

Alexandre Azuma
Cristiano Almeida
Daniele Lyra

Segmentação de clientes com base no uso do cartão de crédito.

Trabalho apresentado ao curso MBA
Business Analytics e Big Data, da
Fundação Getúlio Vargas para a
disciplina de Métodos Matriciais
e Análise de Clusters.

São Paulo

02 de outubro de 2020.

Introdução

O mercado de cartões de crédito e débito cresce no Brasil a cada ano, o país já tem o maior número de cartões de crédito e débito em circulação da América Latina segundo estudo da Minsait-INDRA.

O Banco Central (BC) divulgou que no ano de 2019, segundo levantamento “Estatísticas de Pagamentos de Varejo e de Cartões no Brasil”, existiam 123 milhões de cartões de crédito e 132 milhões de cartões de débitos ativos no País. Gastos com cartão de crédito chegaram a R\$ 1,089 trilhão em transações.

Segundo relatório, divulgada pela Associação das Empresas de Cartões de Crédito e Serviços (ABECS), as compras remotas ganharam destaque durante o período da quarentena com um crescimento de 14% em relação ao segundo trimestre de 2019 atingindo a marca de R\$ 86,7 bilhões de transações com cartão de crédito. O pagamento por aproximação (NFC) cresceu 256% em relação ao segundo trimestre de 2019 e movimentou R\$ 4,3 bilhões nesse segundo trimestre de 2020.

Visto a relevância do mercado do uso de cartão de crédito, optamos por trabalhar com um conjunto de dados de amostra que resume o comportamento de uso do cartão de crédito de cerca de 9.000 titulares de cartões de créditos ativos durante os últimos 06 meses.

Base de Dados

O conjunto de dados que consiste no comportamento de uso do cartão de crédito de clientes com 18 variáveis de recursos comportamentais e 8.950 observações.

Essa base está disponível na plataforma do Kaggle, que é uma subsidiária da Google LLC, sendo uma comunidade online de cientistas de dados e profissionais de aprendizado de máquina.

Fonte: <https://www.kaggle.com/arjunbhasin2013/ccdata>

A seguir o dicionário de dados do conjunto de dados de cartão de crédito:

CUSTID	Identificação do titular do cartão de crédito.
BALANCE	Saldo restante na conta para fazer compras.
BALANCE_FREQUENCY	com que frequência o saldo é atualizado, pontuação entre 0 e 1 (1 = atualizado com frequência, 0 = não atualizado com frequência)
PURCHASES	Valor das compras feitas na conta.
ONEOFF_PURCHASES	Valor máximo da compra realizada de uma só vez.
INSTALLMENTS_PURCHASES	Valor da compra parcelado.
CASH_ADVANCE	Adiantamento em dinheiro dado pelo usuário.
PURCHASES_FREQUENCY	a frequência com que as compras estão sendo feitas, pontuação entre 0 e 1 (1 = compra frequente, 0 = compra não frequente).
ONEOFF_PURCHASES_FREQUENCY	a frequência com que as compras acontecem de uma só vez (1 = compra frequente, 0 = compra não frequente).
PURCHASES_INSTALLMENTS_FREQUENCY	Com que frequência as compras parceladas estão sendo feitas (1 = feito com frequência, 0 = não feito com frequência).
CASH_ADVANCE_FREQUENCY	Com que frequência o dinheiro adiantado está sendo pago.
CASH_ADVANCE_TRX	Número de transações feitas com "Cash in Advance".
PURCHASES_TRX	Número de Transações de compras feitas.
CREDIT_LIMIT	Límite de cartão de crédito para o usuário.
PAYMENTS	Valor do pagamento feito pelo usuário.
MINIMUM_PAYMENTS	Valor mínimo de pagamentos feitos pelo usuário.
PRC_FULL_PAYMENT	Porcentagem do pagamento total pago pelo usuário.
TENURE	Tempo de posse do serviço de cartão de crédito para o usuário.

Análise das Variáveis

Desconsideramos da análise a variável `CUSTID` por representar a identificação do titular do cartão, não sendo significativa para nossa análise. O `CUSTID` é sequencial e para cada cliente só existe uma transação.

Identificamos em apenas duas variáveis (`CREDIT_LIMIT` e `MINIMUM_PAYMENTS`) com campos nulos (NA's). Consideramos tratar esses campos preenchendo com o valor zero (0), como as variáveis consiste no comportamento do usuário do cartão, entendemos que se optarmos por outro tratamento como atribuir um valor médio ou outro qualquer poderíamos estar "inferindo" um determinado tipo de comportamento para o usuário que não seria real.

No tratamento dos outliers identificamos que poderíamos perder muitos elementos, pois há muitos outliers no conjunto de dados. Dado isso, resolvermos tratar os outliers com a técnica de winsorizing atribuindo um valor de referência dado por $Q3 + 1.5 * IQR$.

Algumas variáveis de frequência acreditamos que pode ter ocorrido erro de preenchimento nesse campo. Uma vez que a documentação do conjunto de dados, informa que o valor esperado do campo é um valor binário em que representa a frequência: 1 = atualizado com frequência, 0 = não atualizado com frequência.

Entretanto, esses campos possuem valores diferentes de 1 e 0, devido a isso resolvemos desconsiderar para o modelo. As variáveis desconsideradas foram: *BALANCE_FREQUENCY*, *PURCHASES_FREQUENCY*, *ONEOFF_PURCHASES_FREQUENCY*, *PURCHASES_INSTALLMENTS_FREQUENCY* e *CASH_ADVANCE_FREQUENCY*.

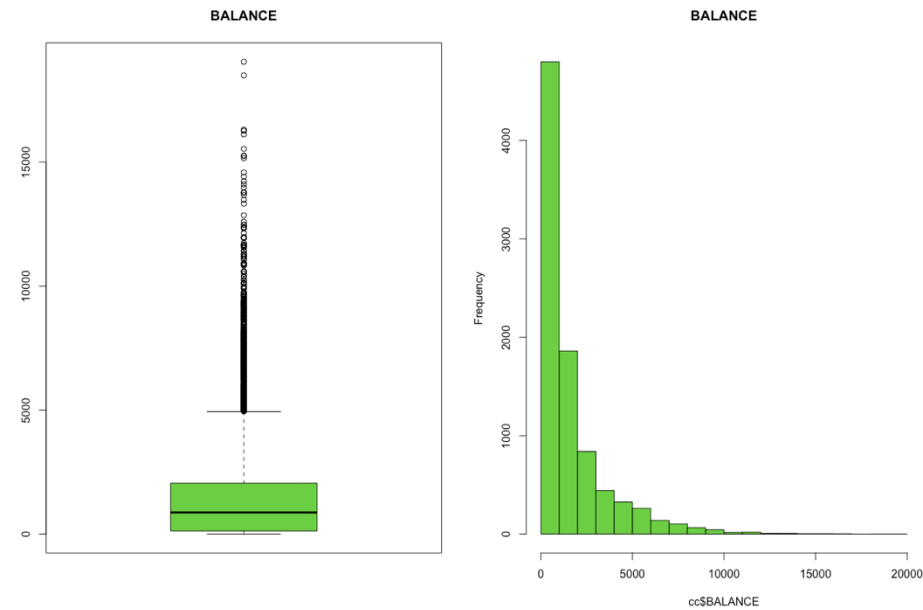
A variável *TENURE* também foi desconsiderada, pois não encontramos na documentação qual seria a medida de unidade dessa variável, se os valores representam as quantidades de meses, ou se é um número atribuído para indicar o mês, os valores começam no 6 e vai até 12.

BALANCE

saldo restante na conta para fazer compras.

Summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0	128.3	873.4	1564.5	2054.1	19043.1



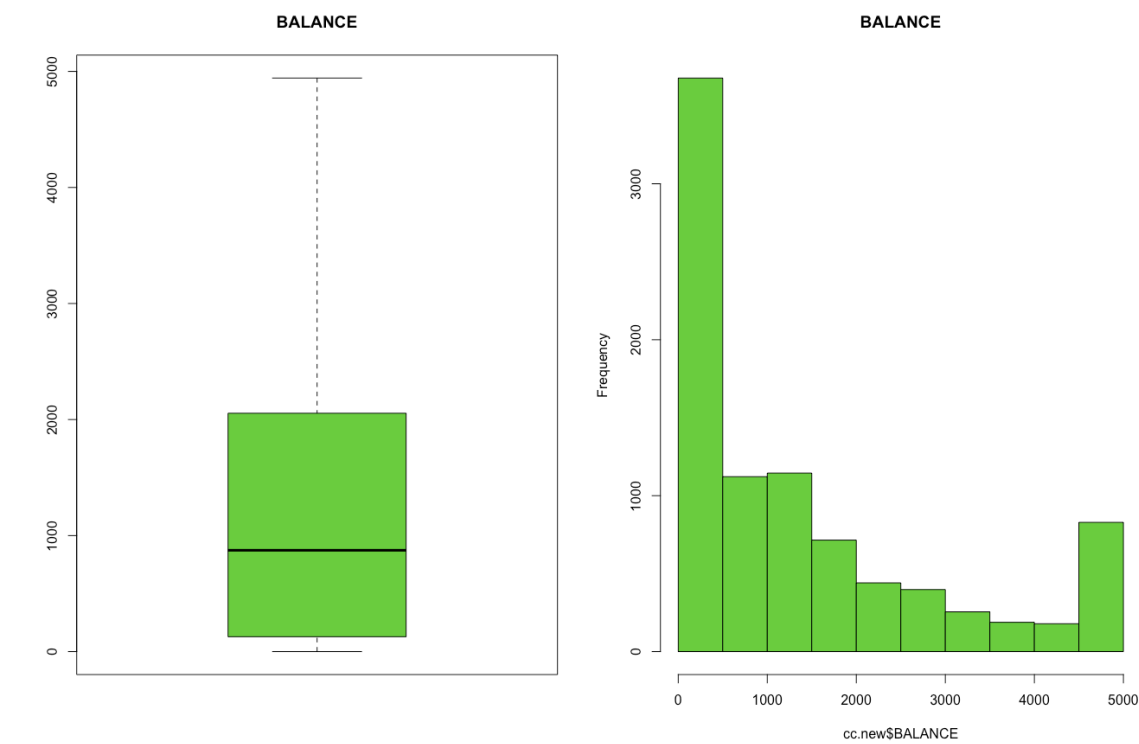
os outlier's representam cerca de 8% do total da base de dados.

BALANCE > 5000	BALANCE <= 5000
[1] 682	[1] 8268

Após o tratamento dos outliers ficou melhor distribuído.

Summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0	128.3	873.4	1390.5	2054.1	4942.9

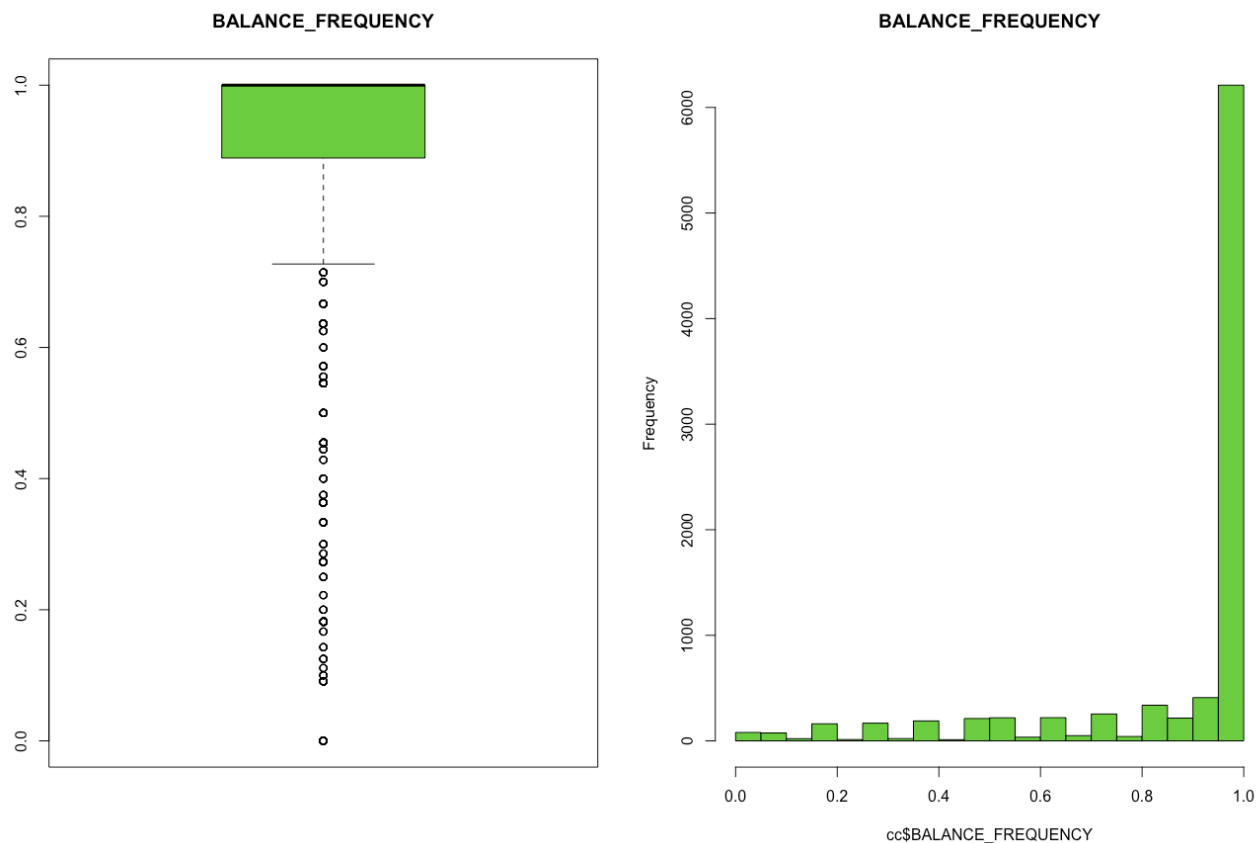


BALANCE_FREQUENCY

Com que frequência o saldo é atualizado, pontuação entre 0 e 1 (1 = atualizado com frequência, 0 = não atualizado com frequência)

Summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0000	0.8889	1.0000	0.8773	1.0000	1.0000



Conforme a documentação, da base essa variável deveria ter um valor binário que representa a frequência, acreditamos que pode ter ocorrido erro de preenchimento nesse campo.

