Churn Prediction

Denis Orina

11/15/2021

Introduction

Objectives

- Predict Churn probabilities of customers and minimize.

Data Source

1. Customer Listing
2. Transactions table
3. Loan Book

Data Analysis

Preliminary Analysis

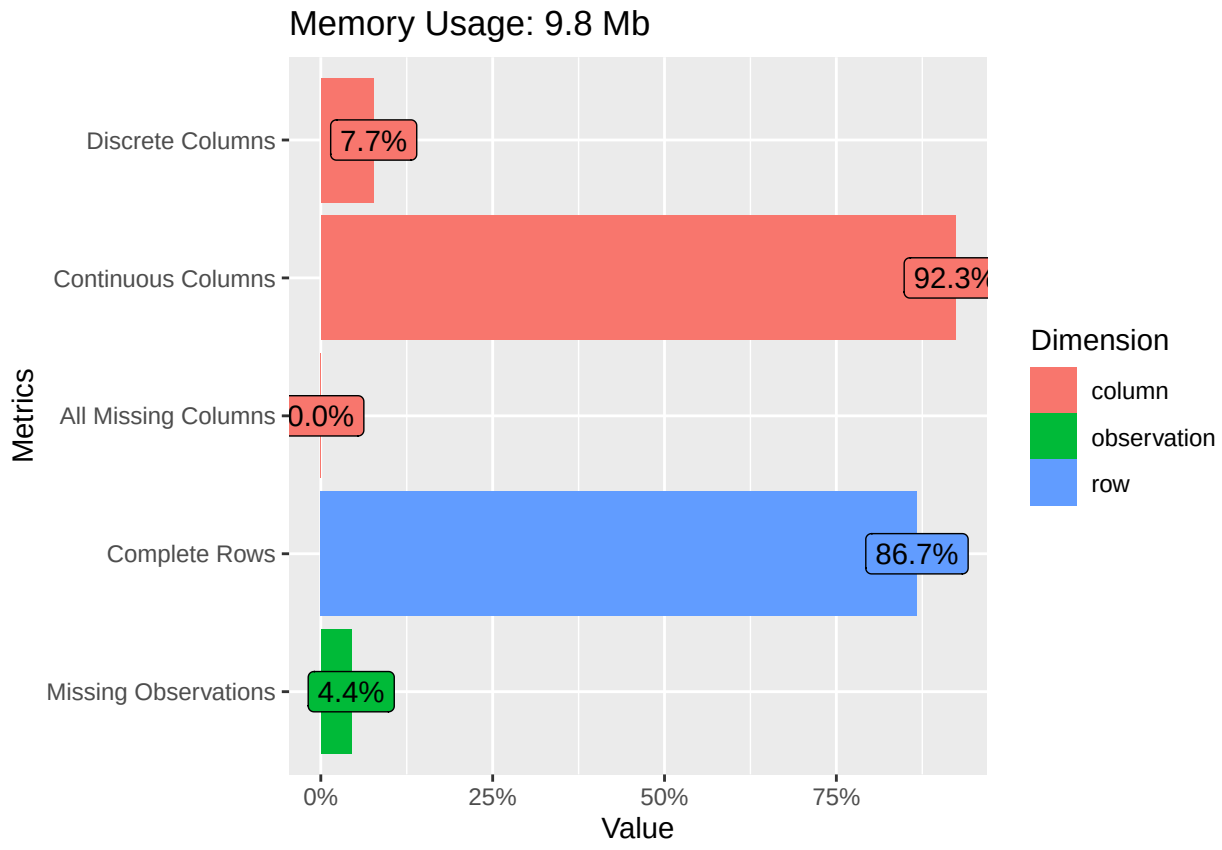
Overview

```
## Warning: The ‘x‘ argument of ‘as_tibble.matrix()‘ must have unique column names if ‘.name_repair‘ is
## Using compatibility ‘.name_repair‘.
```

```
## Warning: Warning: fonts used in ‘flextable‘ are ignored because the ‘pdflatex‘
## engine is used and not ‘xelatex‘ or ‘lualatex‘. You can avoid this warning
## by using the ‘set_flextable_defaults(fonts_ignore=TRUE)‘ command or use a
## compatible engine by defining ‘latex_engine: xelatex‘ in the YAML header of the
## R Markdown document.
```

Variable	Value
rows	32,506
columns	39
discrete_columns	3
continuous_columns	36

Variable	Value
all_missing_columns	0
total_missing_values	56,095
complete_rows	28,191
total_observations	1,267,734




```
## Warning: Warning: fonts used in ‘flextable‘ are ignored because the ‘pdflatex‘
## engine is used and not ‘xelatex‘ or ‘lualatex‘. You can avoid this warning
## by using the ‘set_flextable_defaults(fonts_ignore=TRUE)‘ command or use a
## compatible engine by defining ‘latex_engine: xelatex‘ in the YAML header of the
## R Markdown document.
```

Table 2: Dataset Variables

Column	**Explanation**	**Datatype**
ClientID		
NumberOfProducts		
Tenure		

Column	Explanation	Datatype
No_of_transactions		
No_of_transactions_Deposits		
TotalDepositsMade		
Max_DepositsMade		
Average_DepositsMade		
No_of_transactions_Withdrawals		
WithdrawalMade		
Max_WithdrawalMade		
LastTrxDate		
Days_Since_LastTransaction		
transactions_made_12		
transactions_made_6		
transactions_made_3		
deposits_12		
deposits_6		
deposits_3		
withdrawals_12		
withdrawals_6		
withdrawals_3		
last_deposit_amount		
last_withdrawal_amount		
Max_LoanTaken		
Min_LoanTaken		
ave_LoanTaken		
no_of_LoanTaken		
LastLoanDate		
DaysSince_Last_Loan		
LastLoanTaken		
FirstLoanTaken		
Has_ActiveLoan		
Max_ArrearsDays		
First_Loan_Last_Loan_Ratio		
First_Loan_Last_Loan_diff		
Last_Loan_First_Loan_Ratio		
Trx_Status		

Column	Explanation	Datatype
Churn		

Exploratory Data Analysis

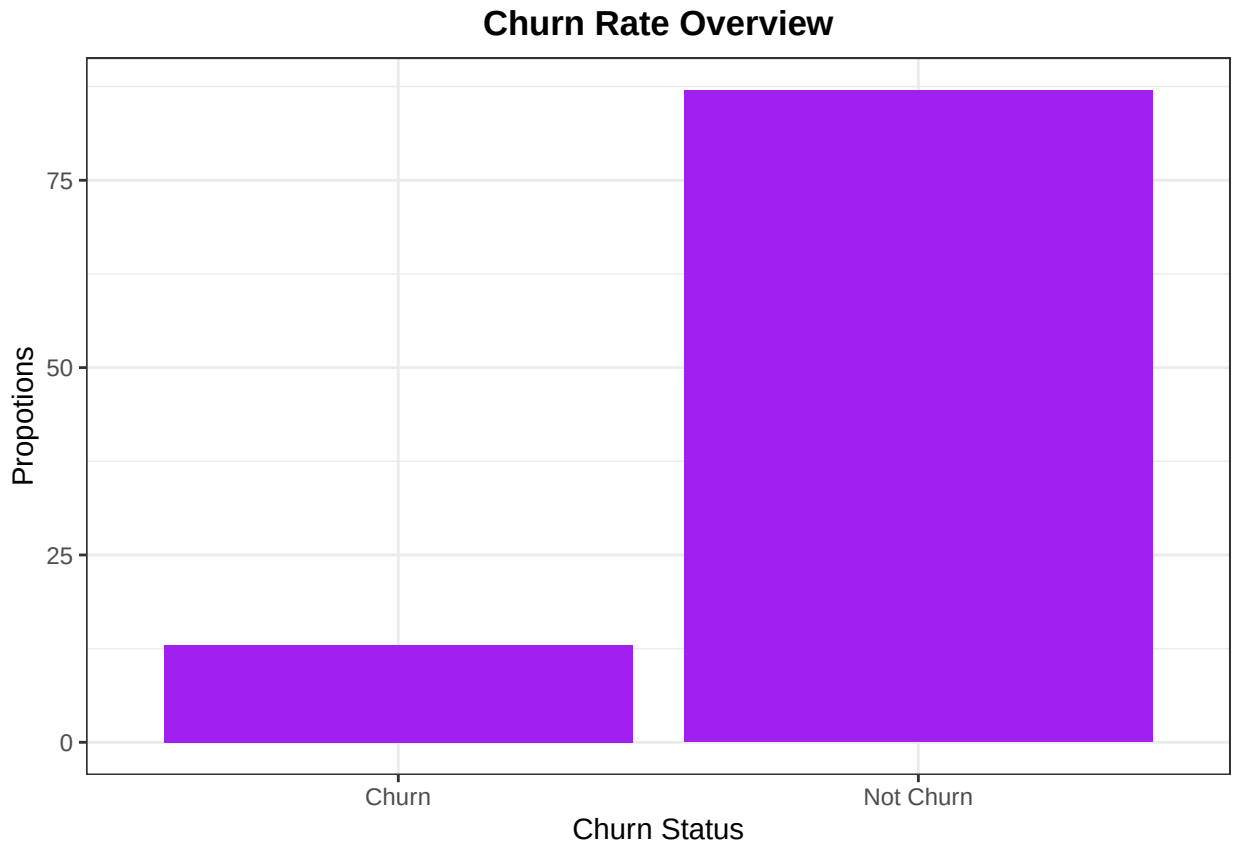
Univariate Distribution

Target Variable

```
## Warning: Warning: fonts used in ‘flextable‘ are ignored because the ‘pdflatex‘
## engine is used and not ‘xelatex‘ or ‘lualatex‘. You can avoid this warning
## by using the ‘set_flextable_defaults(fonts_ignore=TRUE)‘ command or use a
## compatible engine by defining ‘latex_engine: xelatex‘ in the YAML header of the
## R Markdown document.
```

Table 3: Churn Rate

Churn	Count	Propotions
Churn	4,315	13
Not Churn	28,191	87
Total	32,506	100



Observations and inference

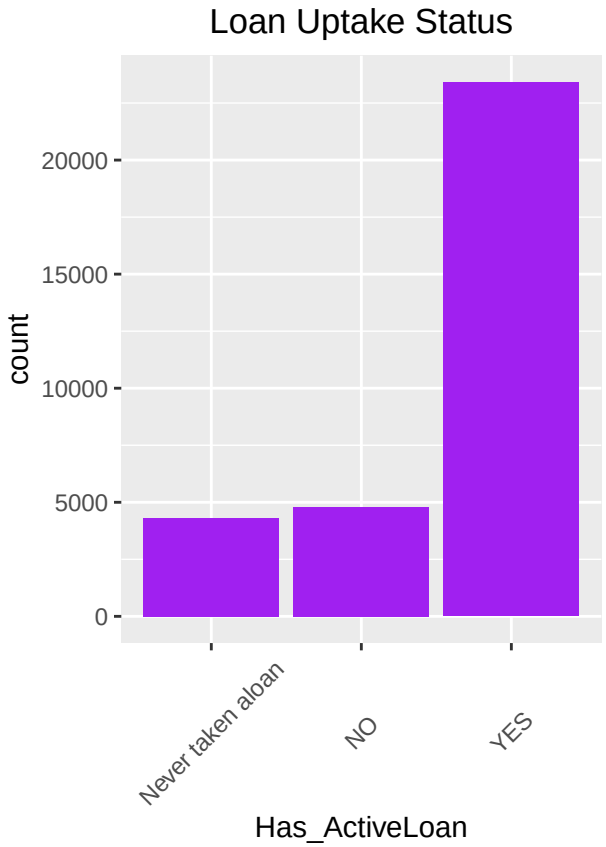
- The churn rate for this institution is very low as compared to the overall client count at 13% and 87% respectively.

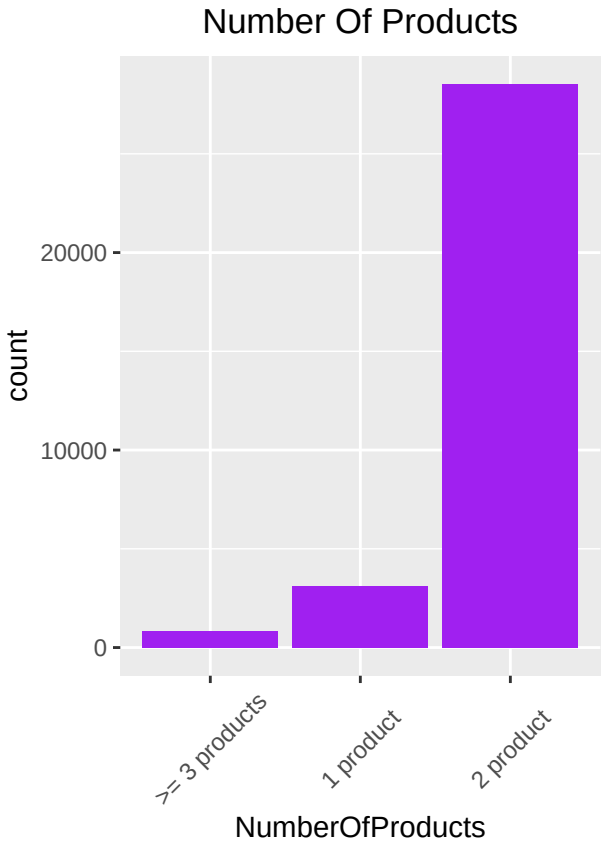
Explanatory variables.

Categorical variables

```
## Warning: Warning: fonts used in ‘flextable‘ are ignored because the ‘pdflatex‘
## engine is used and not ‘xelatex‘ or ‘lualatex‘. You can avoid this warning
## by using the ‘set_flextable_defaults(fonts_ignore=TRUE)‘ command or use a
## compatible engine by defining ‘latex_engine: xelatex‘ in the YAML header of the
## R Markdown document.
```

variable	value	Count	Proportion
Has_ActiveLoan	Never taken aloan	4,315	13 %
Has_ActiveLoan	NO	4,788	15 %
Has_ActiveLoan	YES	23,403	72 %
Total	-	32,506	-
NumberOfProducts	>=3 products	838	3 %
NumberOfProducts	1 product	3,128	10 %
NumberOfProducts	2 products	28,540	88 %
Total	-	32,506	-





Observations and inferences

- Majority of clients have 2 products as compared to the remaining < 50% with more than one product.
- Majority of clients have existing running loans with the institution.

Numerical Variables

```
## Warning: Warning: fonts used in ‘flextable‘ are ignored because the ‘pdflatex‘
## engine is used and not ‘xelatex‘ or ‘lualatex‘. You can avoid this warning
## by using the ‘set_flextable_defaults(fonts_ignore=TRUE)‘ command or use a
## compatible engine by defining ‘latex_engine: xelatex‘ in the YAML header of the
## R Markdown document.
```

value	mean	variance	sd	max	min	0%	100%	25%
Tenure	19.57	10.70	3.27	65.00	0.00	0.00	65.00	20.00
No_of_transactions	22.16	1,419.17	37.67	3,114.00	1.00	1.00	3,114.00	7.00
No_of_transactions_Deposits	6.22	173.35	13.17	939.00	0.00	0.00	939.00	1.00
TotalDepositsMade	40,773.15	30,191,976,859.33	173,758.39	15,869,967.98	0.00	0.00	15,869,967.98	1,120.00
Max_DepositsMade	7,850.02	1,390,436,850.47	37,288.56	4,000,000.00	0.00	0.00	4,000,000.00	1,120.00
Average_DepositsMade	1,507.12	6,958,201.71	2,637.84	275,108.80	0.00	0.00	275,108.80	318.57

value	mean	variance	sd	max	min	0%	100%	25%	
No_of_transactions_Withdrawals	15.58	666.20	25.81	2,161.00	0.00	0.00	2,161.00	6.00	
WithdrawalMade	45,625.00	30,689,245,553.75	175,183.46	16,171,410.72	0.00	0.00	16,171,410.72	2,240.00	1
Max_Withdrawal_made	7,732.10	1,344,203,060.36	36,663.37	4,000,000.00	0.00	0.00	4,000,000.00	1,120.00	
Days_Since_Last_Transaction	115.69	55,963.45	236.57	751.00	2.00	2.00	751.00	185.00	
transactions_made_12	8.27	485.18	22.03	665.00	0.00	0.00	665.00	0.00	
transactions_made_6	3.04	142.77	11.95	579.00	0.00	0.00	579.00	0.00	
transactions_made_3	1.98	45.82	6.77	408.00	0.00	0.00	408.00	0.00	
deposits_12	607.79	906,971,233.26	30,115.96	4,000,000.00	0.00	0.00	4,000,000.00	0.00	
deposits_6	544.88	856,756,427.81	29,270.40	4,000,000.00	0.00	0.00	4,000,000.00	0.00	
deposits_3	538.85	856,685,933.61	29,269.20	4,000,000.00	0.00	0.00	4,000,000.00	0.00	
withdrawals_12	2,844.73	50,513,498.36	7,107.28	700,000.00	0.00	0.00	700,000.00	0.00	
withdrawals_6	2,003.19	27,379,037.20	5,232.50	220,000.00	0.00	0.00	220,000.00	0.00	
withdrawals_3	1,435.58	20,616,978.23	4,540.59	220,000.00	0.00	0.00	220,000.00	0.00	
last_deposit_amount	5,269.24	155,371,068.92	12,464.79	1,434,950.00	0.00	0.00	1,434,950.00	200.00	
last_withdrawal_amount	12,516.80	1,348,495,850.56	36,721.87	2,480,237.00	0.00	0.00	2,480,237.00	2,240.00	
Max_LoanTaken	7,849.72	387,205,799.23	19,677.55	1,200,000.00	590.00	590.00	1,200,000.00	1,120.00	
Min_LoanTaken	7,701.02	546,603,477.05	23,379.55	1,250,000.00	1,100.00	1,100.00	1,250,000.00	1,120.00	
ave_LoanTaken	7,781.98	363,352,280.82	19,061.80	1,200,000.00	992.25	992.25	1,200,000.00	1,120.00	
no_of_LoanTaken	2.78	15.30	3.91	40.00	1.00	1.00	40.00	1.00	
LastLoanDate	44,011.18	44,466.22	210.87	44,510.00	43,753.00	43,753.00	44,510.00	43,882.00	4
DaysSince_Last_Loan	499.82	44,466.22	210.87	758.00	1.00	1.00	758.00	476.00	
LastLoanTaken	8,150.94	429,223,788.71	20,717.72	1,200,000.00	1,100.00	1,100.00	1,200,000.00	1,120.00	
FirstLoanTaken	7,760.17	574,032,255.44	23,958.97	1,300,000.00	590.00	590.00	1,300,000.00	1,120.00	
Max_ArrearsDays	389.85	68,070.84	260.90	615.00	-28.00	-28.00	615.00	16.00	
First_Loan_Last_Loan_Ratio	1.08	1.35	1.16	159.44	0.03	0.03	159.44	1.00	
First_Loan_Last_Loan_Diff	-390.76	203,180,954.87	14,254.16	1,272,500.00	-600,000.00	-600,000.00	1,272,500.00	0.00	
Last_Loan_First_Loan_Ratio	1.14	0.87	0.93	31.00	0.01	0.01	31.00	1.00	
Trx_Status	1.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	

```
## Warning: Removed 4315 rows containing non-finite values (stat_bin).

## Warning: Removed 4315 rows containing non-finite values (stat_boxplot).

## Warning: Removed 4315 rows containing non-finite values (stat_bin).

## Warning: Removed 4315 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 4315 rows containing non-finite values (stat_bin).

## Warning: Removed 4315 rows containing non-finite values (stat_boxplot).

## Warning: Removed 4315 rows containing non-finite values (stat_bin).

## Warning: Removed 4315 rows containing non-finite values (stat_boxplot).

## Warning: Removed 4315 rows containing non-finite values (stat_bin).

## Warning: Removed 4315 rows containing non-finite values (stat_boxplot).

## Warning: Removed 4315 rows containing non-finite values (stat_bin).

## Warning: Removed 4315 rows containing non-finite values (stat_boxplot).

## Warning: Removed 4315 rows containing non-finite values (stat_bin).

## Warning: Removed 4315 rows containing non-finite values (stat_boxplot).

## Warning: Removed 4315 rows containing non-finite values (stat_bin).

## Warning: Removed 4315 rows containing non-finite values (stat_boxplot).

## Warning: Removed 4315 rows containing non-finite values (stat_bin).

## Warning: Removed 4315 rows containing non-finite values (stat_boxplot).

## Warning: Removed 4315 rows containing non-finite values (stat_bin).

## Warning: Removed 4315 rows containing non-finite values (stat_boxplot).
```

