

# Bank Churner

Title: Bank Churner	Objective: To reduce churn rate	Numeric & Categorical variables	Numeric variables - T-test	Total Relationship Count: 3 or less	Month Inactive: 3 or more	Contacts count: 3 or ..
---------------------	---------------------------------	---------------------------------	----------------------------	-------------------------------------	---------------------------	-------------------------



## BANK CHURNER

[ Reduce Churn Rate]

Presented by Seul Lee  
Nov.13.2021

# Bank Churner

Title: Bank Churner	Objective: To reduce churn rate	Numeric & Categorical variables	Numeric variables - T-test	Total Relationship Count: 3 or less	Month Inactive: 3 or more	Contacts count: 3 or ..
---------------------	---------------------------------	---------------------------------	----------------------------	-------------------------------------	---------------------------	-------------------------

## Problem Statement

Churn is a serious issue because it costs a lot for customer to stop doing business. For example, the Harvard Business School report claims that on average, a **5% increase** in customer retention rates results in 25% - 95% increase of profits.

The bank decided to **find a group of customers that is more likely to leave** than others and **offer better marketing service to reduce churn rate**.

This data consists of **10,127** customers mentioning their gender, salary, marital status, credit card limit, etc.

Current churn rate is **16.07%**.

## Churn Rate

Attrition Flag	
Attrited Customer	16.07% 1,627
Existing Customer	83.93% 8,500

# Bank Churner

Title: Bank Churner	Objective: To reduce churn rate	Numeric & Categorical variables	Numeric variables - T-test	Total Relationship Count: 3 or less	Month Inactive: 3 or more	Contacts count: 3 or ..
---------------------	---------------------------------	---------------------------------	----------------------------	-------------------------------------	---------------------------	-------------------------

## Variables in dataset

Divided the variables into numeric and categorical variables.

Numeric Variables	Categorical Variables
Customer Age	Gender
Months on book	Education Level
Total Relationship Count	Marital Status
Months Inactive 12 mon	Income Category
Contacts Count 12 mon	Card Category
Credit Limit	Dependent Count
Total Revolving Bal	
Avg Open To Buy	
Total Amt Chng Q4 Q1	
Total Trans Amt	
Total Trans Ct	
Total Ct Chng Q4 Q1	
Avg Utilization Ratio	

# Bank Churner

Title: Bank Churner	Objective: To reduce churn rate	Numeric & Categorical variables	Numeric variables - T-test	Total Relationship Count: 3 or less	Month Inactive: 3 or more	Contacts count: 3 or more
---------------------	---------------------------------	---------------------------------	----------------------------	-------------------------------------	---------------------------	---------------------------

## T-test

Implemented T-test with numeric variables to establish significance.

\*Attrited customer and existing customer classes are relevelled to 1 and 0 respectively.

**Significant numeric variables are :**  
*Total\_Relationship\_Count / Months\_Inactive\_12\_mon*  
*Contacts\_Count\_12\_mon / Credit\_Limit*  
*Total\_Revolving\_Bal / Avg\_Utilization\_Ratio*  
*Total\_Amt\_Chng\_Q4\_Q1 / Total\_Ct\_Chng\_Q4\_Q1*  
*Total\_Trans\_Amt / Total\_Trans\_Ct*

t-Test: Two-Sample Assuming Equal Variances		
	Contacts_Count_12_mon_0	Contacts_Count_12_mon_1
Mean	2.356352941	2.972341733
Variance	1.169503111	1.18927145
Observations	8500	1627
Pooled Variance	1.17267776	
Hypothesized Mean Difference	0	
df	10125	
t Stat	-21.0206577	
P(T<=t) one-tail	2.34874E-96	
t Critical one-tail	1.645004137	
P(T<=t) two-tail	4.69749E-96	
t Critical two-tail	1.96019831	

