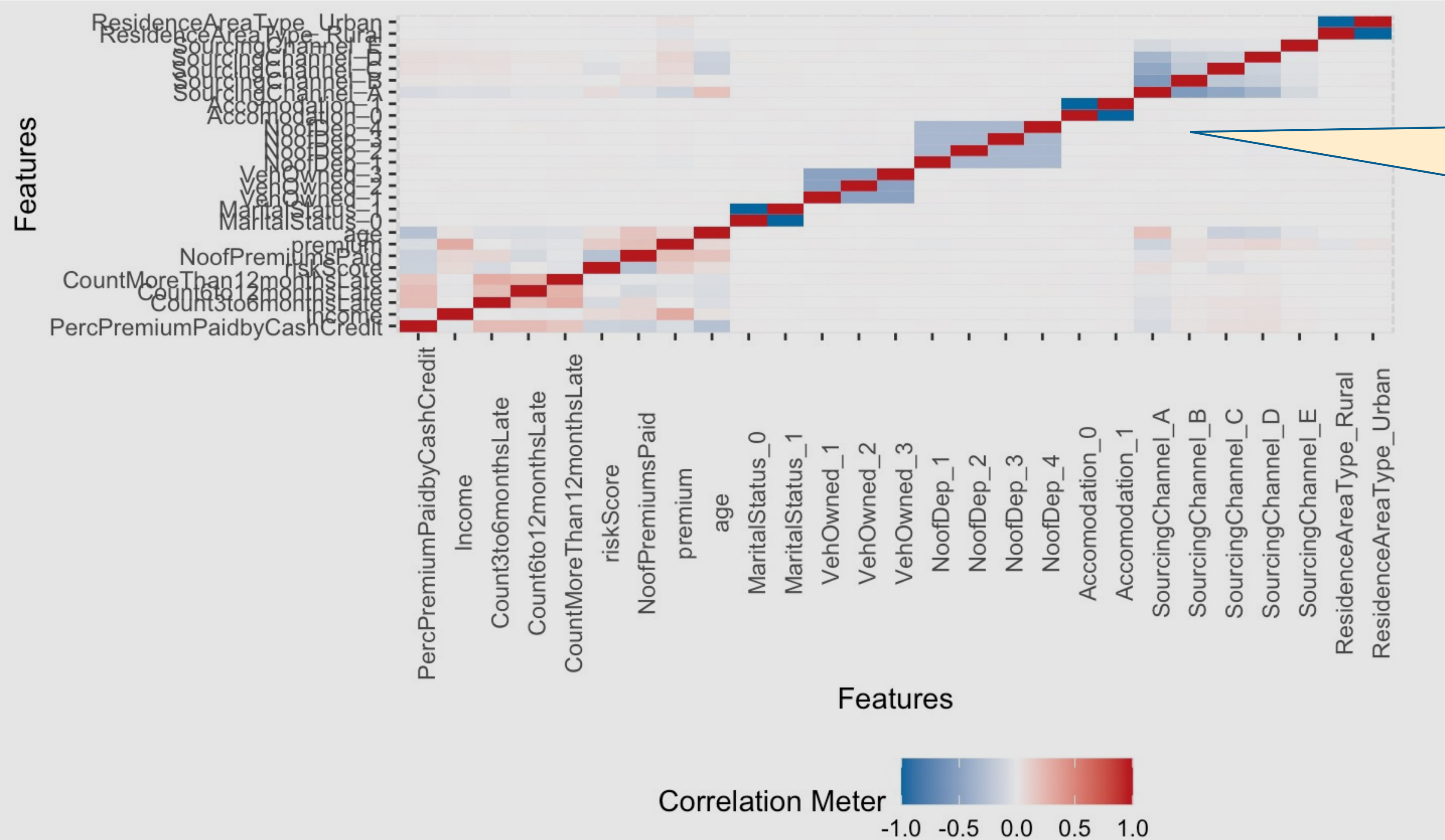
Bi Variant Analysis – old vs new defaulters



Identifying the pattern of the defaulters with the cohorts who have been late in paying their premiums by

- 3 to 6 months
- 6 to 12 months
- More than 12 months

Correlations among all variables



No High Correlation among any variables

Analytical Models

Basic Models

Will apply a Supervised Machine Learning Technique for a Descriptive, Predictive & Prescriptive Analysis

Decision Tree

CART

Supervised Machine Learning Technique

Built a CART model on the train data.
-- create CART model 1 & validate for accuracy.
-- Tuning the model: further tune the model for further accuracy
Model Validation: validate the new model
Model Evaluation: evaluate both the models on the test data & compare their accuracy.
+ Tune the model and prune the tree, if required.
+ Test the data on test set.