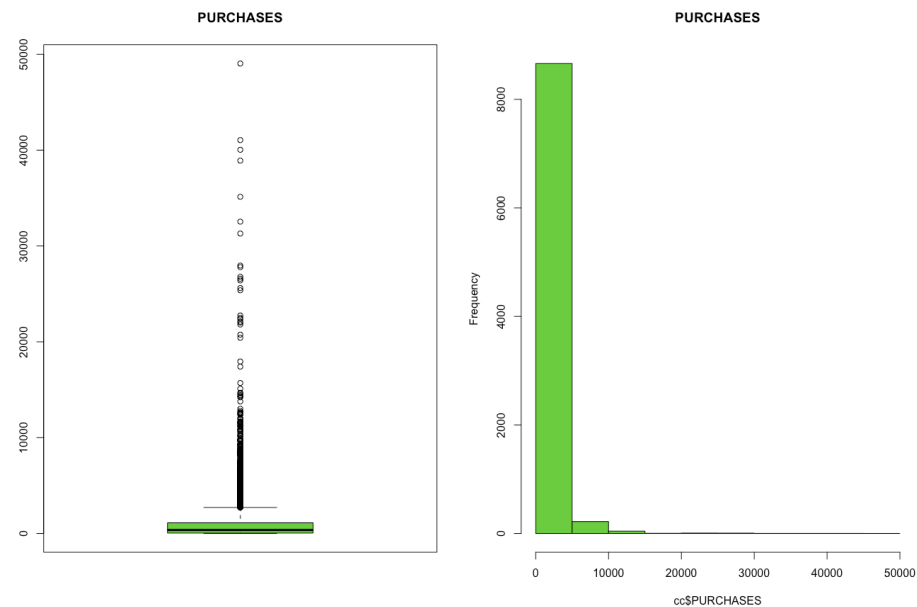
Existem 6.211 elementos com frequência igual a 1 e 80 elementos com frequência 0 (zero). Por isso, não vamos considerar como variável driver.

PURCHASES

valor das compras feitas na conta.

Summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	39.63	361.28	1003.20	1110.13	49039.57



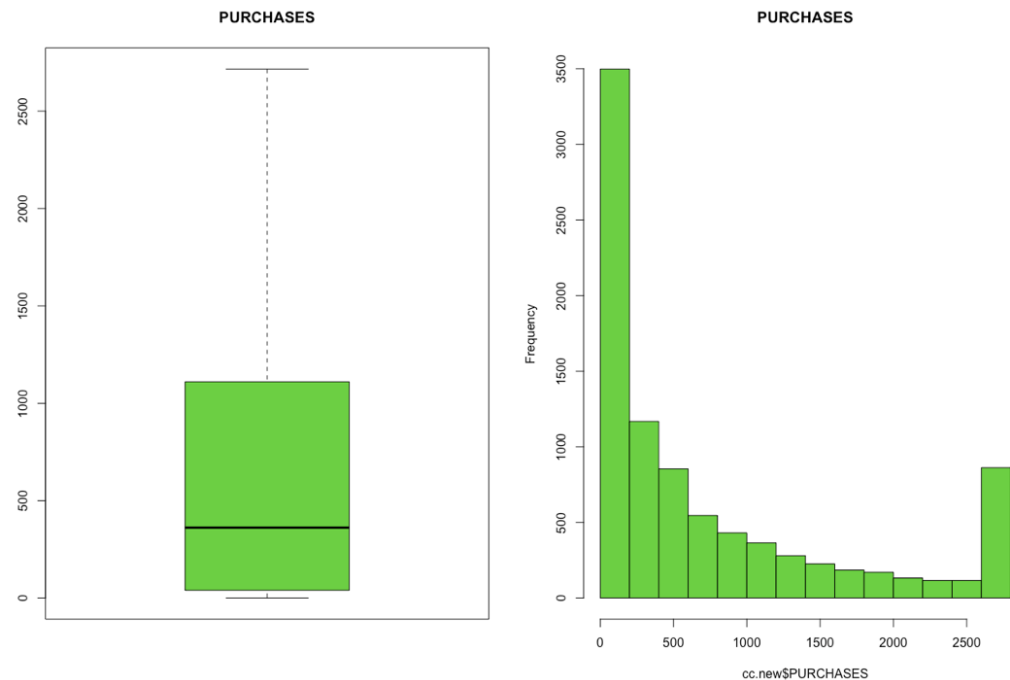
PURCHASES >=5000 [1] 290	PURCHASES <5000 [1] 8660
-----------------------------	-----------------------------

Os outliers representam 290 clientes que possuem compras acima de 5.000 unidades monetárias. A maioria efetua compras com valores abaixo de 5.000 unidades monetárias.

Após o tratamento dos outliers.

Summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	39.63	361.28	737.15	1110.13	2715.87

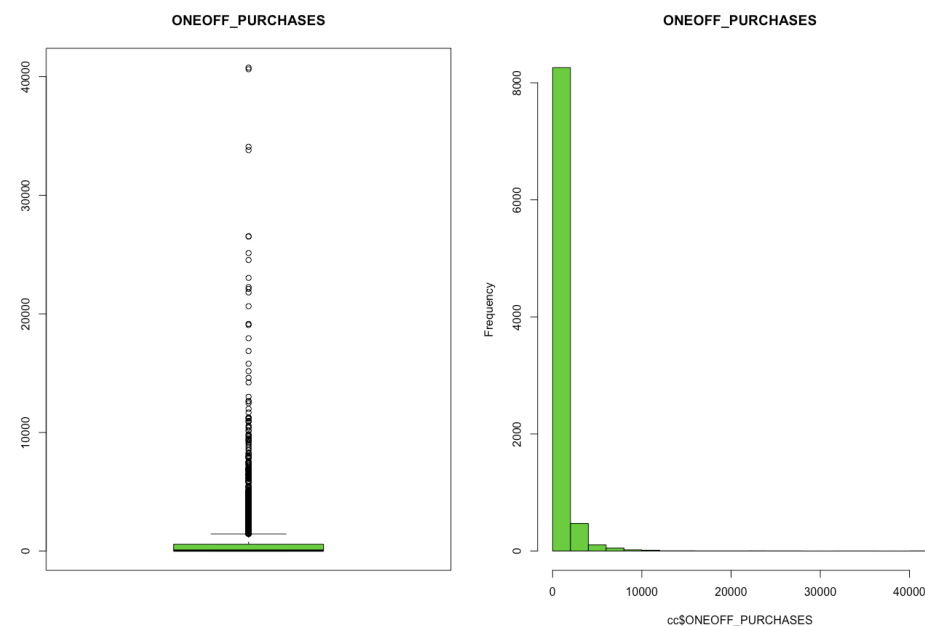


ONEOFF_PURCHASES

valor máximo da compra realizada de uma só vez.

Summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0	0.0	38.0	592.4	577.4	40761.2

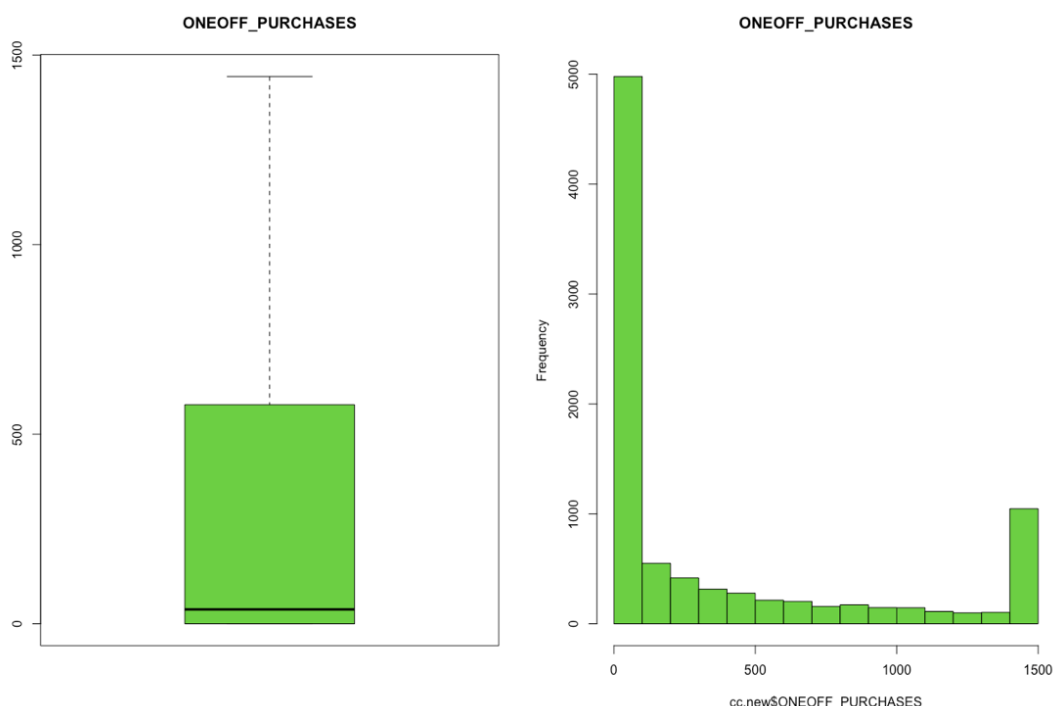


ONEOFF_PURCHASES >=1000 [1] 1522	ONEOFF_PURCHASES <1000 [1] 7428
-------------------------------------	------------------------------------

A maioria dos clientes fazem compras únicas com valor máximo abaixo de 1.000 unidades monetárias. Essa variável tem um comportamento bem parecido com a anterior PURCHASES, a diferença é que nessa mostra o valor máximo que o cliente está disposto a gastar em uma única compra.

Após o tratamento dos outliers.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0	0.0	38.0	354.4	577.4	1443.5

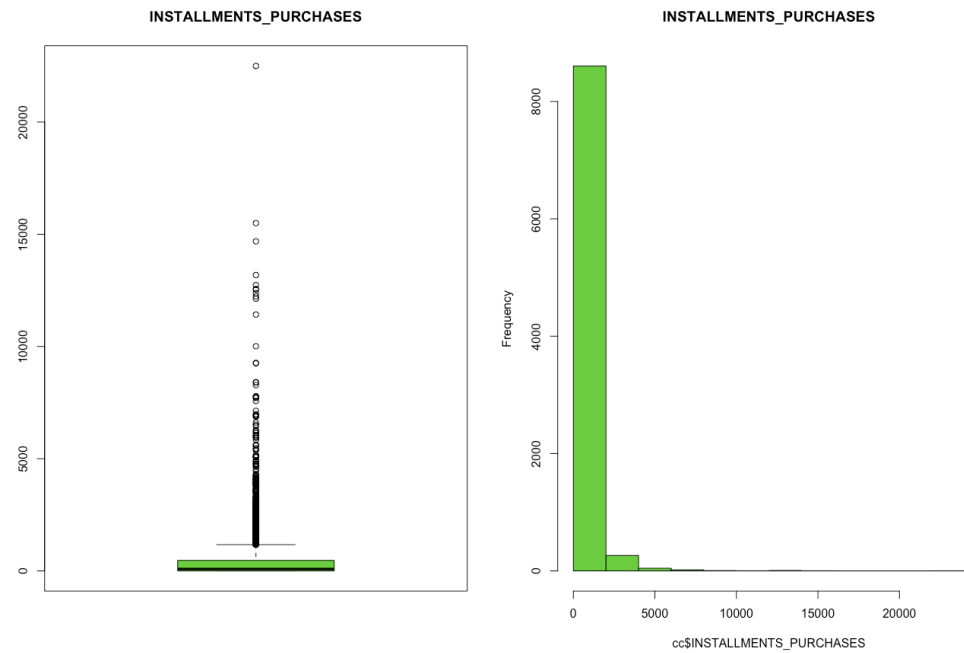


INSTALLMENTS_PURCHASES

valor da compra parcelado.

Summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0	0.0	89.0	411.1	468.6	22500.0

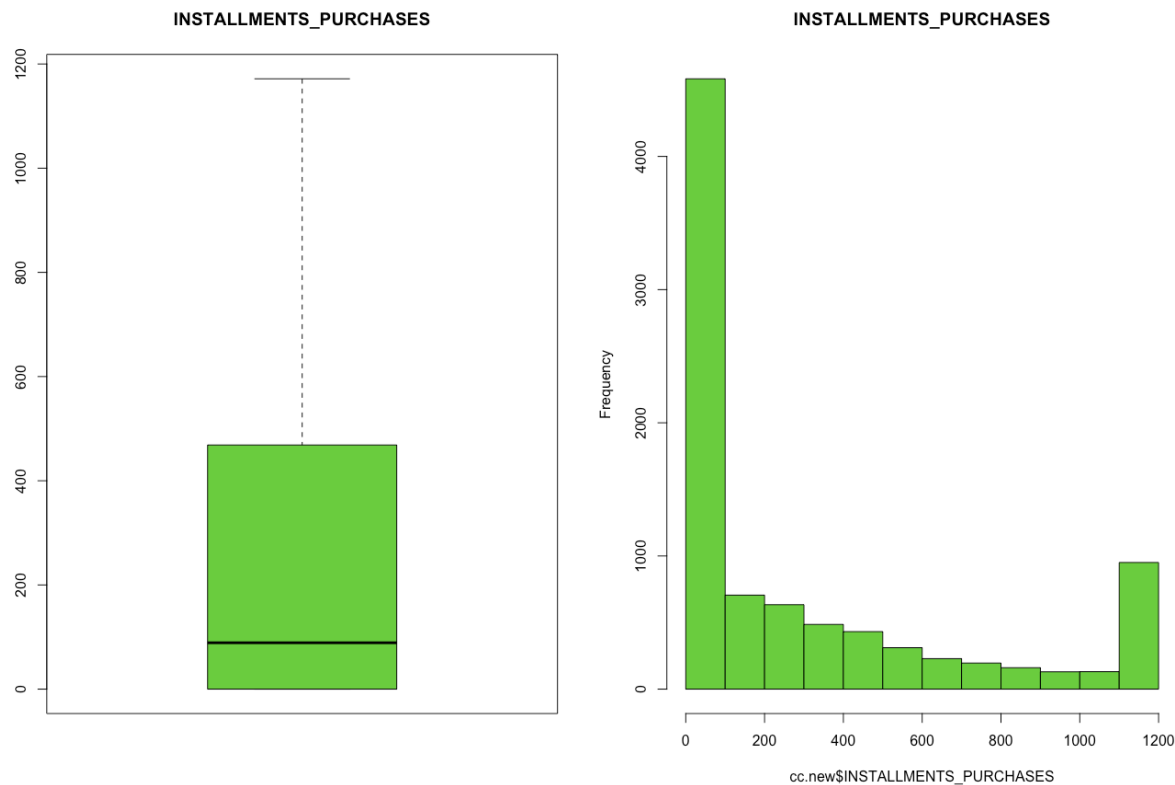


<code>INSTALLMENTS_PURCHASES>=1000</code> [1] 1086	<code>INSTALLMENTS_PURCHASES<1000</code> [1] 7864
--	---

o valor da compra parcela para a maioria é abaixo de 1.000 unidades monetárias.

