

**FUNDAÇÃO GETULIO VARGAS - FGV**

**TRABALHO – Análise de Agrupamentos**

Ana Paula Pudo  
Fábio Monteiro  
Lucas Sena Alves  
Marcos Soares

**MBA Business Analytics e Big Data | Métodos Matriciais e**

**Análise de Clusters**

Prof. Abraham Laredo Sicsú

Brasília  
2021

## Sumário

1. Introdução.....	4
2. Análises das Variáveis .....	4
2.1 Identificação das variáveis drivers e descritivas.....	4
2.1.1 Variáveis Drivers.....	4
2.1.2 Variáveis Descritivas.....	4
2.2 Tratamento da base de dados.....	4
2.3 Análise por variáveis drivers .....	5
2.3.1 DS_SITUACAO_CANDIDATURA.....	5
2.3.2 TP_AGREMIACAO.....	6
2.3.3 ST_REELEICAO.....	7
2.3.4 ST_DECLARAR_BENS.....	8
2.3.5 SG_UF.....	9
2.3.6 DS_CARGO / POR SG_UF.....	9
2.3.7 SG_UF_NASCIMENTO - POR SG_UF.....	10
2.3.8 DS_GRAU_INSTRUCAO.....	12
2.3.9 DS_COR_RACA .....	13
3. Análise por variáveis descritivas.....	14
3.1 DS_ESTADO_CIVIL .....	14
3.4.2 DS_GENERO.....	15
4. Análise de Clusters obtidos pelo método hierárquico aglomerativo.....	16
4.1 Caracterização dos grupos.....	16
4.2 Comparação dos clusters hierárquicos gerados com os drivers.....	17
4.2.1 DS_SITUACAO_CANDIDATURA.....	17
4.2.2 TP_AGREMIACAO.....	17
4.2.3 ST_REELEICAO.....	18
4.2.4 ST_DECLARAR_BENS.....	18
4.2.5 SG_UF.....	18
4.2.6 DS_CARGO.....	19
4.2.7 SG_UF_NASCIMENTO - POR SG_UF.....	19
4.2.8 DS_GRAU_INSTRUCAO.....	20
4.2.9 DS_COR_RACA .....	21
4.3.1 DS_ESTADO_CIVIL.....	21
4.3.2 DS_GENERO.....	22
5. Análise de Clusters obtidos pelo método K-medoid.....	23

<i>5.1 Análise por variáveis drivers</i> .....	23
<i>5.1.1 DS_SITUACAO_CANDIDATURA</i> .....	23
<i>5.1.2 TP_AGREMIACAO</i> .....	23
<i>5.1.3 ST_REELEICAO</i> .....	23
<i>5.1.4 ST_DECLARAR_BENS</i> .....	24
<i>5.1.5 SG_UF</i> .....	24
<i>5.1.6 DS_CARGO</i> .....	24
<i>5.1.7 SG_UF_NASCIMENTO</i> .....	25
<i>5.1.8 DS_GRAU_INSTRUCAO</i> .....	26
<i>5.1.9 DS_COR_RACA</i> .....	26
<i>5.2 Variáveis Descritivas</i> .....	27
<i>5.2.1 DS_ESTADO_CIVIL</i> .....	27
<i>5.2.2 DS_GENERO</i> .....	27
<b>7. conclusões</b> .....	28

## 1. Introdução

O processo para a escolha de representantes federais para cargos nos poderes executivo e legislativo ocorre a cada quatro anos no Brasil. Nos pleitos, candidatos devidamente registrados em partidos políticos disputam as cadeiras de deputados estaduais e federais, senadores, governadores e de Presidente da República. Todas as informações cadastrais desses candidatos são reunidas, processadas e disponibilizadas pelo Tribunal Superior Eleitoral (TSE).

Neste estudo, foram analisados candidatos que disputaram cargos de deputado federal e deputado estadual nos estados da Bahia e de São Paulo.

