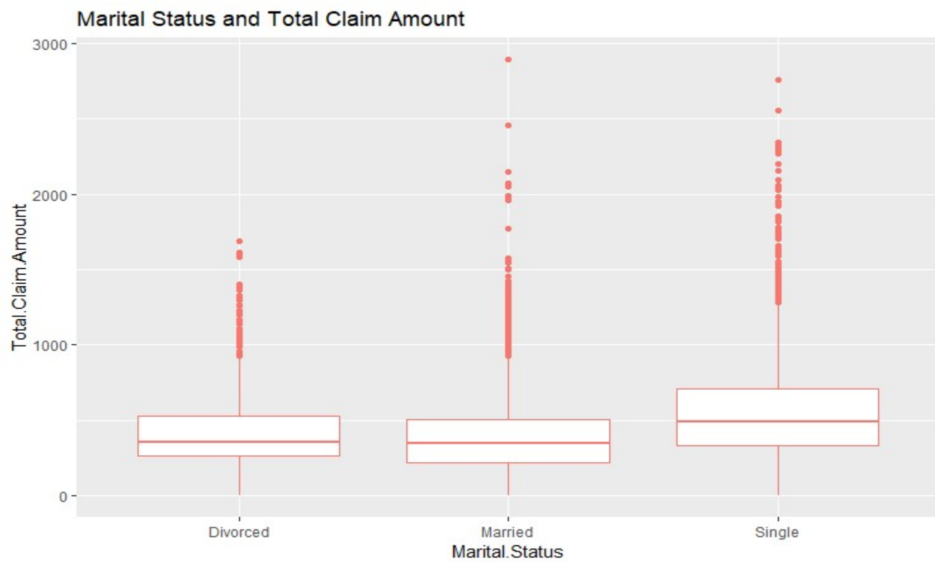
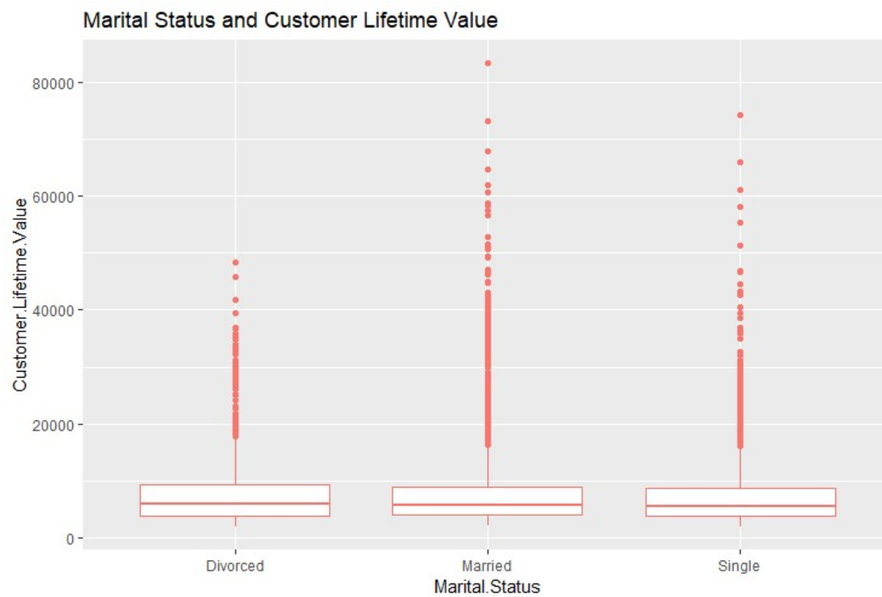
Predicting Marketing Campaign Response

Aditee Bhattarai
Lindiwe Mukurazita
Rajadurga Ganesan
Utsav Pradhan

Variables of Interest

- Response,
- Customer Lifetime Value
- Total Claim Amount
- After an initial data exploration, we believe that “Location Code”, “Employment Status”, “Marital Status” and “Vehicle Class” are important for our analysis.
- We would like to predict the variable “Response” based on the other variables. We would also like to explore and find the best customers to target by predicting “Customer Lifetime Value” and “Total Claim Amount”.

