

## 3 Cpus - oplossingen

June 14, 2021

### 1 CPU's

Lees de gegevens uit `cpus.csv` in als een dataframe. Gebruik de eerste kolom als index.

Hieronder vind je meer informatie over deze dataset:

---

cpus	<i>Performance of Computer CPUs</i>
------	-------------------------------------

---

### Description

A relative performance measure and characteristics of 209 CPUs.

### Usage

cpus

### Format

The components are:

name manufacturer and model.

syct cycle time in nanoseconds.

mmin minimum main memory in kilobytes.

mmax maximum main memory in kilobytes.

cach cache size in kilobytes.

chmin minimum number of channels.

chmax maximum number of channels.

perf published performance on a benchmark mix relative to an IBM 370/158-3.

estperf estimated performance (by Ein-Dor & Feldmesser).

### Source

P. Ein-Dor and J. Feldmesser (1987) Attributes of the performance of central processing units: a relative performance prediction model. *Comm. ACM.* **30**, 308–317.

### References

Venables, W. N. and Ripley, B. D. (2002) *Modern Applied Statistics with S*. Fourth edition. Springer.

- a. Welke kolommen dien je te verwijderen om een Principale-Componenten Analyse te kunnen uitvoeren? Verwijder eventuele kolommen met de verkeerde meetschaal, kolommen die niet zinvol zijn,....

```
<class 'pandas.core.frame.DataFrame'>
```

```
Index: 209 entries, ADVISOR 32/60 to WANG VS 90
```

Data columns (total 8 columns):

#	Column	Non-Null Count	Dtype
0	syct	209 non-null	int64
1	mmin	209 non-null	int64
2	mmax	209 non-null	int64
3	cach	209 non-null	int64
4	chmin	209 non-null	int64
5	chmax	209 non-null	int64
6	perf	209 non-null	int64
7	estperf	209 non-null	int64

dtypes: int64(8)

memory usage: 14.7+ KB

	syct	mmin	mmax	cach	chmin	chmax	perf	estperf
name								
ADVISOR 32/60	125	256	6000	256	16	128	198	199
AMDAHL 470V/7	29	8000	32000	32	8	32	269	253
AMDAHL 470/7A	29	8000	32000	32	8	32	220	253
AMDAHL 470V/7B	29	8000	32000	32	8	32	172	253
AMDAHL 470V/7C	29	8000	16000	32	8	16	132	132

- b. Bekijk de correlaties tussen de variabelen. Zijn de gegevens bruikbaar om er een Principale-Componenten Analyse op toe te passen?

	syct	mmin	mmax	cach	chmin	chmax	perf	\
syct	1.000000	-0.335642	-0.378561	-0.321000	-0.301090	-0.250502	-0.307082	
mmin	-0.335642	1.000000	0.758157	0.534729	0.517189	0.266907	0.794923	
mmax	-0.378561	0.758157	1.000000	0.537990	0.560513	0.527246	0.862994	
cach	-0.321000	0.534729	0.537990	1.000000	0.582245	0.487846	0.662613	
chmin	-0.301090	0.517189	0.560513	0.582245	1.000000	0.548281	0.608903	
chmax	-0.250502	0.266907	0.527246	0.487846	0.548281	1.000000	0.605219	
perf	-0.307082	0.794923	0.862994	0.662613	0.608903	0.605219	1.000000	
estperf	-0.288396	0.819292	0.901202	0.648620	0.610580	0.592156	0.966469	

	estperf
syct	-0.288396
mmin	0.819292
mmax	0.901202
cach	0.648620
chmin	0.610580
chmax	0.592156
perf	0.966469
estperf	1.000000

0.611760899472704

- c. Voer een Principale-Componenten Analyse uit en interpreteer de resultaten.