## 2. Análises das Variáveis

### 2.1 Identificação das variáveis drivers e descritivas

#### *2.1.1 Variáveis Drivers*

- DS\_SITUACAO\_CANDIDATURA
- TP\_AGREMIACAO
- ST\_REELEICAO
- ST\_DECLARAR\_BENS
- SG\_UF
- DS\_CARGO
- SG\_UF\_NASCIMENTO
- DS\_GRAU\_INSTRUCAO
- DS\_COR\_RACA

#### *2.1.2 Variáveis Descritivas*

- DS\_ESTADO\_CIVIL
- DS\_GENERO

### *2.2 Tratamento da base de dados*

Após um tratamento inicial da base de dados original, o data frame ficou com 16 variáveis e 29.145 observações. A partir desta base, optou-se por uma redução do escopo mantendo o número de variáveis, ficando com apenas 4.993 observações, seguindo os seguintes passos:

- a) Restrição para a ABRANGÊNCIA Estadual;
- b) Filtragem para apenas duas Unidades Federativas: *Bahia e São Paulo (tamanho de colégios eleitorais e regiões distintas)*;
- c) Cargos Deputado Estadual e Deputado Federal;
- d) Eliminação variáveis:
  - 3 variáveis com missing values em todos os registros;
  - ABRANGÊNCIA;
  - IDADE\_DATA\_POSSE;

## 2.3 Análise por variáveis drivers

### 2.3.1 DS\_SITUACAO\_CANDIDATURA

A situação da candidatura é definida pela justiça eleitoral, cabe a ela informar se a candidatura está apta ou inapta. 89% dos candidatos mostraram-se aptos para o cargo. O percentual de 11% de inaptos se equivalem nos dois Estados escolhidos nesta base de dados - Bahia e São Paulo.

Tabela 1 - Situação Candidatura

	APTO	INAPTO
	4466	527
	0.894	0.106

Tabela 2 - Situação Candidatura X UF

cand\$DS_SITUACAO_CANDIDATURA	cand\$SG_UF		Row Total
	BA	SP	
APTO	1027	3439	4466
	0.008	0.002	
	0.230	0.770	0.894
	0.897	0.894	
	0.206	0.689	
INAPTO	118	409	527
	0.067	0.020	
	0.224	0.776	0.106
	0.103	0.106	
	0.024	0.082	
Column Total	1145	3848	4993
	0.229	0.771	

### 2.3.2 TP\_AGREMIACAO

A variável agremiação aponta de que forma será a candidatura dos candidatos. A variável apresenta apenas dois níveis de categorização: Coligação e Partido Isolado. Apesar dos totais estarem equilibrados, nota-se que, no Estado da Bahia, apenas 5% estão agremiados em partido isolado. O Estado de São Paulo possui 53% desta camada.

Tabela 3 - Tipo de Agremiação

	COLIGAÇÃO	PARTIDO ISOLADO
	2088	2905
	0.418	0.582

Tabela 4 - Tipo de Agremiação x UF

cand\$TP_AGREMIACAO	cand\$SG_UF		Row Total
	BA	SP	
COLIGAÇÃO	908	1180	2088
	384.680	114.464	
	0.435	0.565	0.418
	0.793	0.307	
	0.182	0.236	
PARTIDO ISOLADO	237	2668	2905
	276.493	82.272	
	0.082	0.918	0.582
	0.207	0.693	
	0.047	0.534	
Column Total	1145	3848	4993
	0.229	0.771	

### 2.3.3 ST\_REELEICAO

No cruzamento das variáveis Reelevação e Cargo 96% dos Deputados Estadual e Federal não estão concorrendo à reeleição. Na comparação entre cargos conforme a tabela 6, 54% são Deputados Estaduais e 42% Deputados Federais

Tabela 5 - Situação de Reelevação

	N	S
	4773	220
	0.956	0.044

Tabela 6 - Situação de Reelevação X Cargo

cand\$ST_REELEICAO	DEPUTADO ESTADUAL	DEPUTADO FEDERAL	Row Total
N	2686	2087	4773
	0.000	0.001	
	0.563	0.437	0.956
	0.956	0.956	
	0.538	0.418	
S	125	95	220
	0.011	0.014	
	0.568	0.432	0.044
	0.044	0.044	
	0.025	0.019	
Column Total	2811	2182	4993
	0.563	0.437	

### 2.3.4 ST\_DECLARAR\_BENS

Esta variável indica se o candidato tem ou não bens a declarar. Em SP e BA, 61% dos candidatos possuem bens para declarar. Optou-se por verificar se o cruzamento da situação da candidatura e a situação da declaração de bens poderia trazer alguns insights, mas este cruzamento não apresentou nenhuma informação relevante.