Após o tratamento dos outliers.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0	0.0	89.0	295.5	468.6	1171.6

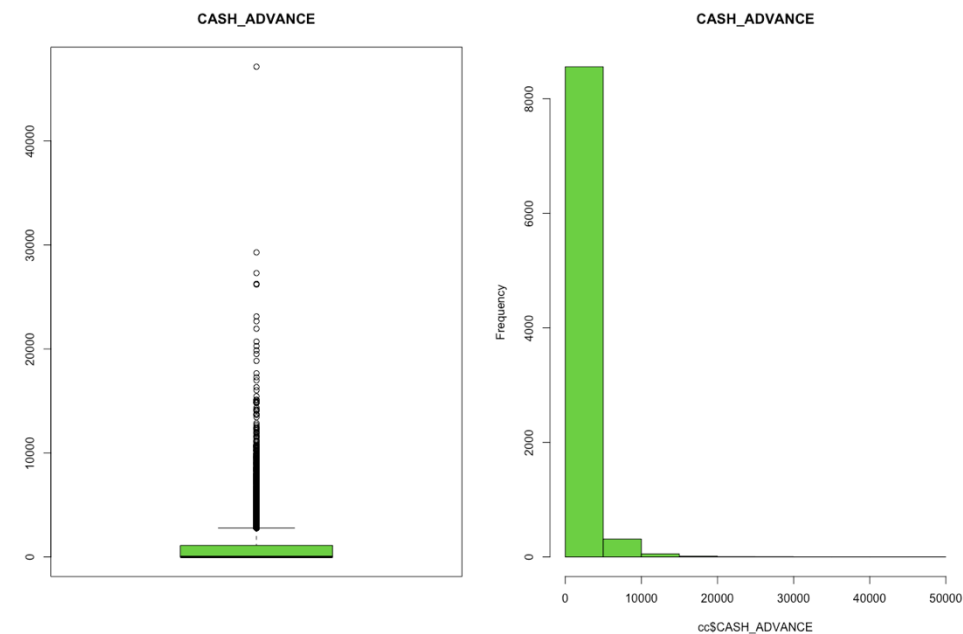


CASH_ADVANCE

Adiantamento em dinheiro dado pelo usuário.

Summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0	0.0	0.0	978.9	1113.8	47137.2



A maioria dos clientes estão dispostos a efetuar um adiantamento em dinheiro de até 5.000 unidades monetárias.

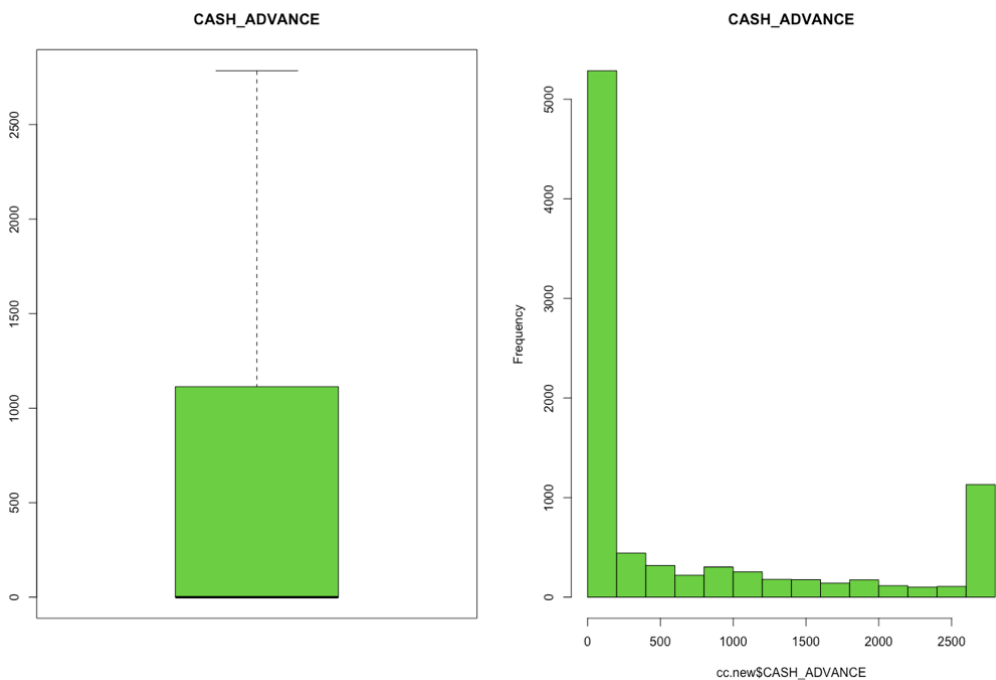
CASH_ADVANCE >=5000	CASH_ADVANCE <5000
[1] 391	[1] 8559

Cerca de 52% da base de clientes não efetuaram adiantamento em dinheiro para o cartão de crédito.

	x	freq
1	0.00000	4628

Após o tratamento dos outliers.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0	0.0	0.0	672.5	1113.8	2784.5

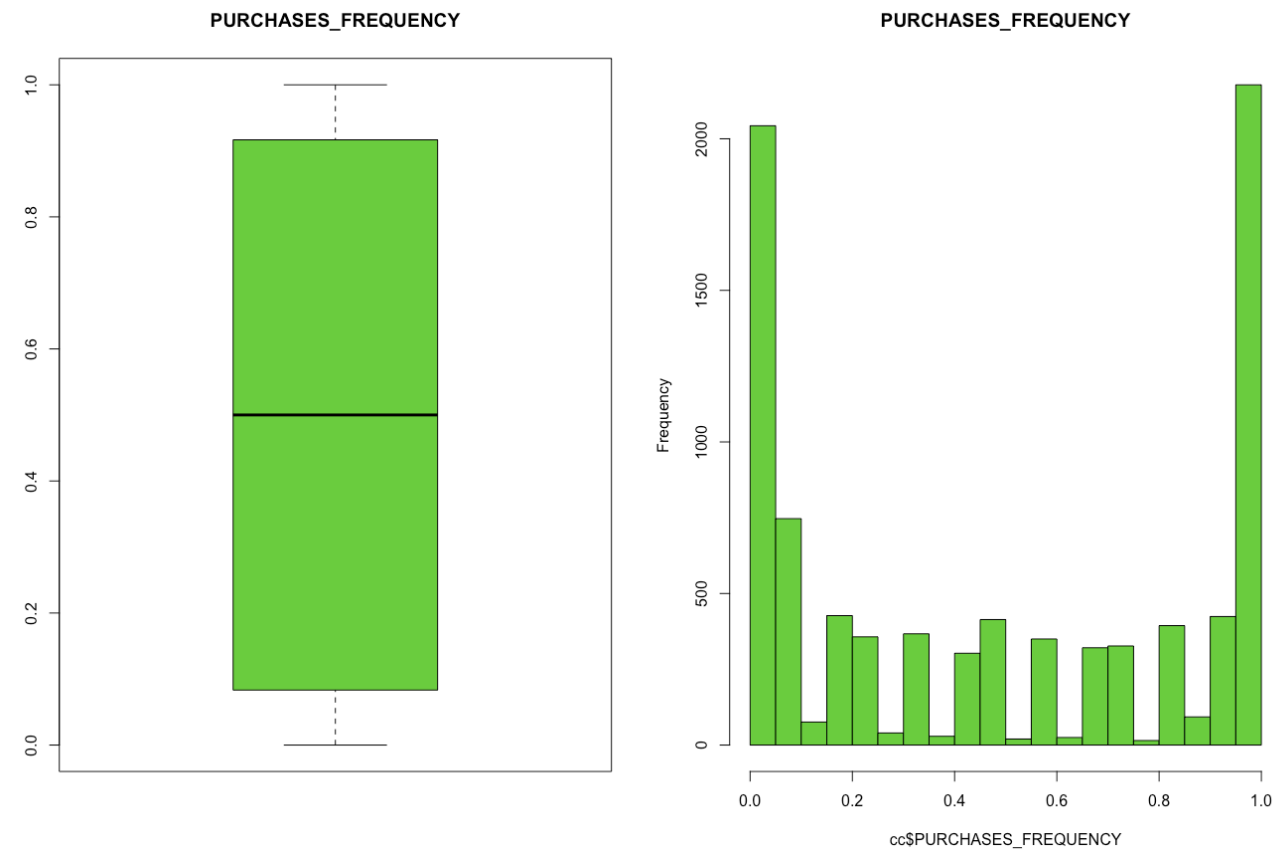


PURCHASES_FREQUENCY

A frequência com que as compras estão sendo feitas, pontuação entre 0 e 1 (1 = compra frequente, 0 = compra não frequente).

Summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00000	0.08333	0.50000	0.49035	0.91667	1.00000



Assim como a BALANCE_FREQUENCY, acreditamos que também houve erro no preenchimento dessa variável, uma vez que também era esperado um valor binário. Também, não vamos considerar como variável driver.

Existem 2043 elementos com frequência igual a 1 e 2178 elementos com frequência 0 (zero), um total de 4.221 elementos preenchidos corretamente e 4.729 com valores no intervalo de 0 a 1, o que a outra metade da base.

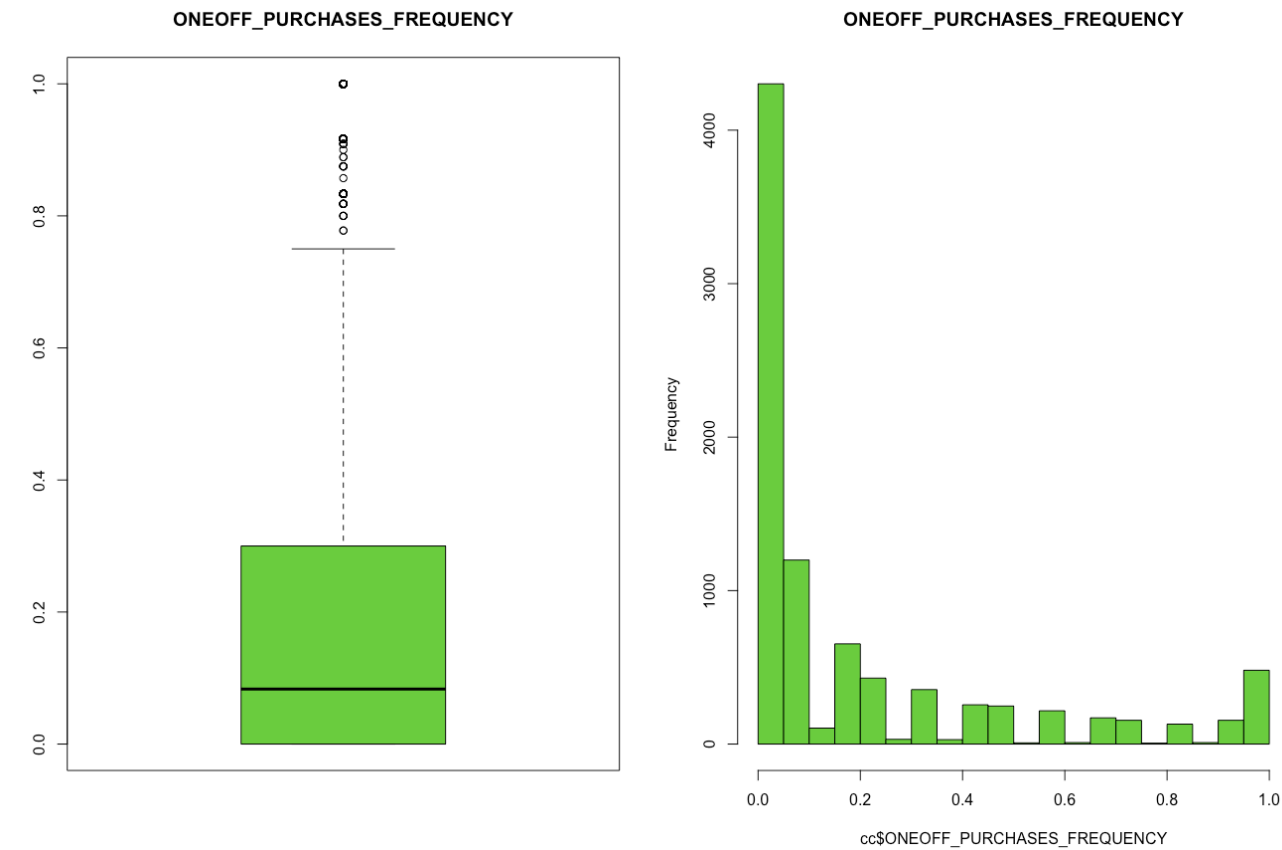
x	freq
0.000000	2043
1.000000	2178

ONEOFF_PURCHASES_FREQUENCY

A frequência com que as compras acontecem de uma só vez (1 = compra frequente, 0 = compra não frequente).

Summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0	0.00000	0.08333	0.20246	0.30000	1.00000



Essa variável também com possível erro de preenchimento, e não iremos considerar para as drivers. Apenas metade da base possui valores entre 0 e 1.

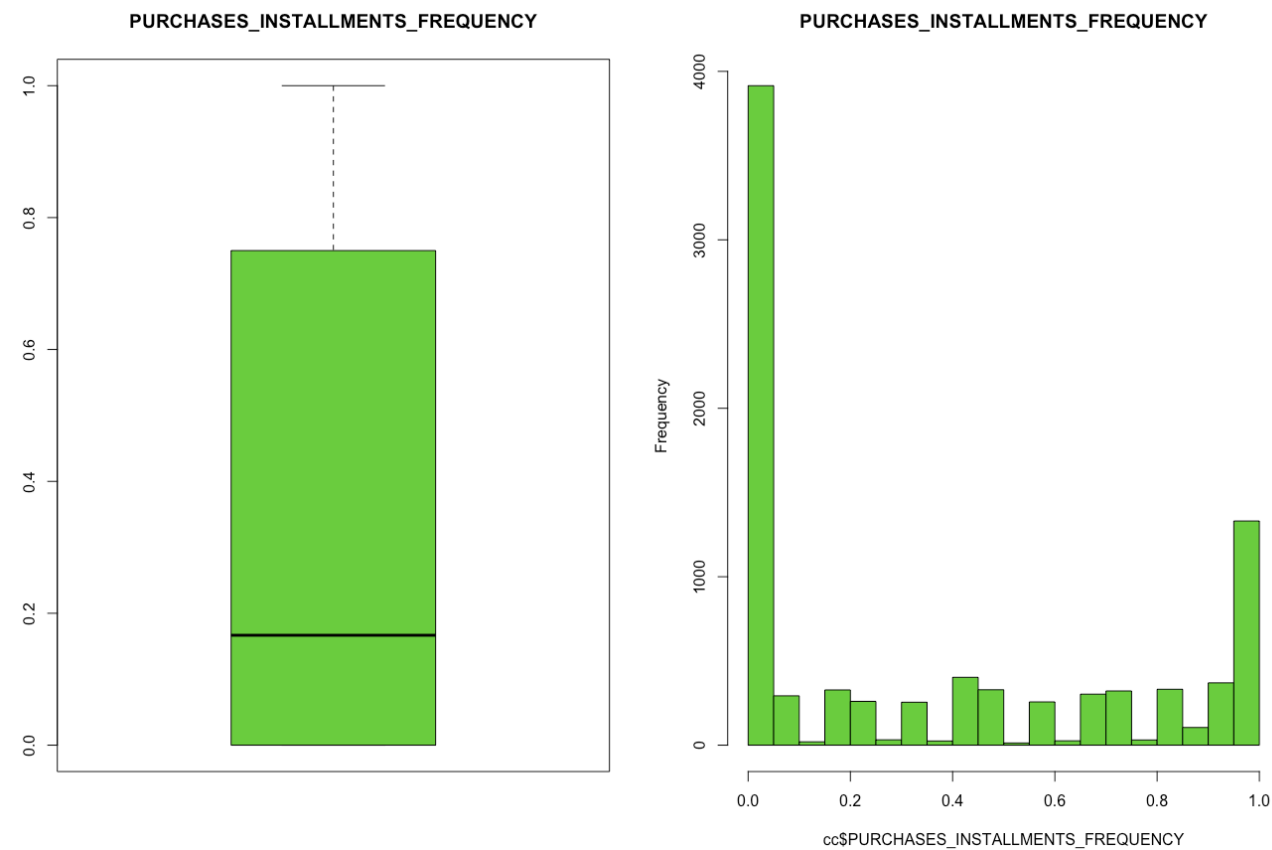
x	freq
0.000000	4302
1.000000	481

PURCHASES_INSTALLMENTS_FREQUENCY

Com que frequência as compras parceladas estão sendo feitas (1 = feito com frequência, 0 = não feito com frequência).

Summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0000	0.0000	0.1667	0.3644	0.7500	1.0000



Mesmo problema de preenchimento das demais, esse tem 60% da base preenchido com os valores corretos e 40% com valores entre 0 e 1. Também não iremos considerar como driver.

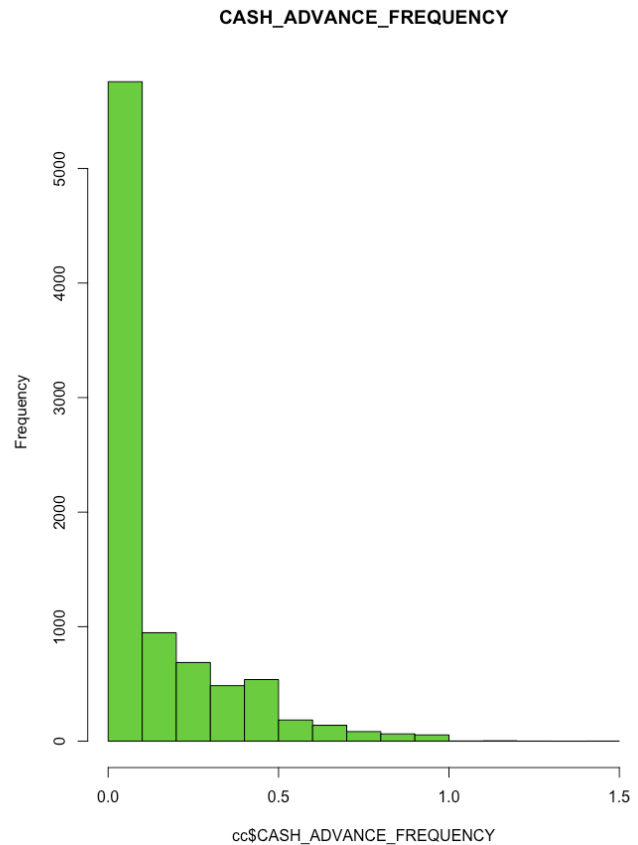
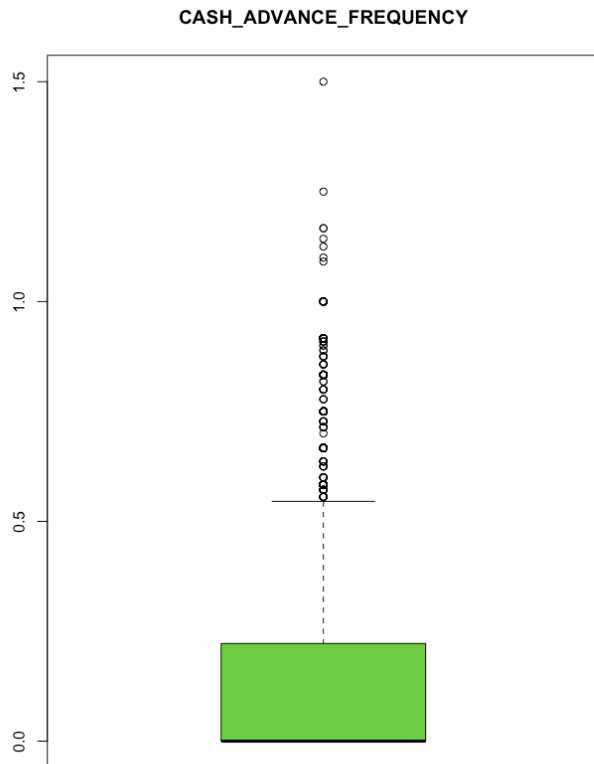
x	freq
0.000000	3915
1.000000	1331

CASH_ADVANCE_FREQUENCY

Com que frequência o dinheiro adiantado está sendo pago.

Summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0000	0.0000	0.0000	0.1351	0.2222	1.5000



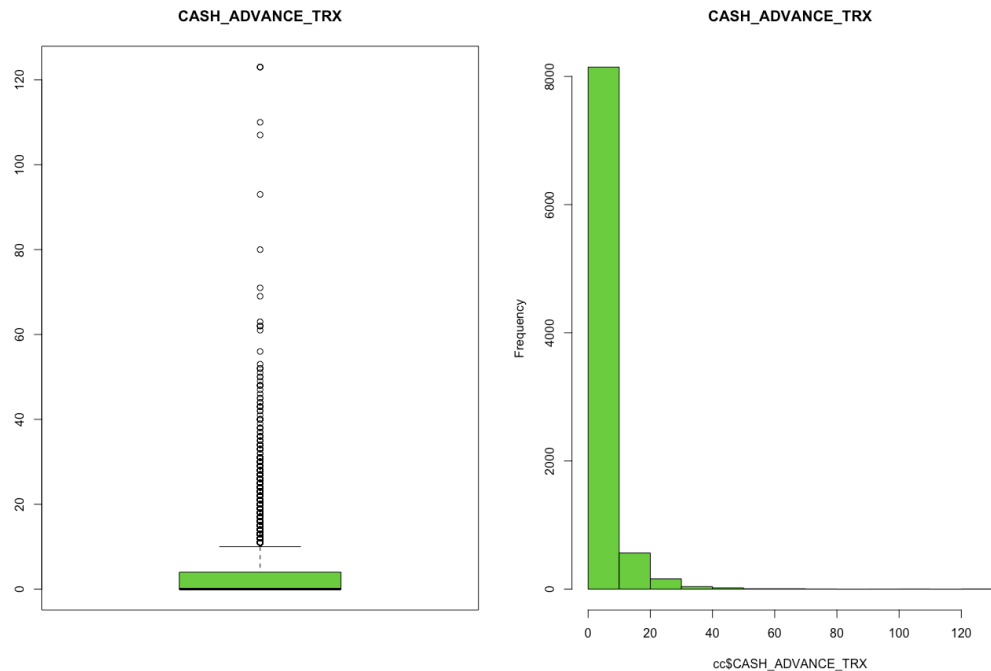
Não temos maiores informações dessa variável para que possamos analisar, pois não é informado se essa frequência em dias, anos, ou se é a quantidade de vezes que o dinheiro está sendo pago. Por isso, não achamos relevante considerar para o cluster.

CASH_ADVANCE_TRX

Número de transações feitas com "Cash in Advance", representa a quantidade de vezes que o cliente fez adiantamento em dinheiro.

Summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000	0.000	0.000	3.249	4.000	123.000

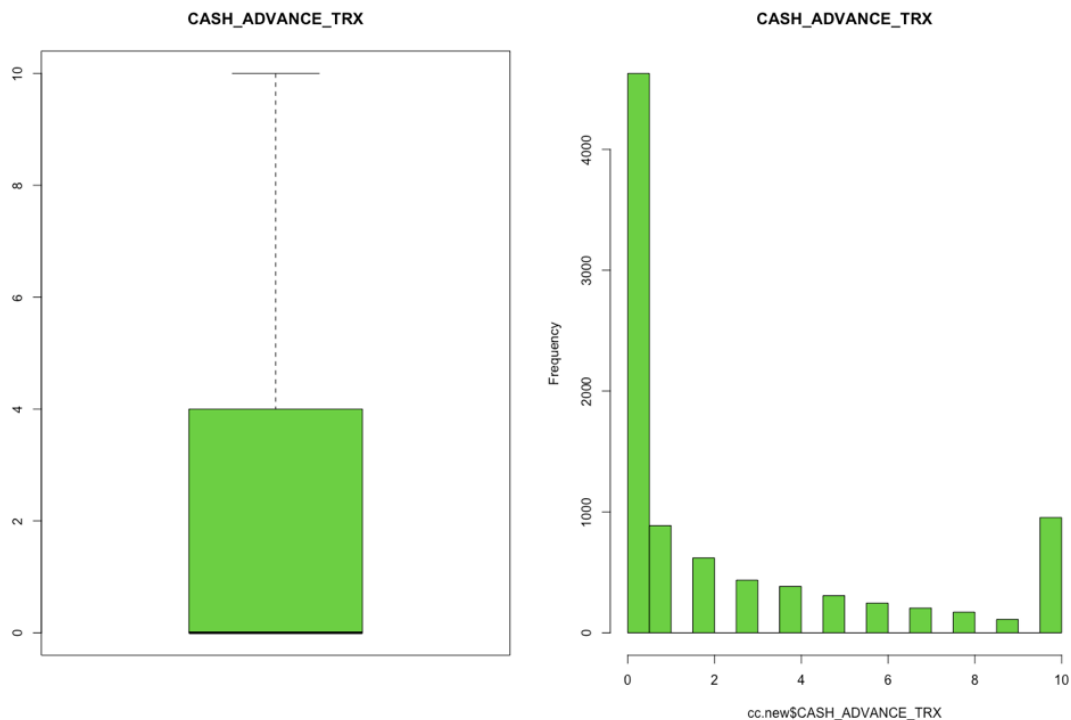


Metade da base não fez nenhuma vez um adiantamento. E a maior parte que fez adiantamento fizeram no máximo em 10 transações.

CASH_ADVANCE_TRX == 0	CASH_ADVANCE_TRX >10
x freq	x freq
1 FALSE 4322 - 2 TRUE 4628	1 FALSE 8146 - 2 TRUE 804

Após o tratamento dos outliers.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000	0.000	0.000	3.249	4.000	123.000

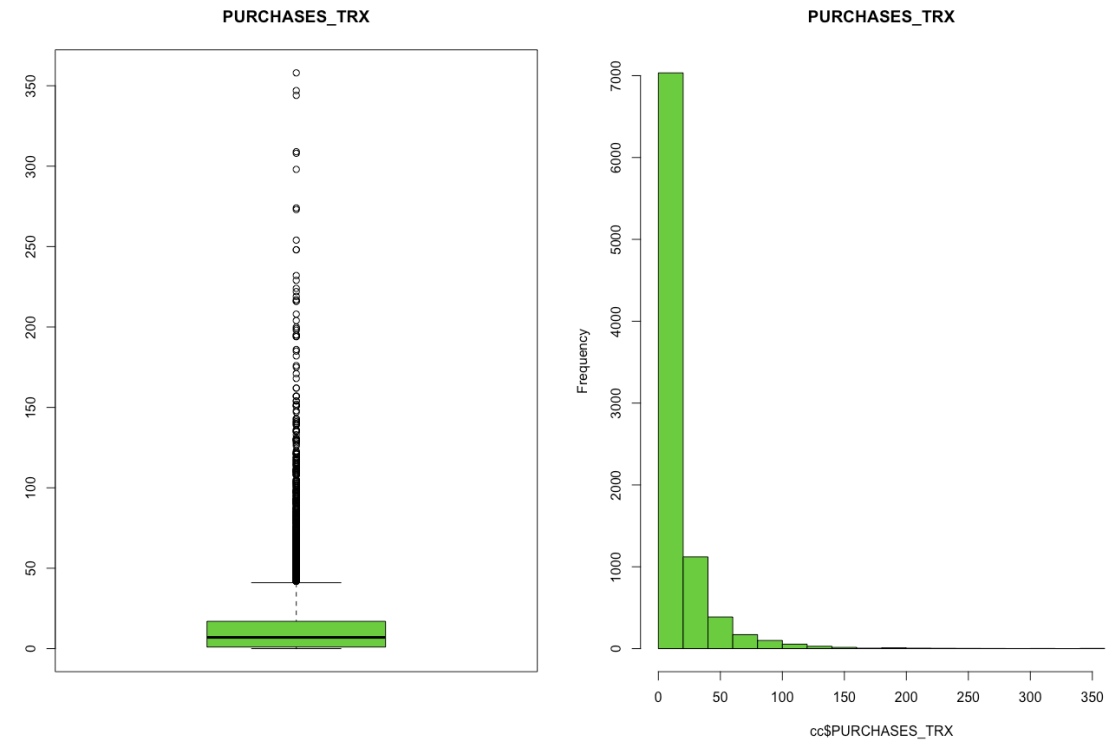


PURCHASES_TRX

Número de Transações de compras feitas.

Summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	1.00	7.00	14.71	17.00	358.00

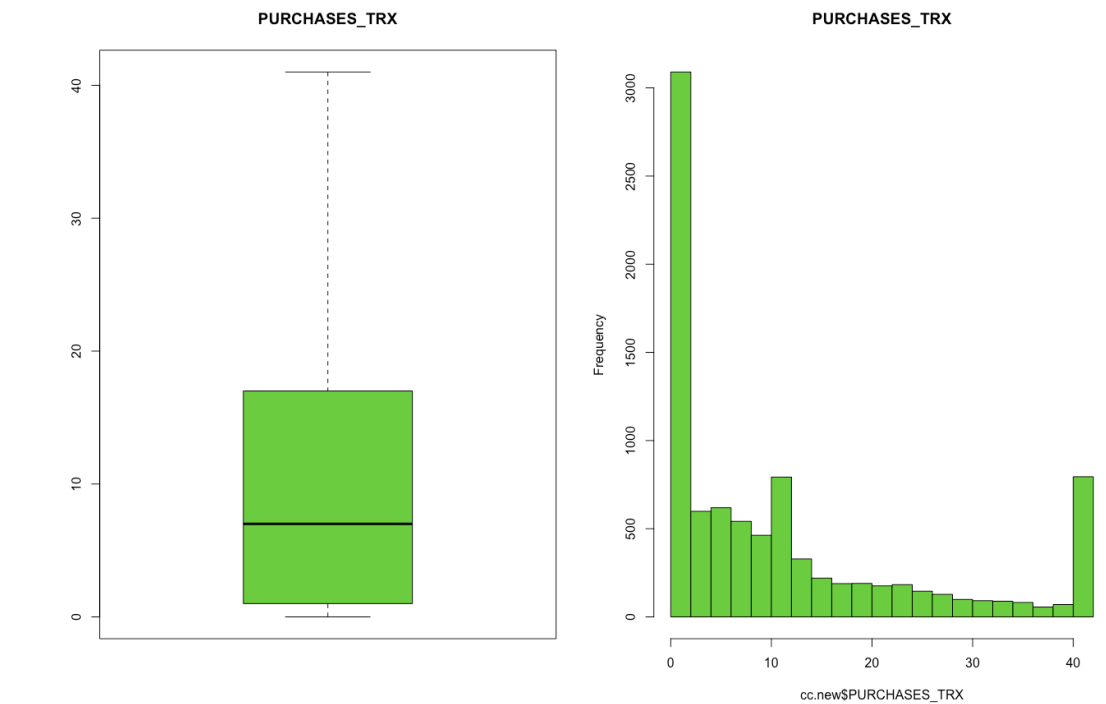


Foram 2.044 clientes não efetuaram transações de compras. E 7.034 efetuaram no máximo 20 transações de compra.

