

# Primera pràctica

Aprenentatge Computacional

Sofia Di Capua  
Marc Bosom

# Objectiu

CNT

Nombre de bicicletes

+

Dies laborables

Nombre de bicicletes

+

Temperatura

# Com ho hem fet

## Adaptació del dataset

### **Reduir el nombre de variables a...**

- No aprofitables
- Baix percentatge d'aparició
- One-hot encoding

## Anàlisi de les dades

- Distribució de les variables
- Estudi de normalitat
- Relació entre les variables
- Estudi de correlació

## Regressió

### **Comparem diferents tipus de regressió:**

- Sense modificar x
- Estandaritzant x
- Normalitzant x
- Regressió polinomial
- Residual errors
- PCA

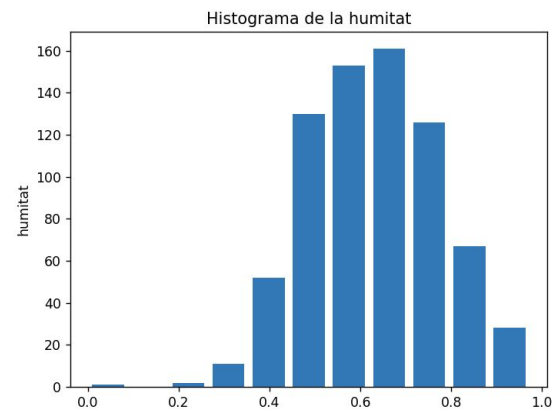
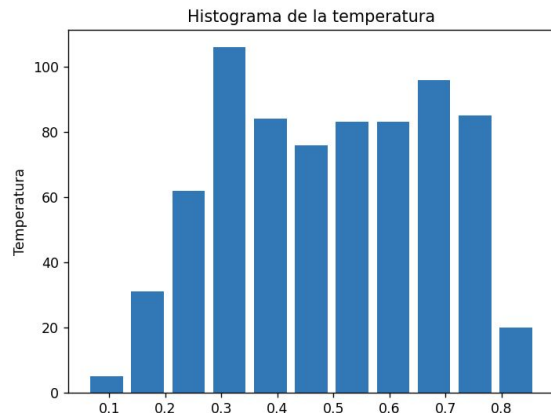
	instant	dteday	season	yr	mnth	holiday	weekday	workingday
0	1	2011-01-01	1	0	1	0	6	0
1	2	2011-01-02	1	0	1	0	0	0
2	3	2011-01-03	1	0	1	0	1	1
3	4	2011-01-04	1	0	1	0	2	1
4	5	2011-01-05	1	0	1	0	3	1
..	...	...	...	..	...	...	...	...
726	727	2012-12-27	1	1	12	0	4	1
727	728	2012-12-28	1	1	12	0	5	1
728	729	2012-12-29	1	1	12	0	6	0
729	730	2012-12-30	1	1	12	0	0	0
730	731	2012-12-31	1	1	12	0	1	1

	weathersit	temp	atemp	hum	windspeed	casual	registered	cnt
0	2	0.344	0.364	0.806	0.160	331	654	985
1	2	0.363	0.354	0.696	0.249	131	670	801
2	1	0.196	0.189	0.437	0.248	120	1229	1349
3	1	0.200	0.212	0.590	0.160	108	1454	1562
4	1	0.227	0.229	0.437	0.187	82	1518	1600
..	...	...	...	...	...	...	...	...
726	2	0.254	0.227	0.653	0.350	247	1867	2114
727	2	0.253	0.255	0.590	0.155	644	2451	3095
728	2	0.253	0.242	0.753	0.124	159	1182	1341
729	1	0.256	0.232	0.483	0.351	364	1432	1796
730	2	0.216	0.223	0.578	0.155	439	2290	2729