	syst	mmin	mmax	cach	chmin	chmax	perf	\
PC1	-0.198791	0.364847	0.399053	0.336254	0.331495	0.297704	0.420988	
PC2	0.916296	0.170859	0.106722	-0.094997	-0.111905	-0.101663	0.192075	
PC3	0.290452	-0.487833	-0.198665	0.199472	0.343205	0.683140	-0.065989	
PC4	-0.138396	-0.145534	0.304876	-0.634908	-0.437152	0.482625	0.133134	
PC5	0.019832	0.127541	0.066404	-0.631699	0.743737	-0.106508	-0.107464	
PC6	-0.022395	0.639887	-0.658019	-0.121283	-0.101343	0.357229	0.037941	
PC7	-0.119701	-0.384236	-0.477340	-0.145419	0.072658	-0.248978	0.652695	
PC8	-0.046309	-0.071918	-0.178038	-0.010546	0.001522	-0.021987	-0.569910	

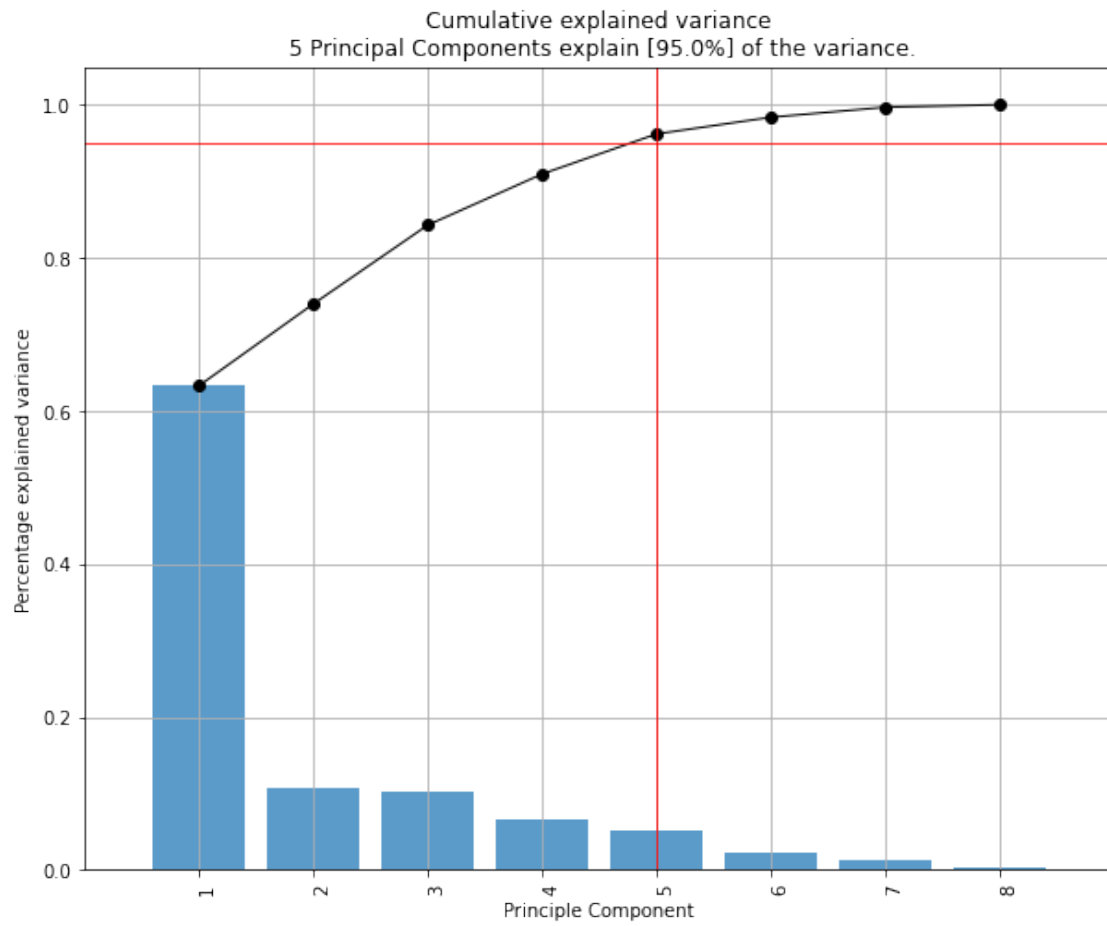
	estperf
PC1	0.423451
PC2	0.225929
PC3	-0.097850
PC4	0.147834
PC5	-0.062048
PC6	-0.054961
PC7	0.309441
PC8	0.797240

```
array([0.63259623, 0.73955709, 0.84258012, 0.90937382, 0.96171572,
       0.98354247, 0.99662992, 1.          ])
```

