



TELCO CUSTOMER CHURN


HEMANT SINGHAL

RISHI MUTNURU

SADHVI ISANAKA

PROBLEM & IMPORTANCE

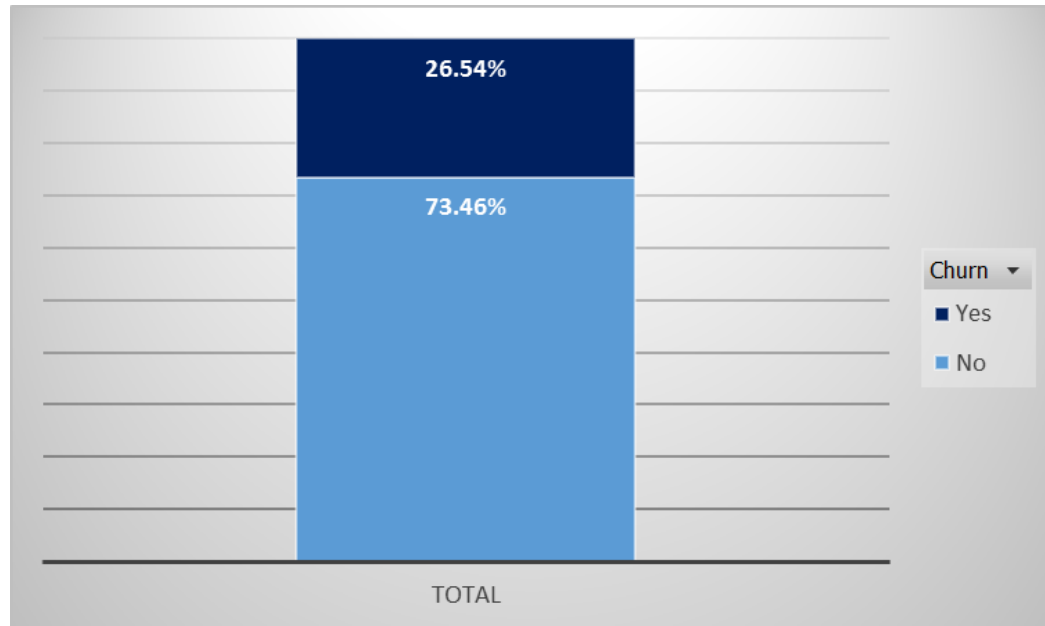


- Definition of Churn rate: It is the percentage of customers who stop their relationship with the company within a defined period.
- Customer churn is a significant issue in any business.
- Cost of retaining an existing customer is far less than acquiring a new one.
- **Goals:** 
 - Predict the potential churn of existing customer – Best predictive (Classification) model with least misclassification rate.
 - Focused marketing strategies - promotional offers, discounts and other value-added services.

DATA DICTIONARY

Features	Description
customerID	Customer ID
gender	Customer gender (female, male)
SeniorCitizen	Whether the customer is a senior citizen or not (1, 0)
Partner	Whether the customer has a partner or not (Yes, No)
Dependents	Whether the customer has dependents or not (Yes, No)
tenure	Number of months the customer has stayed with the company
PhoneService	Whether the customer has a phone service or not (Yes, No)
MultipleLines	Whether the customer has multiple lines or not (Yes, No, No phone service)
InternetService	Customer's internet service provider (DSL, Fiber optic, No)
OnlineSecurity	Whether the customer has online security or not (Yes, No, No internet service)
OnlineBackup	Whether the customer has online backup or not (Yes, No, No internet service)
DeviceProtection	Whether the customer has device protection or not (Yes, No, No internet service)
TechSupport	Whether the customer has tech support or not (Yes, No, No internet service)
StreamingTV	Whether the customer has streaming TV or not (Yes, No, No internet service)
StreamingMovies	Whether the customer has streaming movies or not (Yes, No, No internet service)
Contract	The contract term of the customer (Month-to-month, One year, Two year)
PaperlessBilling	Whether the customer has paperless billing or not (Yes, No)
PaymentMethod	The customer's payment method (Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic))
MonthlyCharges	The amount charged to the customer monthly
TotalCharges	The total amount charged to the customer
Churn	Whether the customer churned or not (Yes or No)

ABOUT DATA

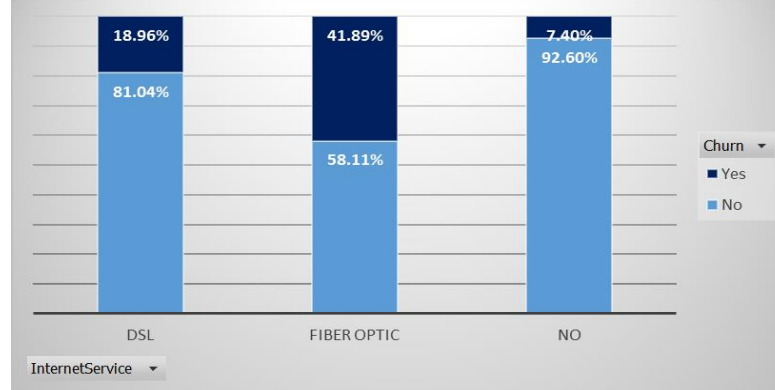
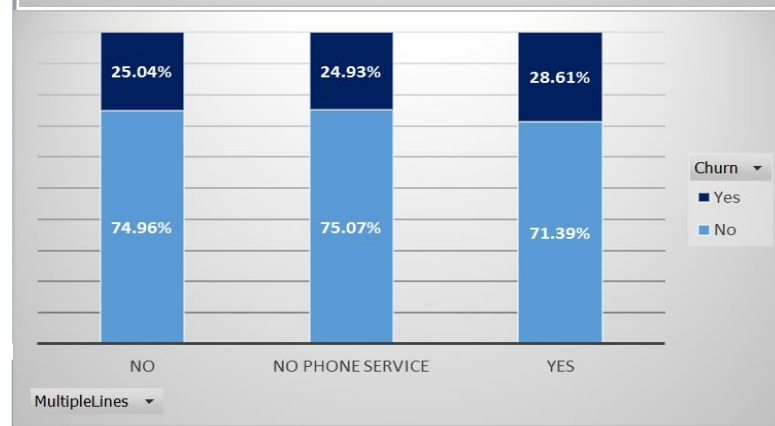
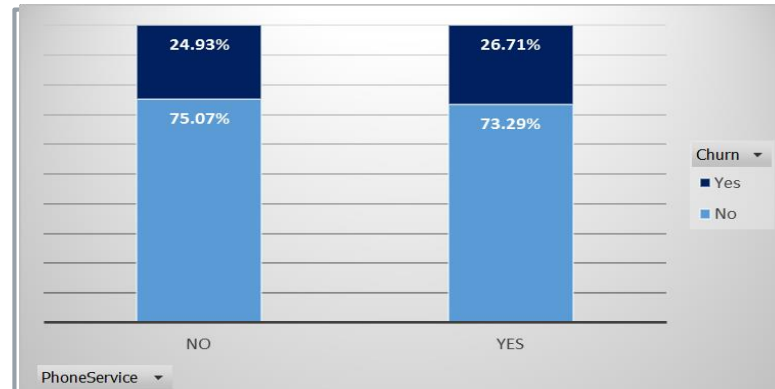
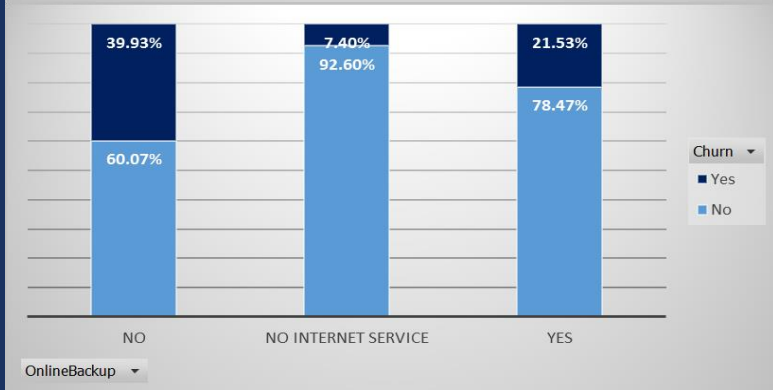
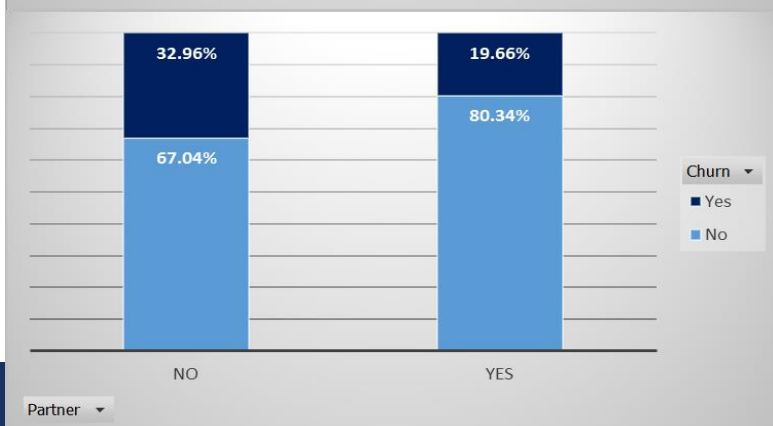
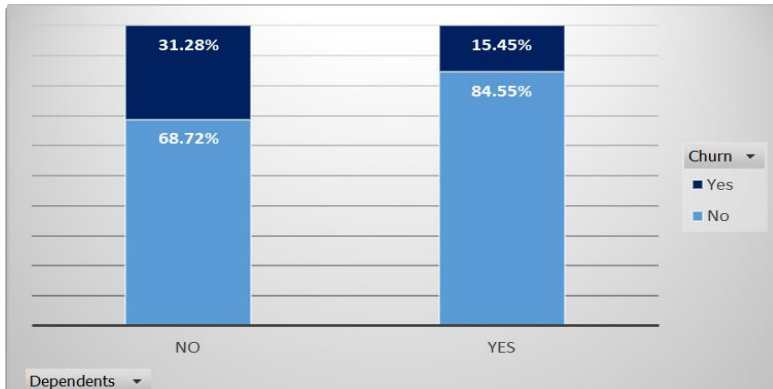
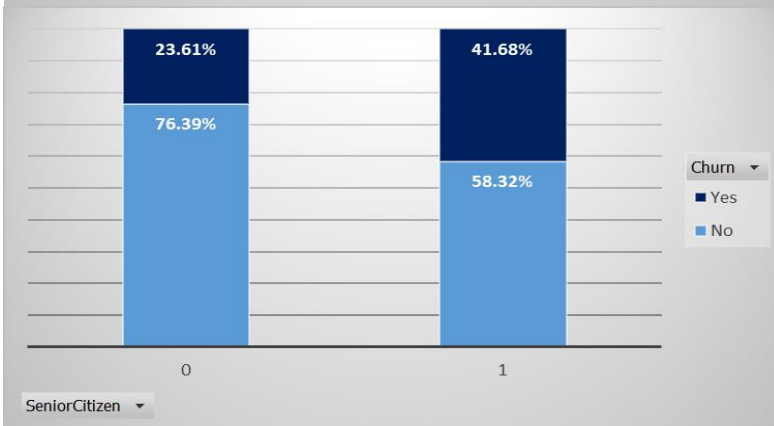