# Adaptació del Dataset

# Anàlisi de les dades

Distribucions

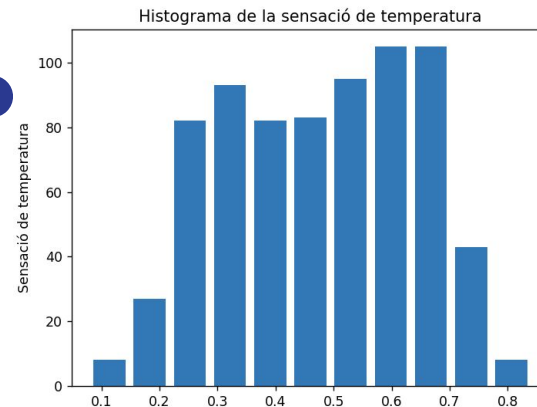
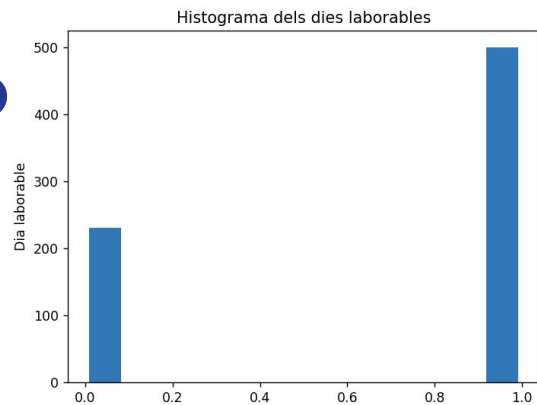


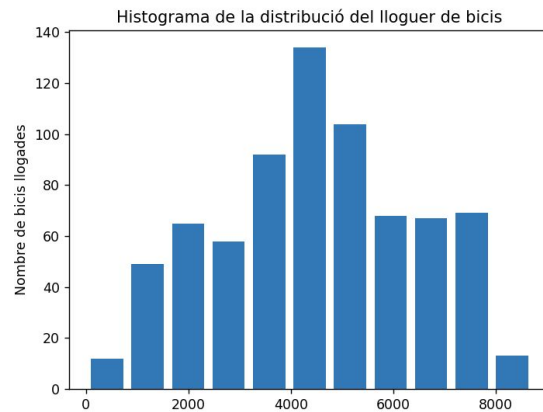
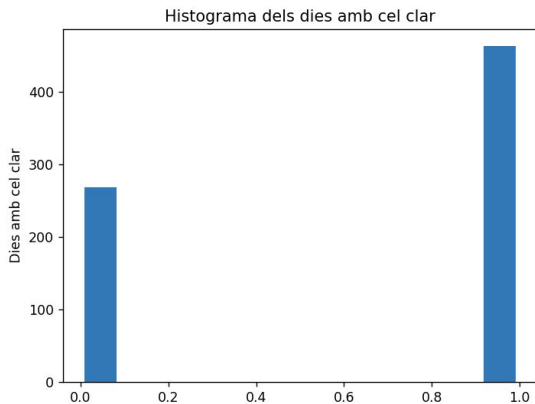
Workingday

temp

atemp

hum



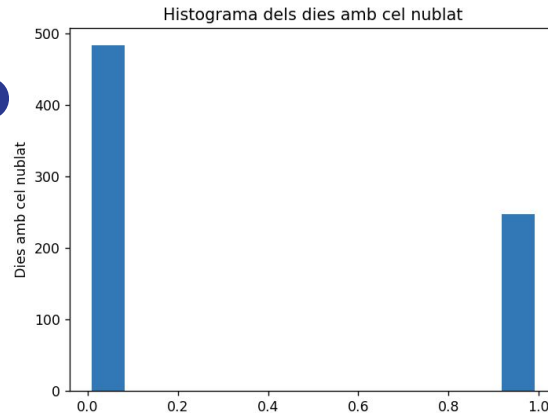
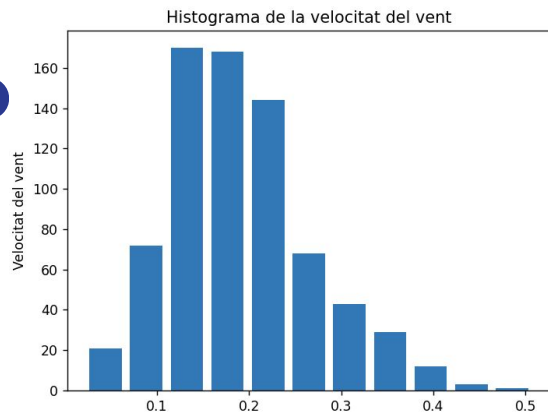