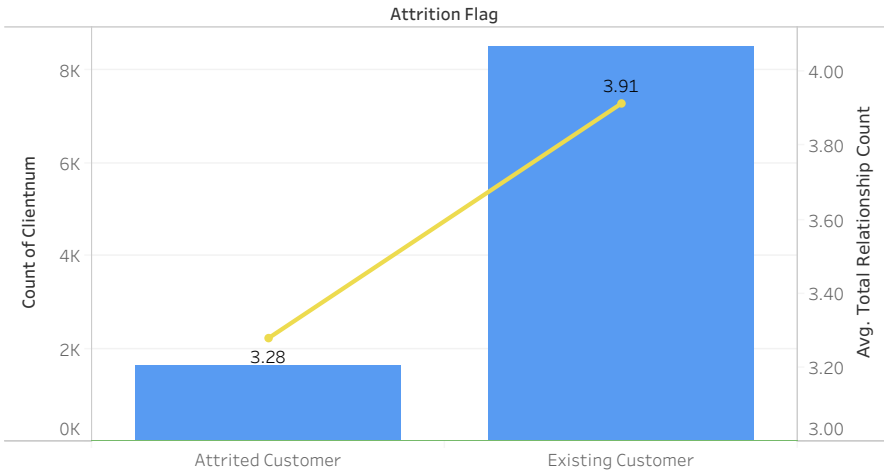
t-Test: Two-Sample Assuming Unequal Variances		
	Total_Trans_Ct_0	Total_Trans_Ct_1
Mean	68.67258824	44.93362016
Variance	525.2810699	212.2391334
Observations	8500	1627
Hypothesized Mean Difference	0	
df	3386	
t Stat	54.14187251	
P(T<=t) one-tail	0	
t Critical one-tail	1.645303771	
P(T<=t) two-tail	0	
t Critical two-tail	1.960664842	

# Bank Churner

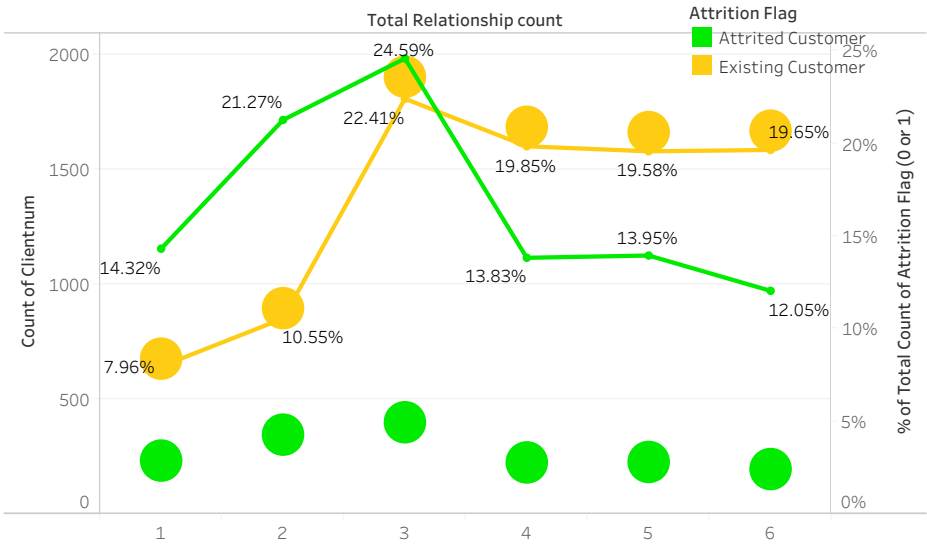
Objective: To reduce churn rate	Numeric & Categoric variables	Numeric variables - T-test	Total Relationship Count: 3 or less	Month Inactive: 3 or more	Contacts count: 3 or more	Avg. Total Amount & Count Change (Q4 o..
---------------------------------	-------------------------------	----------------------------	-------------------------------------	---------------------------	---------------------------	--

## Total Relationship Count

Attrition customer class has **lower** average total relationship count.



Customers with **3 or less** total relationship count has **a higher proportion of attrition**.

