


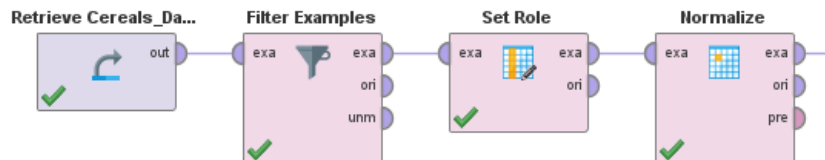
Homework 10.2

Data-preprocessing:

- Remove all cereals with missing values, do normalization and appropriate attribute selection.

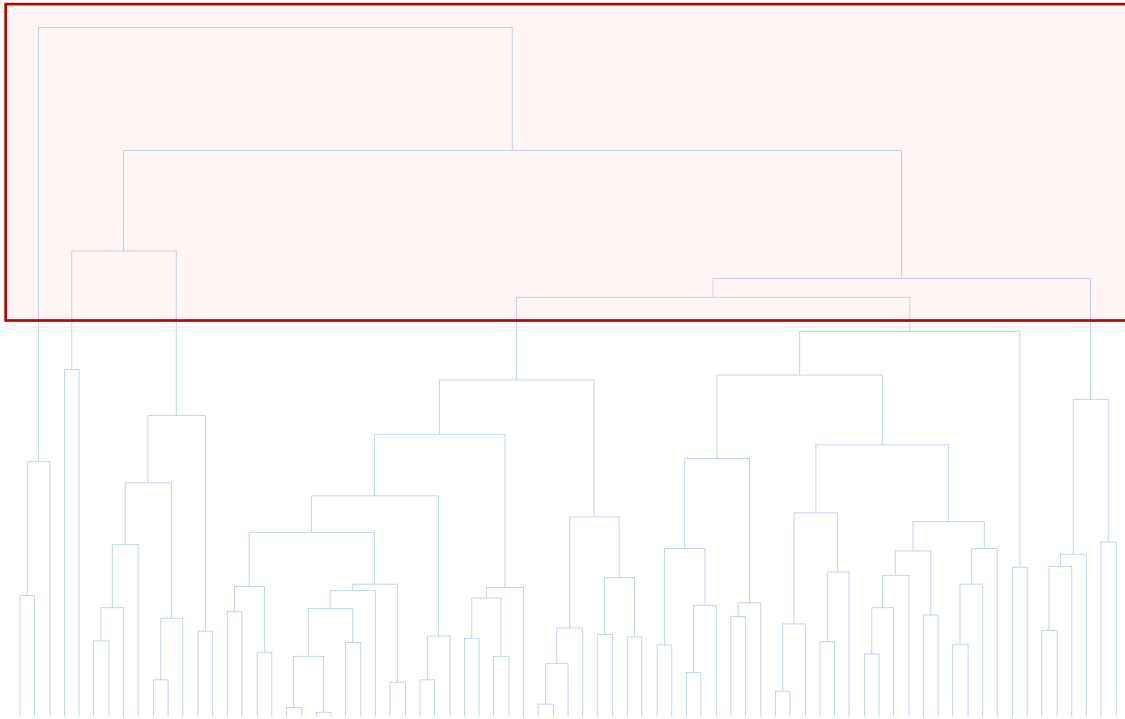
Name	Type	Missing	Statistics			Filter (12 / 12 attributes): <input type="text"/>
✓ Cereals	Polynomial	0	Least Wheaties_Honey_Gold ...	Most 100%_Bran (1)	Values 100%_Bran (1), 100%_Natural_Bran (1), ...[75 more]	
✓ calories	Integer	0	Min 50	Max 160	Average 106.883	
✓ protein	Integer	0	Min 1	Max 6	Average 2.545	
✓ fat	Integer	0	Min 0	Max 5	Average 1.013	
✓ sodium	Integer	0	Min 0	Max 320	Average 159.675	
✓ fiber	Real	0	Min 0	Max 14	Average 2.152	
✓ carbo	Real	0	Min -1	Max 23	Average 14.597	
✓ sugars	Integer	0	Min -1	Max 15	Average 6.922	
✓ potass	Integer	2	Min 15	Max 330	Average 98.667	
✓ vitamins	Integer	0	Min 0	Max 100	Average 28.247	
✓ shelf	Integer	0	Min 1	Max 3	Average 2.208	
✓ rating	Real	0	Min 18.043	Max 93.705	Average 42.666	

 Create Filters: **filters**
Defines the list of filters to apply.



Data Mining Steps:

- Apply Hierarchical algorithm with appropriate distance to determine K number of clusters.



K number of clusters equal 6.

- Apply K-Mean algorithm with K from previous step.

Clustering (2) (k-Means)	
<input checked="" type="checkbox"/> add cluster attribute	①
<input type="checkbox"/> add as label	①
<input type="checkbox"/> remove unlabeled	①
k	6
max runs	10
<input checked="" type="checkbox"/> determine good start values	①
measure types	NumericalMeasures
numerical measure	EuclideanDistance
max optimization steps	100
<input type="checkbox"/> use local random seed	①

Cluster Model

Cluster 0: 18 items
Cluster 1: 3 items
Cluster 2: 18 items
Cluster 3: 20 items
Cluster 4: 10 items
Cluster 5: 6 items
Total number of items: 75