Regression

Logistic Regression

Logistic function to model the conditional probability

In the case of binary classification the probability of defaulting premiums and not defaulting premiums will sum up to 1

classification method

Naïve Bayes

Probability of the given feature vector being associated with a label

the algorithm expects the features to be independent which is not always is the case.

Nearest neighbors

KNN

classify new data points based on similarity measure

Classification is done by a majority vote to its neighbors. The data is assigned to the class which has the nearest neighbors.
As you increase the number of nearest neighbors, the value of k, accuracy might increase.

Analytical Models *Bagging & Boosting*

Process will involve creating a Train & Test Data set from the original data set. The training data set will be used to validate each model and the same will later be used to evaluate the model on the test data set.

Bagging

Random Forest

an ensemble method that trains several decision trees in parallel with bootstrapping followed by aggregation

-- Incase there is no significant improvement in the CART model from the baseline model, we build the Random Forest.
-- Tune the Model
-- Model Validation: validate the new model
-- Model Evaluation: evaluate both the models on the test data & compare their accuracy.

Boosting

Gradient Boosting

machine learning boosting

It relies on the intuition that the best possible next model, when combined with previous models, minimizes the overall prediction error.

Boosting

X G Boost

Efficient to the Gradient Boosting

Extreme Gradient Boosting (XGBoost) is similar to gradient boosting framework but more efficient. It has both linear model solver and tree learning algorithms. What makes it fast is its capacity to do parallel computation on a single machine..

Confusion Matrix

Indication towards extracting "defaulters' & 'non-defaulters

CART MODEL 1	Defaulters	Non-Defaulters
Defaulters	173	826
Non - Defaulters	218	14753

CART MODEL 2	Defaulters	Non-Defaulters
Defaulters	0	999
Non - Defaulters	0	14971

Naïve Bayes	Defaulters	Non-Defaulters
Defaulters	0	3
Non - Defaulters	999	14968

Logistic Regression	Defaulters	Non-Defaulters
Defaulters	122	96
Non - Defaulters	877	14875

KNN	Defaulters	Non-Defaulters
Defaulters	98	901
Non - Defaulters	103	14868

Gradient Boosting	Defaulters	Non-Defaulters
Defaulters	147	852
Non - Defaulters	126	14845

Random Forest 1	Defaulters	Non-Defaulters
Defaulters	159	840
Non - Defaulters	154	14817

Random Forest 1	Defaulters	Non-Defaulters
Defaulters	2	997
Non - Defaulters	0	14971