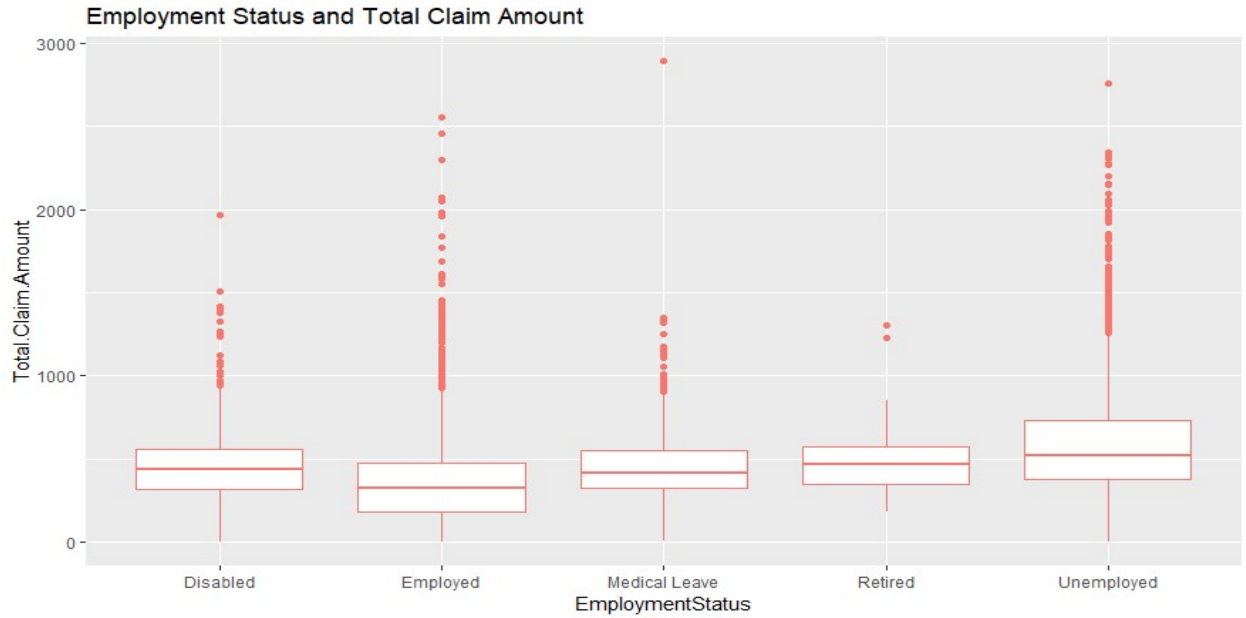




Marital Status

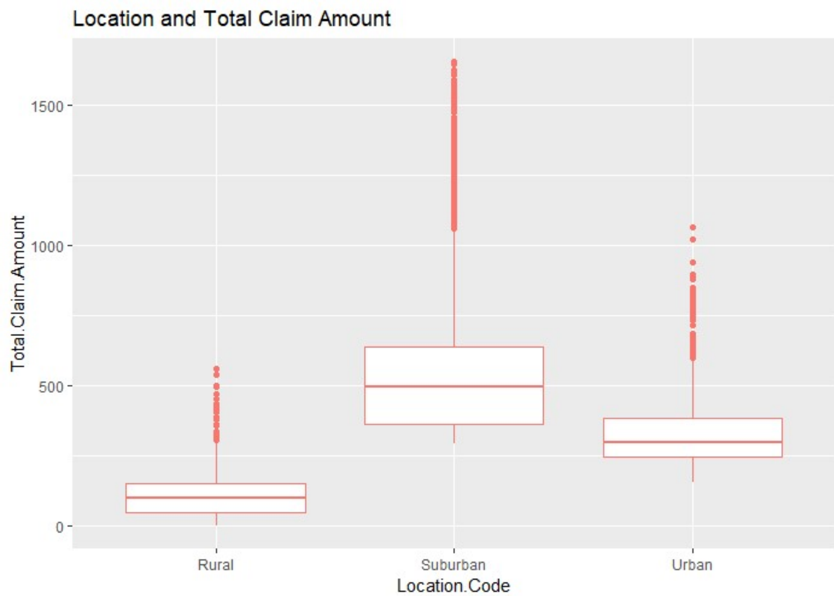
- Singles have high claim amounts than other groups.
- Singles have lowest median and mean lifetime value.
- Singles also don't pay more premium monthly.