Row No.	Cereals	cluster	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	rating
1	100%_Bran	cluster_1	-1.872	1.314	-0.013	-0.358	3.247	-2.229	-0.226	2.575	-0.164	0.954	1.855
2	100%_Natur...	cluster_2	0.662	0.410	3.960	-1.722	-0.076	-1.527	0.226	0.516	-1.280	0.954	-0.608
3	All-Bran	cluster_1	-1.872	1.314	-0.013	1.184	2.831	-1.761	-0.452	3.143	-0.164	0.954	1.212
4	All-Bran_with...	cluster_1	-2.886	1.314	-1.007	-0.240	4.908	-1.527	-1.580	3.285	-0.164	0.954	3.666
5	Apple_Cinna...	cluster_3	0.155	-0.494	0.980	0.235	-0.284	-0.941	0.677	-0.407	-0.164	-1.432	-0.929
6	Apple_Jacks	cluster_3	0.155	-0.494	-1.007	-0.417	-0.491	-0.824	1.580	-0.975	-0.164	-0.239	-0.666
7	Basic_4	cluster_2	1.169	0.410	0.980	0.591	-0.076	0.815	0.226	0.019	-0.164	0.954	-0.390
8	Bran_Cheex	cluster_0	-0.858	-0.494	-0.013	0.472	0.755	0.112	-0.226	0.374	-0.164	-1.432	0.475
9	Bran_Flakes	cluster_0	-0.858	0.410	-1.007	0.591	1.170	-0.356	-0.452	1.297	-0.164	0.954	0.775
10	Cap'nCrunch	cluster_3	0.662	-1.398	0.980	0.709	-0.907	-0.590	1.129	-0.904	-0.164	-0.239	-1.749
11	Cheerios	cluster_0	0.155	3.121	0.980	1.539	-0.076	0.581	-1.355	0.090	-0.164	-1.432	0.593
12	Cinnamon_T...	cluster_3	0.662	-1.398	1.974	0.591	-0.907	-0.356	0.452	-0.762	-0.164	-0.239	-1.622
13	Clusters	cluster_2	0.155	0.410	0.980	-0.240	-0.076	-0.356	0	0.090	-0.164	0.954	-0.149
14	Cocoa_Puffs	cluster_3	0.155	-1.398	-0.013	0.235	-0.907	-0.590	1.355	-0.620	-0.164	-0.239	-1.413

- Explain the meaning of each cluster by going their description.

Attribute	cluster_0	cluster_1	cluster_2	cluster_3	cluster_4	cluster_5
calories	-0.182	-2.210	0.775	0.232	-1.163	0.493
protein	0.159	1.314	0.560	-0.946	0.139	0.108
fat	-0.510	-0.344	1.035	-0.013	-0.709	-0.179
sodium	0.788	0.195	-0.095	0.116	-1.882	0.571
fiber	-0.168	3.662	0.386	-0.699	-0.005	-0.145
carbo	0.880	-1.839	-0.226	-0.473	-0.192	0.854
sugars	-0.815	-0.753	0.477	1.027	-1.129	-0.151
potass	-0.305	3.001	0.733	-0.780	-0.088	-0.040
vitamins	-0.164	-0.164	-0.226	-0.164	-0.833	3.185
shelf	-0.437	0.954	0.822	-0.596	-0.358	0.954
rating	0.348	2.244	-0.257	-1.016	1.382	-0.314

- Cluster_0: ในกลุ่มนี้มี sodium และ carbohydrate ในปริมาณที่สูงกว่าส่วนอื่นๆจึงจัดให้เป็นกลุ่มที่ให้พลังงานสูง
- Cluster_1: ในกลุ่มนี้มี fiber และ potassium ในปริมาณที่สูงกว่าส่วนอื่นๆ และยังมี protein ในปริมาณที่พอเหมาะ รวมทั้งมี carbohydrate และ calories ในปริมาณต่ำจึงจัดให้เป็นกลุ่มที่เหมาะสมกับผู้ที่ต้องการลดน้ำหนักหรือออกกำลังกาย
- Cluster_2: ในกลุ่มนี้มี fat ที่มากกว่ากลุ่มอื่นๆ มี potassium และ calories ในปริมาณที่สูง รวมทั้งมี sugar และ protein จึงจัดให้เป็นกลุ่มที่เหมาะสมสำหรับผู้ที่ต้องการพลังงานจากไขมัน
- Cluster_3: ในกลุ่มนี้มี sugar ในปริมาณที่สูงกว่าส่วนอื่นๆอย่างชัดเจนจึงจัดให้เป็นกลุ่มที่เหมาะสมสำหรับผู้ที่ต้องการความหวานจากน้ำตาล
- Cluster_4: ในกลุ่มนี้มีสารอาหารทุกอย่างในปริมาณที่ไม่แตกต่างกันมากนัก ไม่มีอะไรที่มากจนเกินไปหรือน้อยจนเกินไปจึงจัดให้เป็นกลุ่มธรรมดา
- Cluster_5: ในกลุ่มนี้มี vitamin ในปริมาณที่สูงกว่าส่วนอื่นๆอย่างชัดเจนจึงจัดให้เป็นกลุ่มที่เหมาะสมสำหรับผู้ที่ต้องการวิตามินเสริม

- The elementary public schools would like to choose a set of cereals to include in their daily cafeterias. Every day a different cereal is offered, but all cereals should support a healthy diet. For this goal you are requested to find a cluster of "healthy cereals".

เนื่องจากเด็กในวัยประถมศึกษาเป็นวัยกำลังโตที่ต้องการอาหารที่ให้ protein เป็นจำนวนมากเพื่อเป็นการเสริมสร้างกล้ามเนื้อ ให้พลังงานในจำนวนที่เหมาะสม และมี fiber คอยช่วยในการย่อยอาหาร ดังนั้น cereal ที่เหมาะสมกับเด็กในวัยนี้คือ cereal ใน cluster 1 โดย cereal ใน cluster 1 มีดังนี้