Tabela 7 - Situação da Declaração de bens

	N	S
	1942	3051
	0.389	0.611

Tabela 8 - Situação da Declaração de bens

cand\$ST_DECLARAR_BENS	cand\$DS_SITUACAO_CANDIDATURA		Row Total
	APTO	INAPTO	
N	1629	313	1942
	6.718	56.932	
	0.839	0.161	0.389
	0.365	0.594	
	0.326	0.063	
S	2837	214	3051
	4.276	36.238	
	0.930	0.070	0.611
	0.635	0.406	
	0.568	0.043	
Column Total	4466	527	4993
	0.894	0.106	



### 2.3.5 SG\_UF

SG\_ UF significa a Unidade da Federação em que ocorreu a eleição. Para este trabalho selecionamos 2 UFs de amostra para avaliar o perfil dos candidatos a deputado estadual e federal eleitos. A UF de SP elegeu 43% dos Deputados Estaduais inscritos, enquanto a UF da Bahia 13%. Na esfera federal, cargo de deputado federal, a UF de SP elegeu 34 % e a UF da Bahia 10%.

Tabela 9 - UF da candidatura

	BA	SP
	1145	3848
	0.229	0.771

### 2.3.6 DS\_CARGO / POR SG\_UF

Tabela 10 - Cargos

DEPUTADO ESTADUAL	DEPUTADO FEDERAL
2811	2182
0.563	0.437

Tabela 11 - Cargos X UF

	cand\$SG_UF		
cand\$DS_CARGO	BA	SP	Row Total
DEPUTADO ESTADUAL	642	2169	2811
	0.011	0.003	
	0.228	0.772	0.563
	0.561	0.564	
	0.129	0.434	
DEPUTADO FEDERAL	503	1679	2182
	0.014	0.004	
	0.231	0.769	0.437
	0.439	0.436	
	0.101	0.336	
Column Total	1145	3848	4993
	0.229	0.771	

### 2.3.7 SG\_UF\_NASCIMENTO - POR SG\_UF

Na proposta de conhecer melhor o perfil dos candidatos optamos a fazer um cruzamento das variáveis UF de Nascimento com a UF onde ocorreu a eleição, que no caso desta amostra somente consideramos as UFs de SP e BA. A tabela 12 mostra as informações detalhadas de cada UF. Devido ao histórico de algumas décadas SP, foi o estado que mais recebeu migrantes no Brasil, o que justifica os 77% de candidatos vindos de outras UFs.

Tabela 12 -UF Nascimento X UF Eleitos(as)

cand\$SG_UF_NASCIMENTO	cand\$SG_UF			PI	2	33	35
	BA	SP	Row Total				
AC	0	1	1		4.525	1.346	
	0.229	0.068			0.057	0.943	0.007
	0.000	1.000	0.000		0.002	0.009	
	0.000	0.000			0.000	0.007	
	0.000	0.000					
AL	4	32	36	PR	1	122	123
	2.194	0.653			26.242	7.808	
	0.111	0.889	0.007		0.008	0.992	0.025
	0.003	0.008			0.001	0.032	
	0.001	0.006			0.000	0.024	
AM	3	4	7	RJ	9	56	65
	1.212	0.361			2.340	0.696	
	0.429	0.571	0.001		0.138	0.862	0.013
	0.003	0.001			0.008	0.015	
	0.001	0.001			0.002	0.011	
AP	0	1	1	RN	3	22	25
	0.229	0.068			1.303	0.388	
	0.000	1.000	0.000		0.120	0.880	0.005
	0.000	0.000			0.003	0.006	
	0.000	0.000			0.001	0.004	
BA	1036	176	1212	RO	0	4	4
	2067.588	615.226			0.917	0.273	
	0.855	0.145	0.243		0.000	1.000	0.001
	0.905	0.046			0.000	0.001	
	0.207	0.035			0.000	0.001	
				RR	0	1	1
					0.229	0.068	
					0.000	1.000	0.000
					0.000	0.000	
					0.000	0.000	

CE	7	53	60
	3.321	0.988	
	0.117	0.883	0.012
	0.006	0.014	
	0.001	0.011	
DF	3	8	11
	0.090	0.027	
	0.273	0.727	0.002
	0.003	0.002	
	0.001	0.002	
ES	5	12	17
	0.311	0.093	
	0.294	0.706	0.003
	0.004	0.003	
	0.001	0.002	
GO	2	8	10
	0.037	0.011	
	0.200	0.800	0.002
	0.002	0.002	
	0.000	0.002	
MA	0	27	27
	6.192	1.842	
	0.000	1.000	0.005
	0.000	0.007	
	0.000	0.005	
MG	12	172	184
	21.608	6.430	
	0.065	0.935	0.037
	0.010	0.045	
	0.002	0.034	