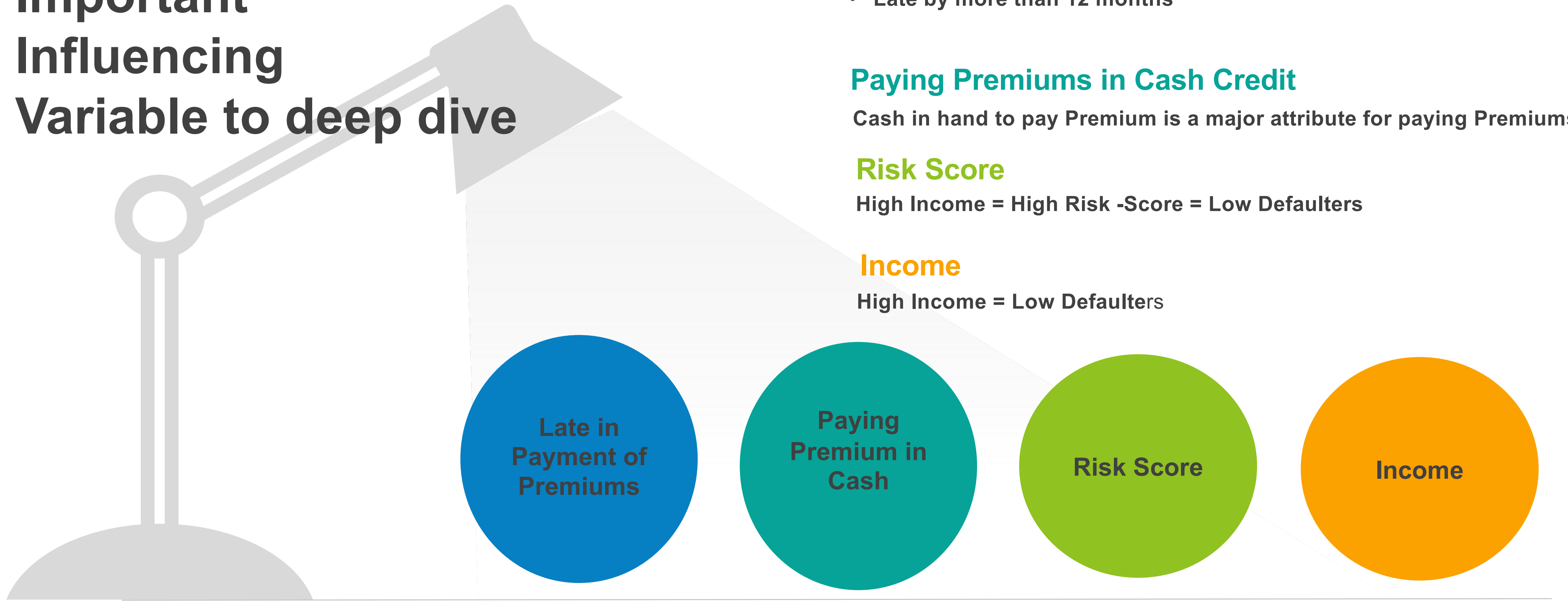
XG Boost	Defaulters	Non-Defaulters
Defaulters	113	886
Non - Defaulters	81	14890

Overview: Models Analysis

Models	Accuracy	Sensitivity	Specificity	Precision	KS	AUC	Gini	Concordance
Cart Model 1	0.93	0.92	0.01	0.92	0.5	0.80	0.03	0.17
Cart Model 2	9.374452e-01	9.374452e-01	0.00	9.374452e-01	0.00	5.0	-2.27	0.00
Random Forest 1	0.93	0.92	.009	47.33	0.50	0.81	0.05	NaN
Random Forest 2	9.375704e-01	9.375704e-01	1.252348e-04	7.48	5.111567e-01	8.177137e-01	1.159270e-02	NaN
Logistic Regression	0.93	0.93	0.007	0.93	0.5	NA	0.03	NaN
Naïve Bayes	0.93	0.93	0.00	0.93	0.51	NA	0.0001	NaN
KNN	0.93	0.93	0.006	73.97	0.39	0.72	0.043	NaN
Gradient Boost	0.93	0.92	0.009	0.92	NA	0.82	NA	NaN
X G Boost	0.93	0.93	0.007	0.93	0.51	0.82	0.02	NaN

Most Important & Influencer variables

Important Influencing Variable to deep dive



Late in Payment of Premiums

- Late by 3 to 6 Months
- Late by 6 to 12 months
- Late by more than 12 months

Paying Premiums in Cash Credit

Cash in hand to pay Premium is a major attribute for paying Premiums

Risk Score

High Income = High Risk -Score = Low Defaulters

Income

High Income = Low Defaulters

Key Findings

01 Looking at customers who are late in paying premiums in all 3 categories. They seem to follow a common pattern. Also, the risk score is a good indicator here as low risk score customer tend to have defaulted on the payments.

02 Paying Premiums in Cash Credit seems like a stumbling block for some customers. Look at solutions to manage this issue by creating options towards offering insurance types and subsequent premiums.