Bivariate Analysis

Descriptives

```
## Warning in min(x, na.rm = T): no non-missing arguments to min; returning Inf

## Warning in max(x, na.rm = T): no non-missing arguments to max; returning -Inf

## Warning in min(x, na.rm = T): no non-missing arguments to min; returning Inf

## Warning in max(x, na.rm = T): no non-missing arguments to max; returning -Inf

## Warning in min(x, na.rm = T): no non-missing arguments to min; returning Inf

## Warning in max(x, na.rm = T): no non-missing arguments to max; returning -Inf
```

```
## Warning in min(x, na.rm = T): no non-missing arguments to min; returning Inf

## Warning in max(x, na.rm = T): no non-missing arguments to max; returning -Inf

## Warning in min(x, na.rm = T): no non-missing arguments to min; returning Inf

## Warning in max(x, na.rm = T): no non-missing arguments to max; returning -Inf

## Warning in min(x, na.rm = T): no non-missing arguments to min; returning Inf

## Warning in max(x, na.rm = T): no non-missing arguments to max; returning -Inf

## Warning in min(x, na.rm = T): no non-missing arguments to min; returning Inf

## Warning in max(x, na.rm = T): no non-missing arguments to max; returning -Inf

## Warning in min(x, na.rm = T): no non-missing arguments to min; returning Inf

## Warning in max(x, na.rm = T): no non-missing arguments to max; returning -Inf

## Warning in min(x, na.rm = T): no non-missing arguments to min; returning Inf

## Warning in max(x, na.rm = T): no non-missing arguments to max; returning -Inf

## Warning in min(x, na.rm = T): no non-missing arguments to min; returning Inf

## Warning in max(x, na.rm = T): no non-missing arguments to max; returning -Inf

## Warning: Warning: fonts used in ‘flextable‘ are ignored because the ‘pdflatex‘
## engine is used and not ‘xelatex‘ or ‘lualatex‘. You can avoid this warning
## by using the ‘set_flextable_defaults(fonts_ignore=TRUE)‘ command or use a
## compatible engine by defining ‘latex_engine: xelatex‘ in the YAML header of the
## R Markdown document.
```

Variable	ChurnStatus	mean	variance	sd	max	min	0%	100%
ave_LoanTaken	Churn	0.00	0.00	0.00	-Inf	Inf	0.00	0.00
ave_LoanTaken	Not_Churn	7,781.98	363,352,280.82	19,061.80	1,200,000.00	992.25	992.25	1,200,000.00
Average_Deposits_Made	Churn	2.77	82.17	9.06	241.00	0.00	0.00	241.00
Average_Deposits_Made	Not_Churn	17.55	726.61	26.96	2,161.00	1.00	1.00	2,161.00
Days_Since_Last_Transaction	Churn	438.64	35,931.92	189.56	751.00	2.00	2.00	751.00
Days_Since_Last_Transaction	Not_Churn	412.18	58,937.98	242.77	694.00	2.00	2.00	694.00
DaysSince_Last_Loan	Churn	0.00	0.00	0.00	-Inf	Inf	0.00	0.00
DaysSince_Last_Loan	Not_Churn	499.82	44,466.22	210.87	758.00	1.00	1.00	758.00

Variable	ChurnStatus	mean	variance	sd	max	min	0%	100%
deposits_12	Churn	1,263.59	3,746,228,103.52	21,206.44	4,000,000.00	0.00	0.00	4,000,000.00
deposits_12	Not_Churn	507.42	472,427,531.02	21,735.40	1,300,000.00	0.00	0.00	1,300,000.00
deposits_3	Churn	1,150.35	3,745,104,674.52	21,197.26	4,000,000.00	0.00	0.00	4,000,000.00
deposits_3	Not_Churn	445.25	414,626,967.69	20,362.39	1,300,000.00	0.00	0.00	1,300,000.00
deposits_6	Churn	1,186.84	3,745,315,100.96	21,198.98	4,000,000.00	0.00	0.00	4,000,000.00
deposits_6	Not_Churn	446.62	414,669,311.59	20,363.43	1,300,000.00	0.00	0.00	1,300,000.00
First_Loan_Last_Loan_diff	Churn	0.00	0.00	0.00	-Inf	Inf	0.00	0.00
First_Loan_Last_Loan_diff	Not_Churn	-390.76	203,180,954.87	14,254.16	1,272,500.00	-600,000.00	-600,000.00	1,272,500.00
First_Loan_Last_Loan_Ratio	Churn	0.00	0.00	0.00	-Inf	Inf	0.00	0.00
First_Loan_Last_Loan_Ratio	Not_Churn	0.98	1.35	1.16	159.44	0.03	0.03	159.44
FirstLoanTaken	Churn	0.00	0.00	0.00	-Inf	Inf	0.00	0.00
FirstLoanTaken	Not_Churn	7,760.17	574,032,255.44	23,958.97	1,300,000.00	590.00	590.00	1,300,000.00
last_deposit_amount	Churn	635.37	480,653,690.81	21,923.82	1,434,950.00	0.00	0.00	1,434,950.00
last_deposit_amount	Not_Churn	5,978.51	101,807,723.22	10,089.98	530,000.00	0.00	0.00	530,000.00
Last_Loan_First_Loan_Ratio	Churn	0.00	0.00	0.00	-Inf	Inf	0.00	0.00
Last_Loan_First_Loan_Ratio	Not_Churn	1.14	0.87	0.93	31.00	0.01	0.01	31.00
last_withdrawal_amount	Churn	625.29	482,871,137.23	21,974.33	1,434,950.00	0.00	0.00	1,434,950.00
last_withdrawal_amount	Not_Churn	14,336.95	1,456,054,672.03	38,158.28	2,480,237.00	0.00	0.00	2,480,237.00
LastLoanDate	Churn	0.00	0.00	0.00	-Inf	Inf	0.00	0.00
LastLoanDate	Not_Churn	44,011.18	44,466.22	210.87	44,510.00	43,753.00	43,753.00	44,510.00
LastLoanTaken	Churn	0.00	0.00	0.00	-Inf	Inf	0.00	0.00
LastLoanTaken	Not_Churn	8,150.94	429,223,788.71	20,717.72	1,200,000.00	1,100.00	1,100.00	1,200,000.00
Max_Arrears_Days	Churn	0.00	0.00	0.00	-Inf	Inf	0.00	0.00
Max_Arrears_Days	Not_Churn	389.85	68,070.84	260.90	615.00	-28.00	-28.00	615.00
Max_Deposit_Made	Churn	210.54	21,685,108.90	4,656.73	275,108.80	0.00	0.00	275,108.80
Max_Deposit_Made	Not_Churn	1,705.58	4,408,030.14	2,099.53	95,998.40	45.23	45.23	95,998.40
Max_Withdrawal_Made	Churn	1,874.72	3,907,725,918.45	62,511.81	4,000,000.00	0.00	0.00	4,000,000.00
Max_Withdrawal_Made	Not_Churn	8,628.64	945,891,749.16	30,755.35	2,000,000.00	1,100.00	1,100.00	2,000,000.00
no_of_LoanTaken	Churn	0.00	0.00	0.00	-Inf	Inf	0.00	0.00
no_of_LoanTaken	Not_Churn	2.78	15.30	3.91	40.00	1.00	1.00	40.00
No_of_transactions	Churn	5.52	177.53	13.32	298.00	1.00	1.00	298.00
No_of_transactions	Not_Churn	24.70	1,560.41	39.50	3,114.00	1.00	1.00	3,114.00
No_of_transactions_Deposits	Churn	2.48	26.37	5.14	121.00	0.00	0.00	121.00
No_of_transactions_Deposits	Not_Churn	6.79	193.38	13.91	939.00	0.00	0.00	939.00
No_of_transactions_Withdrawals	Churn	3,562.32	18,572,693,134.36	136,281.67	8,268,800.00	1.00	1.00	8,268,800.00

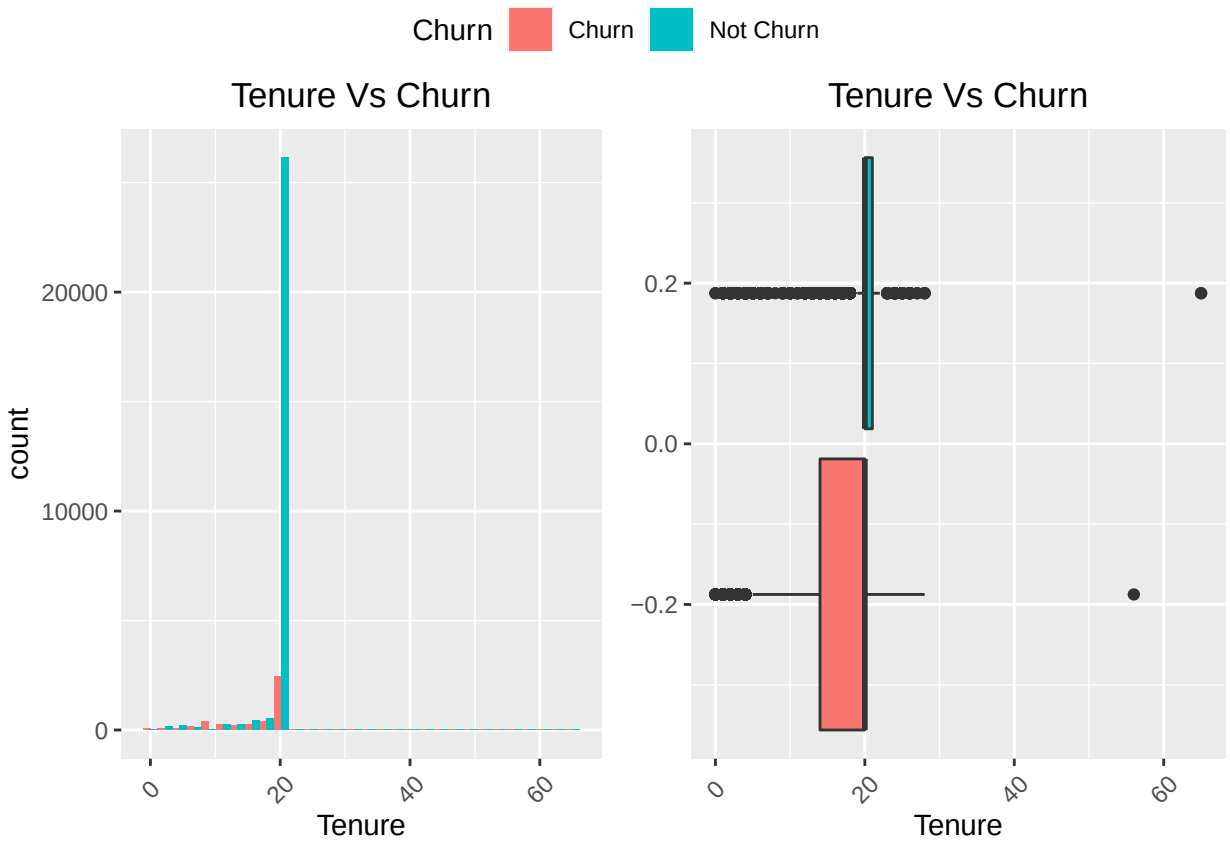
Variable	ChurnStatus	mean	variance	sd	max	min	0%	100%
No_of_transactions_Made	Churn	46,162.63	31,752,359,452.78	178,191.92	15,869,967.98	0.00	0.00	15,869,967.98
Tenure	Churn	16.77	29.27	5.41	56.00	0.00	0.00	56.00
Tenure	Not_Churn	20.00	6.46	2.54	65.00	0.00	0.00	65.00
TotalDeposits_Made	Churn	1,954.86	3,900,852,835.37	62,456.81	4,000,000.00	1.00	1.00	4,000,000.00
TotalDeposits_Made	Not_Churn	8,752.35	1,000,175,910.50	31,625.56	2,000,000.00	0.00	0.00	2,000,000.00
transactions_made_12	Churn	2.37	101.84	10.09	298.00	0.00	0.00	298.00
transactions_made_12	Not_Churn	9.52	537.09	23.18	665.00	0.00	0.00	665.00
transactions_made_3	Churn	0.52	17.65	4.20	126.00	0.00	0.00	126.00
transactions_made_3	Not_Churn	2.21	49.75	7.05	408.00	0.00	0.00	408.00
transactions_made_6	Churn	1.00	49.13	7.01	221.00	0.00	0.00	221.00
transactions_made_6	Not_Churn	4.39	155.58	12.47	579.00	0.00	0.00	579.00
Trx_Status	Churn	1.00	0.00	0.00	1.00	1.00	1.00	1.00
Trx_Status	Not_Churn	1.00	0.00	0.00	1.00	1.00	1.00	1.00
Withdrawal_Made	Churn	4,452.20	6,685,237,766.87	81,763.30	4,126,632.00	0.00	0.00	4,126,632.00
Withdrawal_Made	Not_Churn	51,927.03	34,064,540,710.84	184,565.82	16,171,410.72	1,120.00	1,120.00	16,171,410.72
withdrawals_12	Churn	162.81	113,558,124.91	10,656.37	700,000.00	0.00	0.00	700,000.00
withdrawals_12	Not_Churn	3,255.23	39,597,892.87	6,292.69	220,000.00	0.00	0.00	220,000.00
withdrawals_3	Churn	0.46	927.00	30.45	2,000.00	0.00	0.00	2,000.00
withdrawals_3	Not_Churn	1,655.24	23,409,138.46	4,838.30	220,000.00	0.00	0.00	220,000.00
withdrawals_6	Churn	0.54	947.81	30.79	2,000.00	0.00	0.00	2,000.00
withdrawals_6	Not_Churn	2,309.73	30,861,891.81	5,555.35	220,000.00	0.00	0.00	220,000.00

Categorical Vs Target

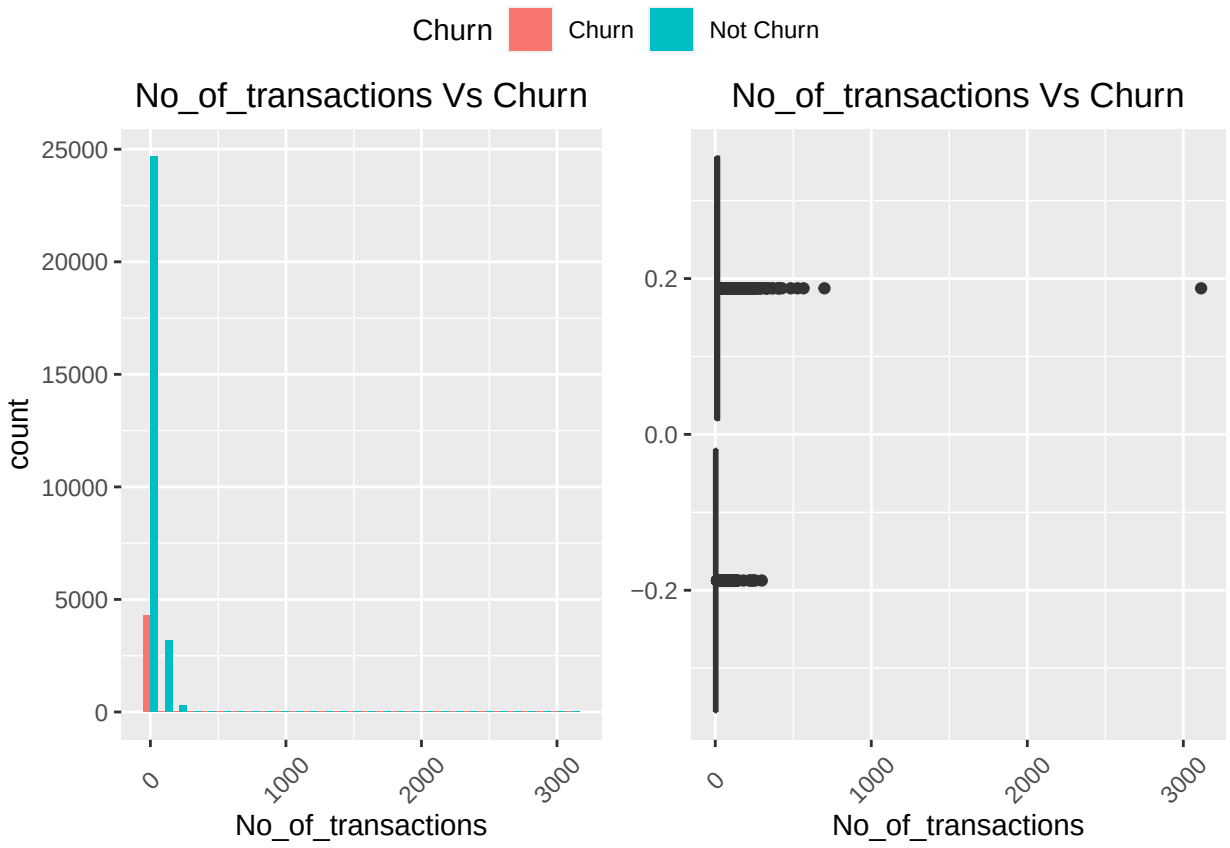
Numerical Vs Target

```
## Warning: Removed 4315 rows containing non-finite values (stat_bin).

## Warning: Removed 4315 rows containing non-finite values (stat_bin).

## Warning: Removed 4315 rows containing non-finite values (stat_boxplot).

## Warning: Removed 4315 rows containing non-finite values (stat_bin).

## Warning: Removed 4315 rows containing non-finite values (stat_bin).

## Warning: Removed 4315 rows containing non-finite values (stat_boxplot).

## Warning: Removed 4315 rows containing non-finite values (stat_bin).

## Warning: Removed 4315 rows containing non-finite values (stat_bin).
```