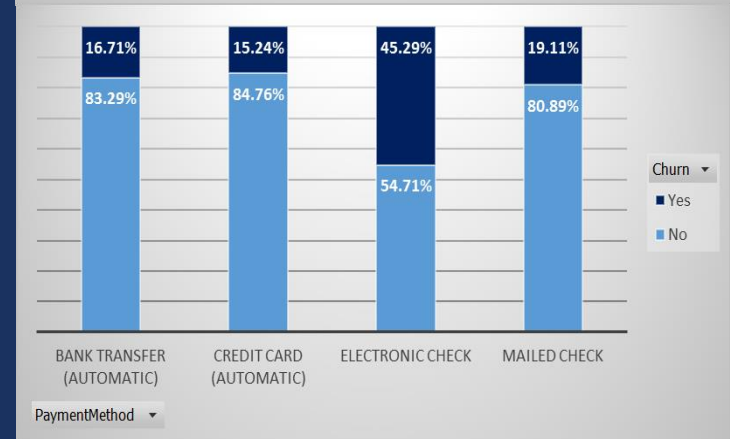
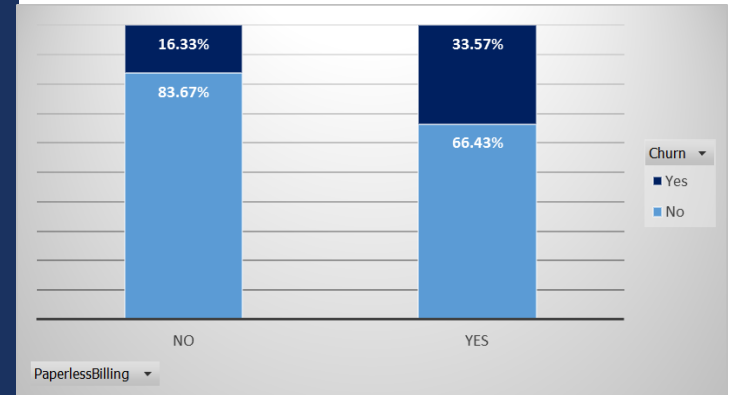
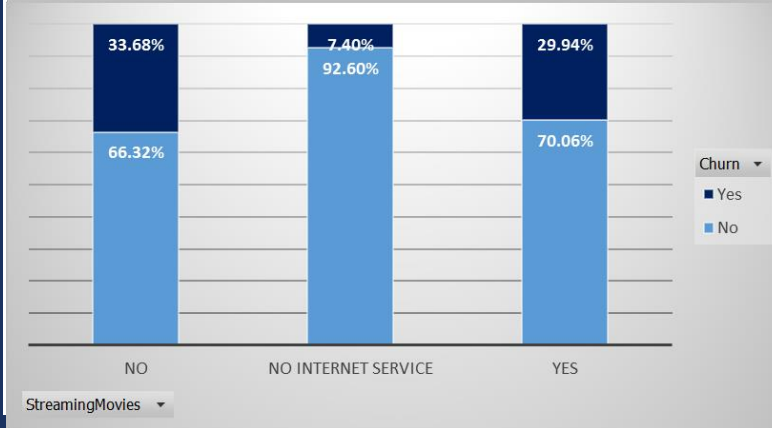
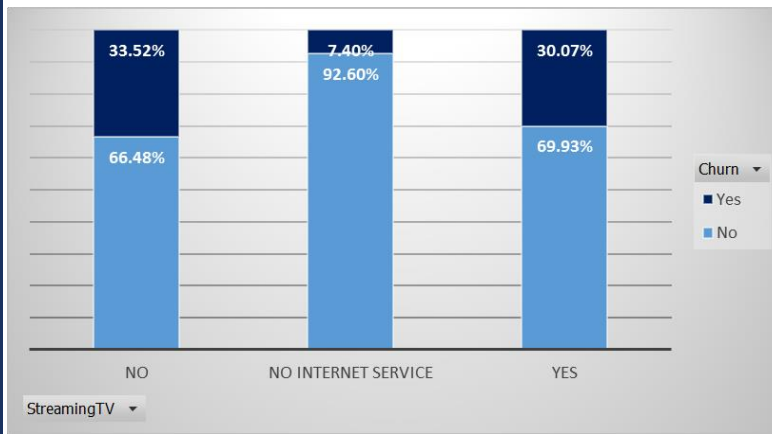
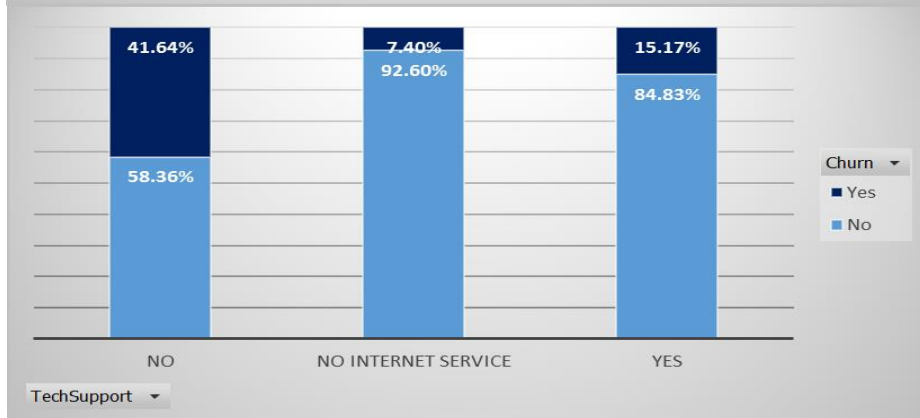
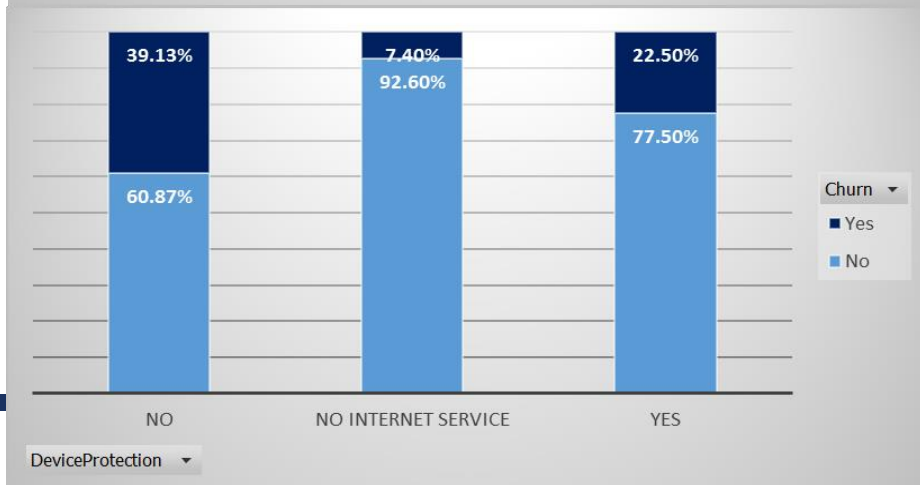
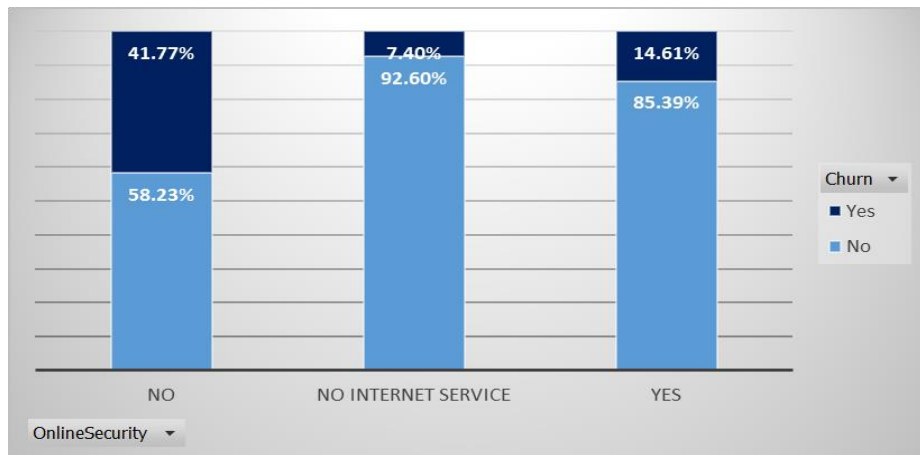
- Source of dataset: www.kaggle.com
- Software used: Frontline Systems Analytic Solver Platform
- Data limitations: Unbalanced dataset
- Desired: Balanced dataset, try several seed values and re-run models using a programming language.