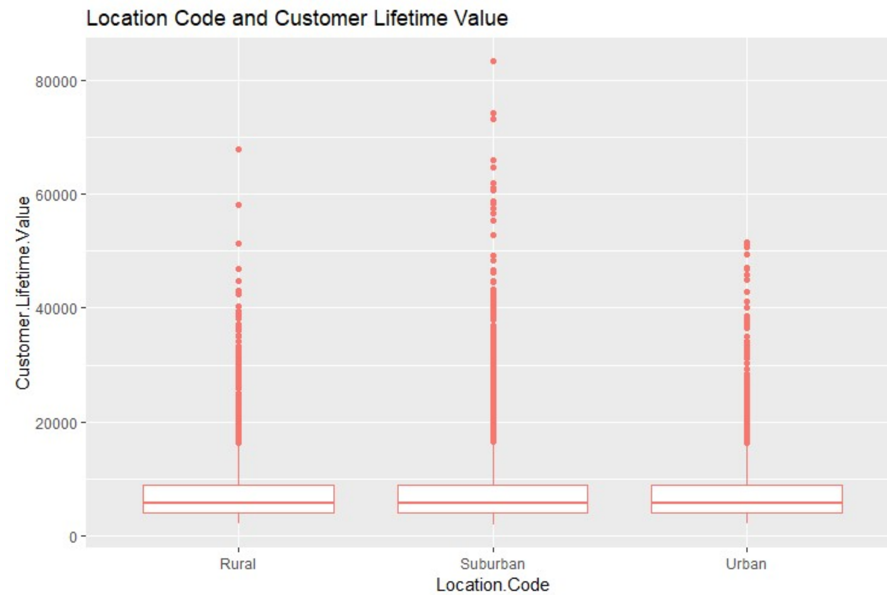
- Having a greater number of policies does not increase claim amount.



Employment Status

- Unemployed people have noticeably higher claim amount. Retired people too
- Lowest claim is from employed people.
- Lifetime value among different groups is not significantly different.
- Premium is about the same