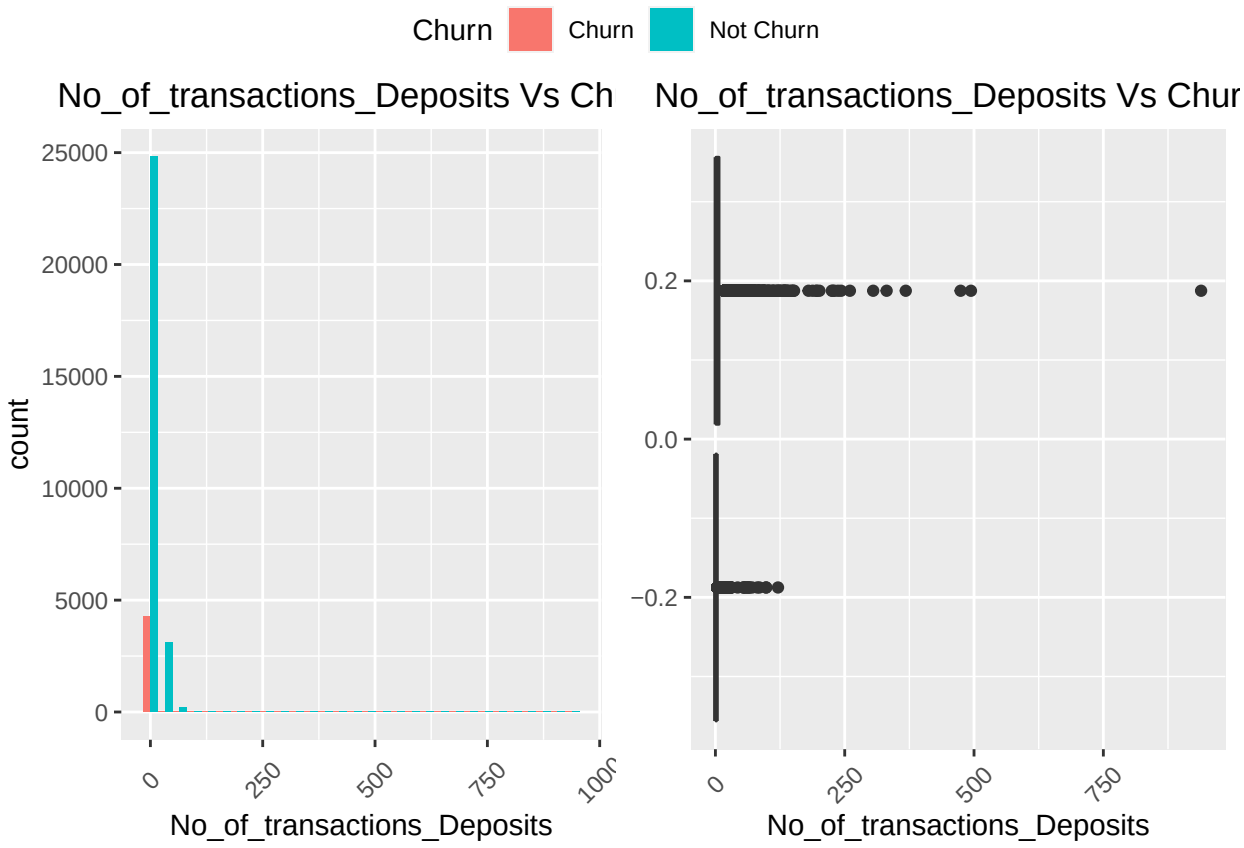
```
## Warning: Removed 4315 rows containing non-finite values (stat_boxplot).

## Warning: Removed 4315 rows containing non-finite values (stat_bin).

## Warning: Removed 4315 rows containing non-finite values (stat_bin).

## Warning: Removed 4315 rows containing non-finite values (stat_boxplot).

## Warning: Removed 4315 rows containing non-finite values (stat_bin).

## Warning: Removed 4315 rows containing non-finite values (stat_bin).

## Warning: Removed 4315 rows containing non-finite values (stat_boxplot).

## Warning: Removed 4315 rows containing non-finite values (stat_bin).

## Warning: Removed 4315 rows containing non-finite values (stat_bin).

## Warning: Removed 4315 rows containing non-finite values (stat_boxplot).

## Warning: Removed 4315 rows containing non-finite values (stat_bin).

## Warning: Removed 4315 rows containing non-finite values (stat_bin).

## Warning: Removed 4315 rows containing non-finite values (stat_boxplot).

## Warning: Removed 4315 rows containing non-finite values (stat_bin).

## Warning: Removed 4315 rows containing non-finite values (stat_bin).

## Warning: Removed 4315 rows containing non-finite values (stat_boxplot).

## Warning: Removed 4315 rows containing non-finite values (stat_bin).

## Warning: Removed 4315 rows containing non-finite values (stat_bin).

## Warning: Removed 4315 rows containing non-finite values (stat_boxplot).

## Warning: Removed 4315 rows containing non-finite values (stat_bin).

## Warning: Removed 4315 rows containing non-finite values (stat_bin).

## Warning: Removed 4315 rows containing non-finite values (stat_boxplot).

## [[1]]
```


Churn
Churn
Not Churn
Tenure Vs Churn
count
20000
10000
0
0
20
40
60
Tenure
Tenure Vs Churn
0.2
0.0
−0.2
0
20
40
60
Tenure

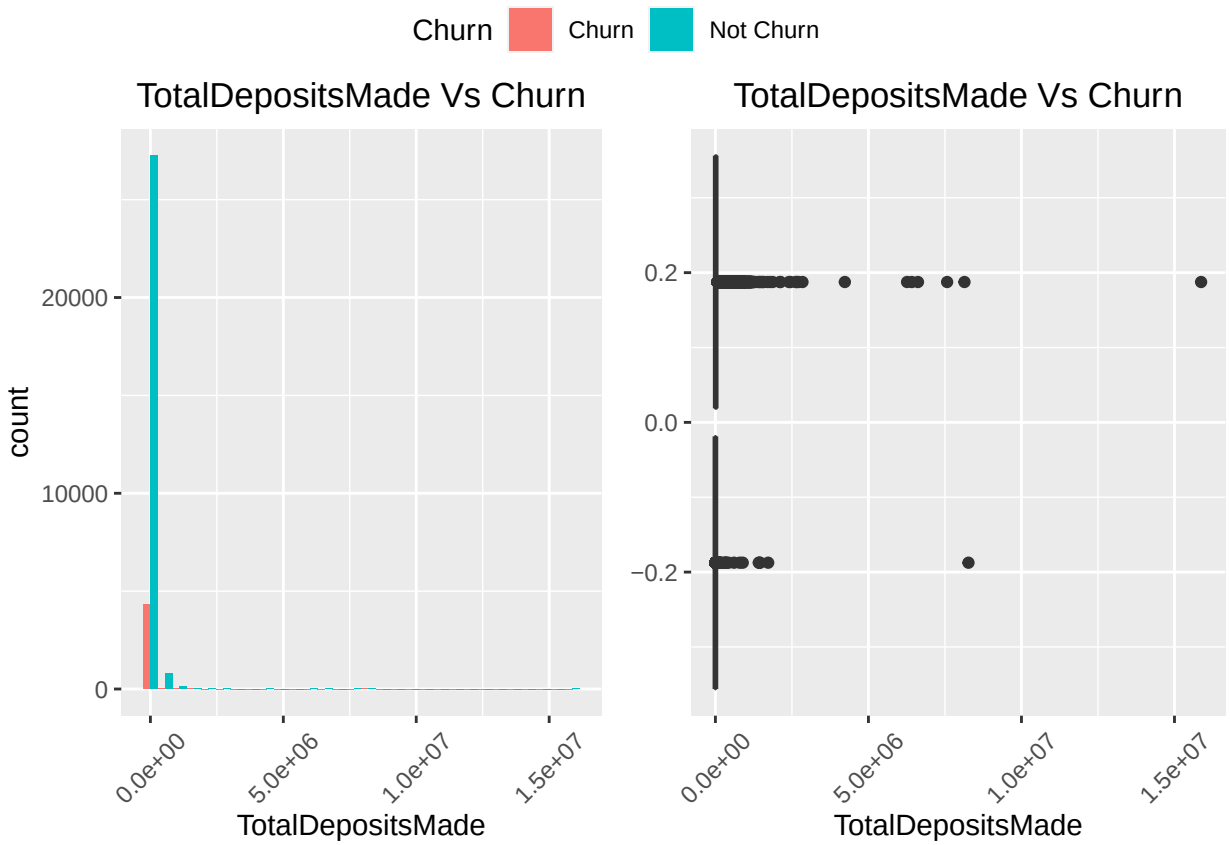
```
##
## [[2]]
```


Churn
Churn
Not Churn
No_of_transactions Vs Churn
count
25000
20000
15000
10000
5000
0
0
1000
2000
3000
No_of_transactions
No_of_transactions Vs Churn
0.2
0.0
−0.2
0
1000
2000
3000
No_of_transactions

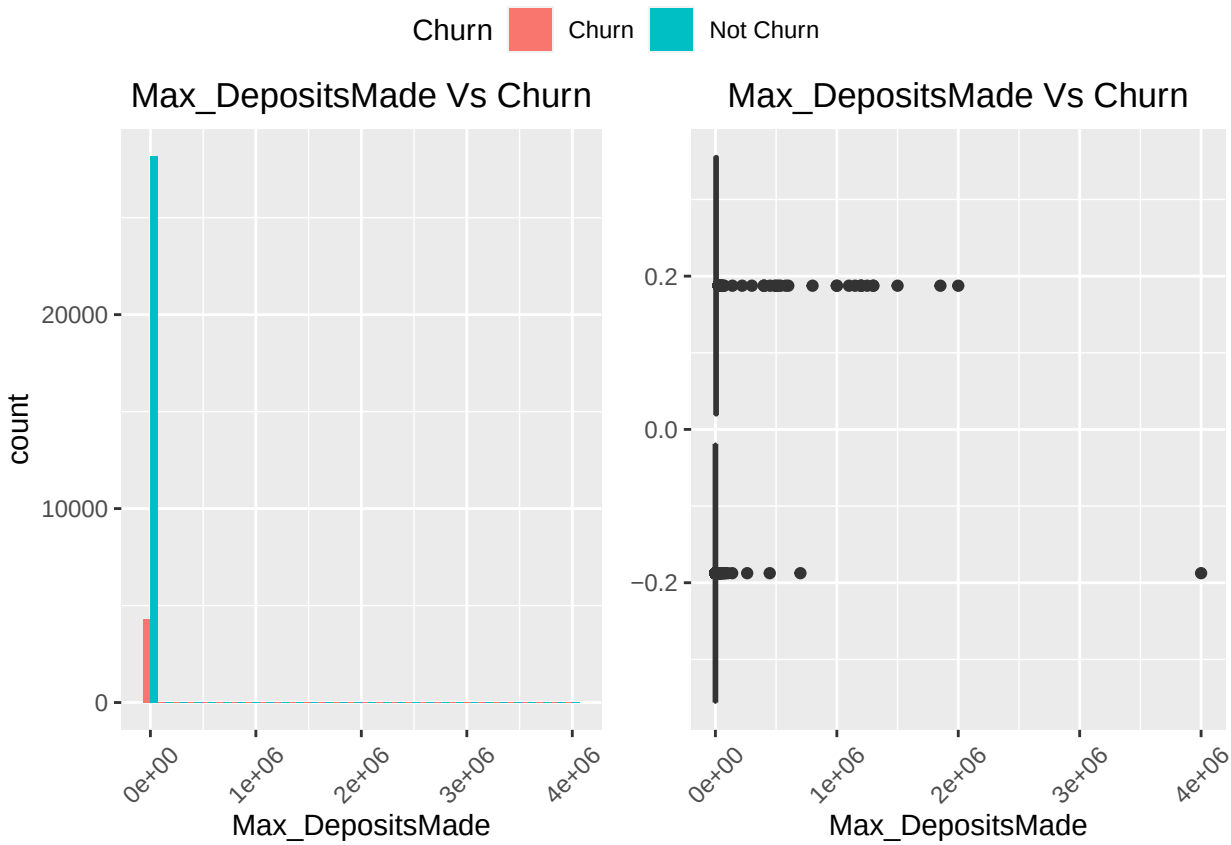
```
##
## [[3]]
```


Churn
Churn
Not Churn
No_of_transactions_Deposits Vs Ch
No_of_transactions_Deposits Vs Chur
count
25000
20000
15000
10000
5000
0
0
250
500
750
1000
No_of_transactions_Deposits
0.2
0.0
−0.2
0
250
500
750
No_of_transactions_Deposits

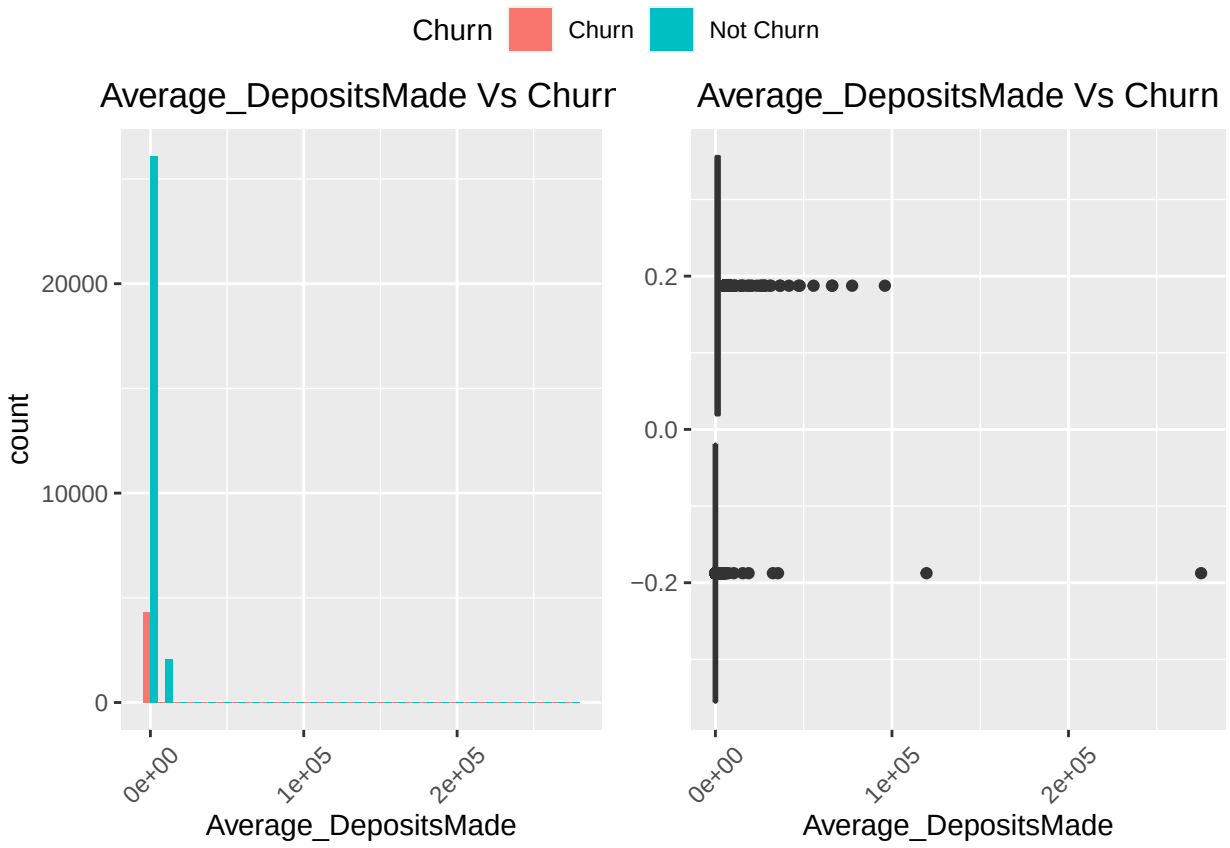
```
##
## [[4]]
```


Churn
Churn
Not Churn
TotalDepositsMade Vs Churn
TotalDepositsMade Vs Churn
count
20000
10000
0
0.2
0.0
−0.2
0.0e+00
5.0e+06
1.0e+07
1.5e+07
TotalDepositsMade
TotalDepositsMade

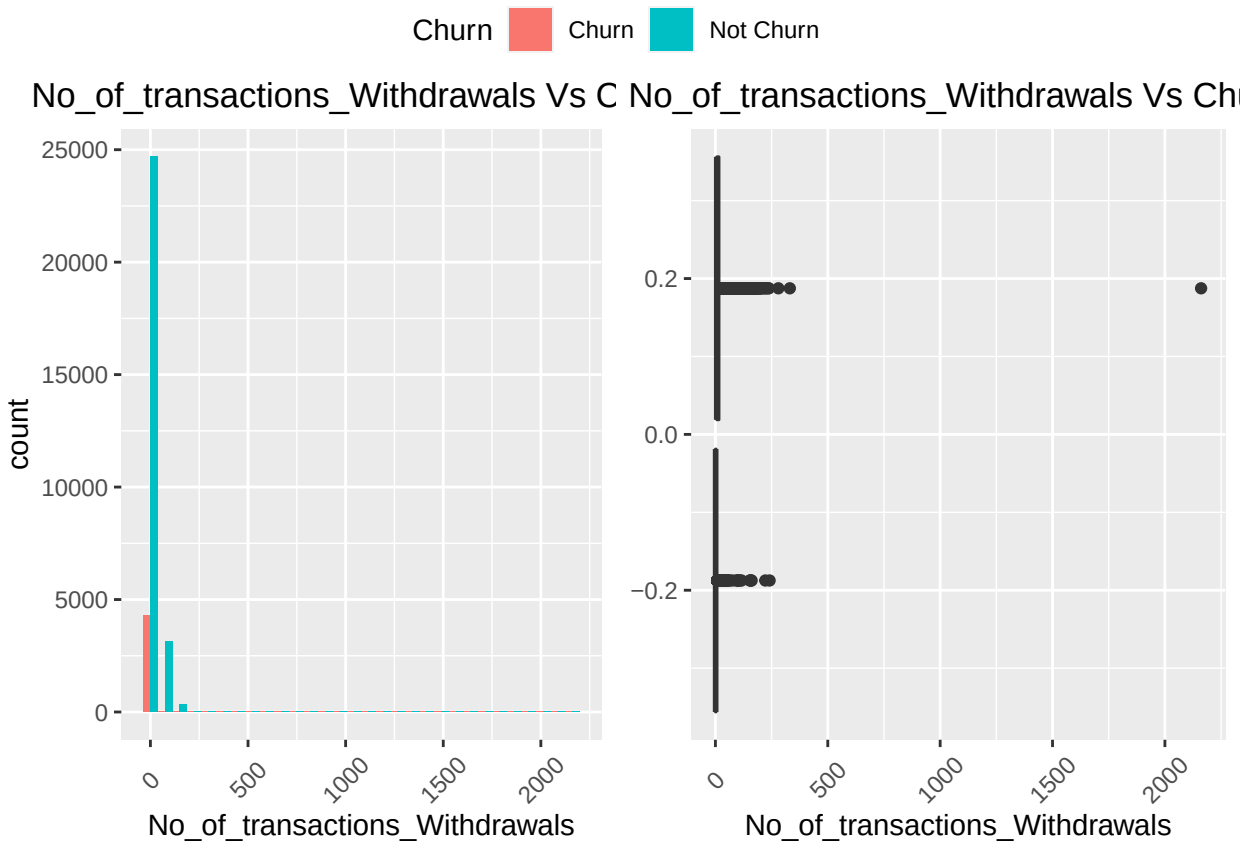
```
##
## [[5]]
```


Churn
Churn
Not Churn
Max_DepositsMade Vs Churn
count
20000
10000
0
0e+00
1e+06
2e+06
3e+06
4e+06
Max_DepositsMade
Max_DepositsMade Vs Churn
0.2
0.0
−0.2
0e+00
1e+06
2e+06
3e+06
4e+06
Max_DepositsMade