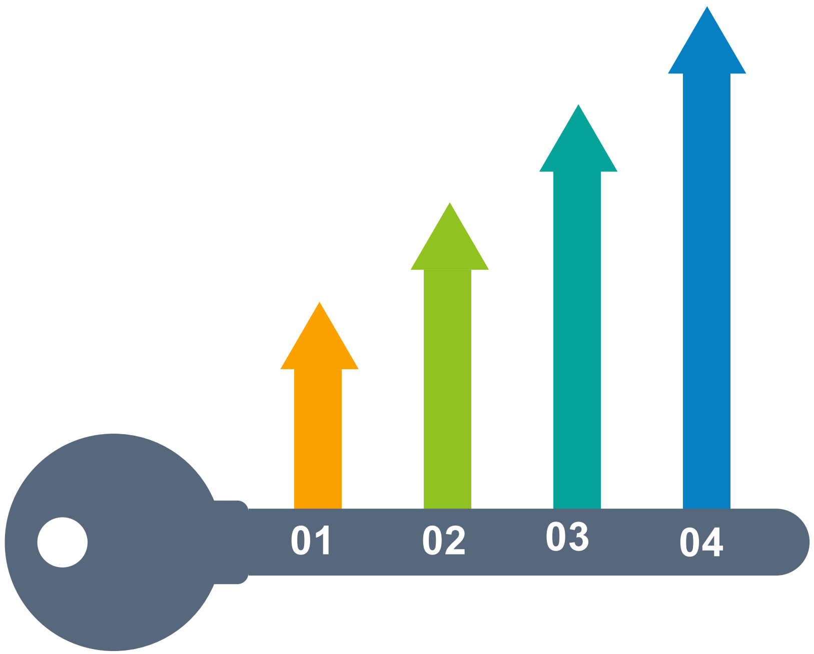
03 Income seems an influencing factor. We comparatively see lesser defaulters amidst the higher income bracket

04 Interactive – Engaging – Incentive based options

05 Flexible, customized, tailored payment options for eligible customers

Recommendation

You can download professional PowerPoint diagrams for free



01

Short term vs Long Term Options

Rather providing a long term offering, what needs to be offered would be a solution which addresses the current issue in hand and something that will rightly fit in their scheme of things.
A comprehensive but simple product which is easy to issue and pick needs to be developed.

02

Is capability to make Payment the only reason to Default?

Convenience, simplified product, easier issuance of insurance could at times solve the lethargy of not prioritizing the payments towards your premiums.
Options for cashless service, credit service, EMI offerings, deferring options (case basis), etc. can be offered to pay premiums timely

03

Solution bases customized offering (Sachet, Bouquet, etc.)

Relevant and simpler options could at times be easier to pick up especially during the challenging times like the pandemic when work, business, jobs and earnings are a challenge.

04

Interactive – Engaging – Incentive based options

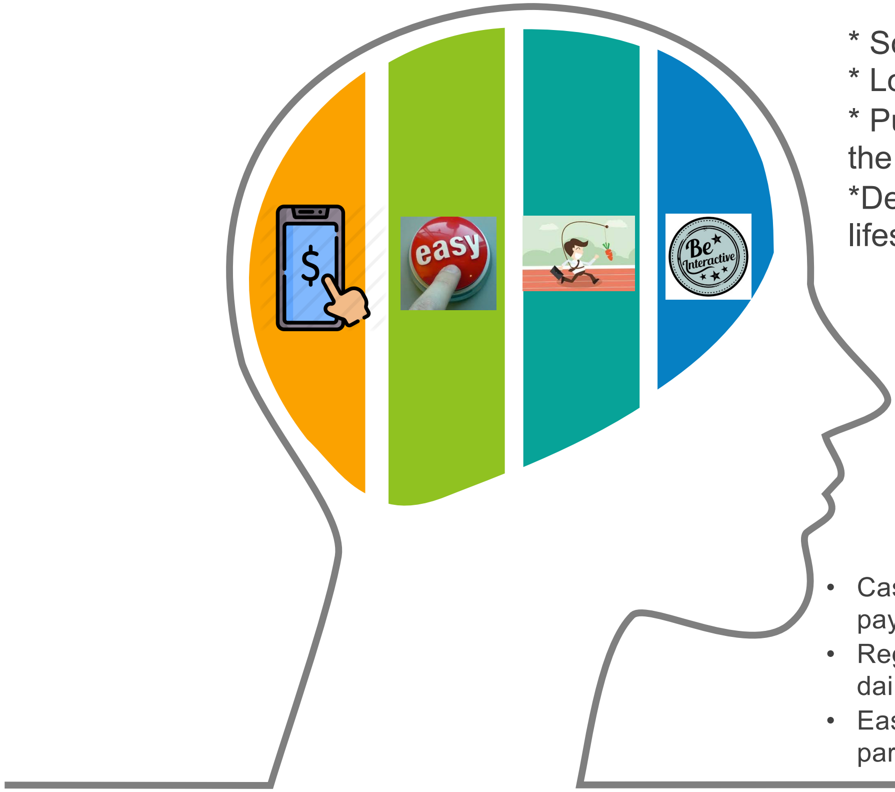
Coming up with easy to adapt and engage programs for the customers by which one can achieve solutions for the insurer and the insured.