PURCHASES_TRX == 0 x freq 1 FALSE 6906- 2 TRUE 2044	PURCHASES_TRX <= 20 x freq 1 FALSE 1916 - 2 TRUE 7034
---	---

Após o tratamento dos outliers.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	1.00	7.00	14.71	17.00	358.00

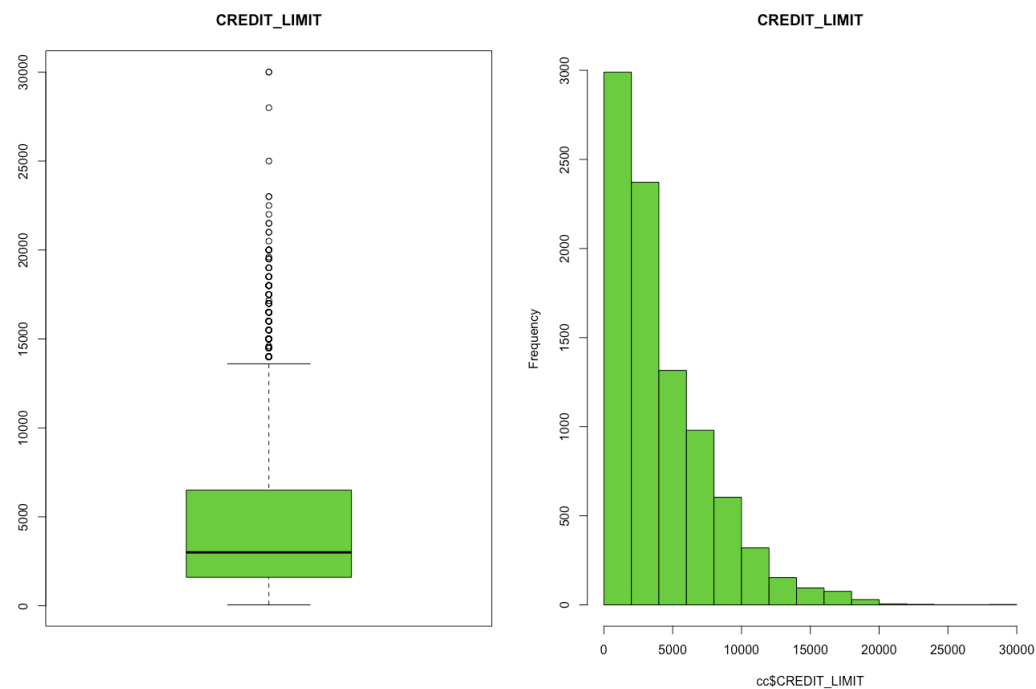


CREDIT_LIMIT

Limite de cartão de crédito para o usuário.

Summary

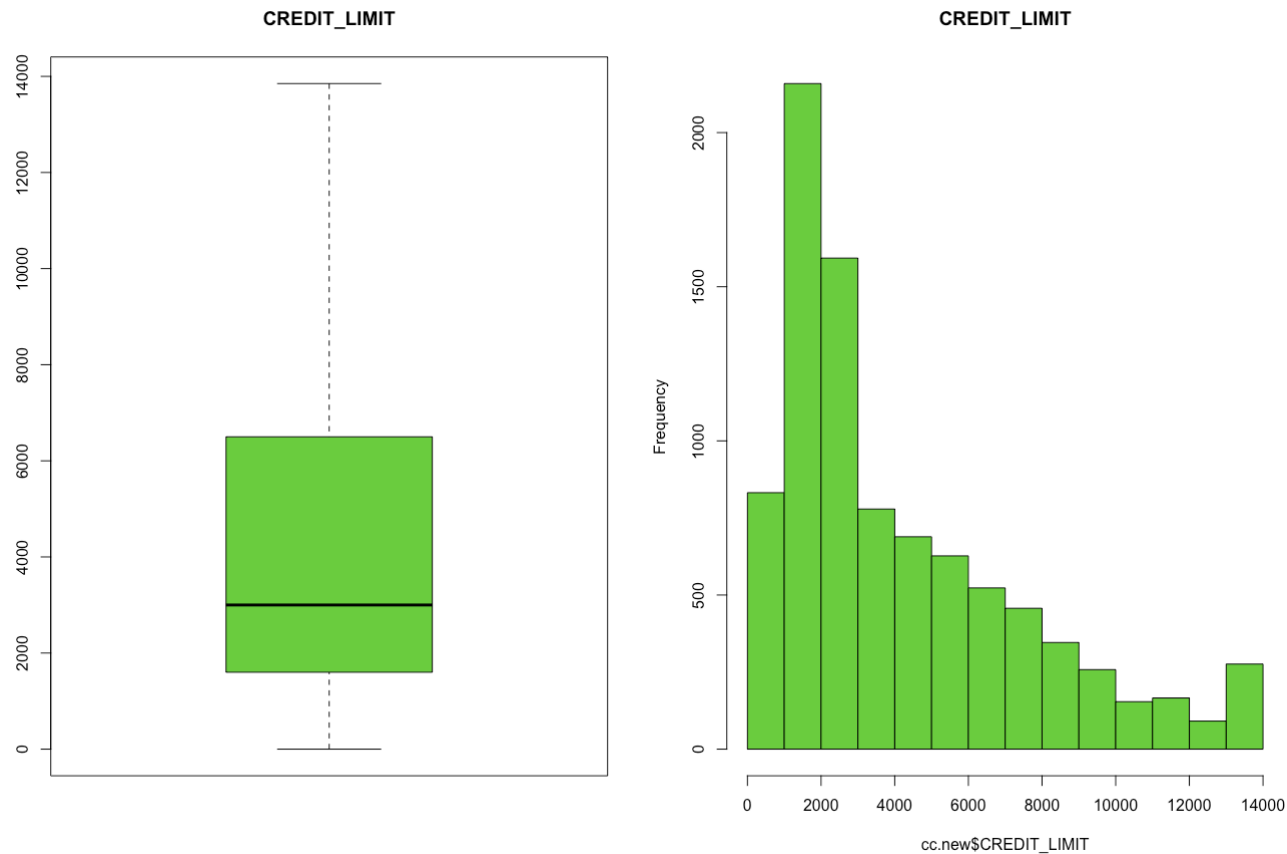
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
50	1600	3000	4494	6500	30000	1



Temos 01 registro com valor NA's, que tratamos preenchendo com o valor zero (0).

Após o tratamento dos outliers.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	1600	3000	4419	6500	13850

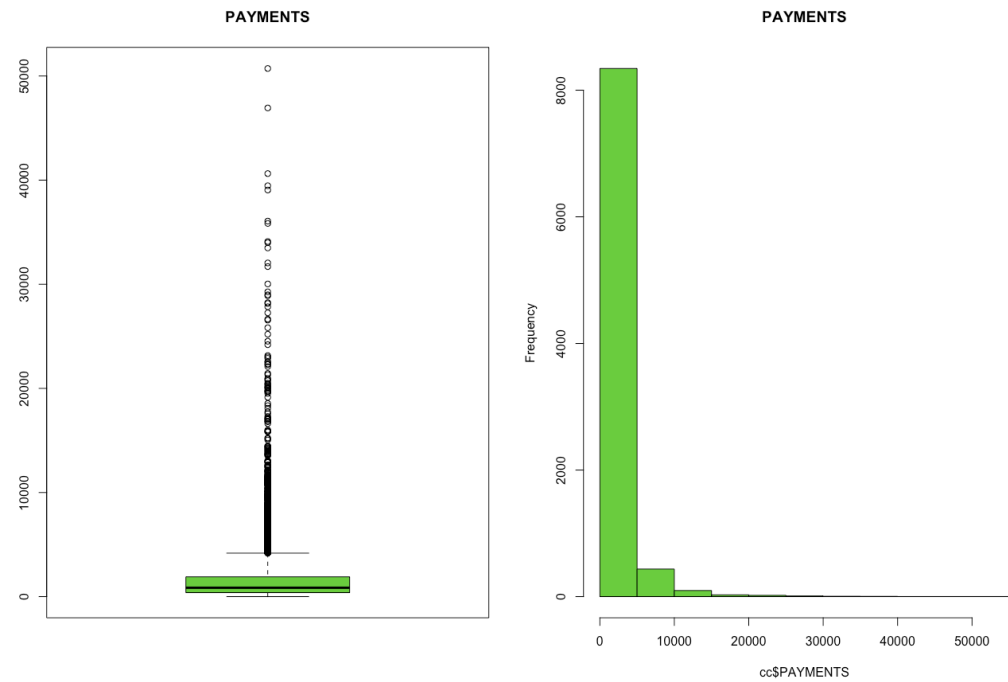


PAYMENTS

valor do pagamento feito pelo usuário.

Summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0	383.3	856.9	1733.1	1901.1	50721.5

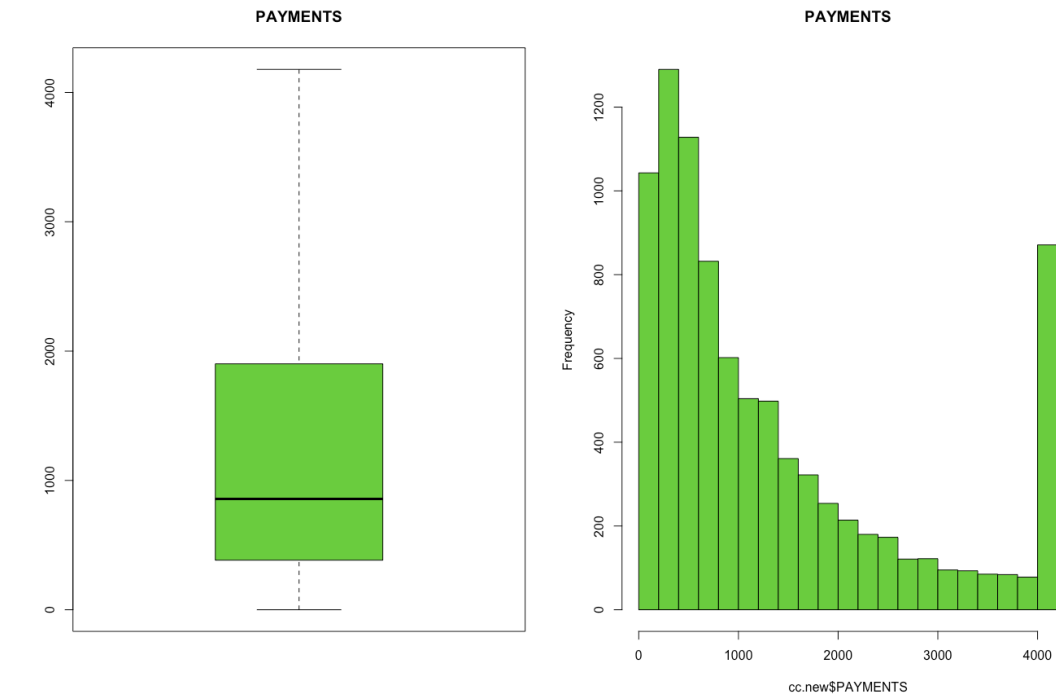


PAYMENTS >= 10000						PAYMENTS <= 5000							
x freq						x freq							
1	FALSE	8779	-	2	TRUE	171	1	FALSE	607	-	2	TRUE	8343

São poucos os clientes que estão dispostos a pagar acima de 10.000 unidades monetárias, a maioria dos clientes efetuam pagamentos até 5.000 unidades monetárias.

Após o tratamento dos outliers.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0	383.3	856.9	1347.1	1901.1	4177.9

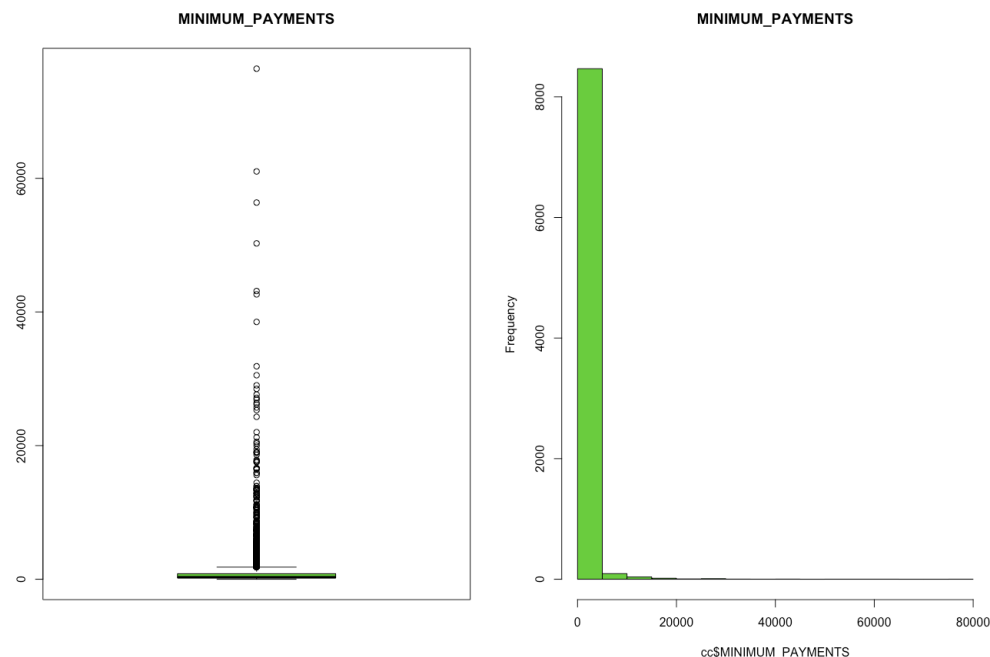


MINIMUM_PAYMENTS

valor mínimo de pagamentos feitos pelo usuário.

Summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
0.02	169.12	312.34	864.21	825.49	76406.21	313

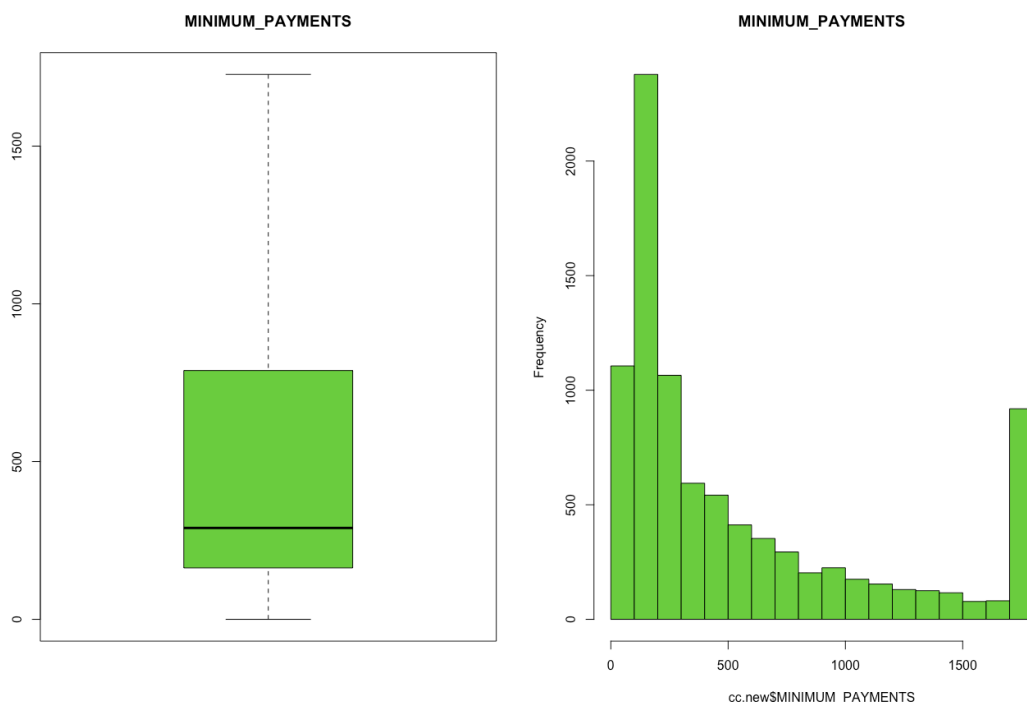


A maioria dos clientes pagaram um valor mínimo até 5.000 unidades monetárias, quando reduzimos para 1.000 unidades monetárias continuamos com a maioria da base.

MINIMUM_PAYMENTS <= 5000			MINIMUM_PAYMENTS <= 1000		
	x	freq		x	freq
1	FALSE	170	1	FALSE	1778
2	TRUE	8467	2	TRUE	6859
3	NA	313	3	NA	313

Após o tratamento dos outliers e campo nulo.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0	163.0	289.6	547.0	788.7	1727.2

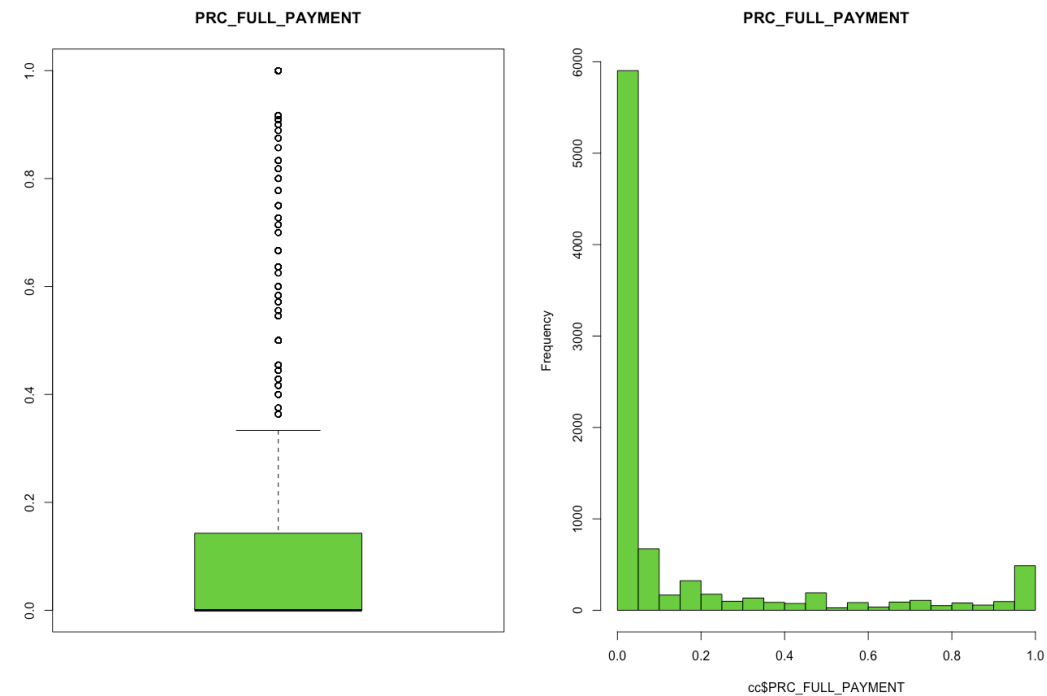


PRC_FULL_PAYMENT

Porcentagem do pagamento total pago pelo usuário.

Summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0000	0.0000	0.0000	0.1537	0.1429	1.0000

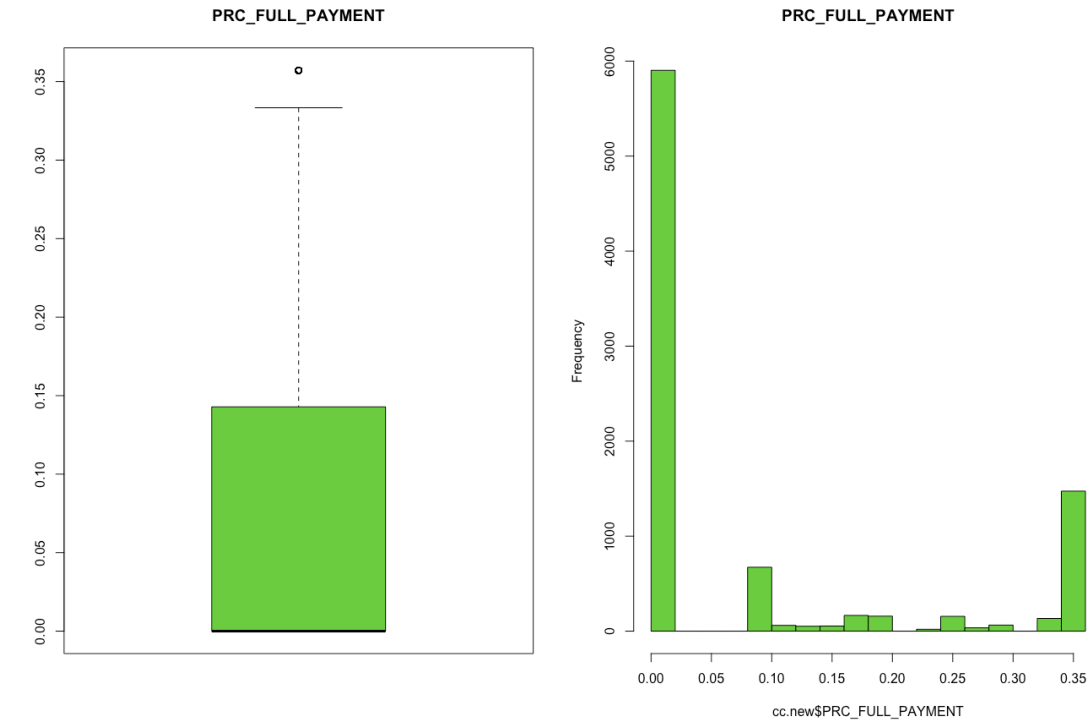


Do total de clientes apenas 488 efetuaram o pagamento total. Mais da metade da base não efetuaram o pagamento total.

PRC_FULL_PAYMENT == 1			PRC_FULL_PAYMENT == 0		
	x	freq		x	freq
1	FALSE	8462	1	FALSE	3047
2	TRUE	488	2	TRUE	5903

Após o tratamento dos outliers

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00000	0.00000	0.00000	0.08723	0.14286	0.35719

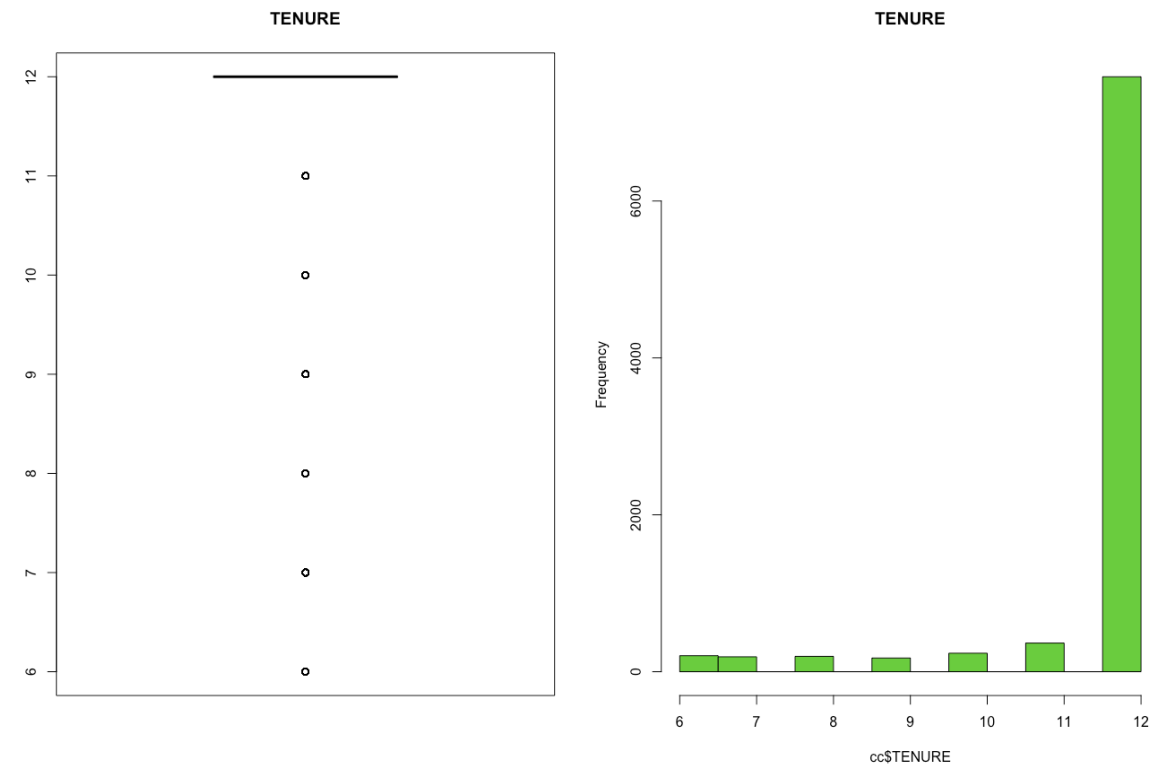


TENURE

Tempo de posse do serviço de cartão de crédito para o usuário.

Summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
6.00	12.00	12.00	11.52	12.00	12.00

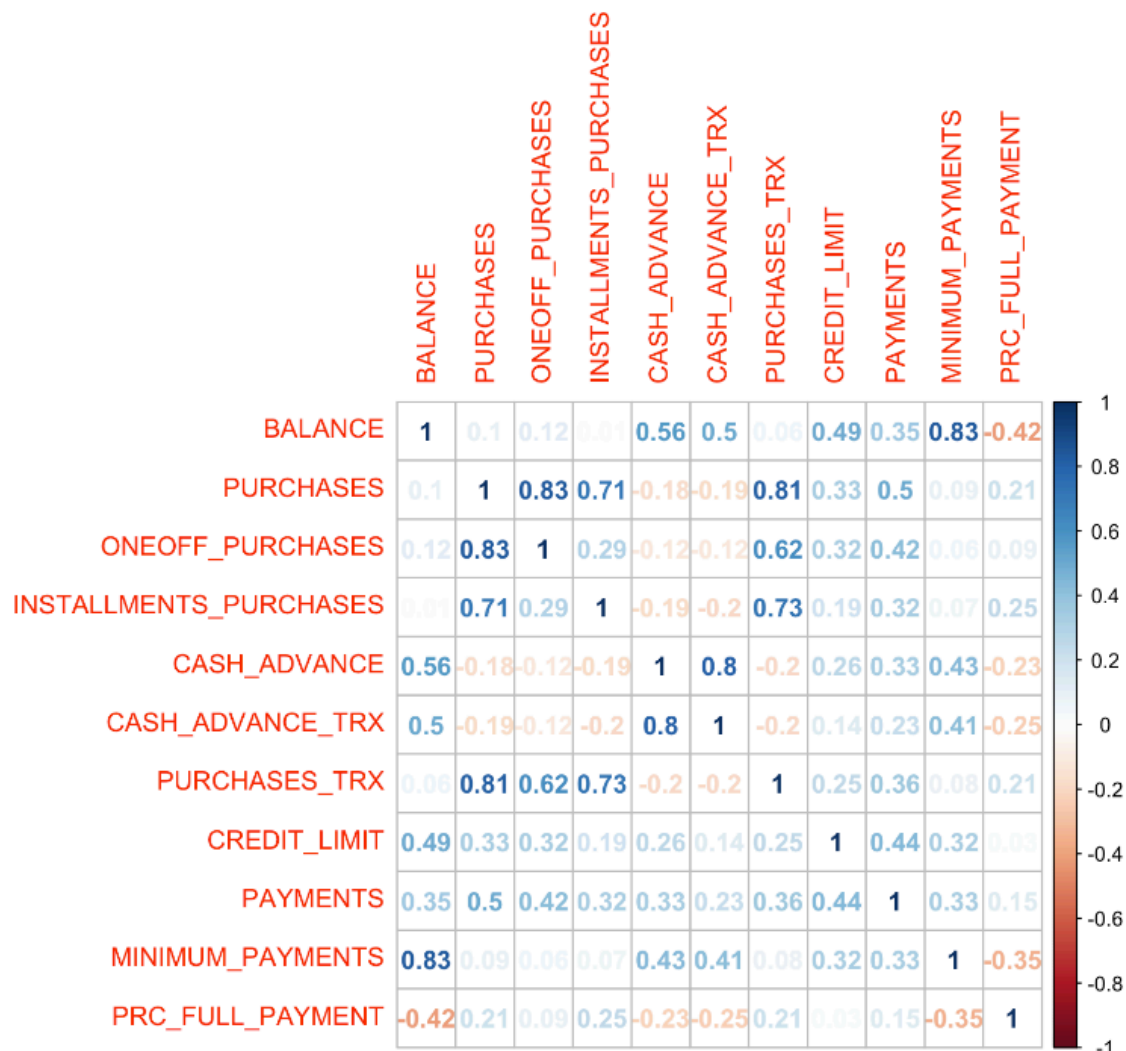


Não encontramos na documentação qual seria a medida de unidade dessa variável, se os valores representam as quantidades de meses, ou se é um número atribuído para indicar o mês. Por isso, resolvemos não utilizar essa variável no modelo.