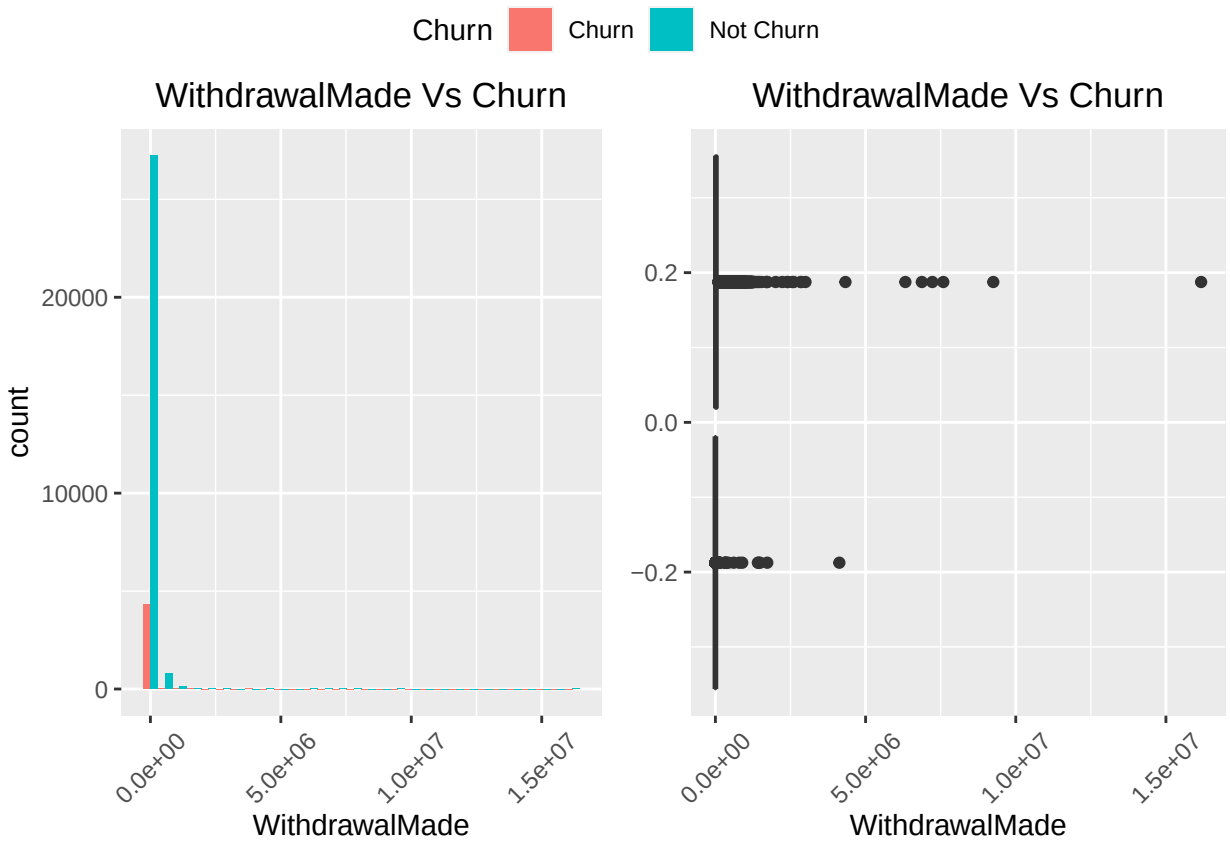
```
##
## [[6]]
```


Churn
Churn
Not Churn
Average_DepositsMade Vs Churr
Average_DepositsMade Vs Churn
count
20000
10000
0
0.2
0.0
−0.2
0e+00
1e+05
2e+05
Average_DepositsMade

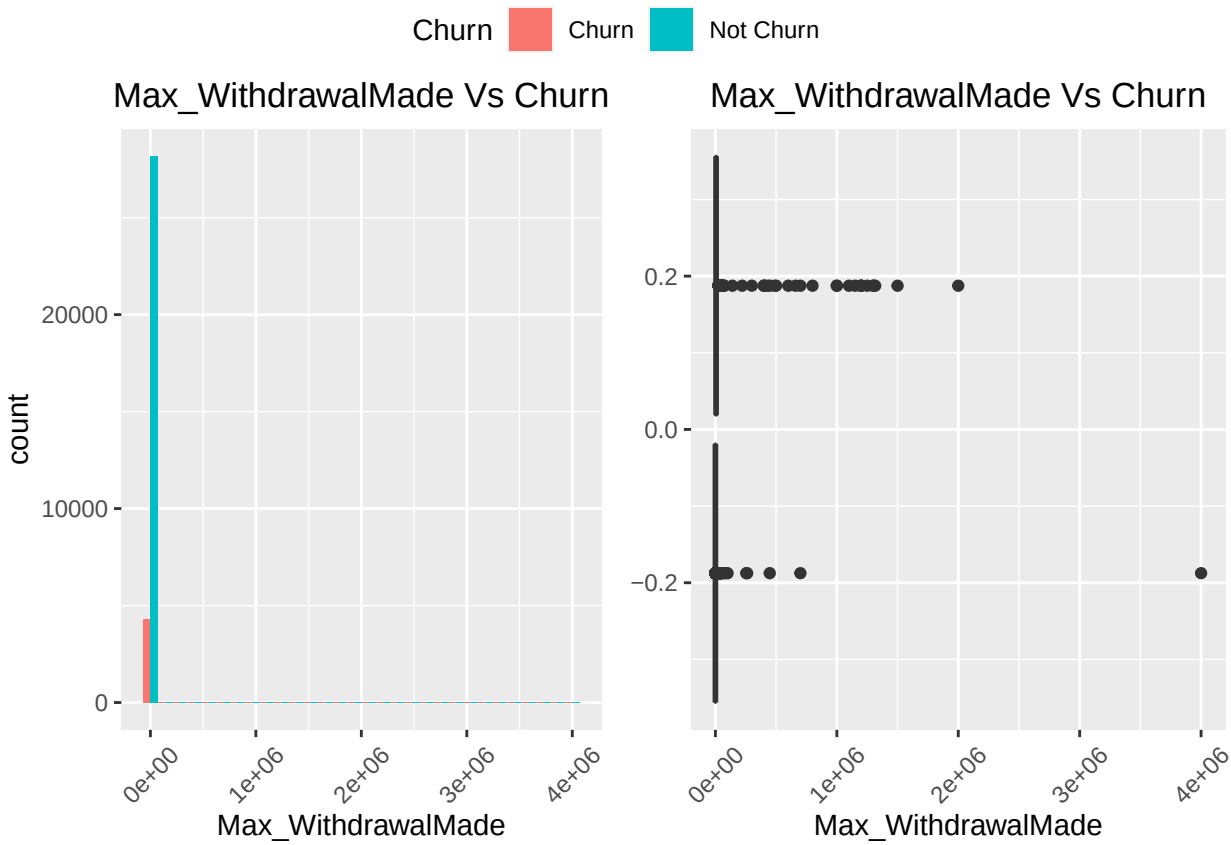
```
##
## [[7]]
```


Churn
Churn
Not Churn
No_of_transactions_Withdrawals Vs C
No_of_transactions_Withdrawals Vs Ch
25000
20000
15000
10000
5000
0
count
0
500
1000
1500
2000
No_of_transactions_Withdrawals
0.2
0.0
−0.2
0
500
1000
1500
2000
No_of_transactions_Withdrawals

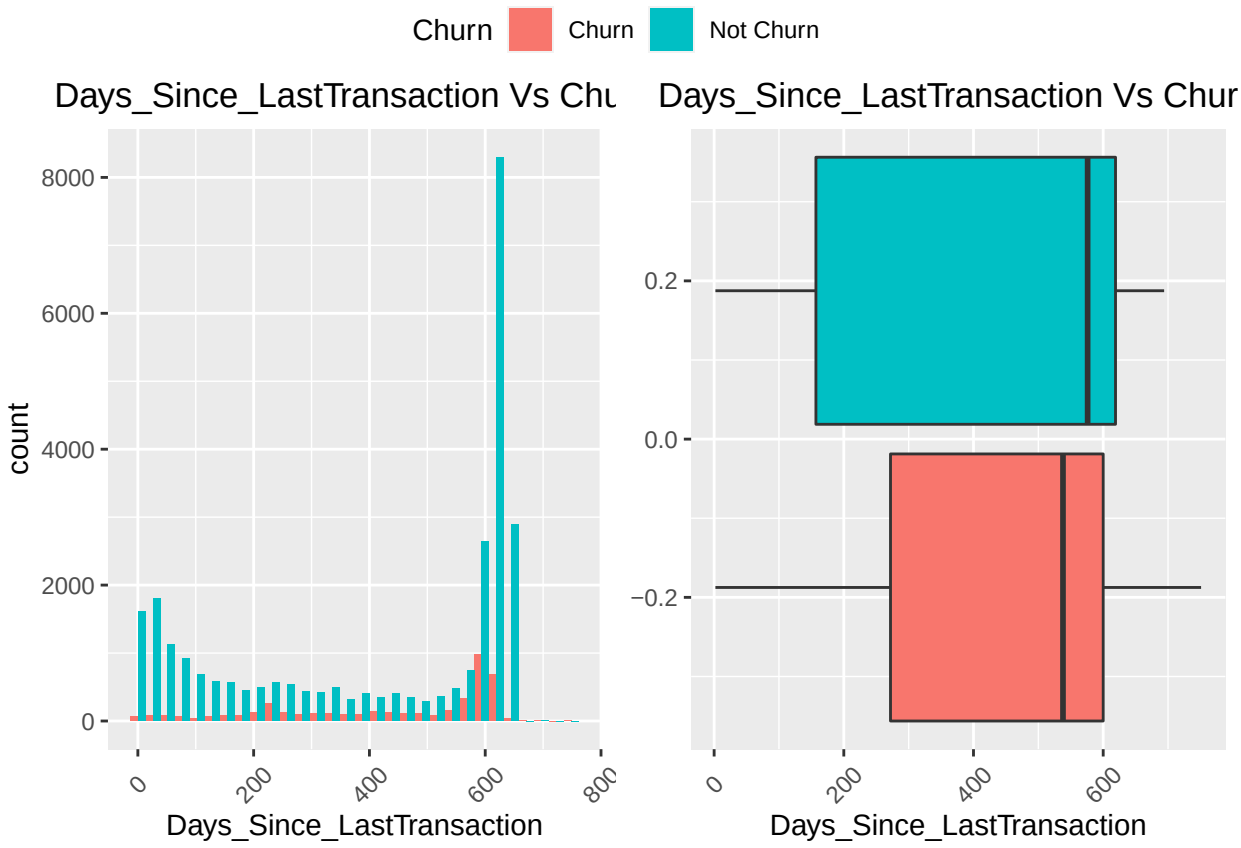
```
##
## [[8]]
```


Churn
Churn
Not Churn
WithdrawalMade Vs Churn
WithdrawalMade Vs Churn
count
20000
10000
0
0.2
0.0
−0.2
0.0e+00
5.0e+06
1.0e+07
1.5e+07
WithdrawalMade
WithdrawalMade

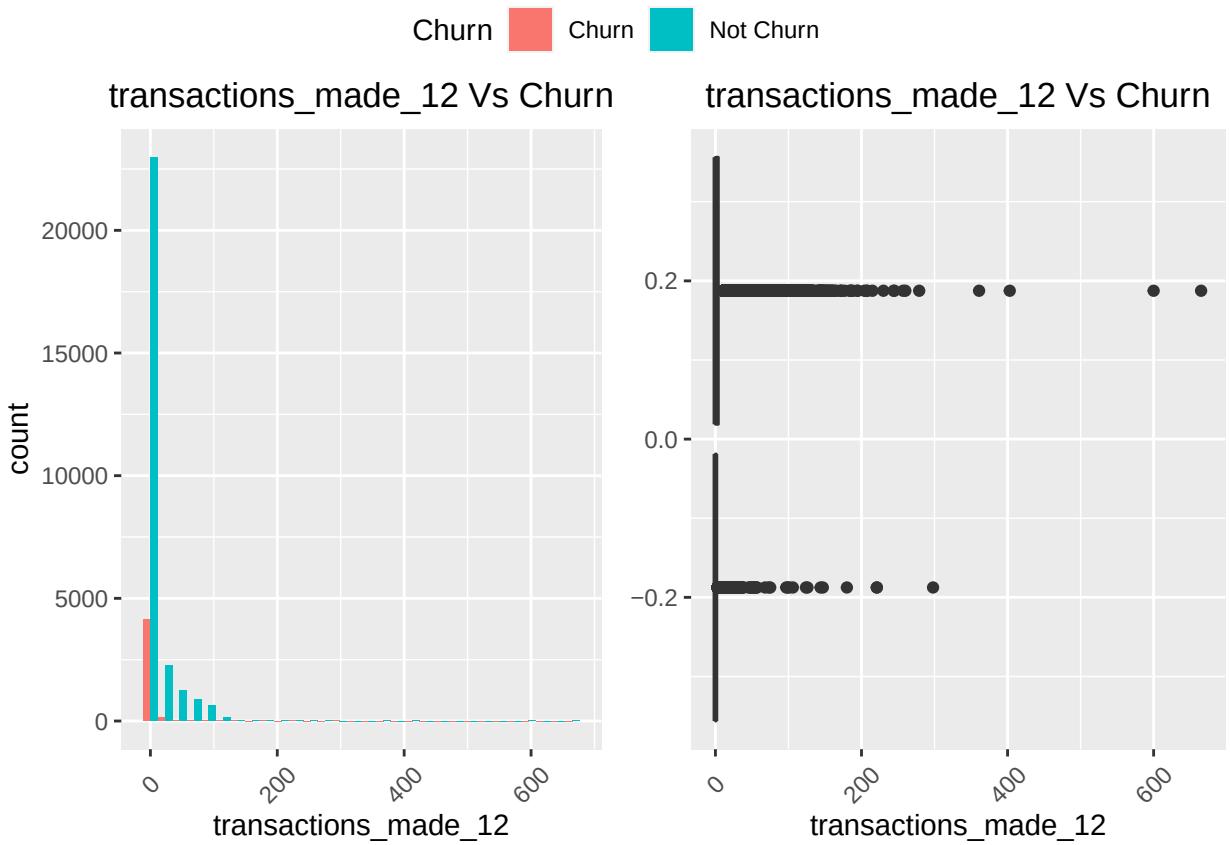
```
##
## [[9]]
```


Churn
Churn
Not Churn
Max_WithdrawalMade Vs Churn
Max_WithdrawalMade Vs Churn
count
20000
10000
0
0.2
0.0
−0.2
0e+00
1e+06
2e+06
3e+06
4e+06
Max_WithdrawalMade
Max_WithdrawalMade

```
##
## [[10]]
```


Churn
Churn
Not Churn
Days_Since_LastTransaction Vs Chu
Days_Since_LastTransaction Vs Chur
count
8000
6000
4000
2000
0
0
200
400
600
800
Days_Since_LastTransaction
0.2
0.0
−0.2
0
200
400
600
Days_Since_LastTransaction

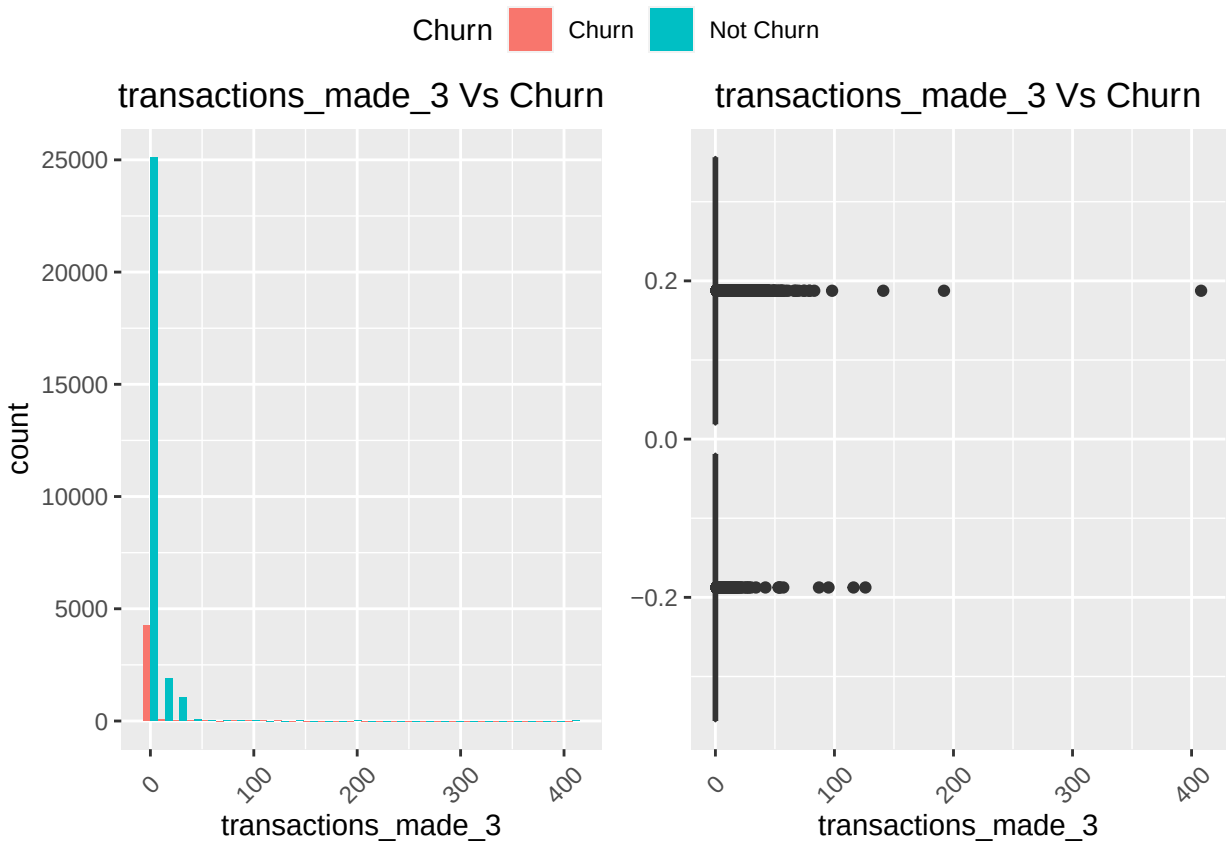
```
##
## [[11]]
```


Churn
Churn
Not Churn
transactions_made_12 Vs Churn
transactions_made_12 Vs Churn
count
20000
15000
10000
5000
0
0.2
0.0
−0.2
0
200
400
600
transactions_made_12
transactions_made_12

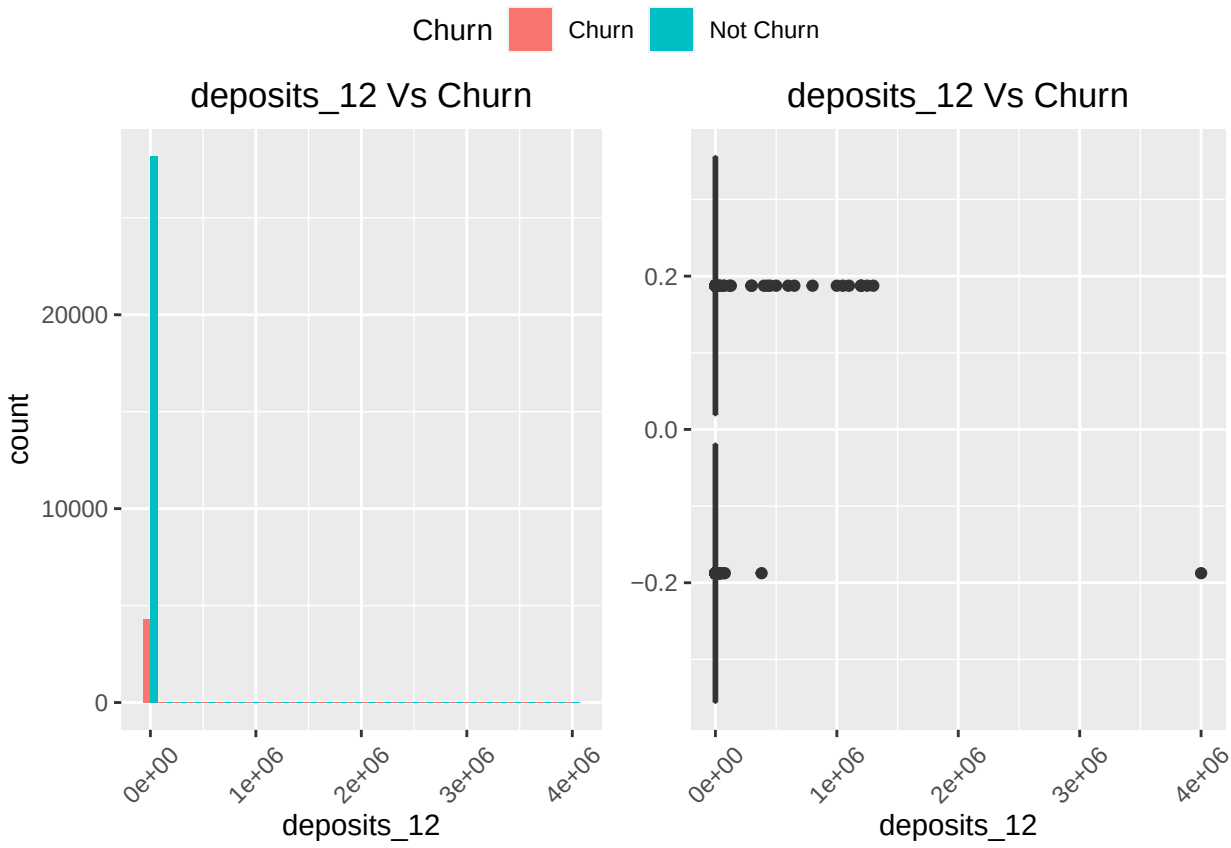
```
##
## [[12]]
```

Churn
Churn
Not Churn
transactions_made_6 Vs Churn
transactions_made_6 Vs Churn
count
25000
20000
15000
10000
5000
0
0.2
0.0
−0.2
0
200
400
600
transactions_made_6
transactions_made_6