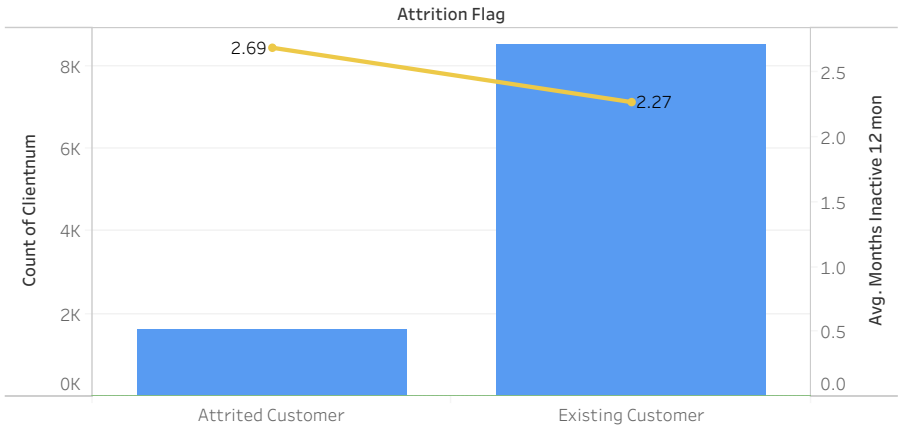


# Bank Churner

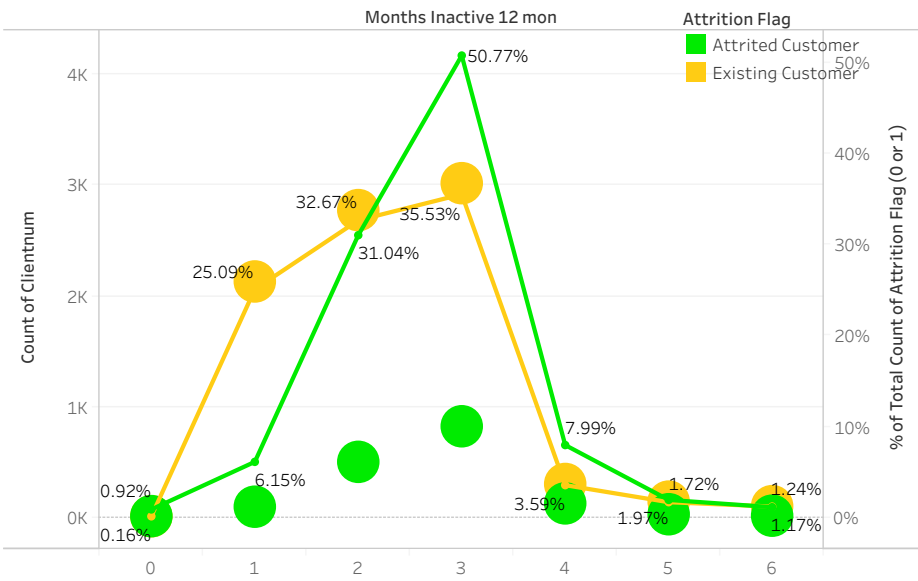
Numeric & Categori	Numeric variables - T-test	Total Relationship Count: 3 or less	Month Inactive: 3 or more	Contacts count: 3 or more	Avg. Total Amount & Count Change (Q4 ove..	Avg. Utilization Ratio: Lower avg.
--------------------	----------------------------	-------------------------------------	---------------------------	---------------------------	--	------------------------------------

## Month Inactive Count

Attrition customer class has **higher** average inactive months count in the past 12 months.



Customers with **3 or more** inactive in the past 12 months have a higher proportion of attrition.

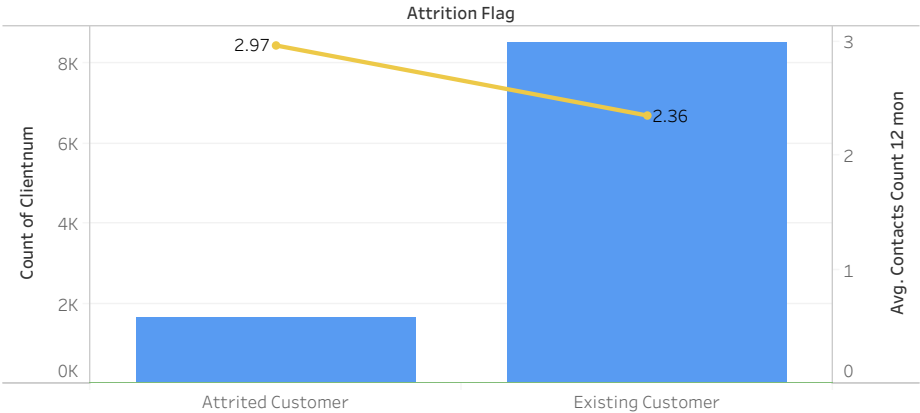


# Bank Churner

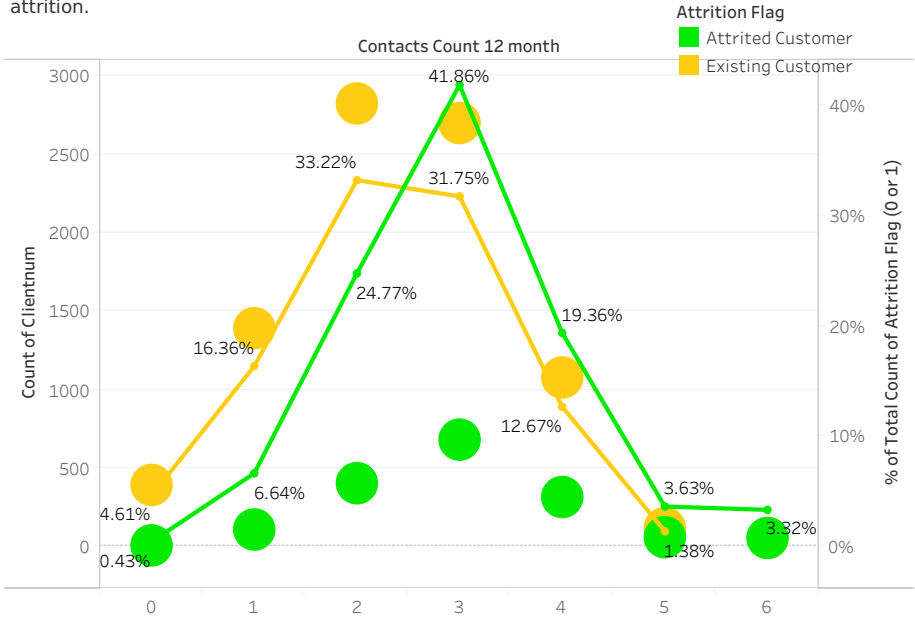
Numeric variables - T-test	Total Relationship Count: 3 or less	Month Inactive: 3 or more	Contacts count: 3 or more	Avg. Total Amount & Count Change (Q4 ove..	Avg. Utilization Ratio: Lower avg.	Transaction Count & Amount Scatter Plot
----------------------------	-------------------------------------	---------------------------	---------------------------	--	------------------------------------	---