	x	freq
1	6	204
2	7	190
3	8	196
4	9	175
5	10	236
6	11	365
7	12	7584

Matriz de Correlação

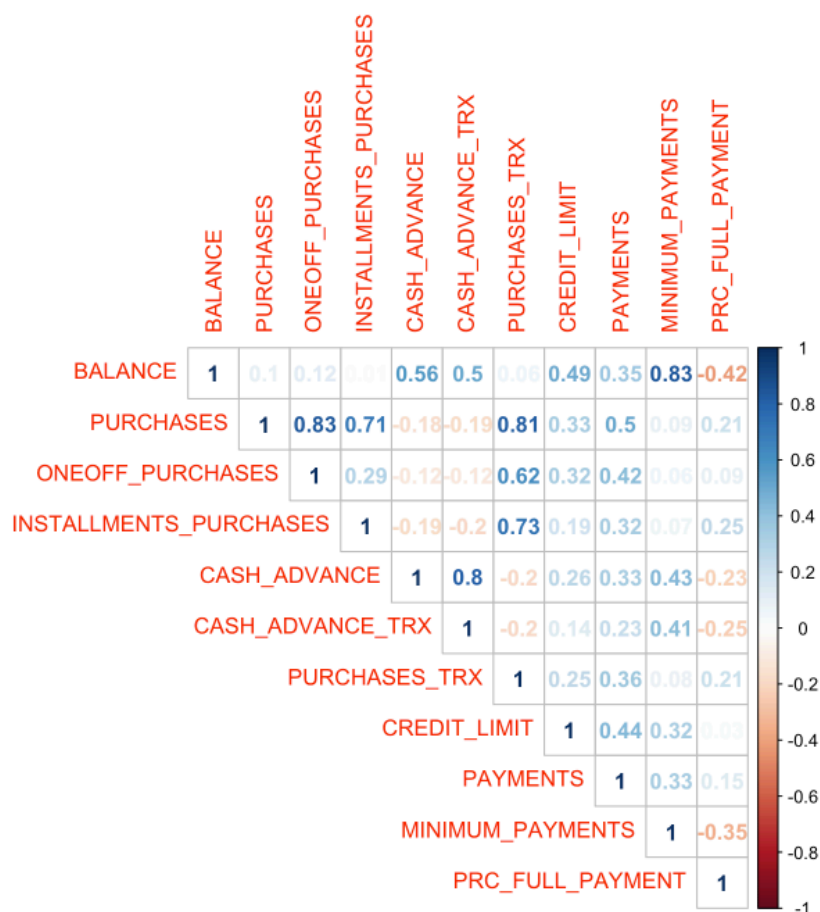
Após a análise das variáveis, resolvemos retirar as variáveis de frequência, o CUST_ID e a TENURE. Ficamos então com 11 variáveis que em seguida fizemos a matriz de correlação e escolhemos as variáveis drivers para o nosso modelo de cluster.



Ordenando os valores para visualizar melhor a correlação. Como observado a variável **PRC_FULL_PAYMENT** que é a porcentagem do pagamento total pago pelo usuário não possui uma correlação positiva significativa com as demais variáveis analisadas, além de não ter correlação nem com a variável de **PAYMENTS** que poderia ser esperado. Logo, iremos desconsiderar para o cluster.

Os valores relacionados a compra por dedução se esperam que estejam relacionados com a compra, e é justamente o que ocorre as variáveis relacionadas a **PURCHASES** possuem uma alta correlação. Iremos considerar todas no modelo, pois queremos identificar se existe uma segmentação de usuários que efetuam mais ou menos compras.

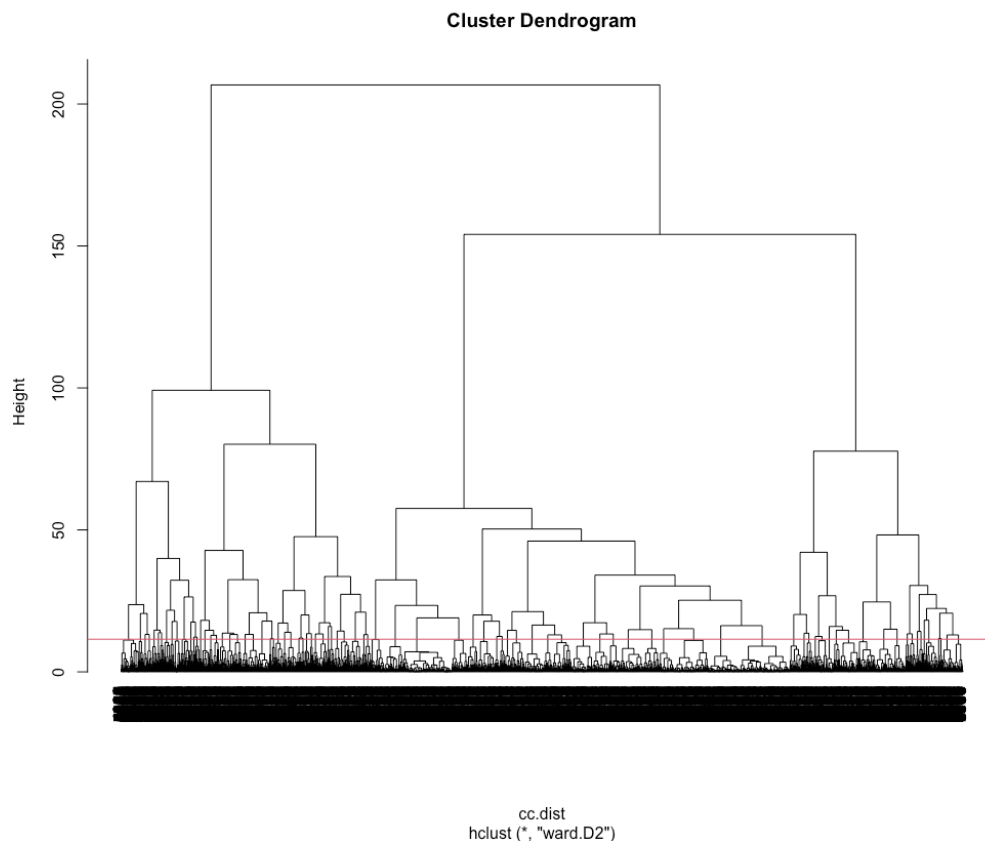
A variável **CASH_ADVANCE_TRX** está altamente relacionada a variável de **CASH_ADVANCE**, o que é esperado já que o número de transações feitas no “Cash in Advance”. Entretanto, ela não tem relação significativa com nenhuma outra variável, e também iremos desconsiderar para o cluster.



Após a análise da correlação escolhemos as 09 seguintes variáveis como drivers: **BALANCE, PURCHASES, ONEOFF_PURCHASES, INSTALMMENTS_PURCHASES, CASH_A DVANCE, PURCHASES_TRX, CREDIT_LIMIT, PAYMENTS e MINIMUM_PAYMENTS.**

Com isso querermos identificar se existe uma segmentação do comportamento do cliente no uso do cartão de crédito, de acordo com suas compras e saldos disponíveis para compra, assim como o limite de crédito disponível.

Cluster Hierárquico Aglomerativo



Pela análise visual do dendrograma escolhemos $k=3$ clusters.

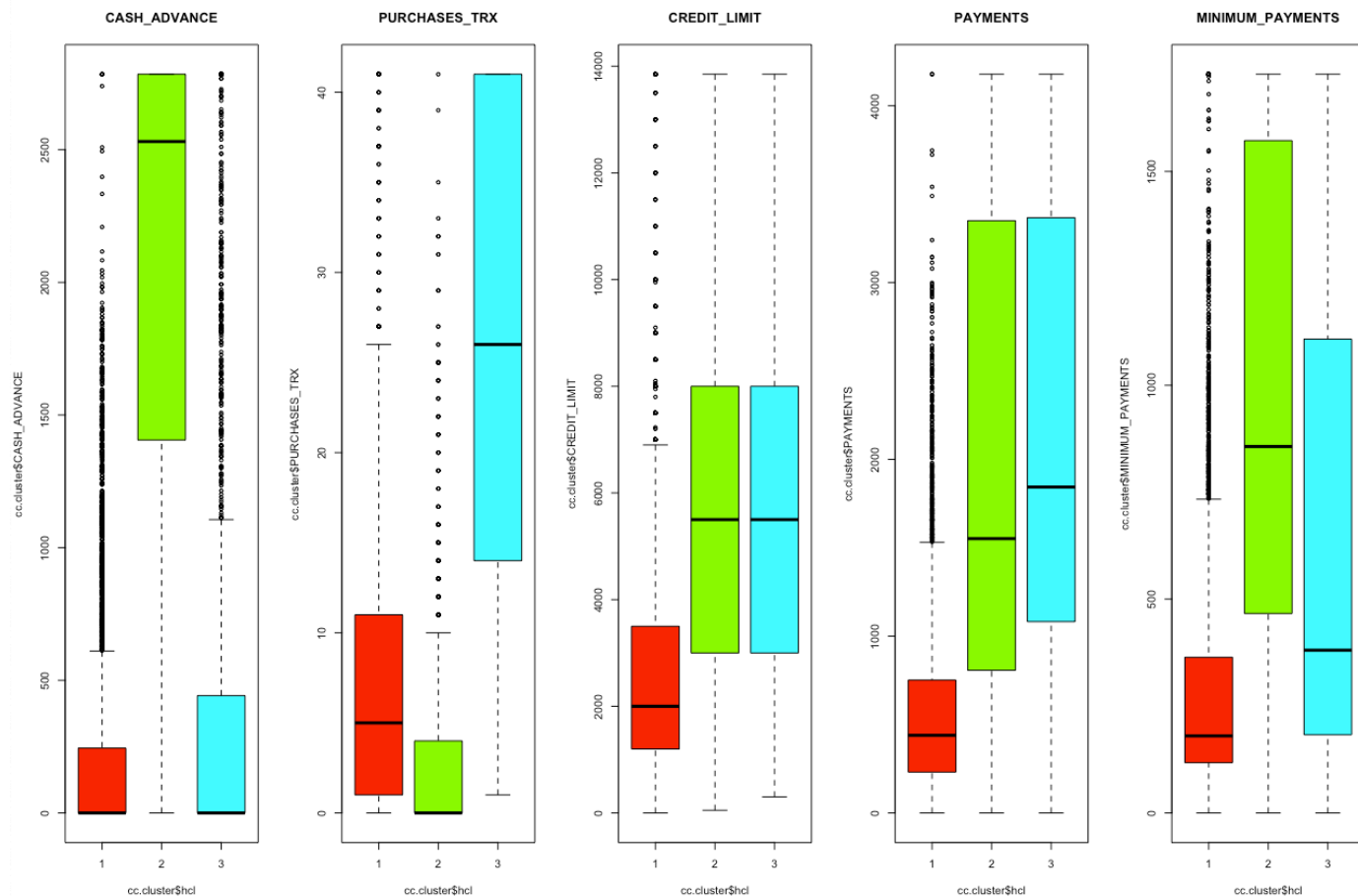
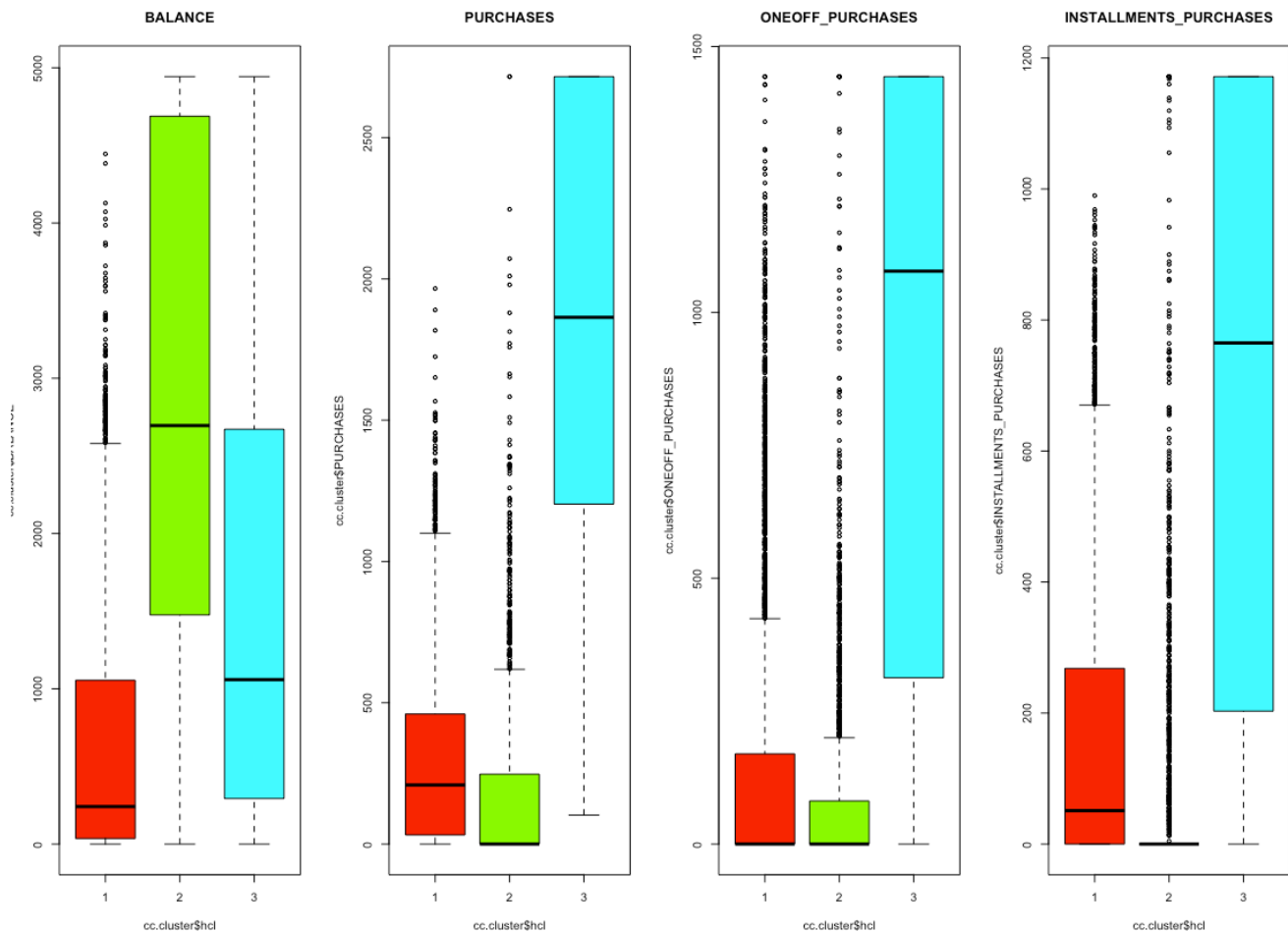
Tabela indicando os elementos dos clusters:

1	2	3
4448	1842	2660

No **cluster 1** temos que os usuários possuem um valor mais baixo de limite de crédito e saldo em conta para fazer compras. Eles parcelam as compras mais vezes, além de efetuarem mais transações de compra com valores pequenos.

No **cluster 2** os usuários possuem o saldo em conta mais alto em comparação aos demais grupos. Praticamente não realizam compras e consequentemente não parcelam suas compras. Usam mais o serviço de adiantamento em dinheiro no cartão de crédito, aproveitando do alto limite de crédito disponível.