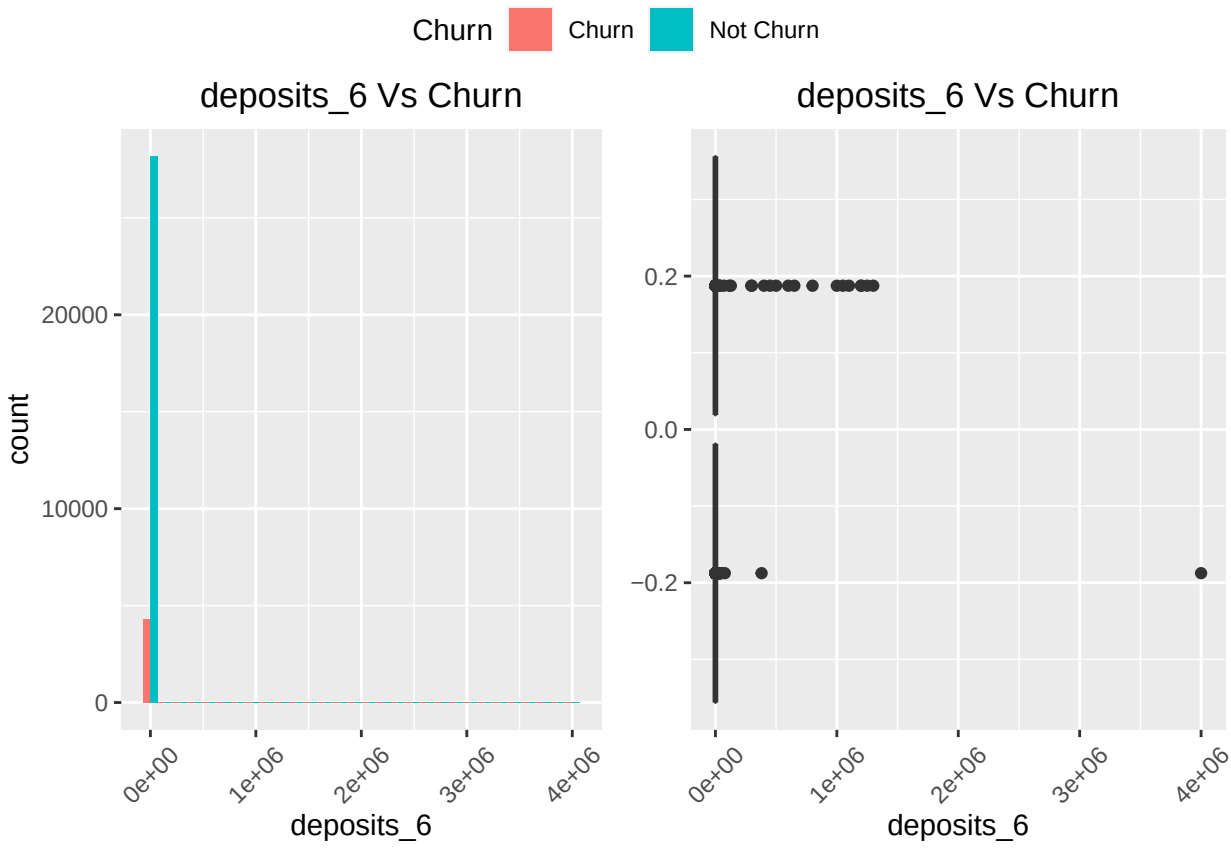
```
##
## [[13]]
```


Churn
Churn
Not Churn
transactions_made_3 Vs Churn
transactions_made_3 Vs Churn
count
25000
20000
15000
10000
5000
0
0.2
0.0
−0.2
0
100
200
300
400
transactions_made_3
transactions_made_3

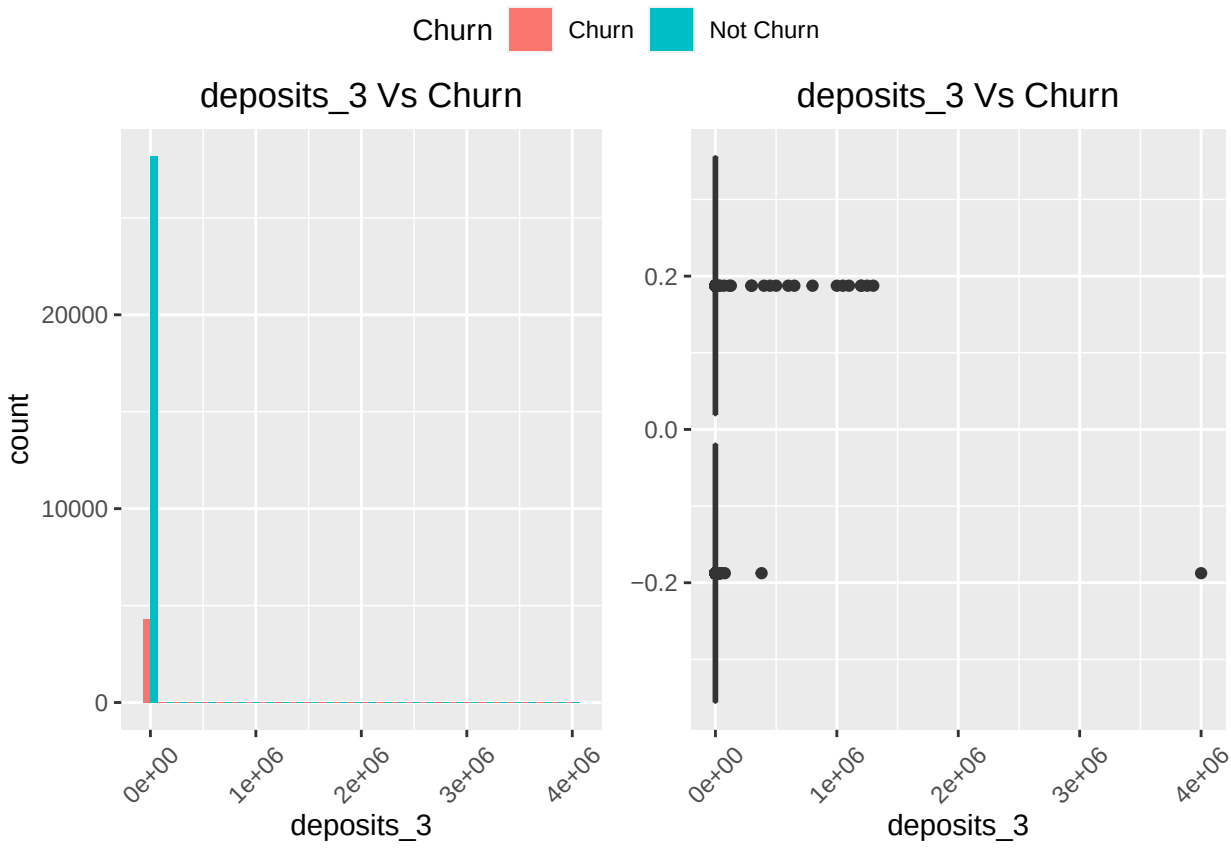
```
##
## [[14]]
```


Churn
Churn
Not Churn
deposits_12 Vs Churn
count
20000
10000
0
0e+00
1e+06
2e+06
3e+06
4e+06
deposits_12
deposits_12 Vs Churn
0.2
0.0
−0.2
0e+00
1e+06
2e+06
3e+06
4e+06
deposits_12

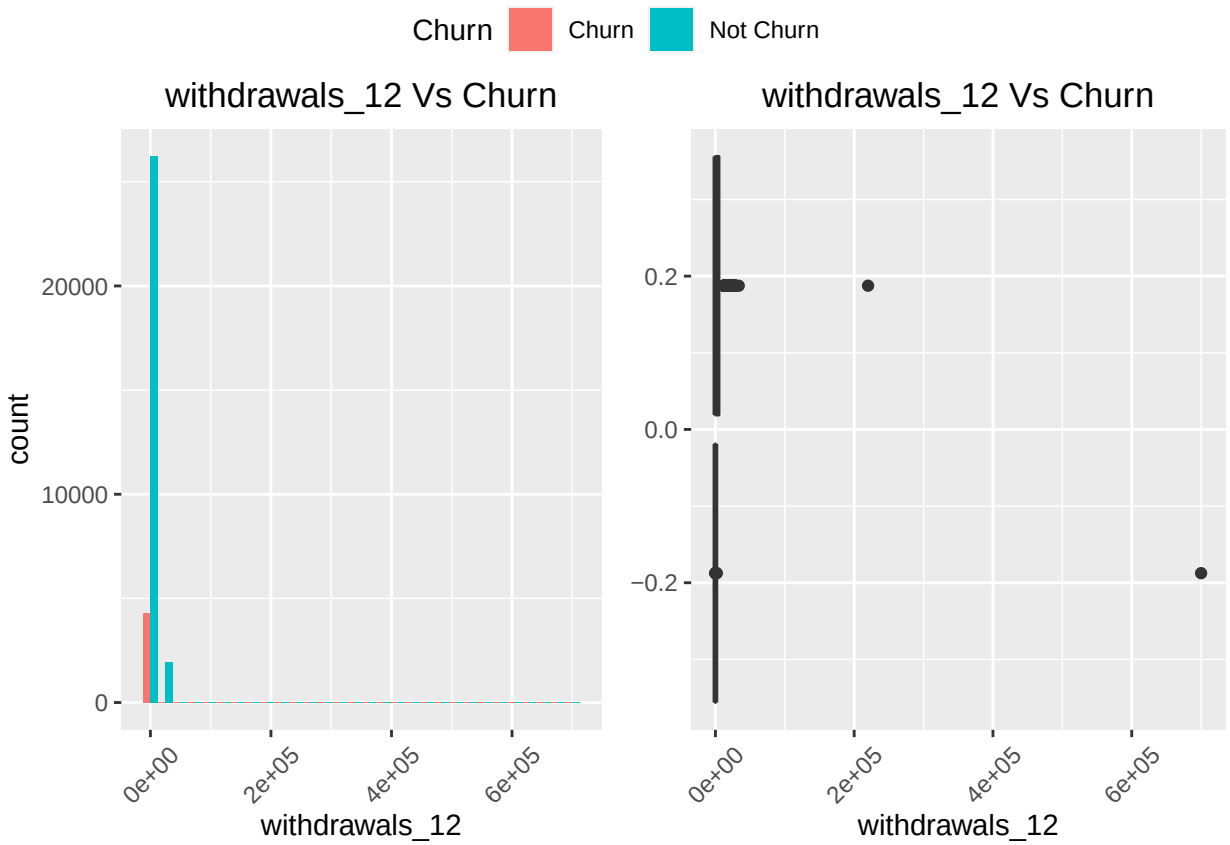
```
##
## [[15]]
```


Churn
Churn
Not Churn
deposits_6 Vs Churn
count
20000
10000
0
0e+00
1e+06
2e+06
3e+06
4e+06
deposits_6
deposits_6 Vs Churn
0.2
0.0
−0.2
0e+00
1e+06
2e+06
3e+06
4e+06
deposits_6

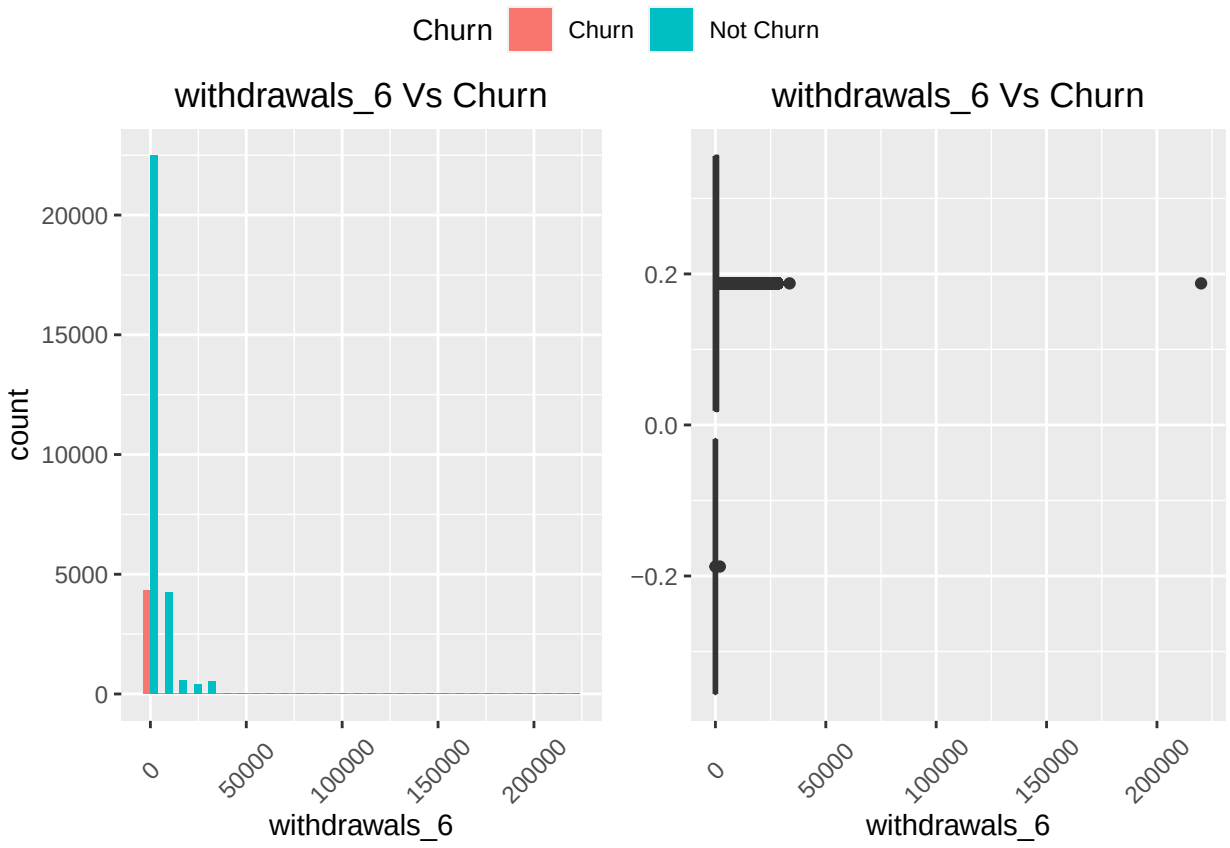
```
##
## [[16]]
```


Churn
Churn
Not Churn
deposits_3 Vs Churn
count
20000
10000
0
0e+00
1e+06
2e+06
3e+06
4e+06
deposits_3
deposits_3 Vs Churn
0.2
0.0
−0.2
0e+00
1e+06
2e+06
3e+06
4e+06
deposits_3

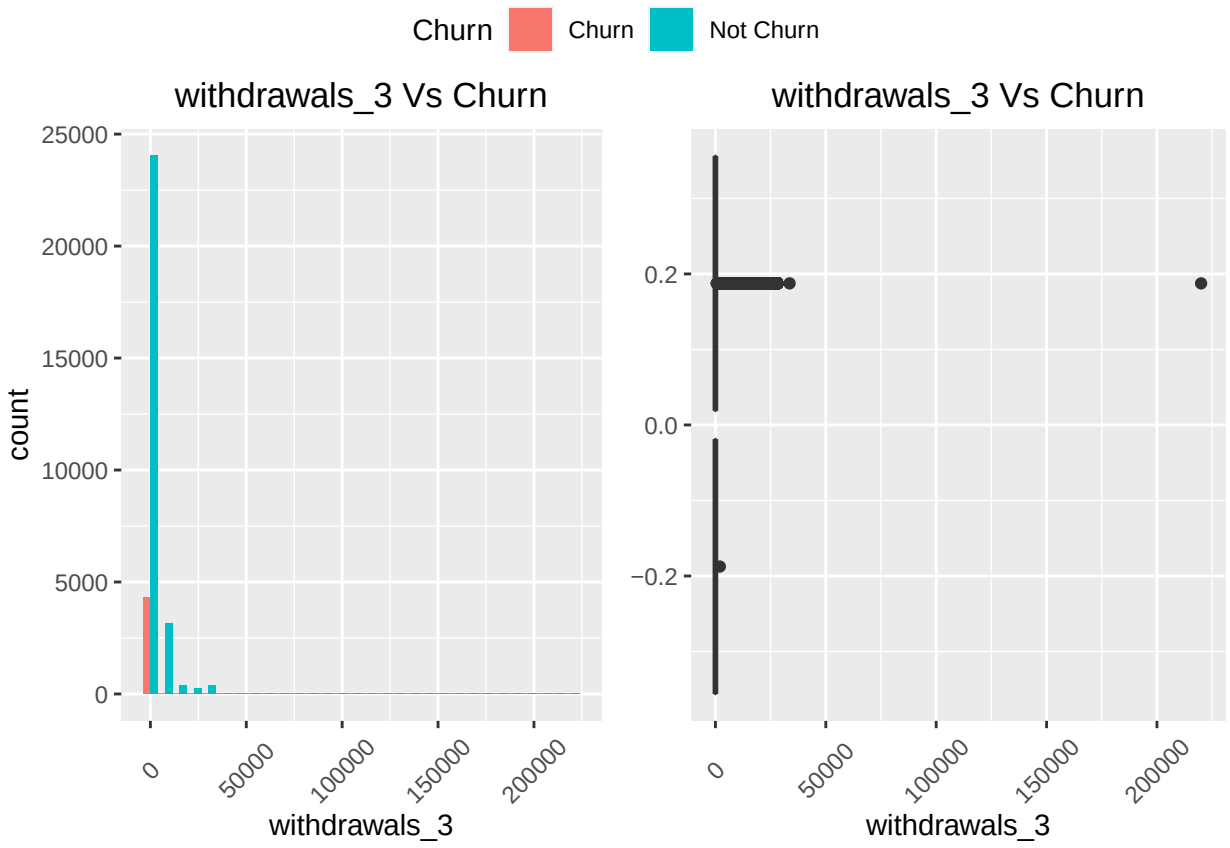
```
##
## [[17]]
```


Churn
Churn
Not Churn
withdrawals_12 Vs Churn
withdrawals_12 Vs Churn
count
20000
10000
0
0.2
0.0
−0.2
0e+00
2e+05
4e+05
6e+05
withdrawals_12
withdrawals_12