RS	2	22	24
	2.230	0.664	
	0.083	0.917	0.005
	0.002	0.006	
	0.000	0.004	
SC	1	25	26
	4.130	1.229	
	0.038	0.962	0.005
	0.001	0.006	
	0.000	0.005	
SE	10	16	26
	2.734	0.814	
	0.385	0.615	0.005
	0.009	0.004	
	0.002	0.003	
SP	20	2870	2890
	623.341	185.480	
	0.007	0.993	0.579
	0.017	0.746	
	0.004	0.575	
TO	0	2	2
	0.459	0.136	
	0.000	1.000	0.000
	0.000	0.001	
	0.000	0.000	
Column Total	1145	3848	4993
	0.229	0.771	

MS	1	17	18
	2.370	0.705	
	0.056	0.944	0.004
	0.001	0.004	
	0.000	0.003	
MT	0	5	5
	1.147	0.341	
	0.000	1.000	0.001
	0.000	0.001	
	0.000	0.001	
PA	1	13	14
	1.522	0.453	
	0.071	0.929	0.003
	0.001	0.003	
	0.000	0.003	
PB	4	46	50
	4.861	1.447	
	0.080	0.920	0.010
	0.003	0.012	
	0.001	0.009	
PE	19	100	119
	2.518	0.749	
	0.160	0.840	0.024
	0.017	0.026	
	0.004	0.020	
PI	2	33	35
	4.525	1.346	
	0.057	0.943	0.007
	0.002	0.009	
	0.000	0.007	

### 2.3.8 DS\_GRAU\_INSTRUCAO

Em uma sequência crescente a maior parte dos candidatos possuem o ensino médio completo e superior completo conforme tabela 13.

Tabela 13 Grau de Instrução

EFC	EFI	EMC	EMI	LÊ E ESC	SC	SI
244	147	1357	162	24	2568	491
0.049	0.029	0.271	0.032	0.005	0.515	0.098

EFC: Ensino Fundamental Completo  
EFI: Ensino Fundamental Incompleto  
EMC: Ensino Médio Completo  
EMI: Ensino Médio Incompleto  
Lê e Esc: Lê e Escreve  
SC: Superior Completo  
SI: Superior Incompleto

### 2.3.9 DS\_COR\_RACA

De acordo com a autodeclaração dos candidatos, 61% são da Cor/Raça Branca, sendo 52% eleitos na UF de SP. Dos 24 % autodeclarados Pardos, 12% estão na UF da Bahia e 13 % em São Paulo.

Tabela 14 Cor e Raça

AMARELA	BRANCA	INDÍGENA	PARDA	PRETA
39	3041	16	1219	678
0.008	0.609	0.003	0.244	0.136

Tabela 15 Cor e Raça x UF Eleitos

cand\$DS_COR_RACA	cand\$SG_UF		Row Total
	BA	SP	
AMARELA	4	35	39
	2.733	0.813	
	0.103	0.897	0.008
	0.003	0.009	
	0.001	0.007	
BRANCA	287	2754	3041
	241.480	71.854	
	0.094	0.906	0.609
	0.251	0.716	
	0.057	0.552	
INDÍGENA	7	9	16
	3.024	0.900	
	0.438	0.562	0.003
	0.006	0.002	
	0.001	0.002	
PARDA	584	635	1219
	331.594	98.668	
	0.479	0.521	0.244
	0.510	0.165	
	0.117	0.127	
PRETA	263	415	678
	74.355	22.125	
	0.388	0.612	0.136
	0.230	0.108	
	0.053	0.083	
Column Total	1145	3848	4993
	0.229	0.771	

3. Análise por variáveis descritivas

As variáveis descritivas deste estudo dizem respeito ao estado civil dos candidatos e ao gênero. As seguintes análises foram realizadas:

3.1 DS\_ESTADO\_CIVIL

Apenas como um aspecto descritivo verificamos que 53% dos candidatos são casados e 31% solteiros.

Tabela 16 Estado Civil

CASADO(A)	DIVORCIADO(A)	SEPARADO(A) JUDICIALMENTE	SOLTEIRO(A)	VIÚVO(A)
2657	651	54	1540	91
0.532	0.130	0.011	0.308	0.018

### 3.4.2 DS\_GENERO

O cenário político brasileiro ainda é dominado por uma maioria masculina - 68%. No cruzamento da variável Gênero com Cargos verificamos que o gênero feminino ocupou 18% no cargo de deputado estadual e 14% para deputado federal. Uma parte ainda tímida dentro da política brasileira.