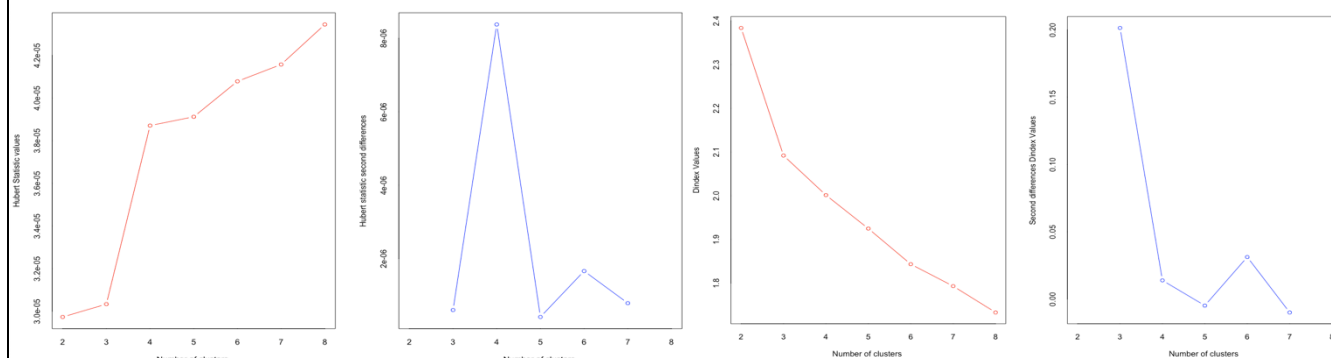
No **cluster 3** os usuários que mais gastam em compras e efetuam compras de uma só vez com valores altos, assim como o valor das suas parcelas também são os mais altos. Diferentemente do cluster 2, eles priorizam as compras ao invés do adiantamento de dinheiro, apesar de terem o mesmo nível de limite de crédito.



Usando o **NbClust** para definir o melhor número de clusters, nos retornou que a maioria dos métodos aplicados indicam que o número ideal de cluster é k=3. O que está de acordo com a nossa escolha feita pela análise do dendograma.

* Among all indices:

- * 4 proposed 2 as the best number of clusters
- * 12 proposed 3 as the best number of clusters
- * 1 proposed 5 as the best number of clusters
- * 1 proposed 6 as the best number of clusters
- * 1 proposed 7 as the best number of clusters
- * 2 proposed 8 as the best number of clusters



Comparação das variáveis descritivas

O cluster 1 e 3 a maioria dos usuários não efetuam transações de adiantamento de dinheiro para o “Cash in Advance”. Ao contrário do cluster 2 que já faz muito uso desse serviço.

	hc1		
CASH_ADVANCE_TRX	1	2	3
0	2767	97	1764
1	532	164	191
2	336	159	125
3	204	153	79
4	158	142	84
5	113	131	64
6	79	122	45
7	51	111	43
8	53	91	27
9	23	67	21
10	132	605	217

Para os três clusters os usuários possuem um tempo de posse do serviço de cartão de crédito de 12 períodos, e acreditamos que esse período é mensal.

	hc1		
TENURE	1	2	3
6	135	48	21
7	125	39	26
8	112	51	33
9	89	53	33
10	122	65	49
11	197	111	57
12	3668	1475	2441

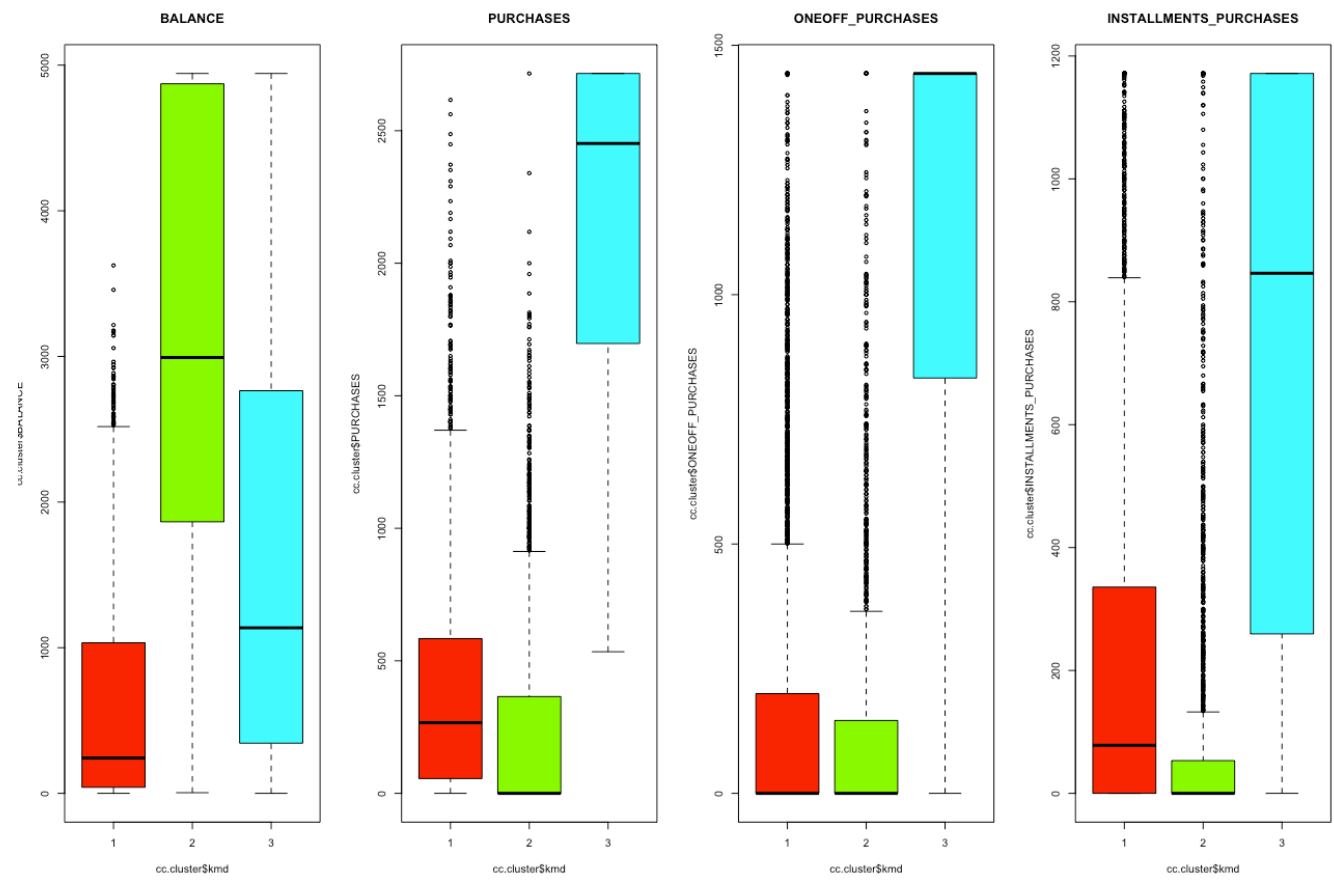
Cluster K-MEDOID

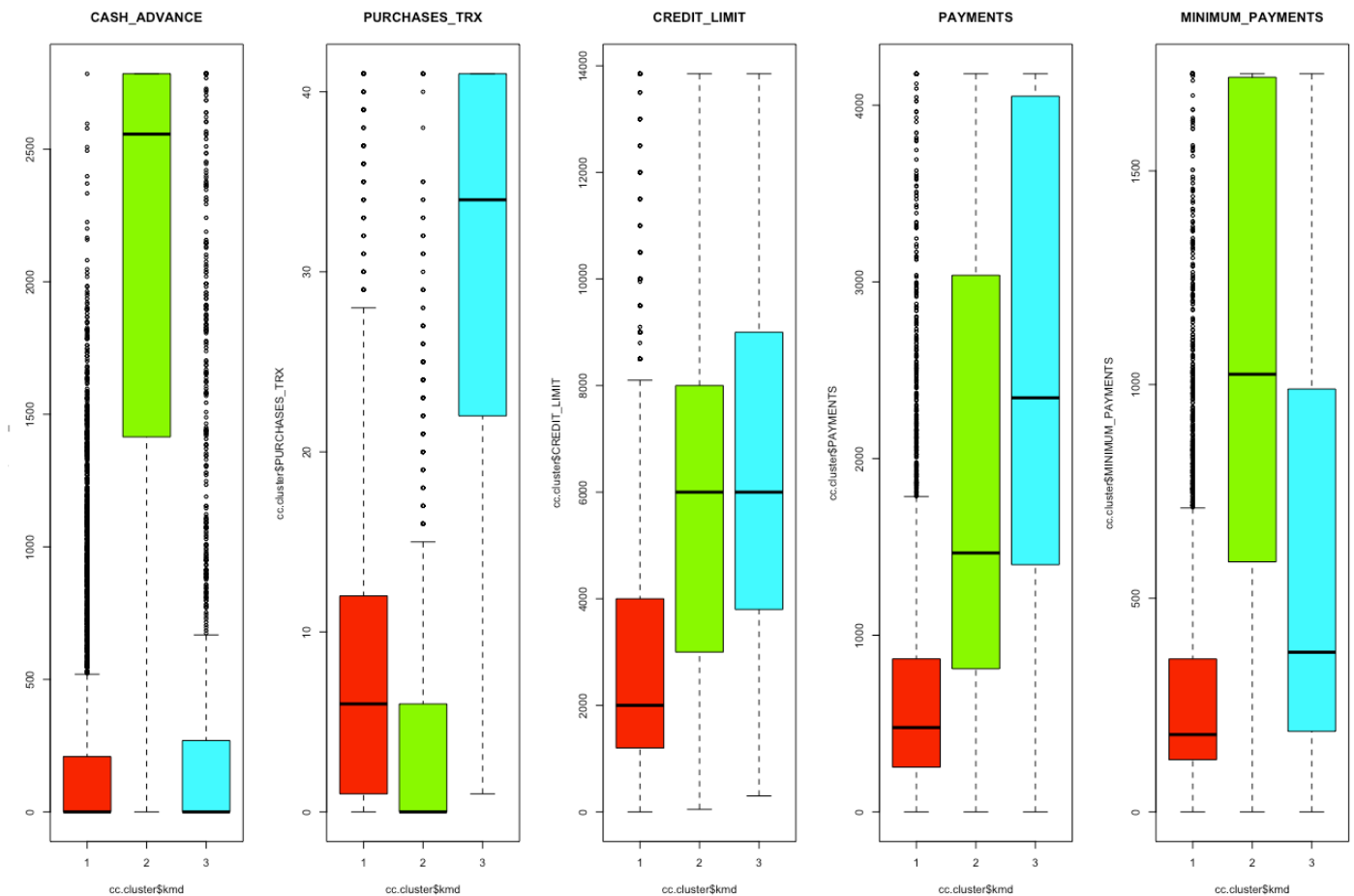
A função pamk selecionou 03 clusters e não difere do número que escolhemos inicialmente e está de acordo com o a seleção do NbCluster.

2	clusters	3585.249
3	clusters	3841.444
4	clusters	3042.592
5	clusters	2693.135

Tabela indicando os elementos dos clusters:

1	2	3
5027	2027	1896





No **cluster 1** os usuários desse grupo possuem o menor valor de saldo em conta em comparação aos demais grupos, assim como efetuam compras com valores inferiores dentro do seu limite de crédito. Realizam compras de uma só vez com valores baixo, pois preferem parcelar as compras. Efetua muitas transações de compras, porém sem comprometer o limite do cartão.

No **cluster 2** os usuários desse grupo possuem alto valores de saldo em conta para fazer compras, porém compram pouco mesmo com um limite de crédito mais alto. Ele utiliza basicamente o cartão de crédito para o serviço de adiantamento de dinheiro sendo seu fluxo de pagamento alto.

No **cluster 3** os usuários desse grupo realizam compras com valores expressivos e priorizam o parcelamento. Entretanto, cumprem com o compromisso dos pagamentos da fatura do cartão. Possuem um limite de crédito diferenciado dos demais grupos, e quase não utilizam o serviço de adiantamento de dinheiro.

Comparação das variáveis descritivas:

O comportamento das variáveis nos dois clusters foram similares, como podemos ver abaixo.

O cluster 1 e 3 a maioria dos usuários não efetuam transações de adiantamento de dinheiro para o “Cash in Advance”. Ao contrário do cluster 2 que já faz muito uso desse serviço.

CASH_ADVANCE_TRX	kmd		
	1	2	3
0	3233	87	1308
1	580	156	151
2	347	181	92
3	219	160	57
4	185	153	46
5	126	149	33
6	79	137	30
7	52	134	19
8	56	96	19
9	23	76	12
10	127	698	129

Para os três clusters os usuários possuem um tempo de posse do serviço de cartão de crédito de 12 períodos, e acreditamos que esse período é mensal.

TENURE	kmd		
	1	2	3
6	147	46	11
7	137	44	9
8	130	51	15
9	110	51	14
10	149	62	25
11	206	120	39
12	4148	1653	1783

Cruzar os resultados dos clusters:

Analisando os dois modelos conseguimos discriminar os eventos, porém houve uma diferença de agrupamento dos elementos. O cluster K-Medoid classificou melhor o cluster 3 capturando suas características de forma homogênea. Devido a isso, ocorreu uma redistribuição equilibrada dos elementos para os demais clusters, o que o diferenciou do cluster hierárquico. Como podemos notar abaixo, os clusters 1 e 2 do K-Medoid possuem um valor maior de elementos que o cluster hierárquico.

cc.cluster\$hc1	cc.cluster\$kmd			Row Total
	1	2	3	
1	4230 0.951	180 0.040	38 0.009	4448 0.497
2	189 0.103	1624 0.882	29 0.016	1842 0.206
3	608 0.229	223 0.084	1829 0.688	2660 0.297
Column Total	5027	2027	1896	8950

Conclusão

Com o conjunto de dados analisado conseguimos obter a segmentação de clientes com base em seu comportamento de uso de cartão de crédito. Foram identificados 03 grupos distintos para definir as estratégias de marketing.

O grupo 1: são usuários que gastam mais e tendem a parcelar mais as compras, apesar do baixo limite de crédito e são bons pagadores. A estratégia para esse grupo poderia ser a revisão dos limites de crédito dando maior poder de compra aos usuários.

O grupo 2: são usuários que utilizam o cartão de crédito para o serviço de adiantamento em dinheiro e com pouca realização de compras. Utilizam o cartão de crédito como mecanismo de aquisição de crédito a curto prazo. A estratégia para esse grupo seria possibilitar maior quantidade de saques de menor valor, possibilitando o usuário sacar mais vezes.

O grupo 3: são usuários com maior poder de compra, pois possuem um maior limite de crédito e suas compras possuem valores expressivos e muito parcelado. O que nos indica que possui um alto poder aquisitivo, são bons compradores e pagadores. A estratégia para esse grupo poderia ser a revisão dos limites de crédito, assim como oferecer um programa de fidelização através de cashback das compras realizadas.