Data Mining

Partitioning Clustering (k-means)

Note on Profiling

17/11/2021

NOVA-IMS

Fernando Lucas Bação

bacao@isegi.unl.pt

http://www.isegi.unl.pt/fbacao

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

1

AGENDA

- Cluster analysis
  - Clustering techniques
    - Partitioning Methods (kmeans and k-meadoids)
    - **A Note on Profiling**

2

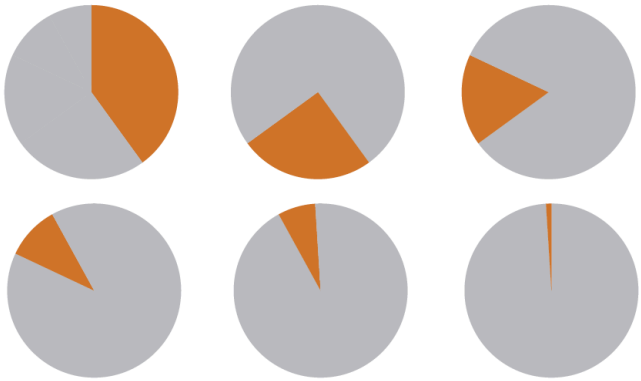
Profiling

3

Clustering

- **Profiling (size of the clusters)**

Part-to-Whole Mini-Pie Charts

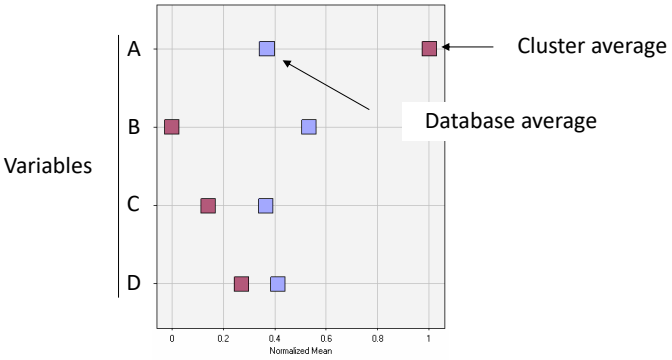


4

NOVA IMS Information Management School

Clustering

- **Profiling (comparing averages)**

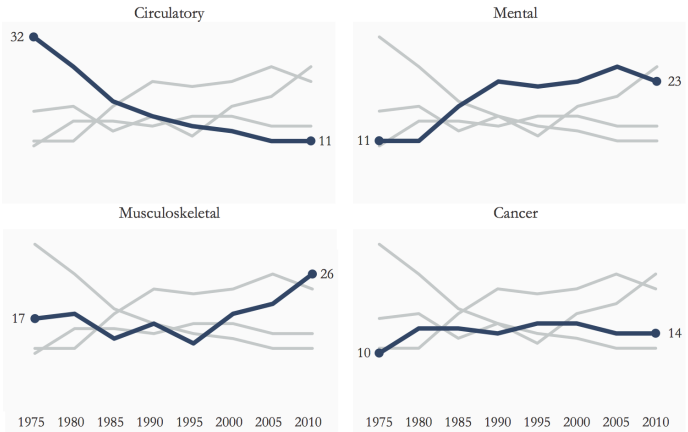


5

NOVA IMS Information Management School

Clustering

- **Profiling (comparing profiles)**



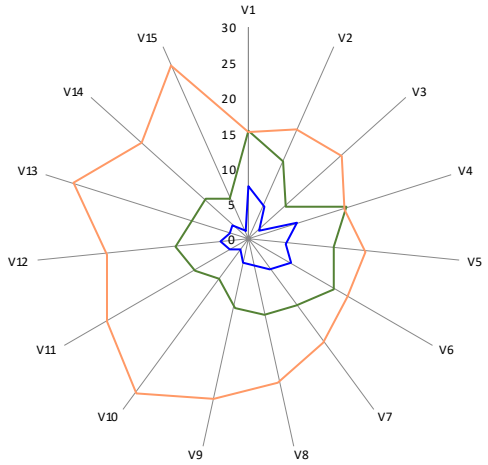
JA Schwabish "An Economist's Guide to Visualizing Data", The Journal of Economic Perspectives, 2014