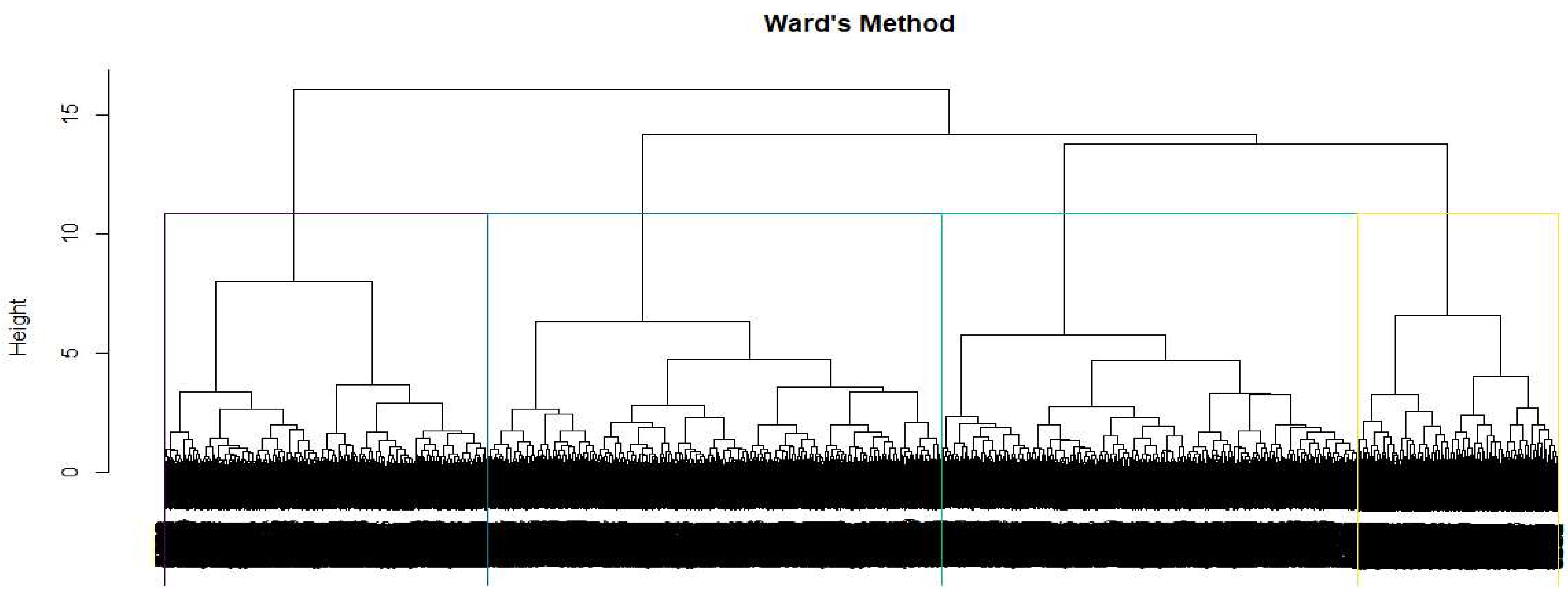




Location Code

- No significant difference in CLV. High difference in claim amount.
- Suburban has significantly higher claim amount. Rural has lowest.
- Suburban pay slightly more premium.