## Contacts Count

Attrition customer class has **higher** average contacts count in the past 12 months.



Customers with **3 or more** contacts count in the past 12 months have a higher proportion of attrition.

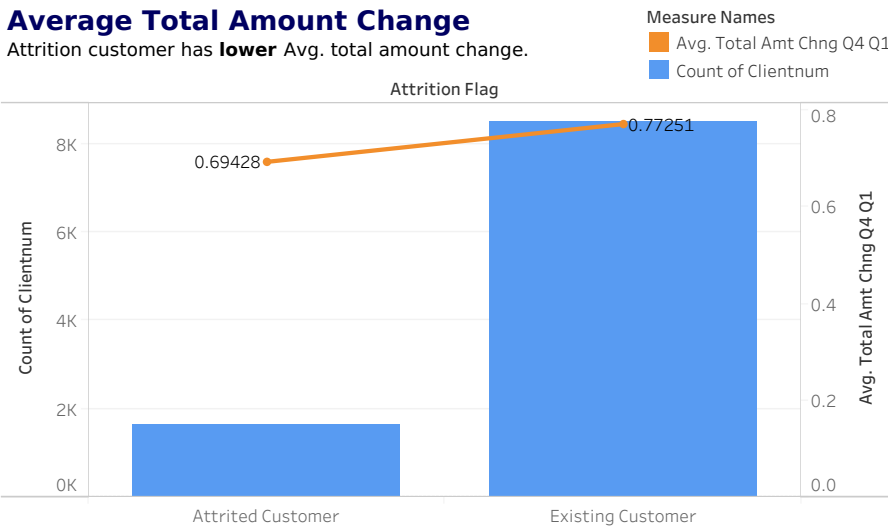


# Bank Churner

Total Relationship Co..	Month Inactive: 3 or more	Contacts count: 3 or more	Avg. Total Amount & Count Change (Q4 ove..	Avg. Utilization Ratio: Lower avg.	Transaction Count & Amount Scatter Plot	Transaction Count Box Plot
-------------------------	---------------------------	---------------------------	--	------------------------------------	---	----------------------------

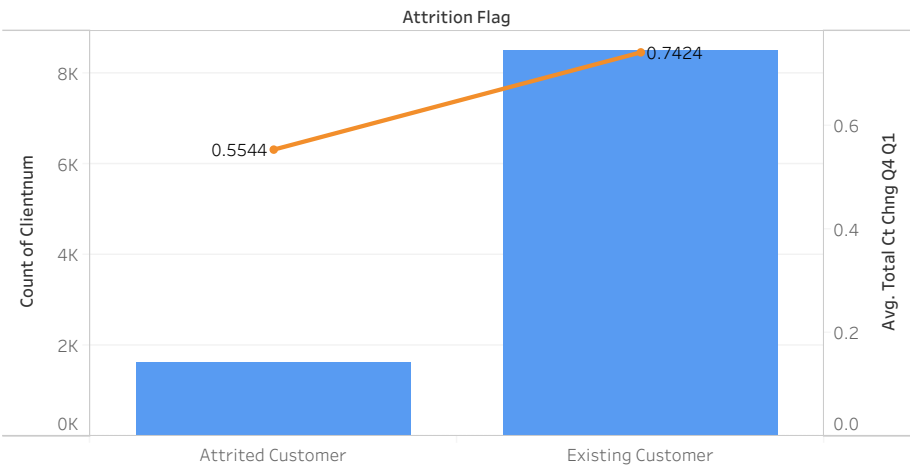
## Average Total Amount Change

Attrition customer has **lower** Avg. total amount change.



## Average Total Count Change

Attrition customer has **lower** Avg. total count change.