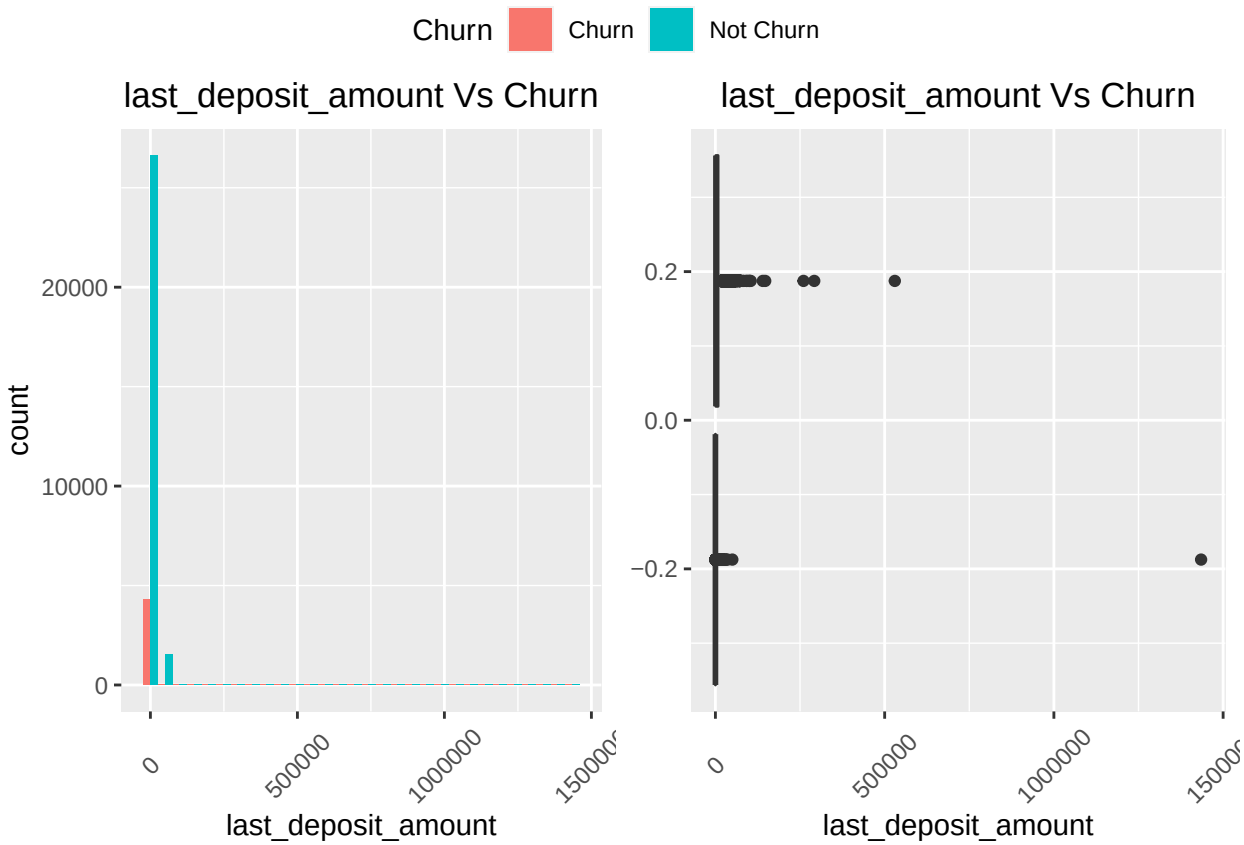
```
##
## [[18]]
```


Churn
Churn
Not Churn
withdrawals_6 Vs Churn
withdrawals_6 Vs Churn
count
20000
15000
10000
5000
0
0.2
0.0
−0.2
0
50000
100000
150000
200000
withdrawals_6
withdrawals_6

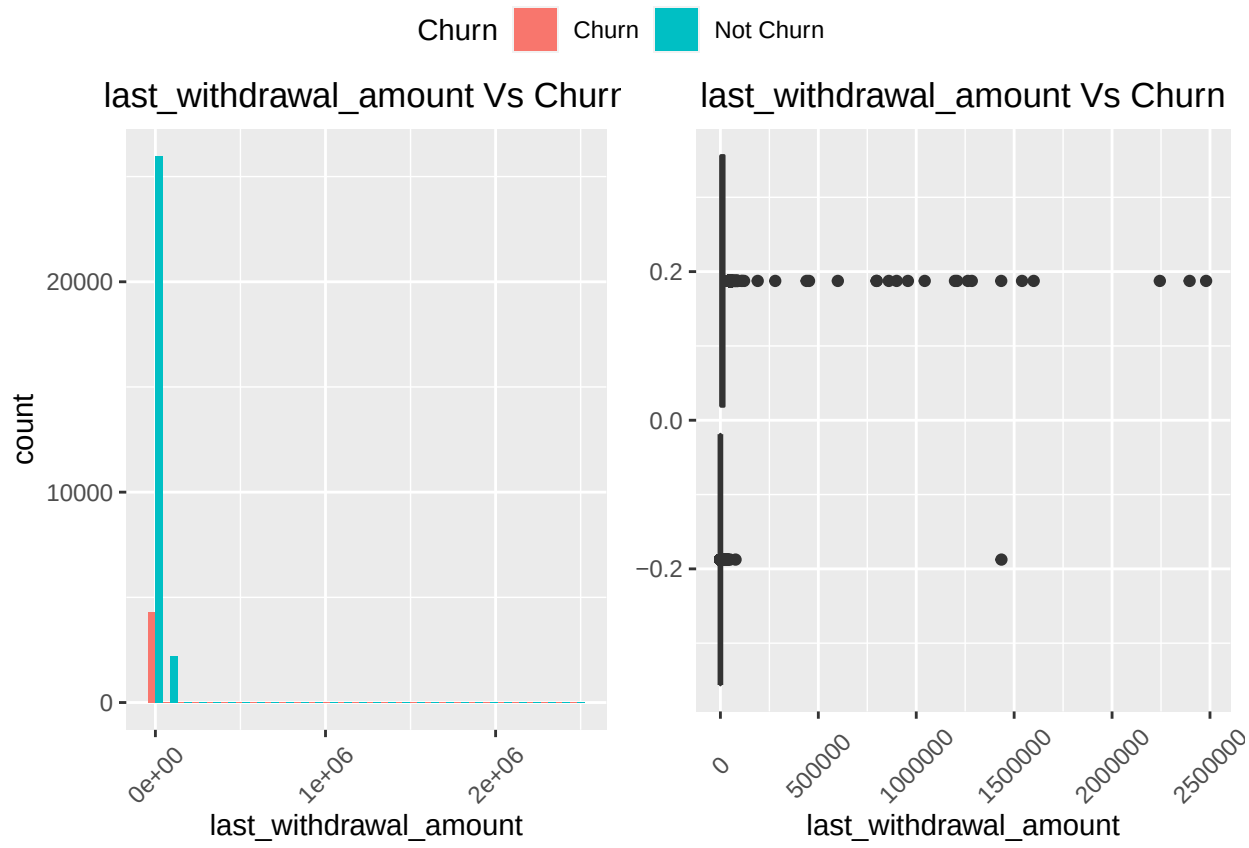
```
##
## [[19]]
```


Churn
Churn
Not Churn
withdrawals_3 Vs Churn
count
25000
20000
15000
10000
5000
0
0
50000
100000
150000
200000
withdrawals_3
withdrawals_3 Vs Churn
0.2
0.0
−0.2
0
50000
100000
150000
200000
withdrawals_3

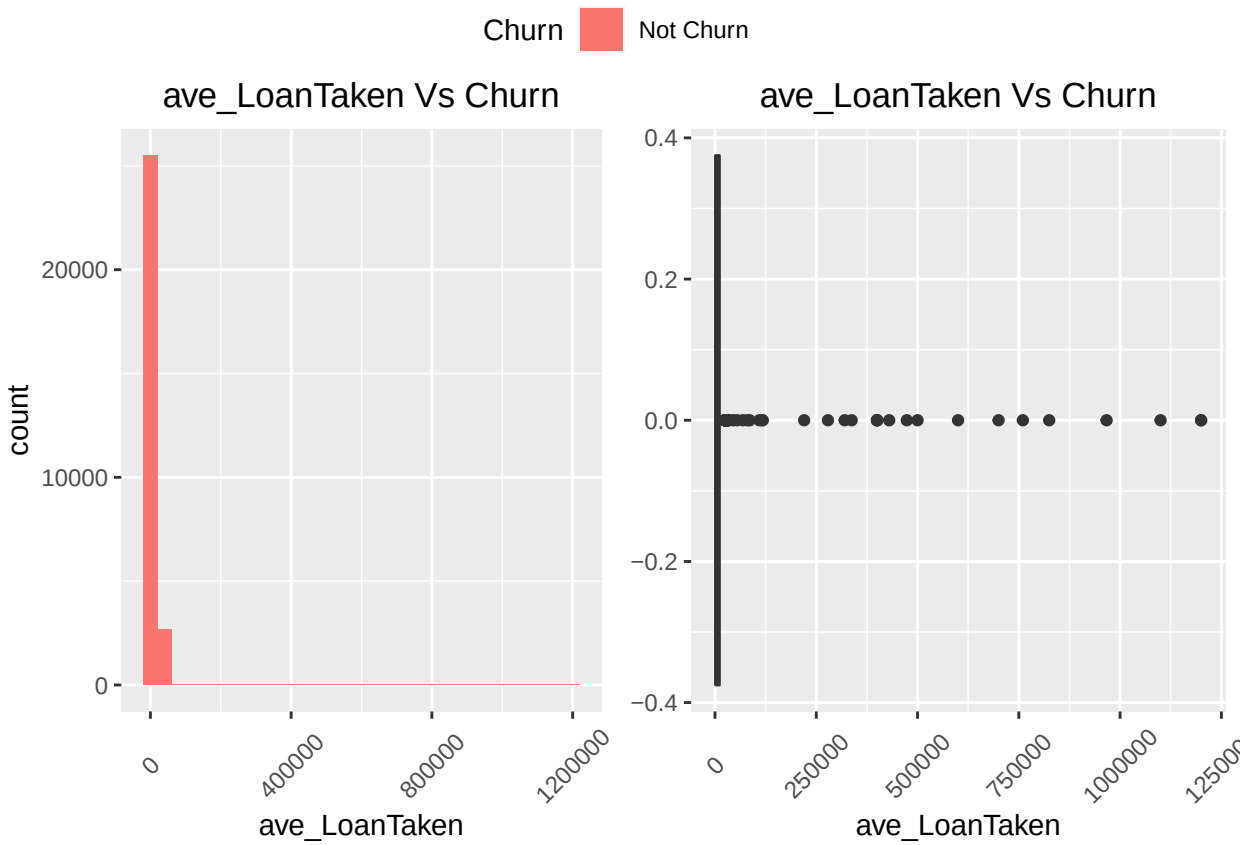
```
##
## [[20]]
```


Churn
Churn
Not Churn
last_deposit_amount Vs Churn
last_deposit_amount Vs Churn
count
20000
10000
0
0.2
0.0
−0.2
0
500000
1000000
0
500000
1000000
last_deposit_amount
last_deposit_amount

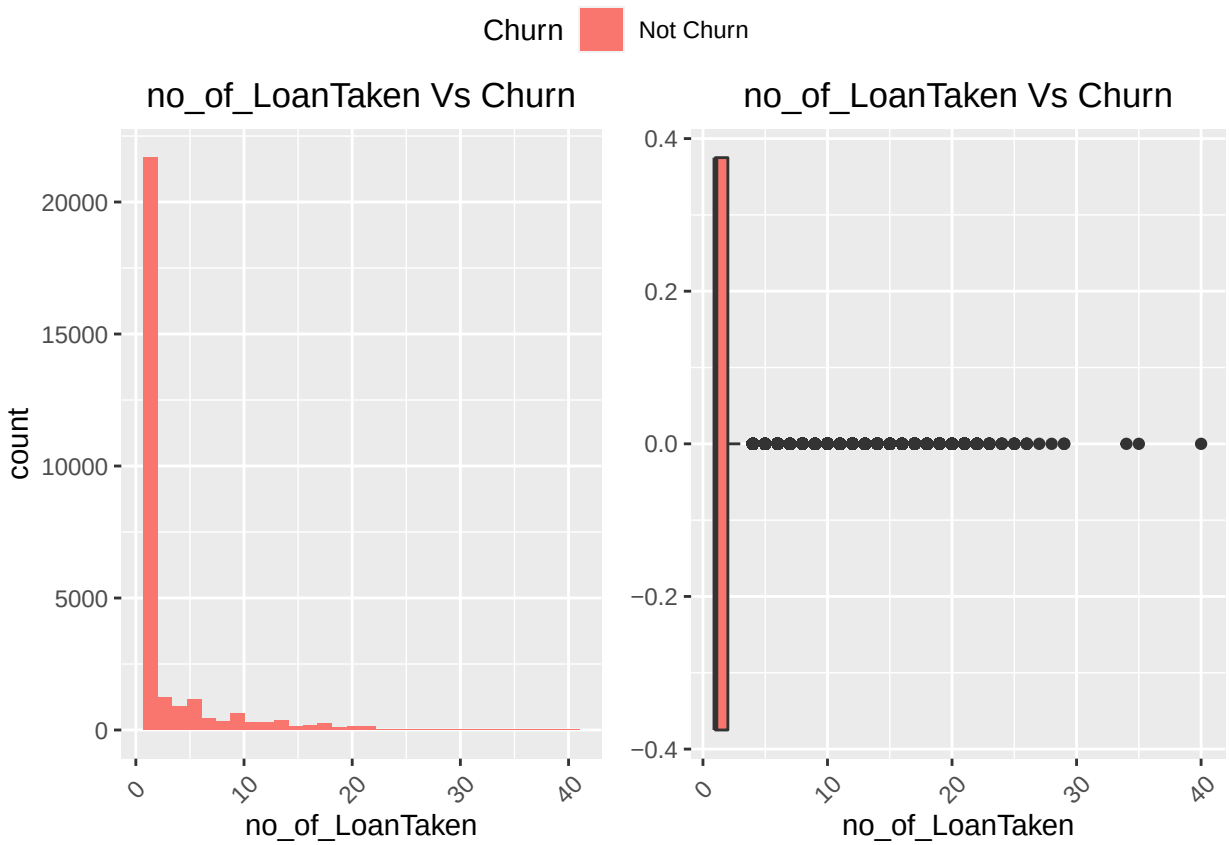
```
##
## [[21]]
```


Churn
Churn
Not Churn
last_withdrawal_amount Vs Churr
last_withdrawal_amount Vs Churn
count
20000
10000
0
0e+00
1e+06
2e+06
last_withdrawal_amount
0.2
0.0
−0.2
0
500000
1000000
1500000
2000000
2500000
last_withdrawal_amount

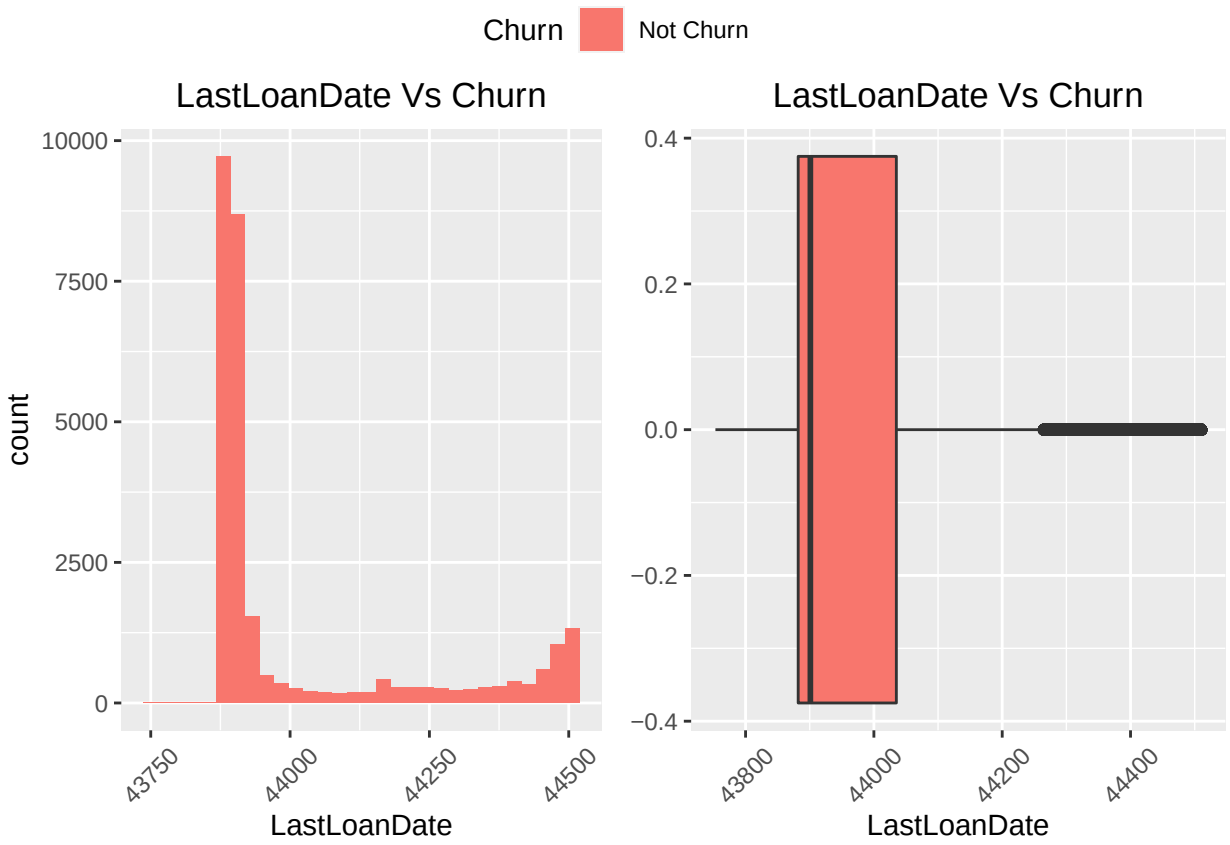
```
##
## [[22]]
```


Churn
Not Churn
ave_LoanTaken Vs Churn
count
20000
10000
0
0
400000
800000
1200000
ave_LoanTaken
ave_LoanTaken Vs Churn
0.4
0.2
0.0
−0.2
−0.4
0
250000
500000
750000
1000000
ave_LoanTaken