We need to keep in mind the current economic fallout due to the pandemic and the impact it has had on jobs/ business/income when we come our with solutions for the payments of the premiums.

Changing the Customer Mindset

outlook towards priority for health – insurance – paying premium

Incentivize customers for having health priorities



* Solutions should be from customer perspective and not a push to simply collect premiums.
* Look at the issues from customers stand-point
* Push to make the customer a healthier person which has advantages for the individual and the insurer
*Develop products, schemes, programs to incentivize those who maintain a regular healthier lifestyle

Cashless alternatives

Convenience to participate

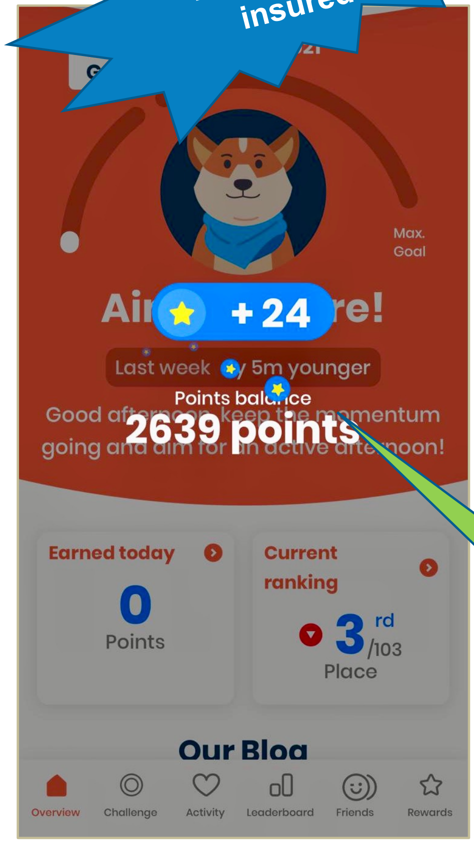
incentives

Interactive

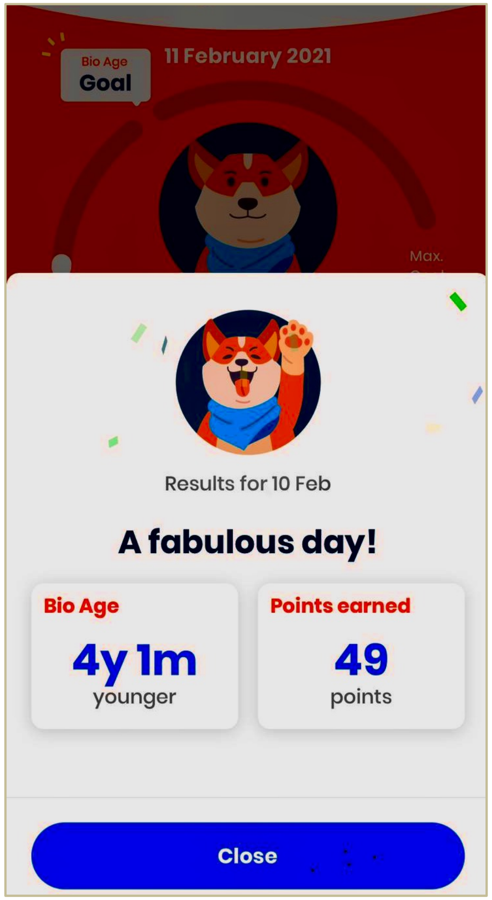
- Cash benefit programs which would include lesser premiums as per healthier lifestyle, deferred period of payments for customers showing a healthier trend.
- Regular incentives on milestones which offers discounts, freebies, add-ons for encouraging fitness on a daily basis.
- Easy to adopt solutions (like loading app on customer’s handset) and interactive options for easy participation to keep customers interested, engaged and encouraged.