Tabela 17 Gênero

	FEMININO	MASCULINO
	1581	3412
	0.317	0.683

Tabela 18 Gênero X Cargo

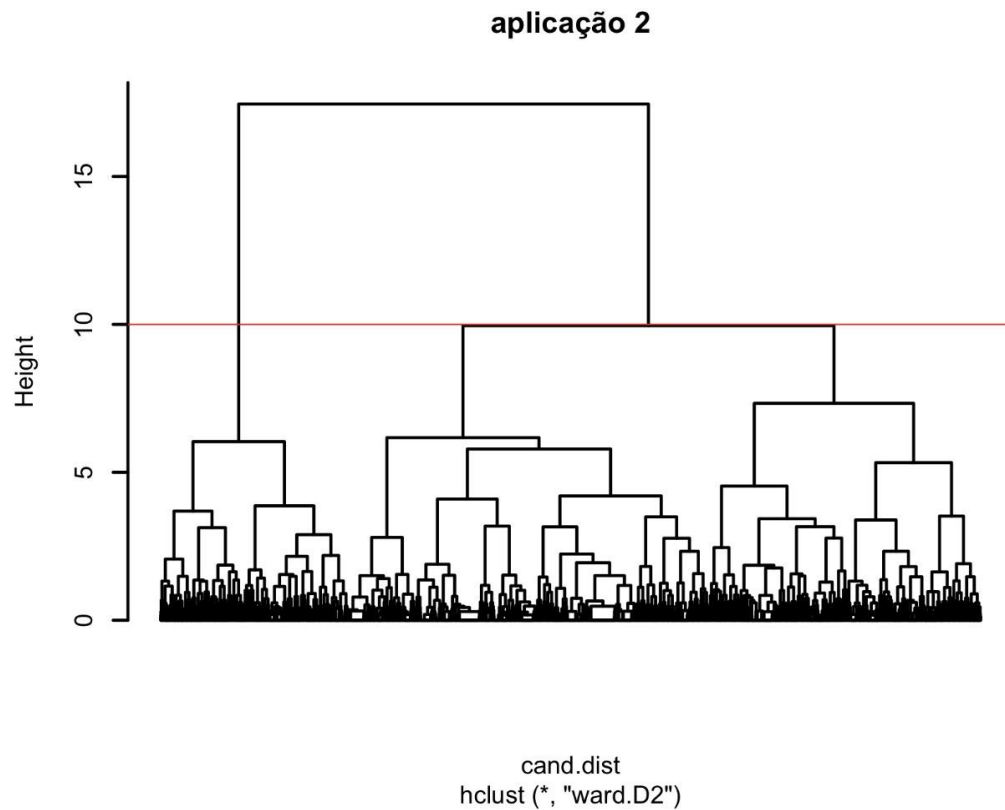
cand\$DS_GENERO	cand\$DS_CARGO		Row Total
	DEPUTADO ESTADUAL	DEPUTADO FEDERAL	
FEMININO	892	689	1581
	0.004	0.005	
	0.564	0.436	0.317
	0.317	0.316	
	0.179	0.138	
MASCULINO	1919	1493	3412
	0.002	0.002	
	0.562	0.438	0.683
	0.683	0.684	
	0.384	0.299	
Column Total	2811	2182	4993
	0.563	0.437	

## 4. Análise de Clusters obtidos pelo método hierárquico aglomerativo

### 4.1 Caracterização dos grupos

Utilizando o método hierárquico aglomerativo, os registros foram agrupados em dois clusters, conforme apresentado no Dendrograma abaixo.

Gráfico 1 - Dendrograma HC





## 4.2 Comparação dos clusters hierárquicos gerados com os drivers

### 4.2.1 DS\_SITUACAO\_CANDIDATURA

cand\$DS_SITUACAO_CANDIDATURA	cand\$hc2		Row Total
	1	2	
APTO	1024	3442	4466
	0.229	0.771	0.894
INAPTO	116	411	527
	0.220	0.780	0.106
Column Total	1140	3853	4993

Com relação ao driver DS\_SITUACAO\_CANDIDATURA o cluster 2 é identificado como candidatos APTOS ao pleito eleitoral, que representa 77% do total de candidatos.

### 4.2.2 TP\_AGREMIACAO

cand\$TP_AGREMIACAO	cand\$hc2		Row Total
	1	2	
COLIGAÇÃO	905	1183	2088
	0.433	0.567	0.418
PARTIDO ISOLADO	235	2670	2905
	0.081	0.919	0.582
Column Total	1140	3853	4993

Com relação ao driver TP\_AGREMIACAO, candidatos cujo partido possui coligação, estão distribuídos de maneira uniforme entre os dois clusters.