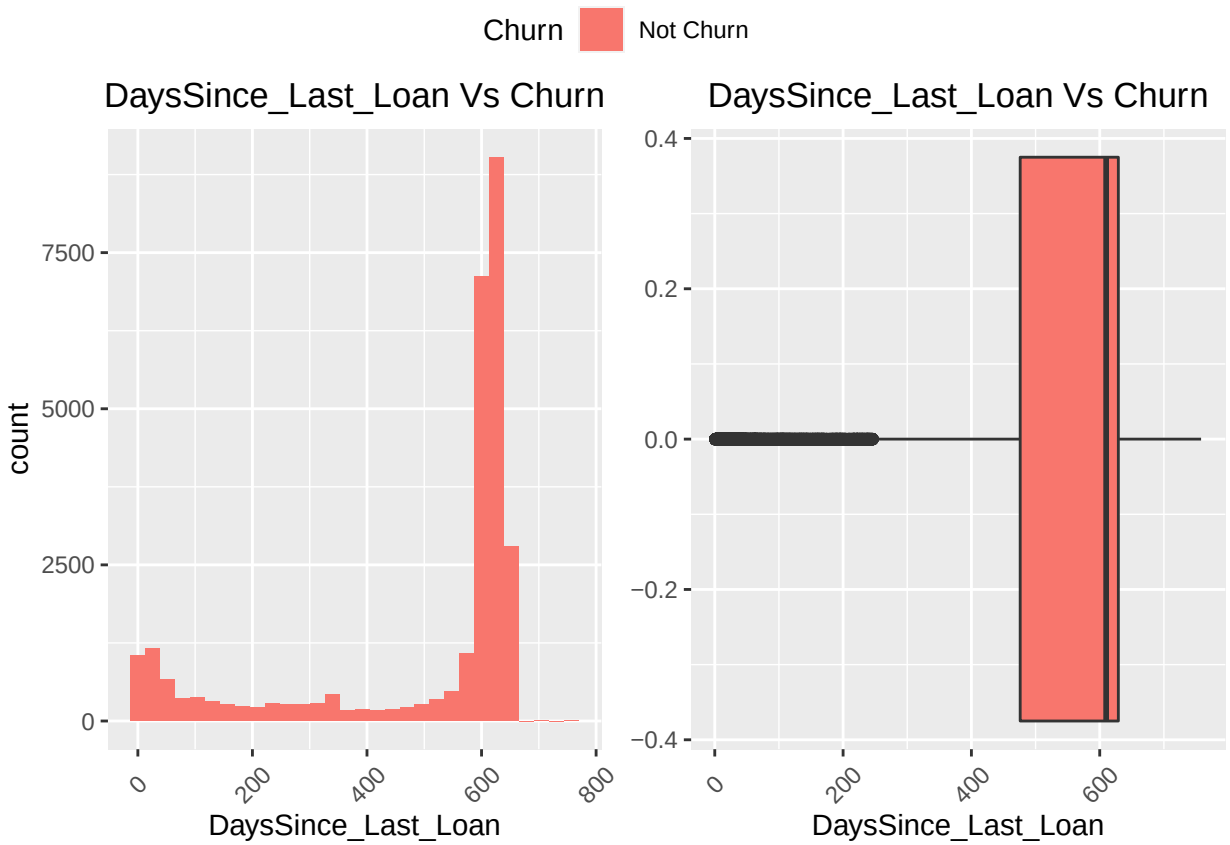
```
##
## [[23]]
```


Churn
Not Churn
no_of_LoanTaken Vs Churn
count
20000
15000
10000
5000
0
0
10
20
30
40
no_of_LoanTaken
no_of_LoanTaken Vs Churn
0.4
0.2
0.0
−0.2
−0.4
0
10
20
30
40
no_of_LoanTaken

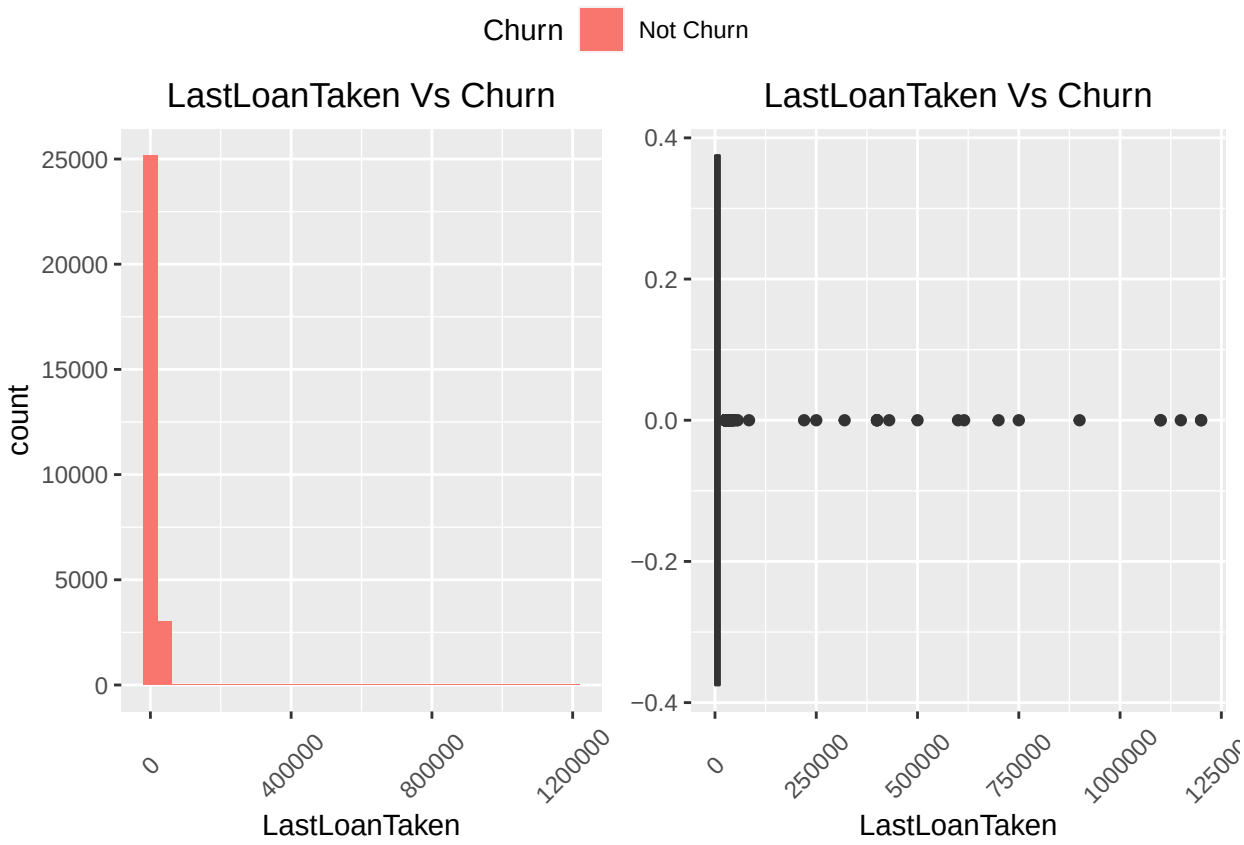
```
##
## [[24]]
```


Churn
Not Churn
LastLoanDate Vs Churn
count
10000
7500
5000
2500
0
43750
44000
44250
44500
LastLoanDate
LastLoanDate Vs Churn
0.4
0.2
0.0
−0.2
−0.4
43800
44000
44200
44400
LastLoanDate

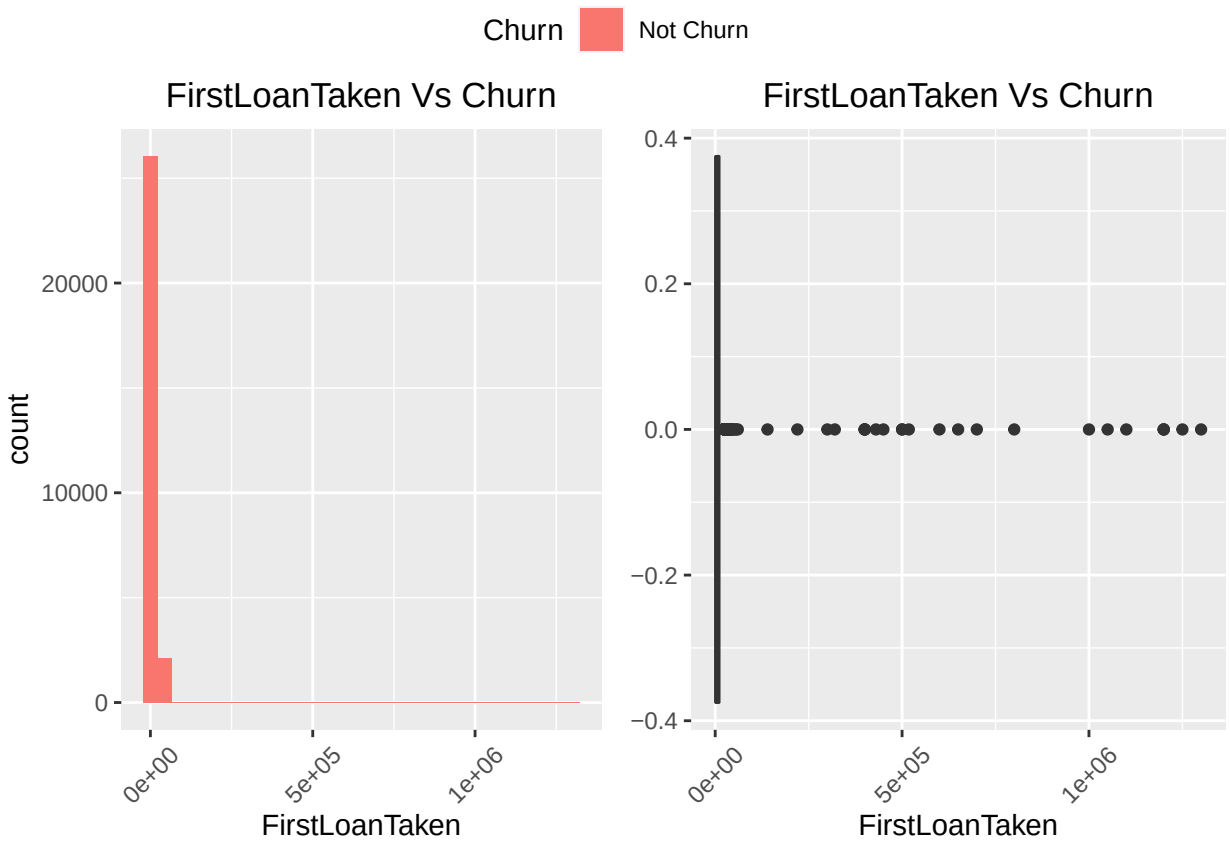
```
##
## [[25]]
```


Churn
Not Churn
DaysSince_Last_Loan Vs Churn
count
7500
5000
2500
0
0
200
400
600
800
DaysSince_Last_Loan
DaysSince_Last_Loan Vs Churn
0.4
0.2
0.0
−0.2
−0.4
0
200
400
600
DaysSince_Last_Loan

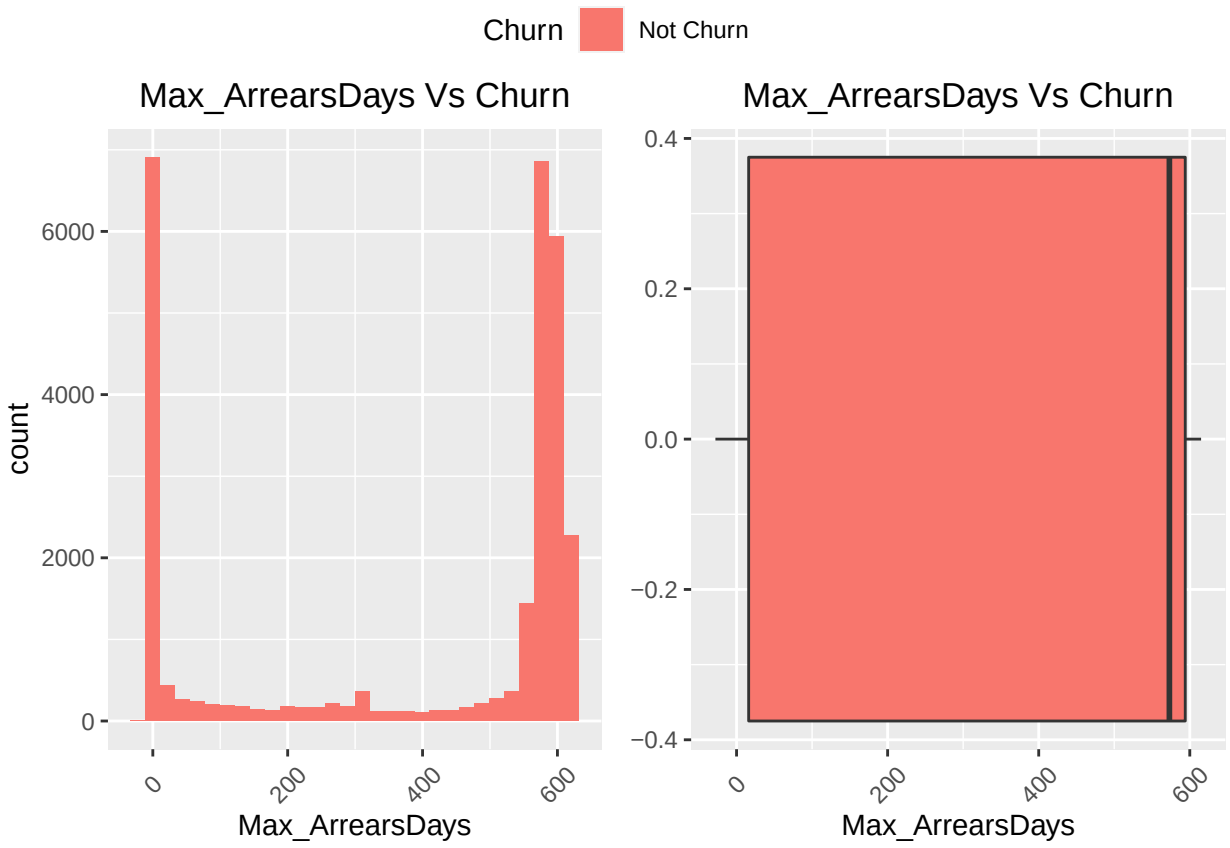
```
##
## [[26]]
```


Churn
Not Churn
LastLoanTaken Vs Churn
LastLoanTaken Vs Churn
count
25000
20000
15000
10000
5000
0
0
400000
800000
1200000
LastLoanTaken
0.4
0.2
0.0
−0.2
−0.4
0
250000
500000
750000
1000000
LastLoanTaken

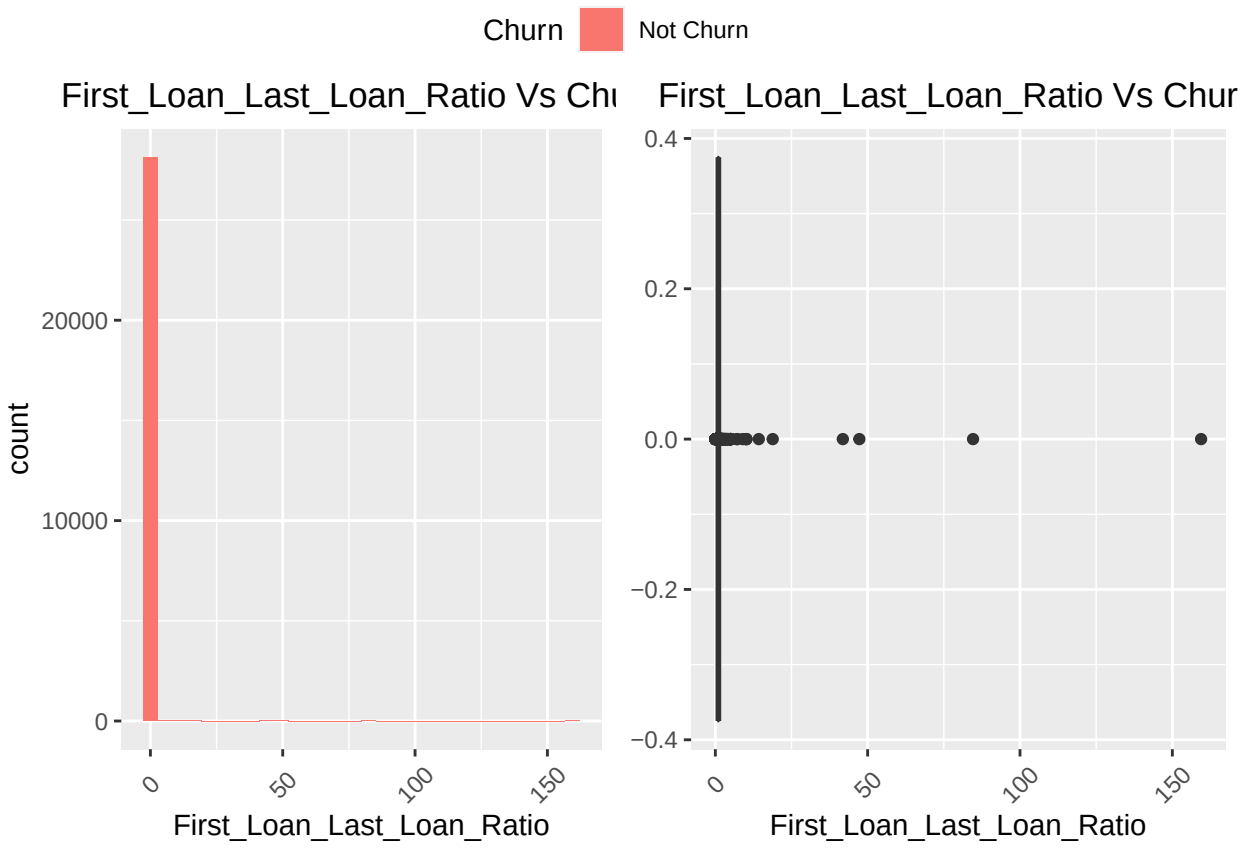
```
##
## [[27]]
```


Churn
Not Churn
FirstLoanTaken Vs Churn
FirstLoanTaken Vs Churn
count
20000
10000
0
0.4
0.2
0.0
−0.2
−0.4
0e+00
5e+05
1e+06
FirstLoanTaken
FirstLoanTaken

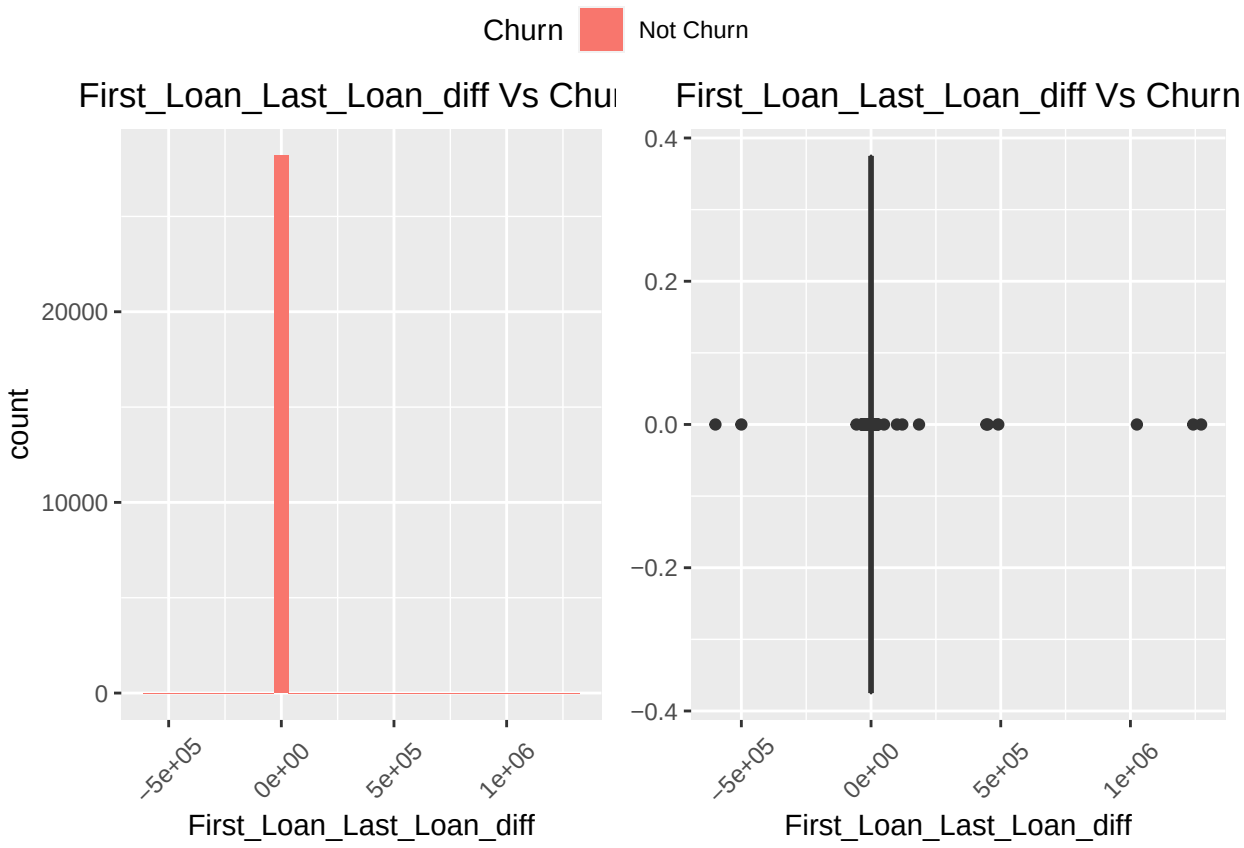
```
##
## [[28]]
```


Churn
Not Churn
Max_ArrearsDays Vs Churn
count
6000
4000
2000
0
0
200
400
600
Max_ArrearsDays
Max_ArrearsDays Vs Churn
0.4
0.2
0.0
−0.2
−0.4
0
200
400
600
Max_ArrearsDays