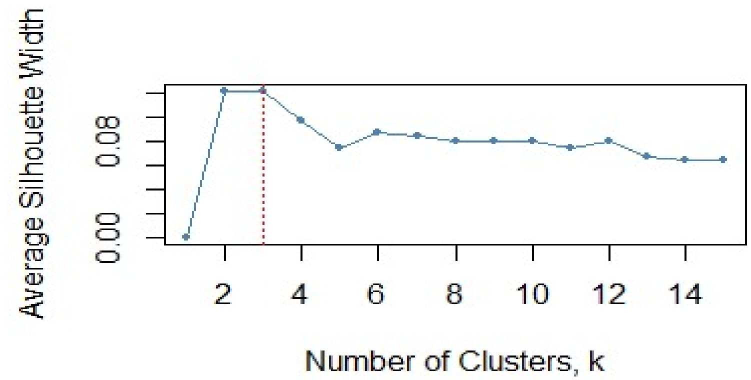
Cluster Analysis - Hierarchical



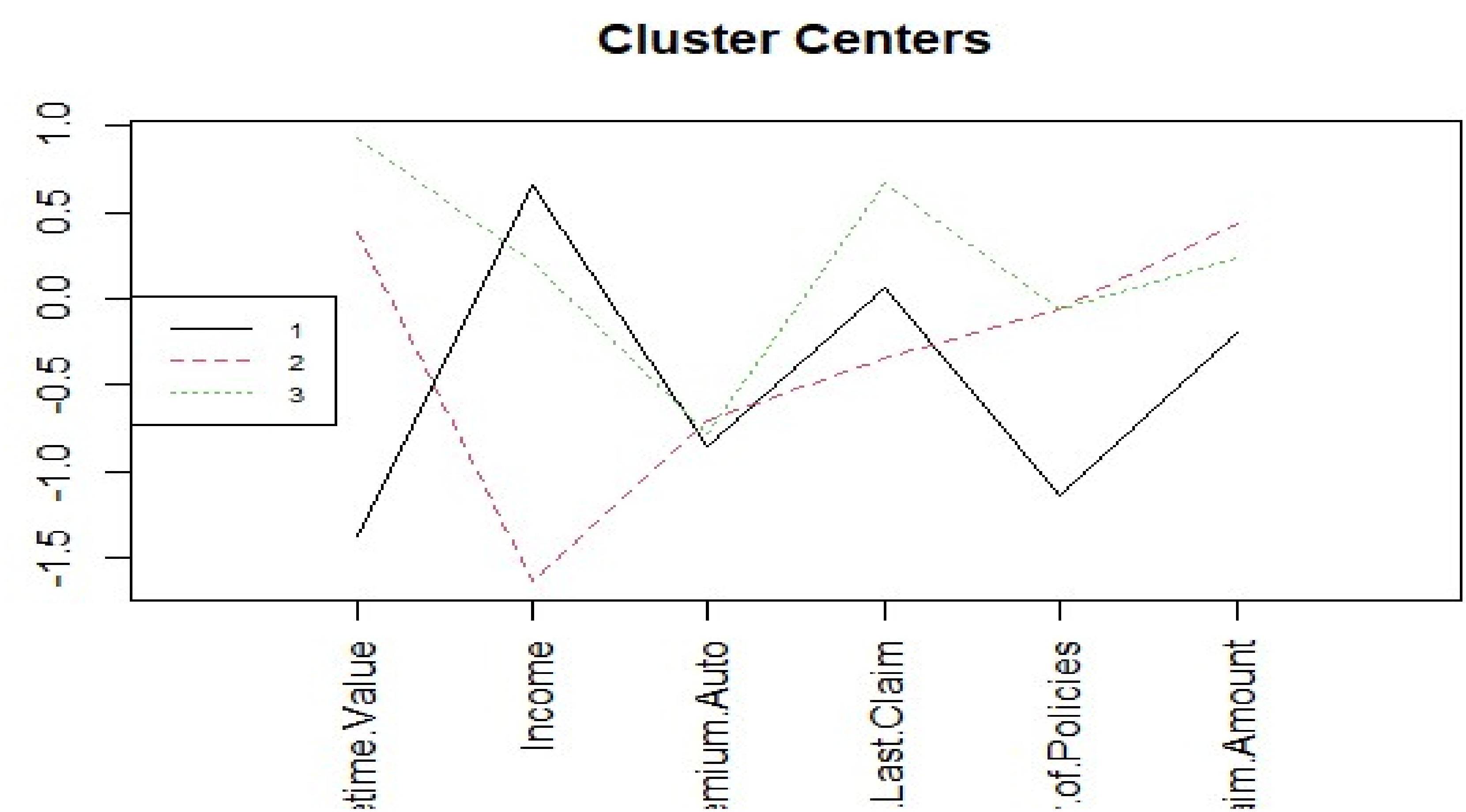
	Employment	Gender	Location	Marital	Renew offer	Response
1	Highest employed,	More females	Suburban	More married	1	0
2	Highest unemployed	Mix	Suburban	More Single	1	0
3	High employed	More Males	Suburban	More Married	1 or 2	6
4	Low employed	Mix	Suburban	More Single	2 and 1	1302

	CUSTOMER LIFETIME VALUE	PREMIUM AMOUNT	TOTAL CLAIM AMOUNT
1	8259.798	91.77	344.1729
2	7501.556	92.6	587.7354
3	8061.936	92	379.9425
4	7857.630	94.1	447.4213

K-
Medoid



	Coverage	Employment	Location	Marital	Offer type	Channel	Response	Customer lifetime Value	Income	Claim Amount
1	Basic	Employed	Sub	Married	1	Agent	No	Low	High	Low
2	Basic	Employed	Sub	Single	1	Agent	No	Medium	Low	High
3	Basic	Employed	Sub	Married	2	Branch	No	High	Medium	Medium



Cluster Centers
1.0
0.5
0.0
-0.5
-1.0
-1.5
1
2
3
etime.Value
Income
emium.Auto
Last.Claim
of.Policies
aim.Amount