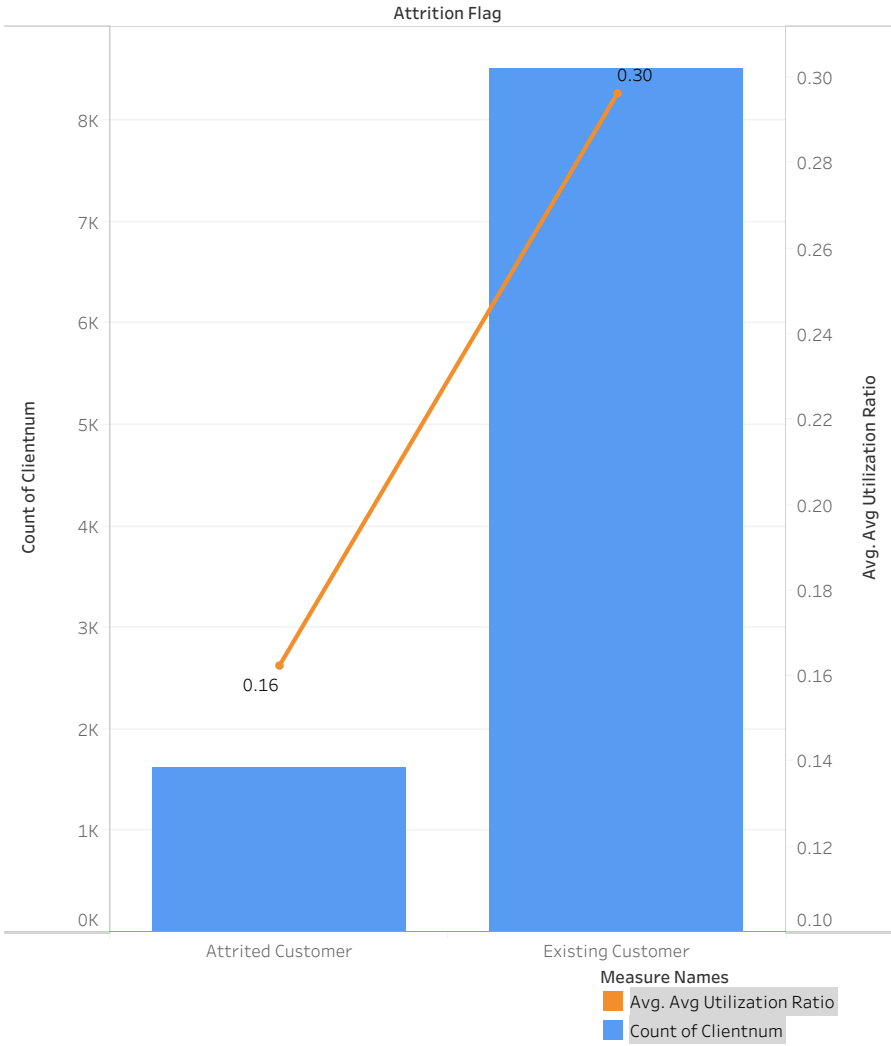




# Bank Churner

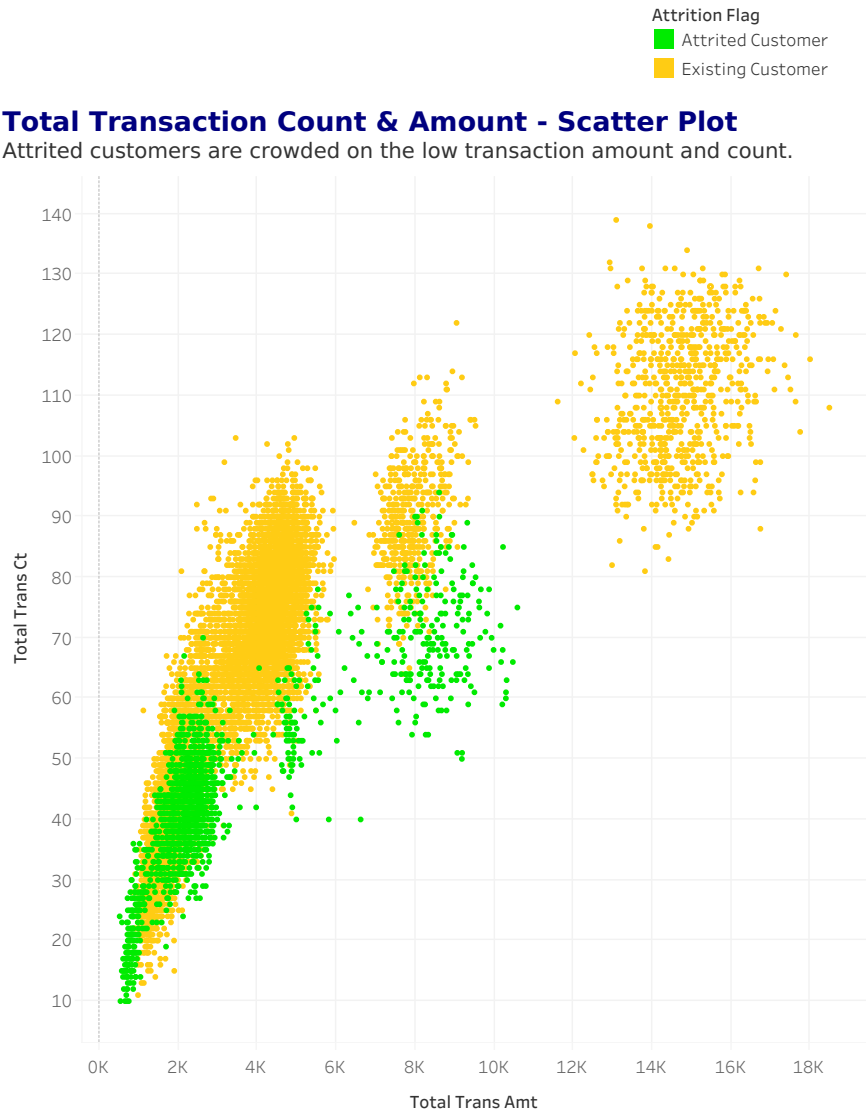
Month Inactive: 3 or more	Contacts count: 3 or more	Avg. Total Amount & Count Change (Q4 ove..	Avg. Utilization Ratio: Lower avg.	Transaction Count & Amount Scatter Plot	Transaction Count Box Plot	Categoric Variables - IV
---------------------------	---------------------------	--	------------------------------------	---	----------------------------	--------------------------

**Avg. Utilization Ratio**  
Attrition customer class has **lower** average utilization ration.



# Bank Churner

Contacts count: 3 or more	Avg. Total Amount & Count Change (Q4 ove..	Avg. Utilization Ratio: Lower avg.	