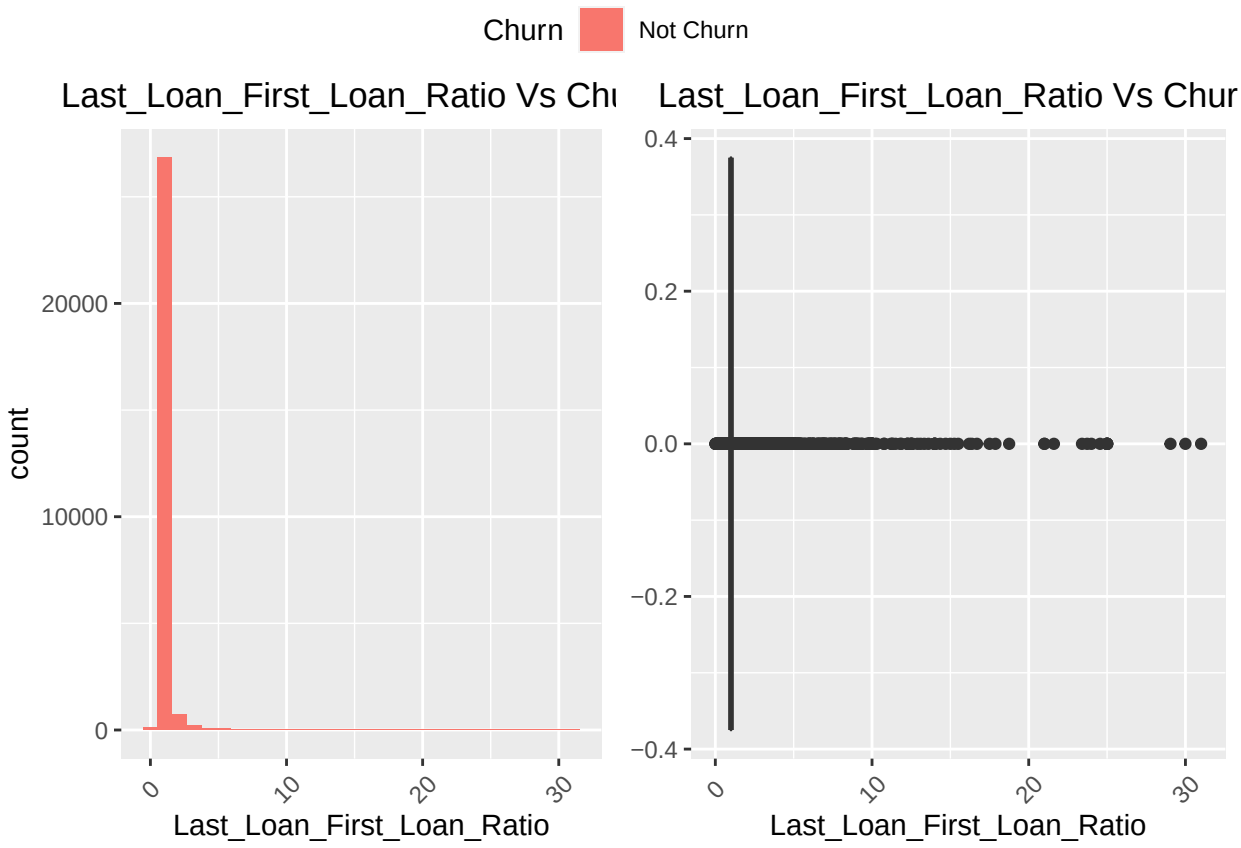
```
##
## [[29]]
```


Churn
Not Churn
First_Loan_Last_Loan_Ratio Vs Chu
First_Loan_Last_Loan_Ratio Vs Chur
count
20000
10000
0
0.4
0.2
0.0
−0.2
−0.4
0
50
100
150
First_Loan_Last_Loan_Ratio
First_Loan_Last_Loan_Ratio

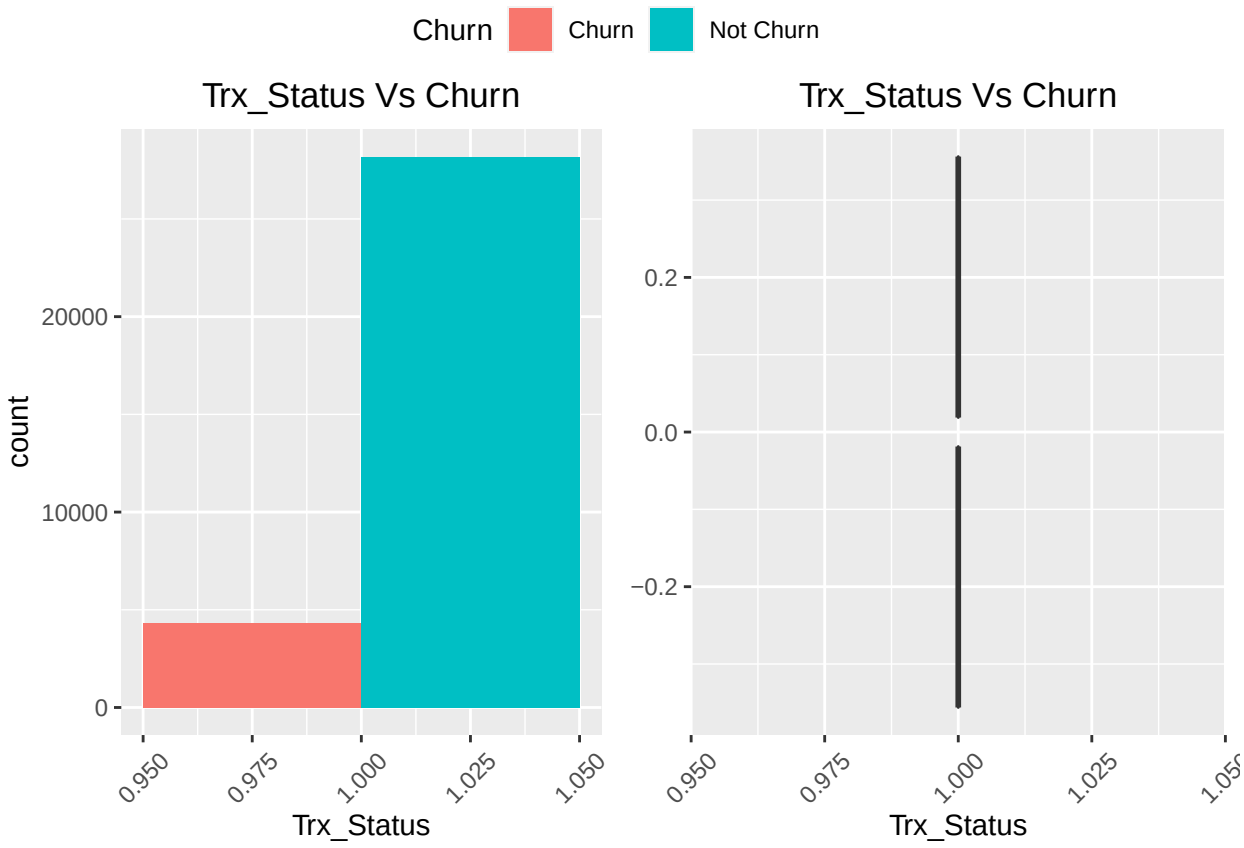
```
##
## [[30]]
```


Churn
Not Churn
First_Loan_Last_Loan_diff Vs Chu
First_Loan_Last_Loan_diff Vs Churn
count
20000
10000
0
0.4
0.2
0.0
−0.2
−0.4
−5e+05
0e+00
5e+05
1e+06
First_Loan_Last_Loan_diff
First_Loan_Last_Loan_diff

```
##
## [[31]]
```

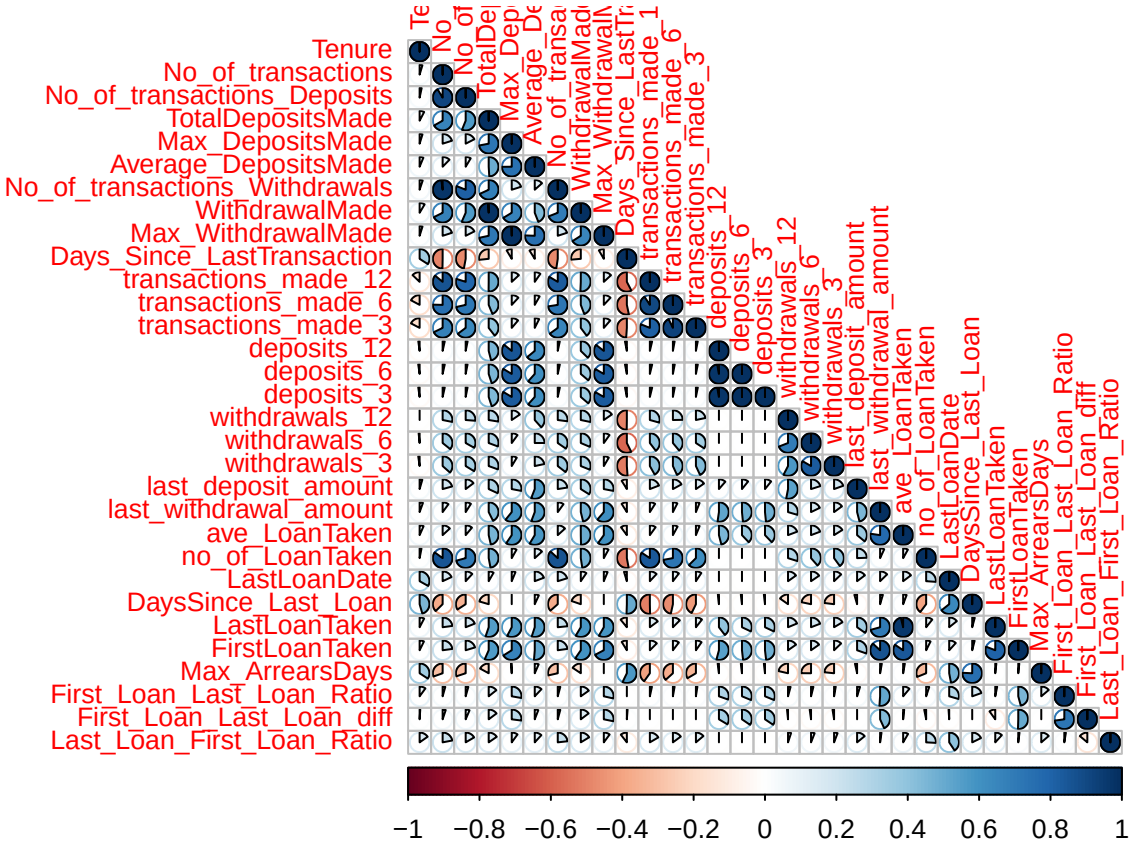

Churn
Not Churn
Last_Loan_First_Loan_Ratio Vs Chu
Last_Loan_First_Loan_Ratio Vs Chur
count
20000
10000
0
0.4
0.2
0.0
−0.2
−0.4
0
10
20
30
Last_Loan_First_Loan_Ratio
Last_Loan_First_Loan_Ratio

```
##
## [[32]]
```



Variable Correlation

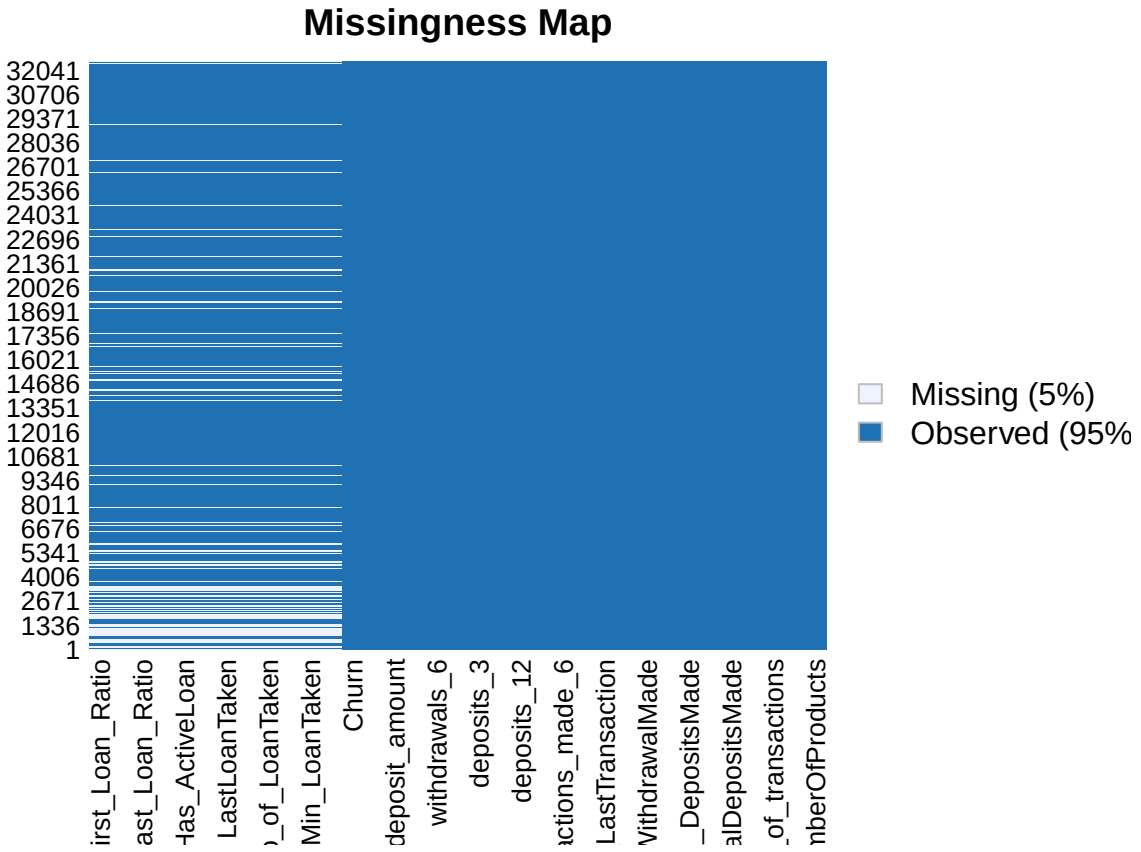
```
## corrplot 0.84 loaded
```


Tenure
No_of_transactions
No_of_transactions_Deposits
TotalDepositsMade
Max_DepositsMade
Average_DepositsMade
No_of_transactions_Withdrawals
WithdrawalMade
Max_WithdrawalMade
Days_Since_LastTransaction
transactions_made_12
transactions_made_6
transactions_made_3
deposits_12
deposits_6
deposits_3
withdrawals_12
withdrawals_6
withdrawals_3
last_deposit_amount
last_withdrawal_amount
ave_LoanTaken
no_of_LoanTaken
LastLoanDate
DaysSince_Last_Loan
LastLoanTaken
FirstLoanTaken
Max_ArrearsDays
First_Loan_Last_Loan_Ratio
First_Loan_Last_Loan_diff
Last_Loan_First_Loan_Ratio
−1
−0.8
−0.6
−0.4
−0.2
0
0.2
0.4
0.6
0.8
1

Data Pre processing

Drop Unncecessary Columns

Dealing with Missing Values



There are so many ways to deal with missing values.For instance in this dataset most variables are missing because of no transaction history.In this case they will be imputed with value zero.

From the above chart,its evident that Loan Related variables are more likely to have a missing value than any other transaction related variable.Such cases would only imply that the said clients do not have any loan transaction history with the financial institution.

Dealing with Skewness and abnormality

As observed from above,the variables were heavily skewed,to correct this ,min max scaling was used.

Train test Split

Upsampling Minority Class

Modeling

Logistic regression

```
## Warning: glm.fit: algorithm did not converge
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: 'tidy.numeric' is deprecated.
## See help("Deprecated")

## Warning: `data_frame()` was deprecated in tibble 1.1.0.
## Please use `tibble()` instead.

## # A tibble: 34 x 2
##    names                          coeff
##    <chr>                          <dbl>
##  1 (Intercept)                   -26.0
##  2 NumberOfProducts                2.14
##  3 Tenure                         -0.340
##  4 No_of_transactions            118.
##  5 No_of_transactions_Deposits   -48.3
##  6 TotalDepositsMade             435.
##  7 Max_DepositsMade              -61.8
##  8 Average_DepositsMade            0.987
##  9 No_of_transactions_Withdrawals -98.8
## 10 WithdrawalMade                -547.
## # ... with 24 more rows

## Confusion Matrix and Statistics
##
##            Reference
## Prediction  Churn Not Churn
##   Churn       863         1
##   Not Churn     0      5637
##
```

```
##                Accuracy : 0.9998
##                  95% CI : (0.9991, 1)
##     No Information Rate : 0.8673
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.9993
##
##  Mcnemar's Test P-Value : 1
##
##             Sensitivity : 1.0000
##             Specificity : 0.9998
##          Pos Pred Value : 0.9988
##          Neg Pred Value : 1.0000
##              Prevalence : 0.1327
##          Detection Rate : 0.1327
##    Detection Prevalence : 0.1329
##       Balanced Accuracy : 0.9999
##
##        'Positive' Class : Churn
##
```

Decision Trees

Model definition

```
##
## Classification tree:
## tree(formula = Churn ~ ., data = trainsmote)
## Variables actually used in tree construction:
## [1] "Max_LoanTaken"
## Number of terminal nodes:  2
## Residual mean deviance:  0 = 0 / 24160
## Misclassification error rate: 0 = 0 / 24164
```

Model evaluation

Random Forest

```
## [1] 1
```

```
## Confusion Matrix and Statistics
##
##            Reference
## Prediction  Churn Not Churn
##   Churn       863         0
##   Not Churn     0      5638
##
##                Accuracy : 1
##                  95% CI : (0.9994, 1)
##     No Information Rate : 0.8673
##     P-Value [Acc > NIR] : < 2.2e-16
##
```

```
##                  Kappa : 1
##
##  Mcnemar's Test P-Value : NA
##
##             Sensitivity : 1.0000
##             Specificity : 1.0000
##          Pos Pred Value : 1.0000
##          Neg Pred Value : 1.0000
##              Prevalence : 0.1327
##          Detection Rate : 0.1327
##    Detection Prevalence : 0.1327
##       Balanced Accuracy : 1.0000
##
##        'Positive' Class : Churn
##
```