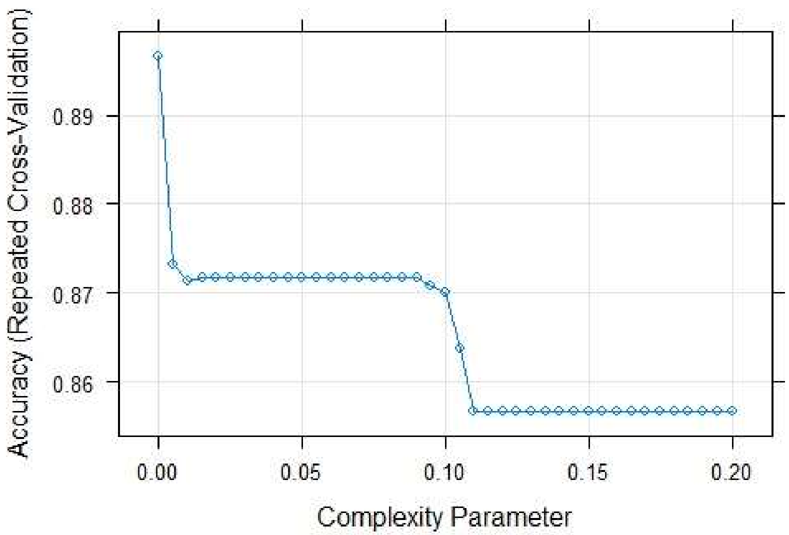
Decision Trees

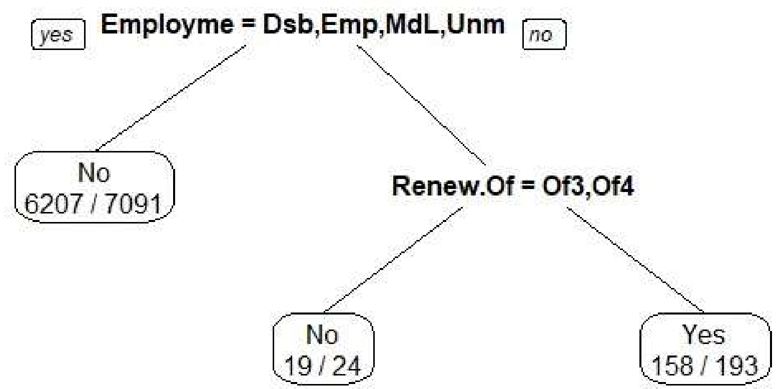
Class imbalance exists in our data set and we see that less than 20% are in the minority class

The top 4 variables of importance before hyperparameter tuning are :-

- Employment status
- Renew offer type
- Customer Lifetime Value
- Monthly Premium Auto

- The true positive rate is 15% which is quite low, Recall and F1 are very low for both training and test data.


Accuracy (Repeated Cross-Validation)
0.89
0.88
0.87
0.86
0.00
0.05
0.10
0.15
0.20
Complexity Parameter


yes
Employme = Dsb,Emp,MdL,Unm
no
No
6207 / 7091
Renew.Of = Of3,Of4
No
19 / 24
Yes
158 / 193

Decision Trees

After hyperparameter tuning, the model improves.

The top 4 variables of importance after tuning are:

Customer Lifetine Value,Tot Claim Amt,Monthly Premium Auto and Income

The values for Sensitivity: Pos Predict Value, Precision, Recall and F1 are much lower for the testing data. This shows that the model performed well for the training data but was not a very good fit for the unseen data.

Naïve Bayes | ANN

	Training	Testing
Accuracy	8.701423e-01	8.729463e-01
Kappa	2.282338e-01	2.453282e-01

	Training	Testing
Accuracy	9.524791e-01	9.086925e-01
Kappa	8.075759e-01	6.451802e-01

- Highly accurate model
- Of the two Models:
  - **Naïve Bayes** fails to correctly predict the classified observations
  - **ANN** does a moderate job in predicting the classified observations
    - Much better predicting power than Naïve

Naïve Bayes | ANN

	Training	Testing
Sensitivity	0.16905444	0.18007663
Specificity	0.98738221	0.98849840
Pos Pred Value	0.69140625	0.72307692
Neg Pred Value	0.87663074	0.87847814
Precision	0.69140625	0.72307692
Recall	0.16905444	0.18007663
F1	0.27168074	0.28834356

	Training	Testing
Sensitivity	0.8417266	0.7397959
Specificity	0.9709905	0.9369139
Pos Pred Value	0.8290523	0.6621005
Neg Pred Value	0.9734780	0.9556522
Precision	0.8290523	0.6621005
Recall	0.8417266	0.7397959
F1	0.8353414	0.6987952

- ANN model can greatly predict actual positive observations that were correctly classified
  - Helps to minimize the losses from inaccurately predicting a favorable response

Insights and Recommendation