Transaction Count & Amount Scatter Plot	Transaction Count Box Plot	Categoric Variables - IV	Logistic Regression: Build a predictive ..
---------------------------	--	------------------------------------	---	----------------------------	--------------------------	--

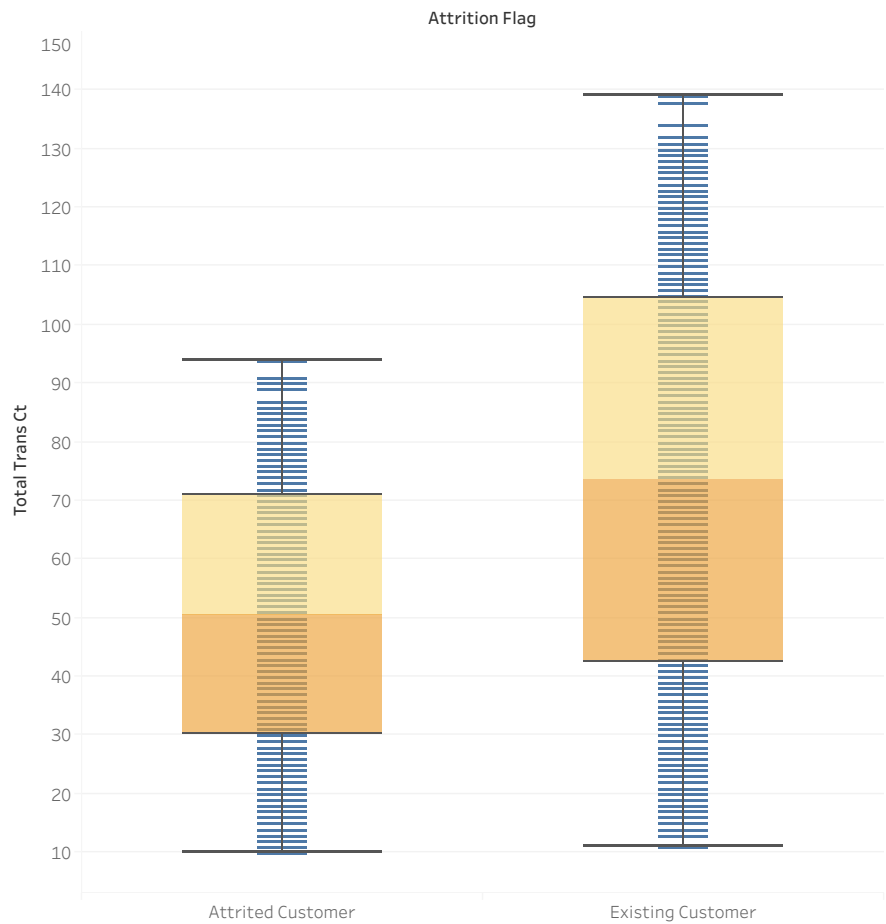


# Bank Churner

Avg. Total Amount & Count ..	Avg. Utilization Ratio: Lower avg.	Transaction Count & Amount Scatter Plot	Transaction Count Box Plot	Categoric Variables - IV	Logistic Regression: Build a predictive mo..	Gain and Lift Chart
------------------------------	------------------------------------	---	----------------------------	--------------------------	--	---------------------