6

NOVA IMS Information Management School

Clustering

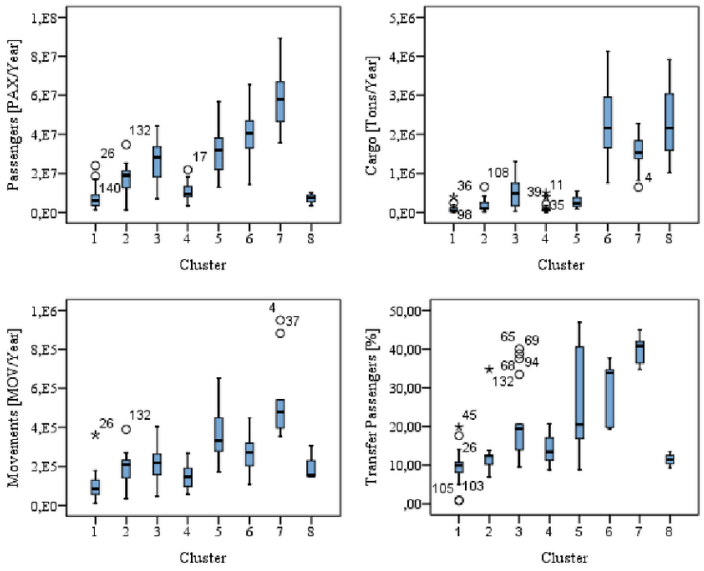
- **Profiling (comparing profiles)**



7

NOVA IMS Information Management School

Clustering

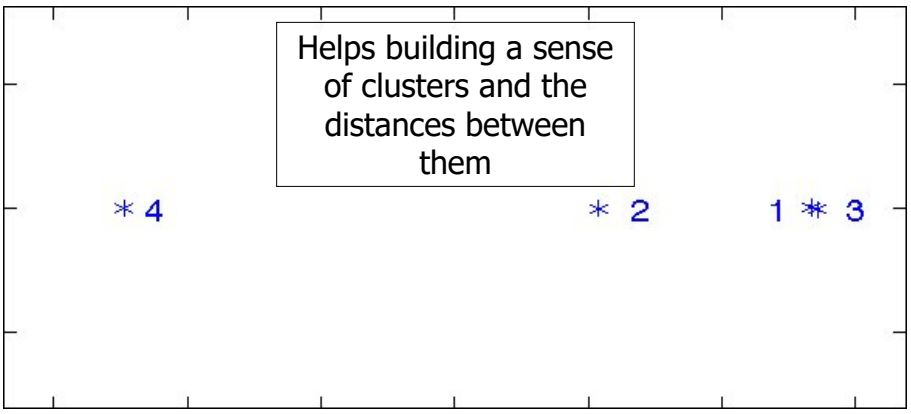


8

NOVA IMS Information Management School

Clustering

- **Profiling (multidimensional scaling)**

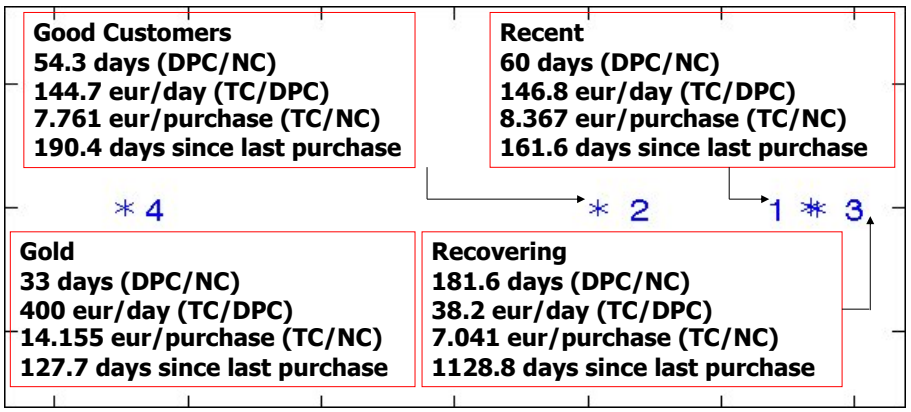


9

NOVA IMS Information Management School

Clustering

- **Profiling (multidimensional scaling)**

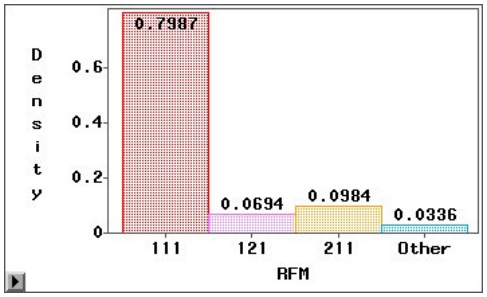


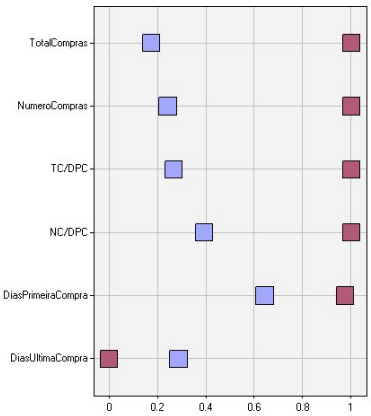
10

NOVA IMS Information Management School

Clustering

- **Profiling (comparing variables not used)**





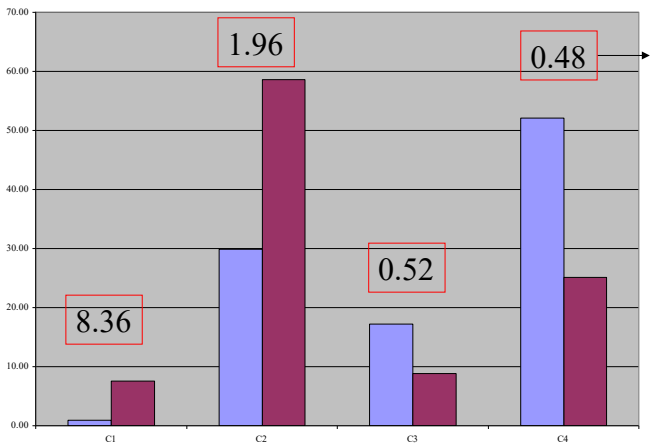
Used and unused variable distribution

11

NOVA IMS Information Management School

Clustering

- **Profiling (leverage)**



Leverage
Ratio Sales(%)/Individuals(%)

12

Clustering

- **Exploring two solutions**

		Sofist_Digital	Variedade Serviço	Preço	Status_PC	ProAtividade	Tempo Internet e RS	Sustentabilidade	Life Status	Propensao_Pfidelizacao	Switching	
k=5	_FREQ_	F1_SD	F2_VS	F3_P	F4_SPC	F5_PA	F6_IRS	F7_CC	F8_LS	F9_FI	F10_AB	
1	222	-0,40	**-0,54**	**-0,31**	-0,29	0,13	**0,65**	**-1,67**	-0,20	-0,09	-0,03	Pouco Envolvidos e pouco sensiveis ao preço
2	463	**0,70**	**0,44**	**0,38**	0,13	**-0,32**	**-0,37**	-0,19	**0,31**	**-0,58**	0,03	Social Customers
3	469	**0,58**	-0,16	0,15	**0,48**	**0,23**	0,00	0,09	**-0,22**	**0,98**	**-0,04**	Sofisticado com Propensao Fidelizacao
4	428	**-1,47**	0,27	0,25	-0,13	0,01	**-0,25**	0,29	0,15	0,15	**0,04**	Info Excluidos
5	420	0,29	-0,30	**-0,68**	**-0,39**	0,02	0,31	**0,69**	-0,16	**-0,56**	-0,02	Sustentaveis
k=6	FREQ	F1_SD	F2_VS	F3_P	F4_SPC	F5_PA	F6_IRS	F7_CC	F8_LS	F9_FI	F10_AB	