Já os candidatos cujo partido não possui coligação, estão agrupados em sua maioria (92%) no cluster 2.

### 4.2.3 ST\_REELEICAO

cand\$ST_REELEICAO	cand\$hc2		Row Total
	1	2	
N	1058 0.222	3715 0.778	4773 0.956
S	82 0.373	138 0.627	220 0.044
Column Total	1140	3853	4993

Com relação ao driver ST\_REELEICAO, o cluster 2 agrupa a maioria proporcional dos candidatos, independentemente da ocupação de um mandato eleitoral vigente.

### 4.2.4 ST\_DECLARAR\_BENS

cand\$ST_DECLARAR_BENS	cand\$hc2		Row Total
	1	2	
N	433 0.223	1509 0.777	1942 0.389
S	707 0.232	2344 0.768	3051 0.611
Column Total	1140	3853	4993

Com relação ao driver ST\_DECLARAR\_BENS, o cluster 2 agrupa a maioria proporcional dos candidatos, independentemente de possuírem bens declarados ou não.

### 4.2.5 SG\_UF

cand\$SG_UF	cand\$hc2		Row Total
	1	2	
BA	1124 0.982	21 0.018	1145 0.229
SP	16 0.004	3832 0.996	3848 0.771
Column Total	1140	3853	4993

Com relação ao driver ST\_UF, o cluster 1 está caracterizado, em sua maioria, por candidatos que disputaram cargos públicos no estado da Bahia, enquanto o cluster 2 agrupa predominantemente candidatos do estado de São Paulo.

#### 4.2.6 DS\_CARGO

	cand\$hc2		
cand\$DS_CARGO	1	2	Row Total
DEPUTADO ESTADUAL	647	2164	2811
	0.230	0.770	0.563
DEPUTADO FEDERAL	493	1689	2182
	0.226	0.774	0.437
Column Total	1140	3853	4993

Com relação ao driver DS\_CARGO, o cluster 2 agrupa a maioria dos candidatos, independentemente da ocupação de um mandato eleitoral vigente.

#### 4.2.7 SG\_UF\_NASCIMENTO - POR SG\_UF

cand\$SG_UF_NASCIMENTO	SUMMARY		
	1	2	Row Total
AC	0	1	1
	0.000	1.000	0.000
AL	3	33	36
	0.083	0.917	0.007
AM	3	4	7
	0.429	0.571	0.001
AP	0	1	1
	0.000	1.000	0.000
BA	1046	166	1212
	0.863	0.137	0.243
CE	7	53	60
	0.117	0.883	0.012
DF	3	8	11
	0.273	0.727	0.002
ES	5	12	17
	0.294	0.706	0.003
GO	2	8	10
	0.200	0.800	0.002
MA	0	27	27
	0.000	1.000	0.005
MG	11	173	184
	0.060	0.940	0.037
MS	1	17	18
	0.056	0.944	0.004
MT	0	5	5
	0.000	1.000	0.001
PA	0	14	14
	0.000	1.000	0.003
PB	4	46	50
	0.080	0.920	0.010
PE	16	103	119
	0.134	0.866	0.024
PI	2	33	35
	0.057	0.943	0.007
PR	1	122	123
	0.008	0.992	0.025
RJ	7	58	65
	0.108	0.892	0.013
RN	3	22	25
	0.120	0.880	0.005
RO	0	4	4
	0.000	1.000	0.001
RR	0	1	1
	0.000	1.000	0.000
RS	1	23	24
	0.042	0.958	0.005

SC	1	25	26
	0.038	0.962	0.005
SE	9	17	26
	0.346	0.654	0.005
SP	15	2875	2890
	0.005	0.995	0.579
TO	0	2	2
	0.000	1.000	0.000
Column Total	1140	3853	4993

Percebe-se aqui uma repetição dos clusters e, portanto, esta análise não gera informações relevantes para o estudo.