## Total Transaction Count - Box Plot

**Attrited Customer**  
Min : 10    Max : 94    Median : 50.5  
**Existing Customer**  
Min : 11    Max : 139    Median : 73.5



# Bank Churner

Avg. Utilization Ratio: Lower avg.	Transaction Count & Amount Scatter Plot	Transaction Count Box Plot	Categoric Variables - IV	Logistic Regression: Build a predictive mo..	Gain and Lift Chart	Summary: The predictive model p..
------------------------------------	---	----------------------------	--------------------------	--	---------------------	-----------------------------------

**Categoric Variables - Information Value**  
Tested Infomation Value to establish significance of categoric variables.  
All the categoric variables are useless for prediction.

Education Level								
Row Labels	Attrited C	Existing C	Total	Churn rate	Attrited %	Existing %	Population%	IV
College	154	859	1013	15%	9%	10%	10%	0.000419522
Doctorate	95	356	451	21%	6%	4%	4%	0.005484962
Graduate	487	2641	3128	16%	30%	31%	31%	0.00042478
High School	306	1707	2013	15%	19%	20%	20%	0.00083596
Post-Graduate	92	424	516	18%	6%	5%	5%	0.000835485
Uneducated	237	1250	1487	16%	15%	15%	15%	1.3238E-05
Unknown	256	1263	1519	17%	16%	15%	15%	0.000501407
Grand Total	1627	8500	10127	16%	100%	100%	100%	0.008515354

Card Category								
Row Labels	Attrited C	Existing C	Total	Churn rate	Attrited %	Existing %	Population%	IV
Blue	1519	7917	9436	16%	93%	93%	93%	5.22995E-06
Gold	21	95	116	18%	1%	1%	1%	0.000249179
Platinum	5	15	20	25%	0%	0%	0%	0.00072581
Silver	82	473	555	15%	5%	6%	5%	0.000519758
Grand Total	1627	8500	10127	16%	100%	100%	100%	0.001499977

Card Category								
Row Labels	Attrited C	Existing C	Total	Churn rate	Attrited %	Existing %	Population%	IV
Blue	1519	7917	9436	16%	93%	93%	93%	5.22995E-06
Gold	21	95	116	18%	1%	1%	1%	0.000249179
Platinum	5	15	20	25%	0%	0%	0%	0.00072581
Silver	82	473	555	15%	5%	6%	5%	0.000519758
Grand Total	1627	8500	10127	16%	100%	100%	100%	0.001499977

Income Category								
Row Labels	Attrited C	Existing C	Total	Churn rate	Attrited %	Existing %	Population%	IV
\$120K +	126	601	727	17%	8%	7%	7%	0.000613194
\$40K - \$60K	271	1519	1790	15%	17%	18%	18%	0.000854291
\$60K - \$80K	189	1213	1402	13%	12%	14%	14%	0.005461549
\$80K - \$120K	242	1293	1535	16%	15%	15%	15%	7.58425E-05
Less than \$40K	612	2949	3561	17%	38%	35%	35%	0.002361414
Unknown	187	925	1112	17%	11%	11%	11%	0.000333976
Grand Total	1627	8500	10127	16%	100%	100%	100%	0.009700267

# Bank Churner

Avg. Utilization R..	Transaction Count & Amount Scatter Plot	Transaction Count Box Plot	Categoric Variables - IV	Logistic Regression: Build a predictive mo..	Gain and Lift Chart	Summary: The predictive model perf..
----------------------	---	----------------------------	--------------------------	--	---------------------	--------------------------------------

## Logistic Regression

Combine significant variables and come up with equation building a predictive model.