Recommendation Example

Advantageous for both insurer & the insured

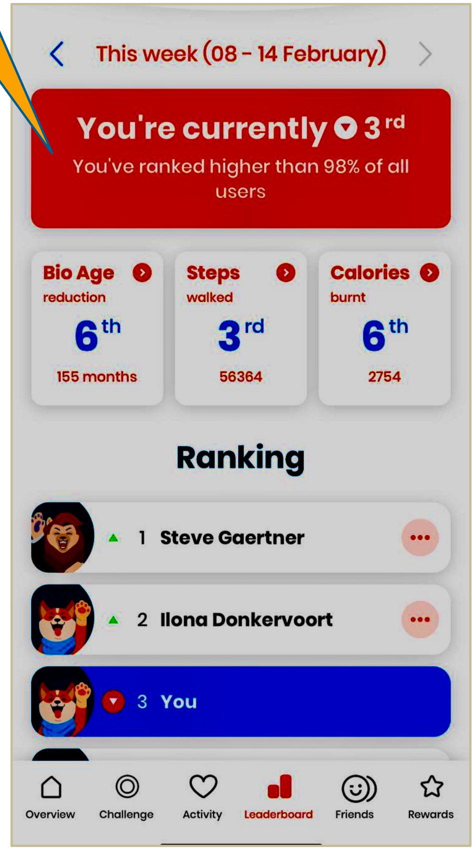


A Vitality Fitness App to keep the user engaged with regards to his/her physical activities like walk, jog, sleep, etc.



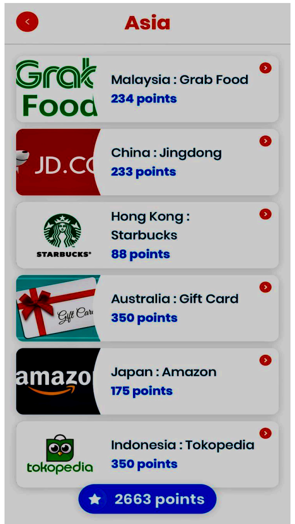
Incentivize the user with a lesser premium amount by basis of their bio age fitness level

the rewards and incentives on a regular basis is a significant gain plus staying in great shape keeps you positive to manage life



Recommendation Example

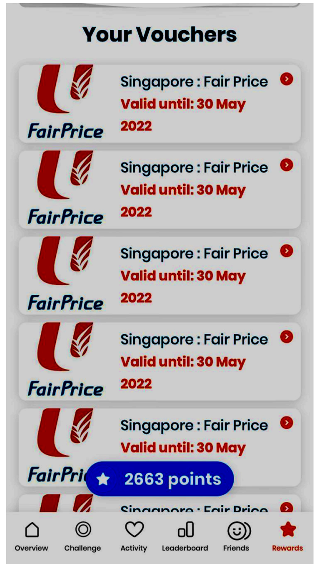
Variety of benefits ranging from daily refreshments, daily rides, vouchers for e-shopping



<u>Benefits * Incentives * Rewards * Recognition * Discounts * Deferred payment options</u>

- Deferred period options for payment of premium**
- More active and healthy = more attractive offerings on premiums**
- a healthier customer = benefits from other financial & professional service**

All stakeholders together create a 'value prop" with customer at the center



Groceries shopping vouchers. The more you stay fit – the more discounts you earn

Thankyou

… end of presentation