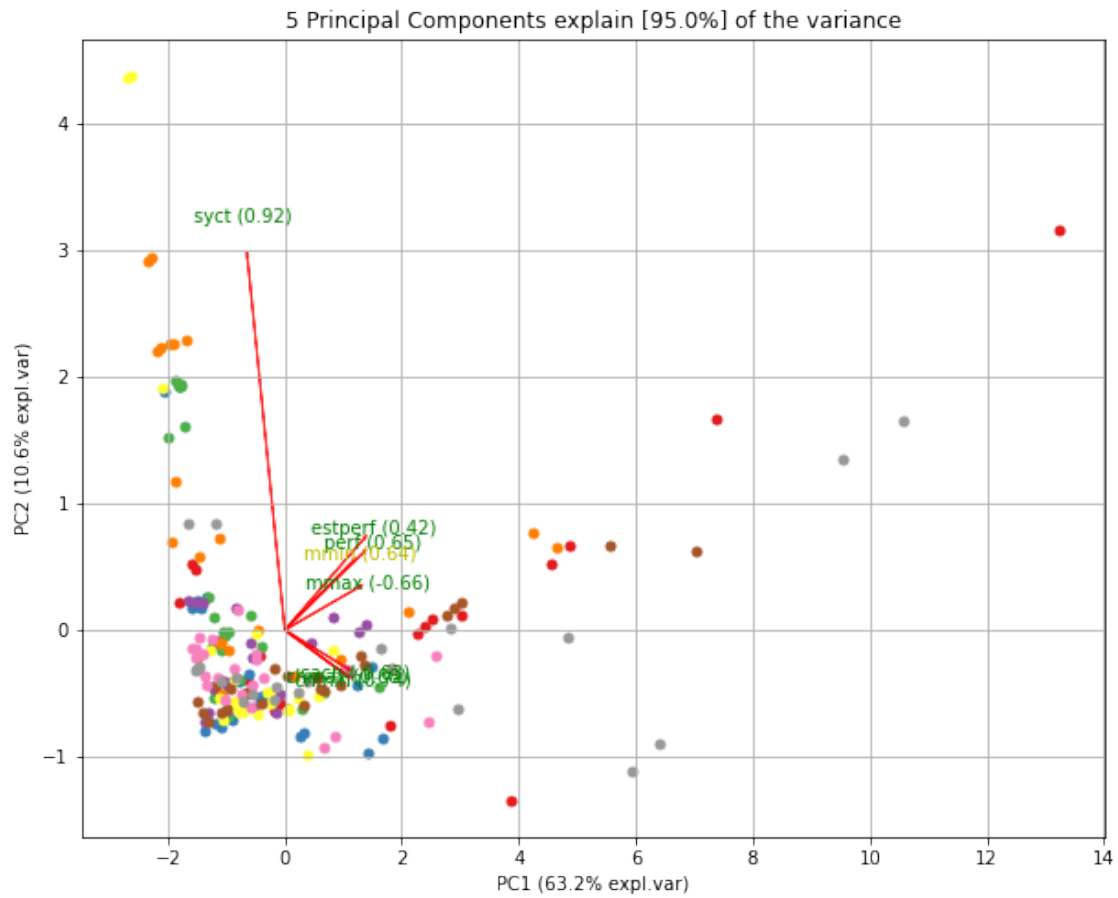
- d. Zijn er een of meerdere variabelen die uitgesproken doorweegt in het bepalen van de eerste hoofdcomponent? Zo ja welke?

	PC	feature	loading	type
0	PC1	estperf	0.423451	best
1	PC2	syst	0.916296	best
2	PC3	chmax	0.683140	best
3	PC4	cach	-0.634908	best
4	PC5	chmin	0.743737	best
5	PC6	mmax	-0.658019	best
6	PC7	perf	0.652695	best
7	PC8	estperf	0.797240	best
8	PC6	mmin	0.639887	weak

- e. Maak een plot en biplot



<Figure size 432x288 with 0 Axes>



- f. Op basis van de plot, hoeveel hoofdcomponenten zou je weerhouden wanneer je de dataset zou 'reducen'?