	Coefficients	andard Err	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	0.901868012	0.020588	43.80524	0	0.861511	0.942225	0.861511	0.942224837
Total_Relationship_Count	-0.044310603	0.002067	-21.4387	9.4E-100	-0.04836	-0.04026	-0.04836	-0.04025917
Months_Inactive_12_mon	0.044669653	0.002966	15.05925	1.06E-50	0.038855	0.050484	0.038855	0.050484113
Contacts_Count_12_mon	0.041803019	0.002751	15.19805	1.35E-51	0.036411	0.047195	0.036411	0.047194645
Total_Amt_Chng_Q4_Q1	-0.069834001	0.014851	-4.70217	2.61E-06	-0.09895	-0.04072	-0.09895	-0.040722235
Total_Trans_Amt	2.85766E-05	1.57E-06	18.2437	3.44E-73	2.55E-05	3.16E-05	2.55E-05	3.1647E-05
Total_Trans_Ct	-0.009119931	0.00022	-41.4393	0	-0.00955	-0.00869	-0.00955	-0.008688533
Total_Ct_Chng_Q4_Q1	-0.305434882	0.013797	-22.1378	4.4E-106	-0.33248	-0.27839	-0.33248	-0.27839011
Avg_Utilization_Ratio	-0.157258245	0.011016	-14.2754	8.66E-46	-0.17885	-0.13566	-0.17885	-0.135664621

	Coefficients
Intercept	0.901868012
Total_Relationship_Count	-0.044310603
Months_Inactive_12_mon	0.044669653
Contacts_Count_12_mon	0.041803019
Total_Amt_Chng_Q4_Q1	-0.069834001
Total_Trans_Amt	2.85766E-05
Total_Trans_Ct	-0.009119931
Total_Ct_Chng_Q4_Q1	-0.305434882
Avg_Utilization_Ratio	-0.157258245

# Bank Churner

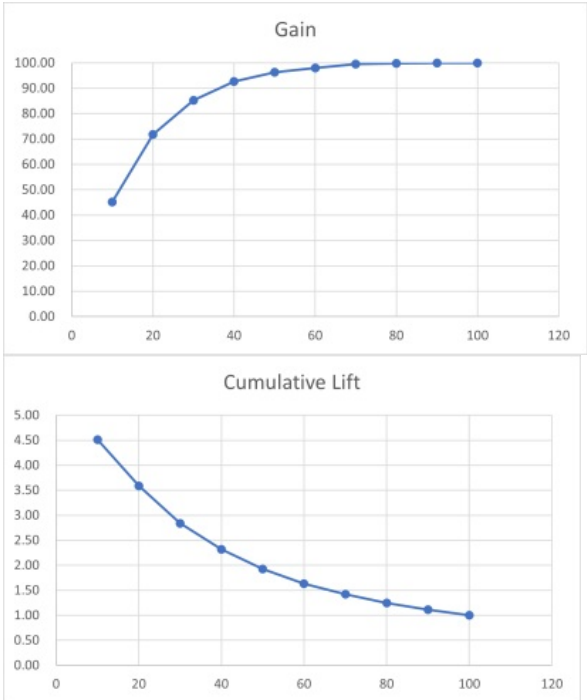
Avg. Utilization R..	Transaction Count & Amount Scatter Plot	Transaction Count Box Plot	Categoric Variables - IV	Logistic Regression: Build a predictive mo..	Gain and Lift Chart	Summary: The predictive model perf..
----------------------	---	----------------------------	--------------------------	--	---------------------	--------------------------------------

## Gain and Lift Charts

Validate the predictive model with classification evaluation.

These gain and lift charts prove that the predictive model performances better than random selection.

% of data sets	Decile	Number of Cases	Number of Responses	Cumulative Responses	% of events	Gain	Cumulative Lift
10	0.9-1	1012.7	734	734	45.11	45.11	4.51
20	0.8 ~ 0.899	1012.7	434	1168	26.67	71.79	3.59
30	0.7~0.799	1012.7	218	1386	13.40	85.19	2.84
40	0.6~0.699	1012.7	122	1508	7.50	92.69	2.32
50	0.5~0.599	1012.7	59	1567	3.63	96.31	1.93
60	0.4~0.499	1012.7	28	1595	1.72	98.03	1.63
70	0.3~0.399	1012.7	25	1620	1.54	99.57	1.42
80	0.2~0.299	1012.7	4	1624	0.25	99.82	1.25
90	0.1~0.199	1012.7	3	1627	0.18	100.00	1.11
100	0~0.099	1012.7	0	1627	0.00	100.00	1.00
Total		10127	1627				



# Bank Churner

Avg. Utilization R..	Transaction Count & Amount Scatter Plot	Transaction Count Box Plot	Categoric Variables - IV	Logistic Regression: Build a predictive mo..	Gain and Lift Chart	Summary: The predictive model perf..
----------------------	---	----------------------------	--------------------------	--	---------------------	--------------------------------------

## Summary

**When selecting 10% of the records based on the predictive model, we can expect 4.51 times the total number of targets found by randomly selecting 10%-of-file without a model.**

**With this predictive model, the bank is able to have higher chance to find the customer group that is most likely drop off the card service.**