DATA PREPROCESSING

- Missing data: For tenure = 0, TotalCharges were empty. Instead of deleting those records, we made: For tenure = 0, TotalCharges = MonthlyCharges
- Used Data Mining-One Hot Encoding to convert the text (yes/no) into binary values and created necessary dummy variables.
- We took few records out and kept them in separate Excel tab called “NewData”. This is to confirm our TestScore Metrics. Although, it is not needed.
 - Tried with 4 new records
 - Tried with 20 new records

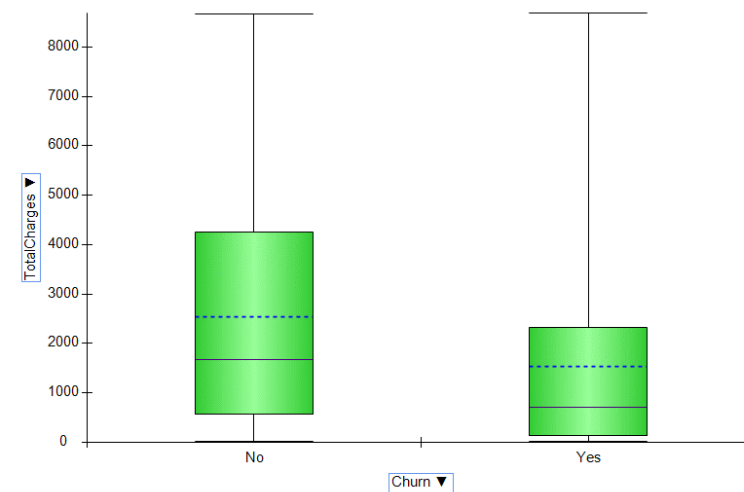
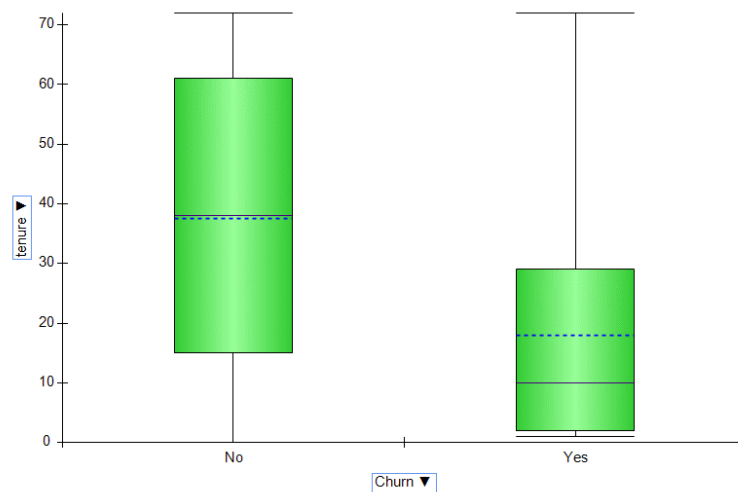
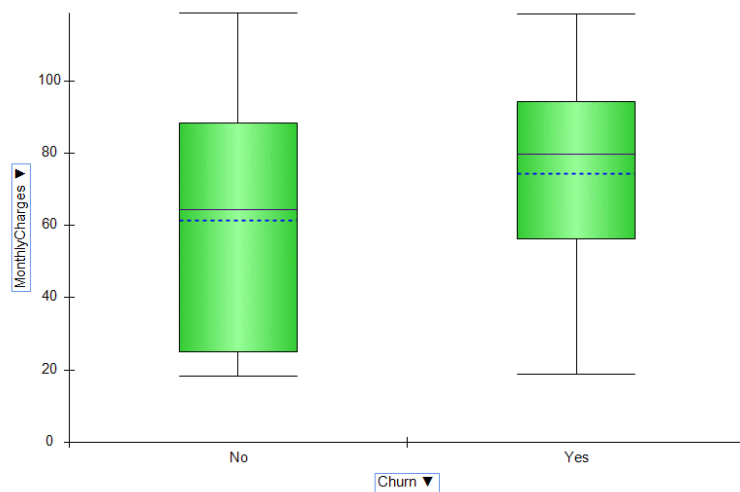
DATA EXPLORATION (CAT VS CAT)