#### 4.2.8 DS\_GRAU\_INSTRUCAO

cand\$DS_GRAU_INSTRUCAO	cand\$hc2		Row Total
	1	2	
ENSINO FUNDAMENTAL COMPLETO	32	212	244
	0.131	0.869	0.049
ENSINO FUNDAMENTAL INCOMPLETO	36	111	147
	0.245	0.755	0.029
ENSINO MÉDIO COMPLETO	386	971	1357
	0.284	0.716	0.272
ENSINO MÉDIO INCOMPLETO	46	116	162
	0.284	0.716	0.032
LÊ E ESCREVE	4	20	24
	0.167	0.833	0.005
SUPERIOR COMPLETO	528	2040	2568
	0.206	0.794	0.514
SUPERIOR INCOMPLETO	108	383	491
	0.220	0.780	0.098
Column Total	1140	3853	4993

Com relação ao driver DS\_GRAU\_INSTRUCAO, o cluster 2 agrupa a maioria dos candidatos, independentemente do nível de escolaridade deles.

#### 4.2.9 DS\_COR\_RACA

cand\$DS_COR_RACA	cand\$hc2		Row Total
	1	2	
AMARELA	3	36	39
	0.077	0.923	0.008
BRANCA	278	2763	3041
	0.091	0.909	0.609
INDÍGENA	7	9	16
	0.438	0.562	0.003
PARDA	584	635	1219
	0.479	0.521	0.244
PRETA	268	410	678
	0.395	0.605	0.136
Column Total	1140	3853	4993

Com relação ao driver DS\_COR\_RACA, o cluster 2 agrupa a maioria dos candidatos, independentemente da sua cor de pele.

### 4.3 - Variáveis Descritivas

#### 4.3.1 DS\_ESTADO\_CIVIL

cand\$DS_ESTADO_CIVIL	cand\$hc2		Row Total
	1	2	
CASADO(A)	551	2106	2657
	0.207	0.793	0.532
DIVORCIADO(A)	137	514	651
	0.210	0.790	0.130
SEPARADO(A) JUDICIALMENTE	7	47	54
	0.130	0.870	0.011
SOLTEIRO(A)	430	1110	1540
	0.279	0.721	0.308
VIÚVO(A)	15	76	91
	0.165	0.835	0.018
Column Total	1140	3853	4993

Com relação ao driver DS\_ESTADO\_CIVIL, o cluster 2 agrupa a maioria dos candidatos, independentemente da sua cor de pele.

#### 4.3.2 DS\_GENERO

cand\$DS_GENERO	cand\$hc2		Row Total
	1	2	
FEMININO	368 0.232	1218 0.768	1586 0.317
MASCULINO	777 0.227	2643 0.773	3420 0.683
Column Total	1145	3861	5006

Com relação ao driver DS\_GENERO, o cluster 2 agrupa a maioria dos candidatos, independentemente do seu sexo.

## 5. Análise de Clusters obtidos pelo método K-medoid

Número de clusters selecionados pelo método k-medoid = 2

### 5.1 Análise por variáveis drivers

#### 5.1.1 DS\_SITUACAO\_CANDIDATURA

cand\$DS_SITUACAO_CANDIDATURA	cand\$cmd		Row Total
	1	2	
APTO	1284	3182	4466
	0.288	0.712	0.894
INAPTO	153	374	527
	0.290	0.710	0.106
Column Total	1437	3556	4993

#### 5.1.2 TP\_AGREMIACAO

cand\$TP_AGREMIACAO	cand\$cmd		Row Total
	1	2	
COLIGAÇÃO	1160	928	2088
	0.556	0.444	0.418
PARTIDO ISOLADO	277	2628	2905
	0.095	0.905	0.582
Column Total	1437	3556	4993

#### 5.1.3 ST\_REELEICAO

cand\$ST_REELEICAO	cand\$cmd		Row Total
	1	2	
N	1349	3424	4773
	0.283	0.717	0.956
S	88	132	220
	0.400	0.600	0.044
Column Total	1437	3556	4993

#### 5.1.4 ST\_DECLARAR\_BENS

cand\$ST_DECLARAR_BENS	cand\$km		Row Total
	1	2	
N	567	1375	1942
	0.292	0.708	0.389
S	870	2181	3051
	0.285	0.715	0.611
Column Total	1437	3556	4993

#### 5.1.5 SG\_UF

cand\$SG_UF	cand\$km		Row Total
	1	2	
BA	1135	10	1145
	0.991	0.009	0.229
SP	302	3546	3848
	0.078	0.922	0.771
Column Total	1437	3556	4993

#### 5.1.6 DS\_CARGO

cand\$DS_CARGO	cand\$km		Row Total
	1	2	
DEPUTADO ESTADUAL	790	2021	2811
	0.281	0.719	0.563
DEPUTADO FEDERAL	647	1535	2182
	0.297	0.703	0.437
Column Total	1437	3556	4993