1	241	0,12	**-0,30**	-0,09	-0,20	-0,04	**1,34**	**-1,29**	-0,03	-0,13	-0,24	Pouco Envolvidos
2	134	-0,16	0,11	0,02	-0,15	-0,10	0,13	0,10	0,00	0,07	**2,84**	Switchers
3	285	0,43	**0,48**	**-1,28**	-0,16	**0,52**	**-0,28**	0,05	0,04	**0,23**	-0,07	Exigentes
4	426	**-1,42**	0,22	**0,30**	-0,20	-0,03	**-0,27**	0,10	0,01	**0,21**	-0,27	Info Excluidos
5	414	0,39	**-0,77**	0,20	**0,92**	0,19	**-0,23**	0,15	**0,21**	0,15	-0,16	Status/Conservadores
6	502	**0,63**	**0,30**	**0,34**	**-0,37**	**-0,38**	-0,10	**0,35**	**-0,19**	**-0,39**	-0,24	Social Customers

Clustering

- **Exploring two solutions with different k**

K=7 VS K=8	Pouco envolvidos	Switchers	Informados / Social Customers	Fidelizáveis Sofisticados	Info Excluidos	Status	Proativos	Desprendido	Grand Total
Pouco envolvidos	71		23	13	45	1	1	49	**203**
Switchers	2	106	4	4	2	1		3	**122**
Informados / Social Customers	4		219	68	2	20	1	64	**378**
Fidelizáveis Sofisticados	72		105	191	4	1	2	69	**444**
Status/Conservadores	7		5	39	13	157		42	**263**
Proativos	5		15	20	3		153	19	**215**
Info Excluidos			12	4	309	1	1	50	**377**
Grand Total	**161**	**106**	**383**	**339**	**378**	**181**	**158**	**296**	**2002**

K=6 VS K=8	Pouco envolvidos	Switchers	Informados / Social Customers	Fidelizáveis Sofisticados	Info Excluidos	Status	Proativos	Desprendido	Grand Total
Pouco Envolvidos	146		16	24	8	17	7	23	**241**
Switchers	2	105	6	11	1	2	4	3	**134**
Exigentes	1	1	26	56	13	27	60	101	**285**
Info Excluidos				22	350	17	11	26	**426**
Status/Conservadores	3		13	177	5	98	58	60	**414**
Social Customers	9		322	49	1	20	18	83	**502**
Grand Total	**161**	**106**	**383**	**339**	**378**	**181**	**158**	**296**	**2002**

NOVA IMS Information Management School

Clustering

- **Consolidation of two solutions**