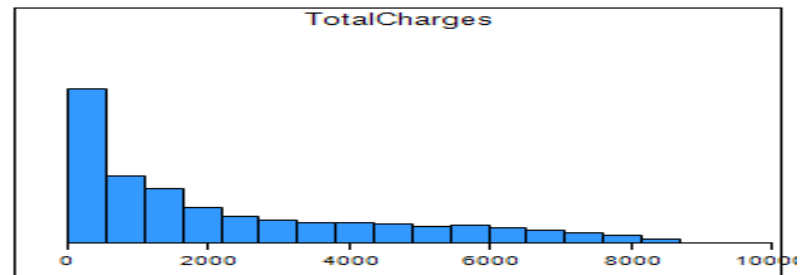
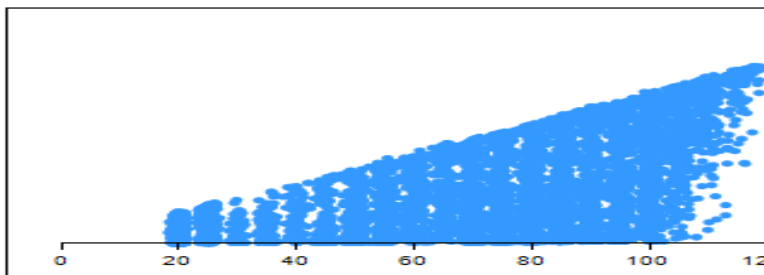
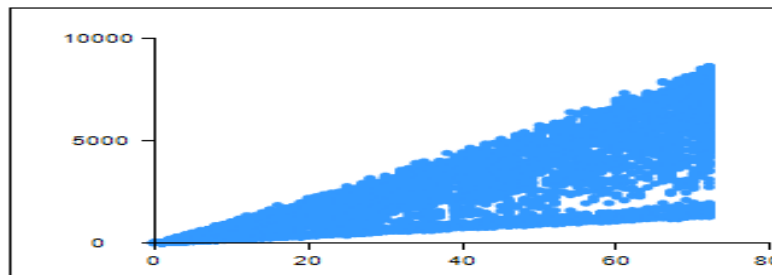
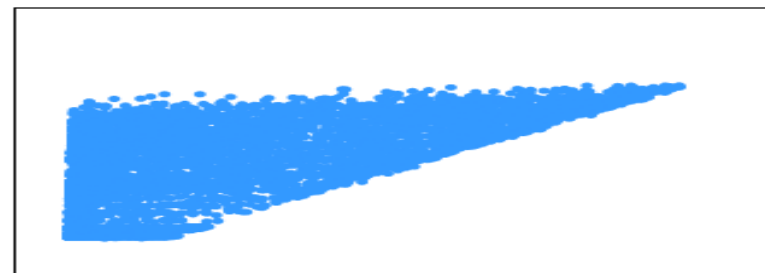
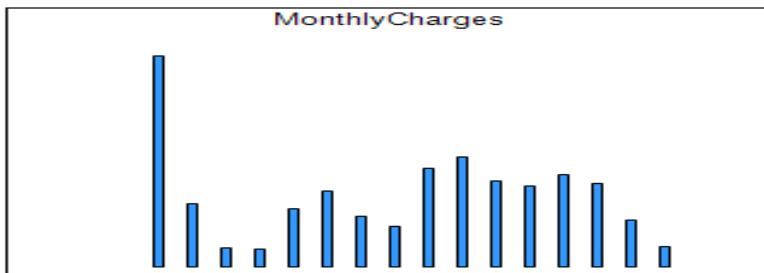
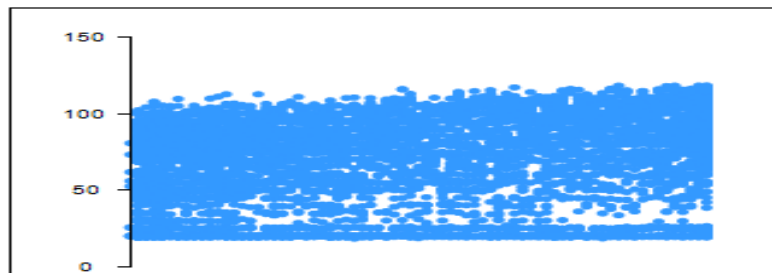
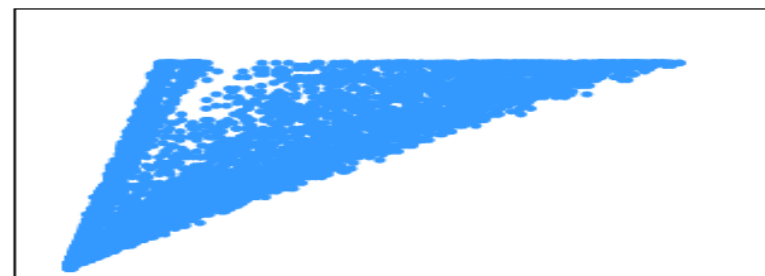
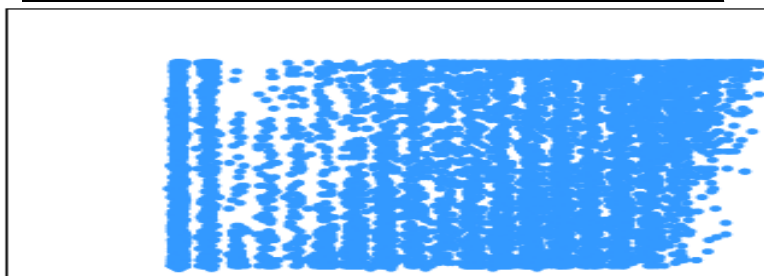
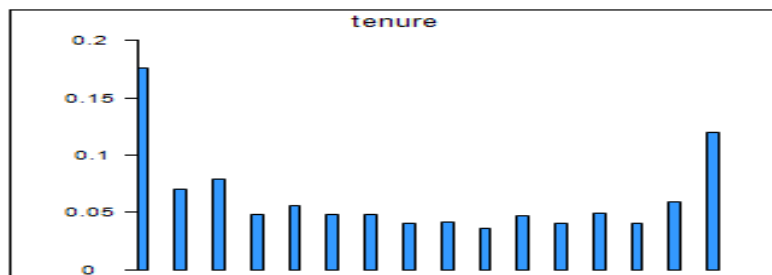
DATA EXPLORATION (CONT.)

CATEGORICAL VS CONTINUOUS VARIABLES

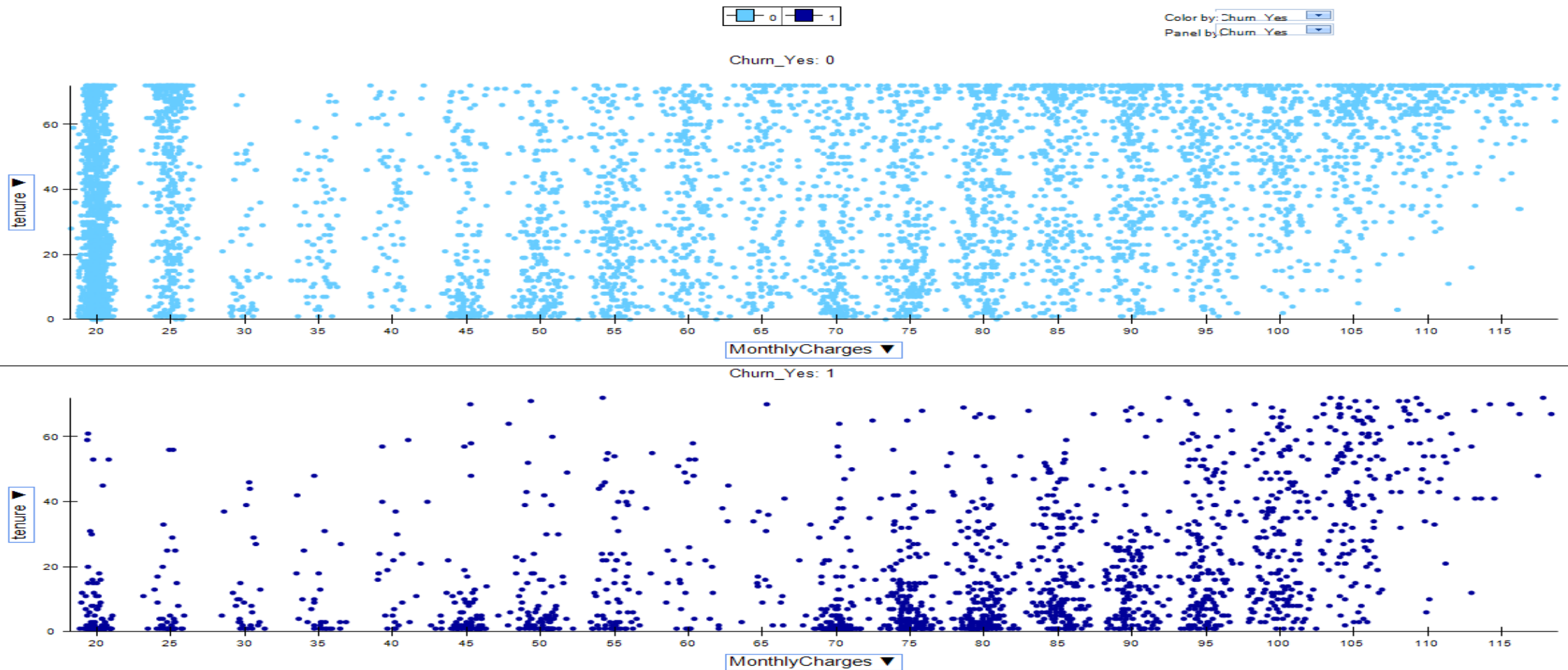


CORRELATION MATRIX (CONTINUOUS VARIABLES)

	<i>tenure</i>	<i>MonthlyCharges</i>	<i>TotalCharges</i>
<i>tenure</i>	1		
<i>MonthlyCharges</i>	0.247899856	1	
<i>TotalCharges</i>	0.826164213	0.651181985	1