## 5.1.7 SG\_UF\_NASCIMENTO

cand\$SG_UF_NASCIMENTO	cand\$km		Row Total
	1	2	
AC	0 0.000	1 1.000	1 0.000
AL	8 0.222	28 0.778	36 0.007
AM	2 0.286	5 0.714	7 0.001
AP	1 1.000	0 0.000	1 0.000
BA	1128 0.931	84 0.069	1212 0.243
CE	10 0.167	50 0.833	60 0.012
DF	3 0.273	8 0.727	11 0.002
ES	7 0.412	10 0.588	17 0.003
GO	3 0.300	7 0.700	10 0.002
MA	3 0.111	24 0.889	27 0.005
MG	31 0.168	153 0.832	184 0.037
MS	3 0.167	15 0.833	18 0.004
MT	0 0.000	5 1.000	5 0.001
PA	4 0.286	10 0.714	14 0.003
PB	7 0.140	43 0.860	50 0.010
PE	39 0.328	80 0.672	119 0.024
PI	5 0.143	30 0.857	35 0.007
PR	10 0.081	113 0.919	123 0.025
RJ	14 0.215	51 0.785	65 0.013
RN	4 0.160	21 0.840	25 0.005
RO	0 0.000	4 1.000	4 0.001
RR	0 0.000	1 1.000	1 0.000
RS	2 0.083	22 0.917	24 0.005
SC	2 0.077	24 0.923	26 0.005
SE	13 0.500	13 0.500	26 0.005
SP	138 0.048	2752 0.952	2890 0.579
TO	0 0.000	2 1.000	2 0.000
Column Total	1437	3556	4993

### 5.1.8 DS\_GRAU\_INSTRUCAO

cand\$DS_GRAU_INSTRUCAO	cand\$cmd		Row Total
	1	2	
ENSINO FUNDAMENTAL COMPLETO	57 0.234	187 0.766	244 0.049
ENSINO FUNDAMENTAL INCOMPLETO	49 0.333	98 0.667	147 0.029
ENSINO MÉDIO COMPLETO	480 0.354	877 0.646	1357 0.272
ENSINO MÉDIO INCOMPLETO	58 0.358	104 0.642	162 0.032
LÊ E ESCRIVE	4 0.167	20 0.833	24 0.005
SUPERIOR COMPLETO	640 0.249	1928 0.751	2568 0.514
SUPERIOR INCOMPLETO	149 0.303	342 0.697	491 0.098
Column Total	1437	3556	4993

### 5.1.9 DS\_COR\_RACA

cand\$DS_COR_RACA	cand\$cmd		Row Total
	1	2	
AMARELA	5 0.128	34 0.872	39 0.008
BRANCA	293 0.096	2748 0.904	3041 0.609
INDÍGENA	8 0.500	8 0.500	16 0.003
PARDA	832 0.683	387 0.317	1219 0.244
PRETA	299 0.441	379 0.559	678 0.136
Column Total	1437	3556	4993

## 5.2 Variáveis Descritivas

### 5.2.1 DS\_ESTADO\_CIVIL

cand\$DS_ESTADO_CIVIL	cand\$cmd		Row Total
	1	2	
CASADO(A)	717	1940	2657
	0.270	0.730	0.532
DIVORCIADO(A)	166	485	651
	0.255	0.745	0.130
SEPARADO(A) JUDICIALMENTE	11	43	54
	0.204	0.796	0.011
SOLTEIRO(A)	518	1022	1540
	0.336	0.664	0.308
VIÚVO(A)	25	66	91
	0.275	0.725	0.018
Column Total	1437	3556	4993

### 5.2.2 DS\_GENERO

cand\$DS_GENERO	cand\$cmd		Row Total
	1	2	
FEMININO	459	1122	1581
	0.290	0.710	0.317
MASCULINO	978	2434	3412
	0.287	0.713	0.683
Column Total	1437	3556	4993

## 6. Comparação e Validação entre os agrupamentos pelos métodos Hierárquico Aglomerativo e K-Medoid

cand\$hc2	cand\$cmd		Row Total
	1	2	
1	1128	12	1140
2	309	3544	3853
Column Total	1437	3556	4993

## 7. Conclusões

Os dois métodos, hierárquico e kmedoid, obtiveram praticamente o mesmo resultado de agrupamento, tanto na quantidade de clusters identificados, quanto na distribuição das observações em cada grupo, com relação às variáveis drivers, obtendo o índice de validação(*corrected.rand*) entre os métodos de 0.99%, resultando em uma coesão entre os agrupamentos.