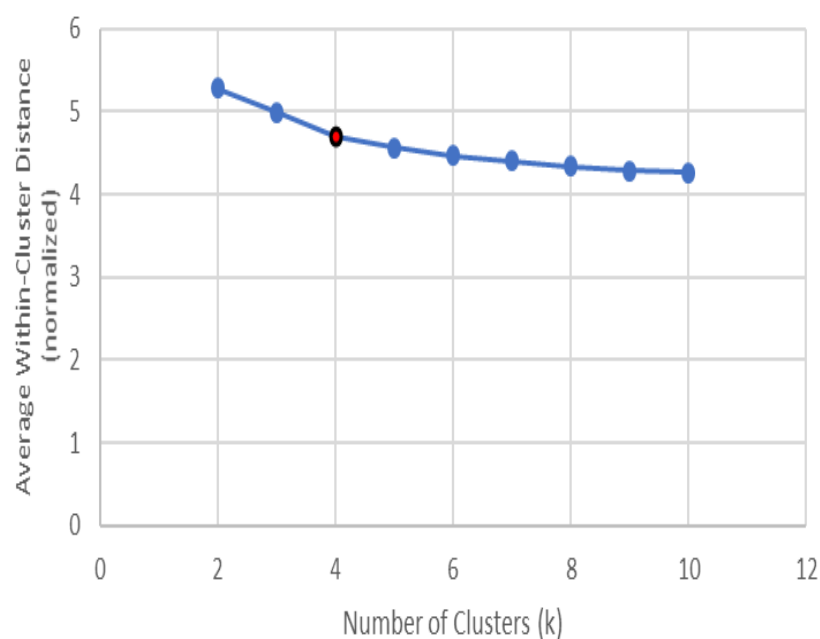


TENURE VS MONTHLY CHARGES VS CHURN



SEGMENTATION (UNSUPERVISED LEARNING: K- MEANS CLUSTERING)

Elbow Chart



	Fiber+1/2 yr contract	No internet	Fiber+month-to-month contract	DSL+no phone service
Average of TotalCharges	5,043.96	662.60	1,265.06	1,500.81
Average of MonthlyCharges	91.45	21.08	74.37	42.02
Average of tenure	54.70	30.54	16.50	31.84
Average of PhoneService_Yes	1.00	1.00	1.00	0.00
Average of Contract_Month-to-month	0.20	0.34	0.93	0.55
Average of TechSupport_No	0.38	0.00	0.83	0.58
Average of OnlineSecurity_No	0.42	0.00	0.81	0.58
Average of DeviceProtection_No	0.28	0.00	0.78	0.55
Average of OnlineBackup_No	0.30	0.00	0.75	0.58
Average of PaperlessBilling_Yes	0.67	0.29	0.71	0.57
Average of StreamingTV_No	0.27	0.00	0.68	0.58
Average of StreamingMovies_No	0.26	0.00	0.68	0.56
Average of InternetService_Fiber optic	0.60	0.00	0.67	0.00
Average of MultipleLines_No	0.28	0.78	0.59	0.00
Average of PaymentMethod_Electronic check	0.26	0.08	0.54	0.33
Average of gender_Male	0.50	0.51	0.50	0.52
Average of Churn_Yes	0.13	0.07	0.48	0.25
Average of MultipleLines_Yes	0.72	0.22	0.41	0.00
Average of InternetService_DSL	0.40	0.00	0.33	1.00
Average of Partner_Yes	0.69	0.48	0.33	0.46
Average of StreamingMovies_Yes	0.74	0.00	0.32	0.44
Average of StreamingTV_Yes	0.73	0.00	0.32	0.42
Average of OnlineBackup_Yes	0.70	0.00	0.25	0.42
Average of SeniorCitizen	0.18	0.03	0.22	0.15
Average of DeviceProtection_Yes	0.72	0.00	0.22	0.45
Average of OnlineSecurity_Yes	0.58	0.00	0.19	0.42
Average of PaymentMethod_Mailed check	0.10	0.49	0.18	0.23
Average of Dependents_Yes	0.37	0.42	0.18	0.30
Average of TechSupport_Yes	0.62	0.00	0.17	0.42
Average of PaymentMethod_Bank transfer (automatic)	0.32	0.22	0.14	0.21
Average of PaymentMethod_Credit card (automatic)	0.32	0.22	0.13	0.23
Average of Contract_One year	0.38	0.24	0.06	0.21
Average of Contract_Two year	0.42	0.42	0.01	0.24
Average of MultipleLines_No phone service	0.00	0.00	0.00	1.00
Average of InternetService_No	0.00	1.00	0.00	0.00
Average of OnlineSecurity_No internet service	0.00	1.00	0.00	0.00
Average of OnlineBackup_No internet service	0.00	1.00	0.00	0.00
Average of DeviceProtection_No internet service	0.00	1.00	0.00	0.00
Average of TechSupport_No internet service	0.00	1.00	0.00	0.00
Average of StreamingTV_No internet service	0.00	1.00	0.00	0.00
Average of StreamingMovies_No internet service	0.00	1.00	0.00	0.00

Accuracy for the test set:

k-NN

Testing: Classification Summary

Confusion Matrix		
Actual\Predicted	0	1
0	875	162
1	174	196

Error Report			
Class	# Cases	# Errors	% Error
0	1037	162	15.62199
1	370	174	47.02703
Overall	1407	336	23.8806

Metrics	
Metric	Value
Accuracy (#correct)	1071
Accuracy (%correct)	76.1194
Specificity	0.84378
Sensitivity (Recall)	0.52973
Precision	0.547486
F1 score	0.538462
Success Class	1
Success Probability	0.5

Logistic Regression

Testing: Classification Summary

Confusion Matrix		
Actual\Predicted	0	1
0	933	104
1	183	187

Error Report			
Class	# Cases	# Errors	% Error
0	1037	104	10.02893
1	370	183	49.45946
Overall	1407	287	20.39801

Metrics	
Metric	Value
Accuracy (#correct)	1120
Accuracy (%correct)	79.60199
Specificity	0.899711
Sensitivity (Recall)	0.505405
Precision	0.642612
F1 score	0.565809
Success Class	1
Success Probability	0.5

Classification Tree (Best Pruned)

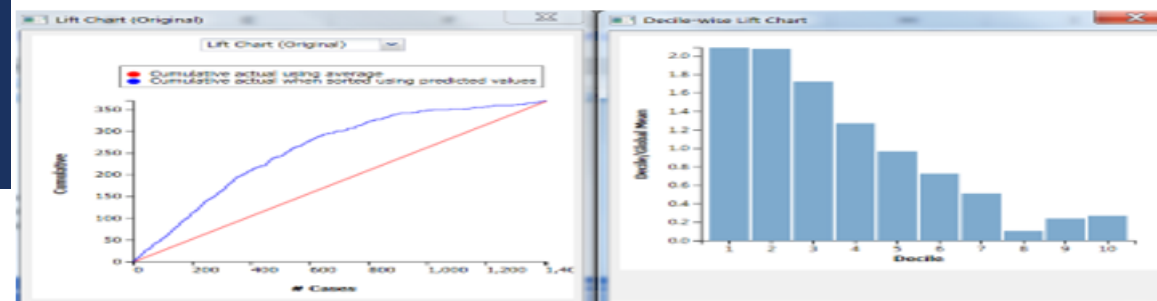
Testing: Classification Summary

Confusion Matrix		
Actual\Predicted	0	1
0	900	137
1	185	185

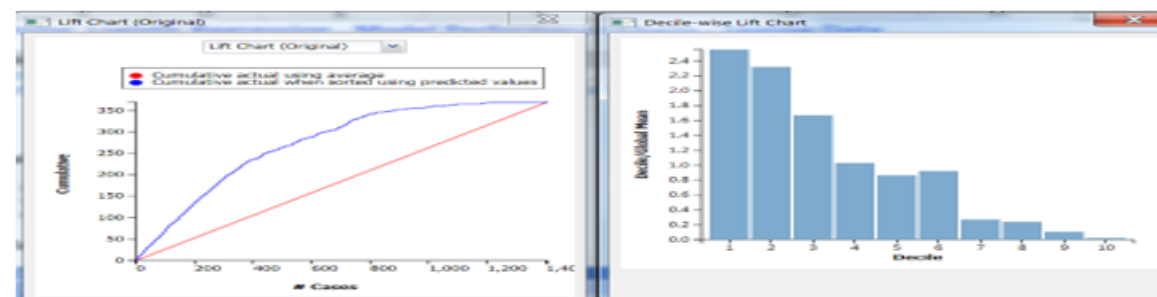
Error Report			
Class	# Cases	# Errors	% Error
0	1037	137	13.21119
1	370	185	50
Overall	1407	322	22.88557

Metrics	
Metric	Value
Accuracy (#correct)	1085
Accuracy (%correct)	77.11443
Specificity	0.867888
Sensitivity (Recall)	0.5
Precision	0.574534
F1 score	0.534682
Success Class	1
Success Probability	0.5

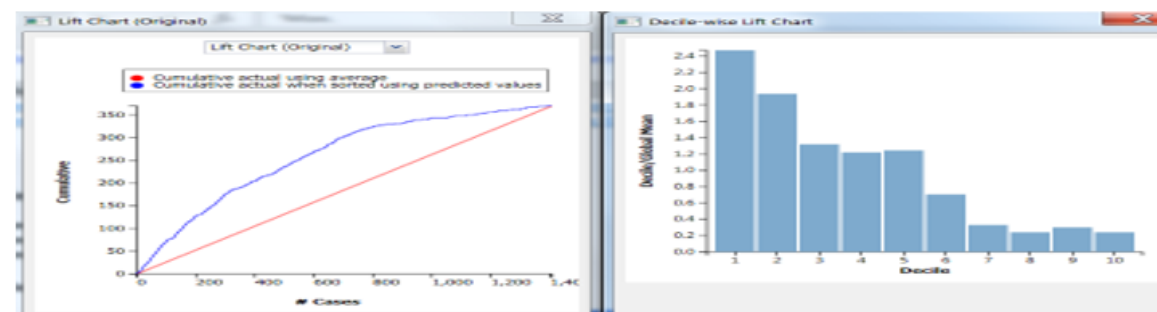
k-NN



Logistic Regression



Classification Tree (Best-Pruned)



MODEL PERFORMANCES (SUPERVISED LEARNING)

Ensembles k-NN (Bagging)

Testing: Classification Summary

Confusion Matrix		
Actual \ Predicted	0	1
0	813	224
1	196	174

Error Report			
Class	# Cases	# Errors	% Error
0	1037	224	21.60077
1	370	196	52.97297
Overall	1407	420	29.85075

Metrics	
Metric	Value
Accuracy (#correct)	987
Accuracy (%correct)	70.14925
Specificity	0.783992
Sensitivity (Recall)	0.47027
Precision	0.437186
F1 score	0.453125
Success Class	1
Success Probability	0.5

Ensembles Logistic Regression (Bagging)

Testing: Classification Summary

Confusion Matrix		
Actual \ Predicted	0	1
0	929	108
1	182	188

Error Report			
Class	# Cases	# Errors	% Error
0	1037	108	10.41466
1	370	182	49.18919
Overall	1407	290	20.61123

Metrics	
Metric	Value
Accuracy (#correct)	1117
Accuracy (%correct)	79.38877
Specificity	0.895853
Sensitivity (Recall)	0.508108
Precision	0.635135
F1 score	0.564565
Success Class	1
Success Probability	0.5

Ensembles Classification Tree (Bagging)

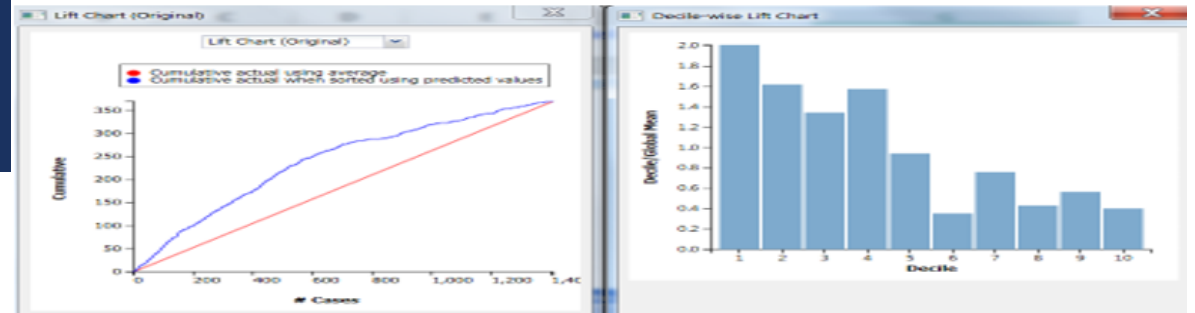
Testing: Classification Summary

Confusion Matrix		
Actual \ Predicted	0	1
0	892	145
1	171	199

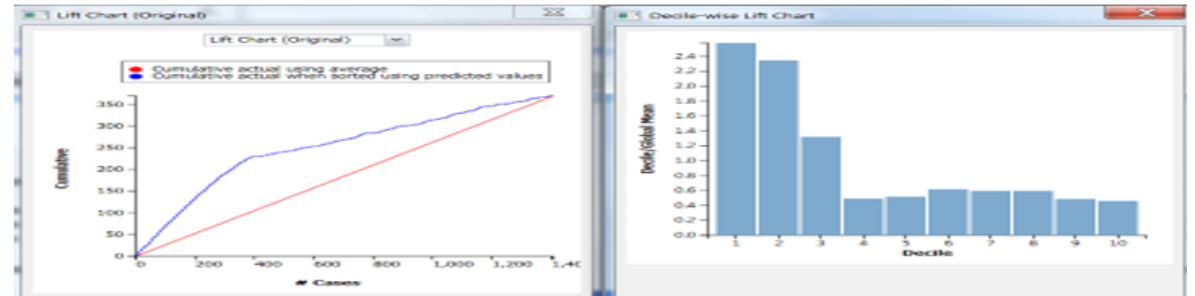
Error Report			
Class	# Cases	# Errors	% Error
0	1037	145	13.98264
1	370	171	46.21622
Overall	1407	316	22.45913

Metrics	
Metric	Value
Accuracy (#correct)	1091
Accuracy (%correct)	77.54087
Specificity	0.860174
Sensitivity (Recall)	0.537838
Precision	0.578488
F1 score	0.557423
Success Class	1
Success Probability	0.5

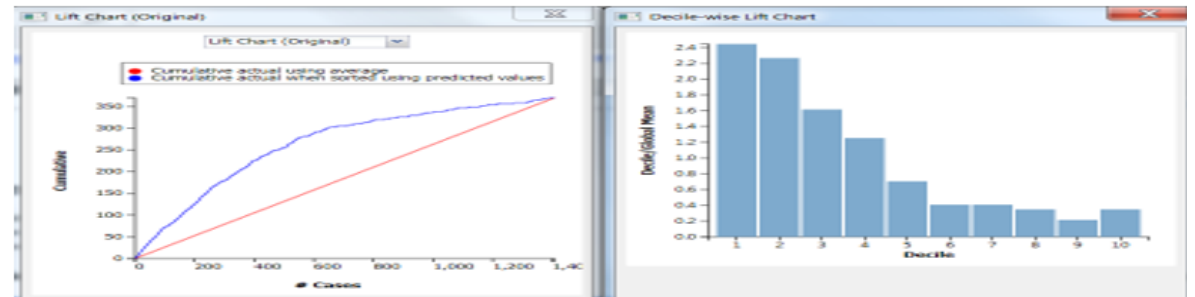
Ensembles k-NN Tree (Bagging)



Ensembles Logistic Regression (Bagging)



Ensembles Classification Tree (Bagging)



Ensembles k-NN (Boosting)

Testing: Classification Summary

Confusion Matrix		
Actual \ Predicted	0	1
0	823	214
1	206	164

Error Report			
Class	# Cases	# Errors	% Error
0	1037	214	20.63645
1	370	206	55.67568
Overall	1407	420	29.85075

Metrics	
Metric	Value
Accuracy (#correct)	987
Accuracy (%correct)	70.14925
Specificity	0.793635
Sensitivity (Recall)	0.443243
Precision	0.433862
F1 score	0.438503
Success Class	1
Success Probability	0.5

Ensembles Logistic Regression (Boosting)

Testing: Classification Summary

Confusion Matrix		
Actual \ Predicted	0	1
0	924	113
1	191	179

Error Report			
Class	# Cases	# Errors	% Error
0	1037	113	10.89682
1	370	191	51.62162
Overall	1407	304	21.60625

Metrics	
Metric	Value
Accuracy (#correct)	1103
Accuracy (%correct)	78.39375
Specificity	0.891032
Sensitivity (Recall)	0.483784
Precision	0.613014
F1 score	0.540785
Success Class	1
Success Probability	0.5

Ensembles Classification Tree (Boosting)

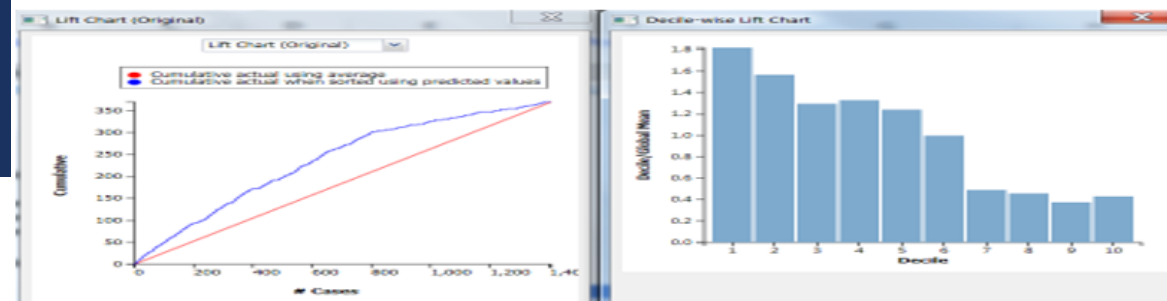
Testing: Classification Summary

Confusion Matrix		
Actual \ Predicted	0	1
0	882	155
1	183	187

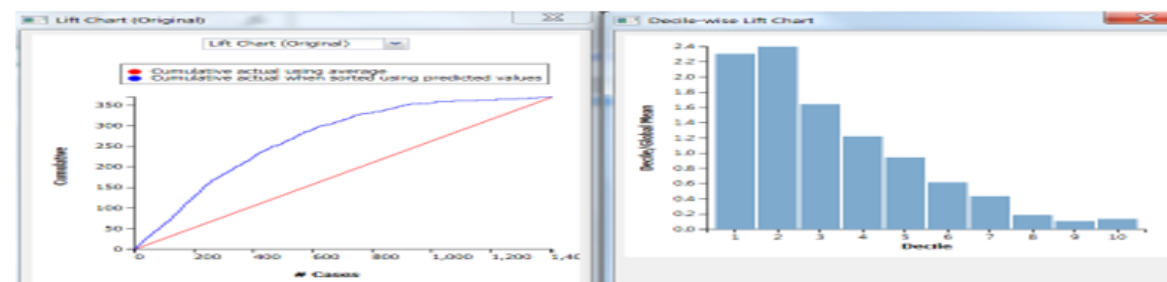
Error Report			
Class	# Cases	# Errors	% Error
0	1037	155	14.94696
1	370	183	49.45946
Overall	1407	338	24.02274

Metrics	
Metric	Value
Accuracy (#correct)	1069
Accuracy (%correct)	75.97726
Specificity	0.85053
Sensitivity (Recall)	0.505405
Precision	0.546784
F1 score	0.525281
Success Class	1
Success Probability	0.5

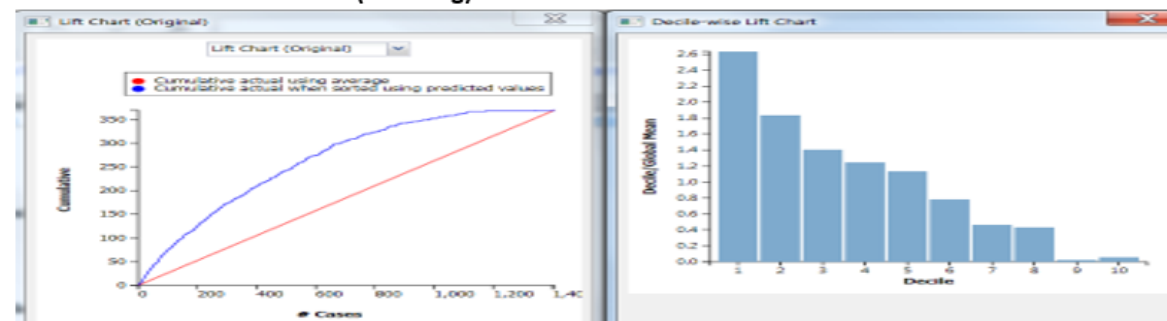
Ensembles k-NN Tree (Boosting)



Ensembles Logistic Regression (Boosting)



Ensembles Classification Tree (Boosting)



MODEL PERFORMANCES (SUPERVISED LEARNING) (CONT.)

Ensembles Classification Tree (Random Forest)

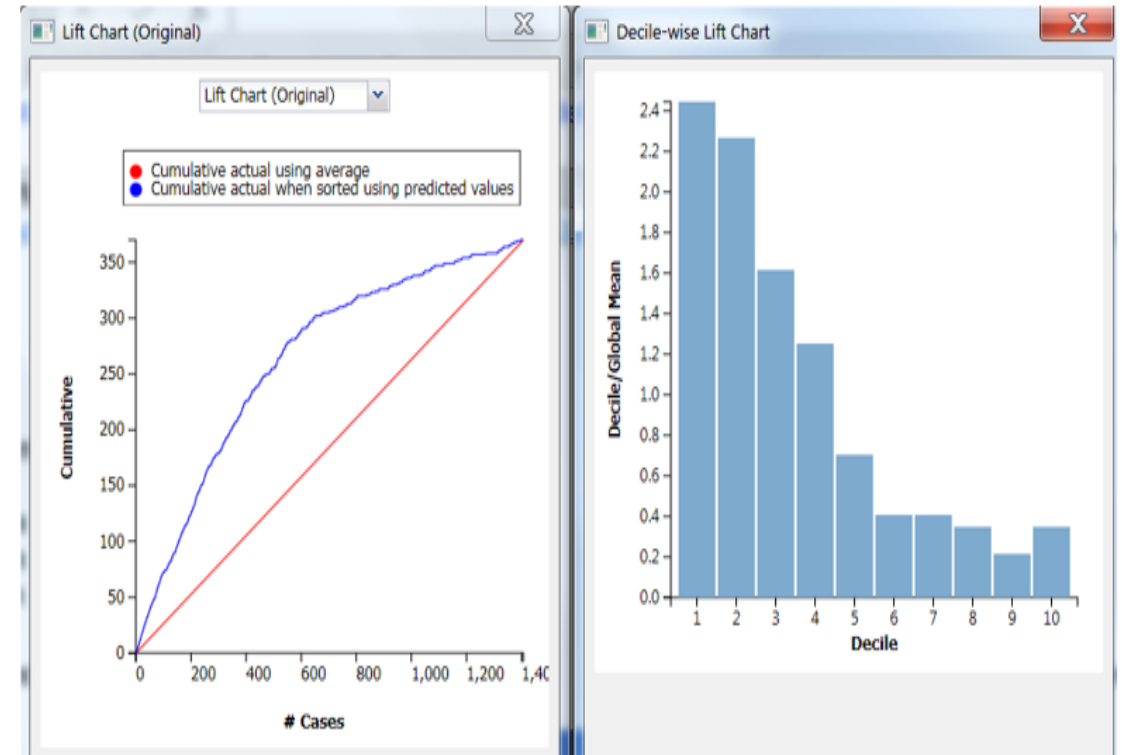
Testing: Classification Summary

Confusion Matrix			
Actual\Predicted	0	1	
0	892	145	
1	171	199	

Error Report			
Class	# Cases	# Errors	% Error
0	1037	145	13.98264
1	370	171	46.21622
Overall	1407	316	22.45913

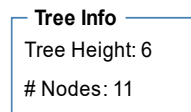
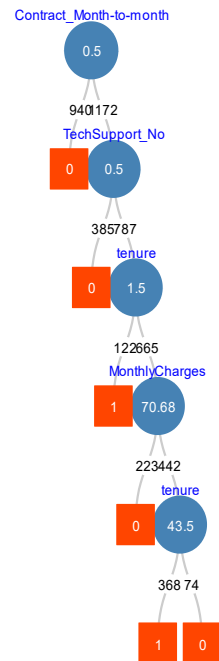
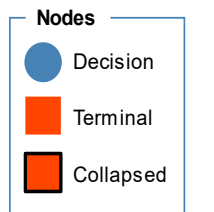
Metrics	
Metric	Value
Accuracy (#correct)	1091
Accuracy (%correct)	77.54087
Specificity	0.860174
Sensitivity (Recall)	0.537838
Precision	0.578488
F1 score	0.557423
Success Class	1
Success Probability	0.5

Ensembles Classification Tree (Random Forest)

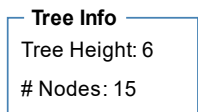
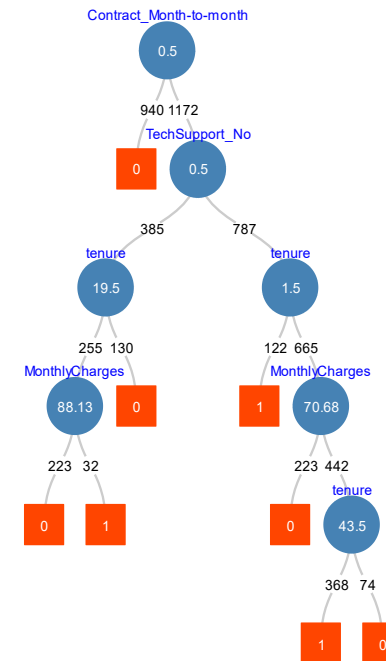
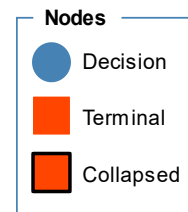


CLASSIFICATION TREE

Best-Pruned Tree



Best-Pruned and Min-Error Tree



NON-ENSEMBLES VS ENSEMBLES AND THE WINNER IS.....LOGISTIC REGRESSION!!!!



	Errors on test			Default Seed = 12345			Seed = 11111		
	Overall error rate	F1 score	New Data (4 Records)	Overall error rate	F1 score	New Data (20 Records)	Overall error rate	F1 score	New Data (20 Records)
NON-ENSEMBLES									
Classification Tree (best-pruned)	22.885	0.535	Correctly predicted	23.077	0.532	Wrongly predicted 3 records (2 Class 1 + 1 Class 0)	19.943	0.529	Wrongly predicted 2 records (Class 1)
k-NN with k=9	23.881	0.538	Correctly predicted	23.718	0.542	Wrongly predicted 5 records (2 Class 1 + 3 Class 0)	23.504	0.549	Wrongly predicted 6 records (2 Class 1 + 4 Class 0)
Logistic Regression (Explanatory) With Best-Subset 14 features	20.398	0.566	Correctly predicted	20.513	0.564	Wrongly predicted 4 records (2 Class 1 + 2 Class 0)	18.447	0.616	Wrongly predicted 4 records (2 Class 1 + 2 Class 0)
Logistic Regression (Predictive)	20.469	0.565	Correctly predicted	20.513	0.564	Wrongly predicted 4 records (2 Class 1 + 2 Class 0)	18.518	0.615	Wrongly predicted 5 records (2 Class 1 + 3 Class 0)
ENSEMBLES									
Classification Tree + Bagging	22.459	0.557	Wrongly predicted 1 record (class 1)	21.795	0.578	Wrongly predicted 4 records (2 Class 1 + 2 Class 0)	19.587	0.604	Wrongly predicted 3 records (Class 1)
Classification Tree + Boosting	24.023	0.525	Correctly predicted	23.932	0.504	Wrongly predicted 4 records (2 Class 1 + 2 Class 0)	21.724	0.555	Wrongly predicted 5 records (2 Class 1 + 3 Class 0)
Classification Tree + Random Forest	22.459	0.557	Wrongly predicted 1 record (class 1)	21.866	0.552	Wrongly predicted 4 records (2 Class 1 + 2 Class 0)	19.444	0.607	Wrongly predicted 4 records (2 Class 1 + 2 Class 0)
k-NN with k=9 + Bagging	29.851	0.453	Excel Froze	29.843	0.451	Excel Froze	28.134	0.493	Excel Froze
k-NN with k=9 + Boosting	29.851	0.438	Excel Froze	29.843	0.452	Excel Froze	29.416	0.474	Excel Froze
Logistic Regression + Bagging	20.611	0.564	Correctly predicted	20.442	0.567	Wrongly predicted 4 records (2 Class 1 + 2 Class 0)	18.732	0.617	Wrongly predicted 5 records (2 Class 1 + 3 Class 0)
Logistic Regression + Boosting	21.606	0.541	Correctly predicted	22.08	0.548	Wrongly predicted 5 records (2 Class 1 + 3 Class 0)	19.088	0.607	Wrongly predicted 6 records (2 Class 1 + 4 Class 0)

Coefficients

Predictor	Estimate	Confidence Interval: Lower	Confidence Interval: Upper	Odds	Standard Error	Chi2-Statistic	P-Value
Intercept	1.7193768	0.607560995	2.831192537	5.581049	0.567263368	9.186977002	0.002437
tenure	-0.0317816	-0.037805114	-0.025757987	0.968718	0.003073303	106.9398122	4.59E-25
MonthlyCharges	-0.0244473	-0.039413602	-0.009481009	0.975849	0.007636006	10.25013319	0.001367
SeniorCitizen	0.3002219	0.070256315	0.530187396	1.350158	0.117331514	6.547194818	0.010505
MultipleLines_Yes	0.4137356	0.169333517	0.658137667	1.512457	0.124697228	11.008602	0.000907
InternetService_DSL	-1.3439863	-1.86750151	-0.820471117	0.260804	0.267104498	25.31787718	4.86E-07
OnlineSecurity_Yes	-0.3928511	-0.643325663	-0.14237644	0.675129	0.127795518	9.449843088	0.002112
TechSupport_Yes	-0.3971648	-0.66243849	-0.131891146	0.672223	0.135346197	8.610917411	0.003342
StreamingTV_No internet service	-2.7933282	-3.659848817	-1.926807592	0.061217	0.442110477	39.91924499	2.65E-10
StreamingTV_Yes	0.5481276	0.2762437	0.820011598	1.730011	0.138718849	15.61321941	7.77E-05
StreamingMovies_Yes	0.5788997	0.311009496	0.846789871	1.784074	0.136681179	17.93862673	2.28E-05
Contract_One year	-0.773316	-1.070306948	-0.476325122	0.46148	0.15152876	26.04497169	3.34E-07
Contract_Two year	-1.3644462	-1.845173156	-0.883719312	0.255522	0.245273345	30.94654344	2.65E-08
PaperlessBilling_Yes	0.4515634	0.243084878	0.660041903	1.570766	0.106368542	18.02233401	2.18E-05
PaymentMethod_Electronic check	0.3066281	0.113738244	0.499517924	1.358836	0.098414992	9.707364458	0.001835

Training: Classification Summary

Confusion Matrix			
Actual\Predicted	0	1	
0	2338	256	
1	413	513	

Error Report			
Class	# Cases	# Errors	% Error
0	2594	256	9.868928
1	926	413	44.60043
Overall	3520	669	19.00568

Metrics	
Metric	Value
Accuracy (#correct)	2851
Accuracy (%correct)	80.99432
Specificity	0.901311
Sensitivity (Recall)	0.553996
Precision	0.6671
F1 score	0.60531
Success Class	1
Success Probability	0.5

Validation: Classification Summary

Confusion Matrix			
Actual\Predicted	0	1	
0	1384	156	
1	261	311	

Error Report			
Class	# Cases	# Errors	% Error
0	1540	156	10.12987
1	572	261	45.62937
Overall	2112	417	19.74432

Metrics	
Metric	Value
Accuracy (#correct)	1695
Accuracy (%correct)	80.25568
Specificity	0.898701
Sensitivity (Recall)	0.543706
Precision	0.665953
F1 score	0.598653
Success Class	1
Success Probability	0.5

Testing: Classification Summary

Confusion Matrix			
Actual\Predicted	0	1	
0	933	104	
1	183	187	

Error Report			
Class	# Cases	# Errors	% Error
0	1037	104	10.02893
1	370	183	49.45946
Overall	1407	287	20.39801

Metrics	
Metric	Value
Accuracy (#correct)	1120
Accuracy (%correct)	79.60199
Specificity	0.899711
Sensitivity (Recall)	0.505405
Precision	0.642612
F1 score	0.565809
Success Class	1
Success Probability	0.5

SUMMARY OF ANALYSIS

Higher churn rate on following attributes

Attribute	Priority
Month-to-month Contract	1
Low tenure	1
No TechSupport	2
Higher than average MonthlyCharges	2
No OnlineSecurity	3
No DeviceProtection	4
No OnlineBackup	5
PaperlessBilling	6
No StreamingTV	7
No StreamingMovies	8
Fiber optic InternetService	9
Electronic check PaymentMethod	10
SeniorCitizen	11
No Partners	12
No Dependents	13
Lower than average TotalCharges	14

- Simplified Customer Lifetime Value (CLV) ~ Average of TotalCharges per cluster. Shows “Fiber + month-to-month contract” has significant CLV and should not be ignored.

■ Recommendations:

- Provide promotional offers or discounts on additional services to this segment
- Incentivize customers and persuade them to switch from monthly contract to long-term contract.



Q&A

THANK YOU