Windspeed

Clear

Mist&Cloudy

CNT



## Test de normalitat de Shapiro-Wilk

Dades	Resultats	Conclusió
temp	stat=0.966, p=0.000	No Gausiana
atemp	stat=0.974, p=0.000	No Gausiana
hum	stat=0.993, p=0.002	No Gausiana
windspeed	stat=0.971, p=0.000	No Gausiana
cnt	stat=0.980, p=0.000	No Gausiana

## Test de normalitat d'Agostino

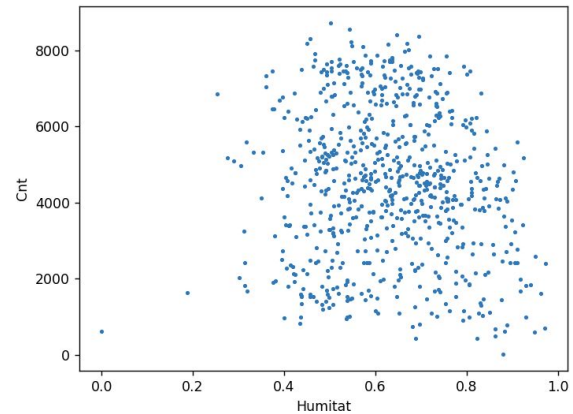
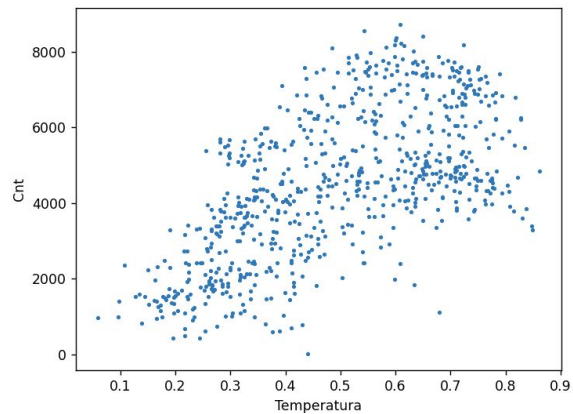
Dades	Resultats	Conclusió
temp	stat=294.297, p=0.000	No Gausiana
atemp	stat=144.352, p=0.000	No Gausiana
hum	stat=0.683, p=0.711	Gausiana
windspeed	stat=51.425, p=0.000	No Gausiana
cnt	stat=62.708, p=0.000	No Gausiana

# Estudi numèric de la normalitat



# Anàlisi de les dades

Correlació

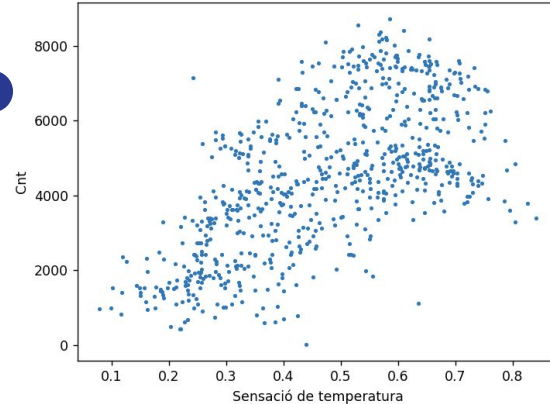
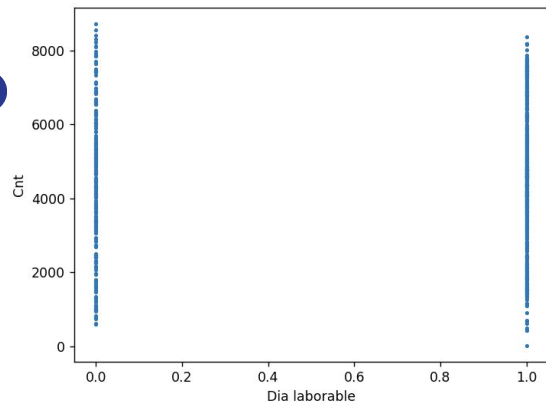


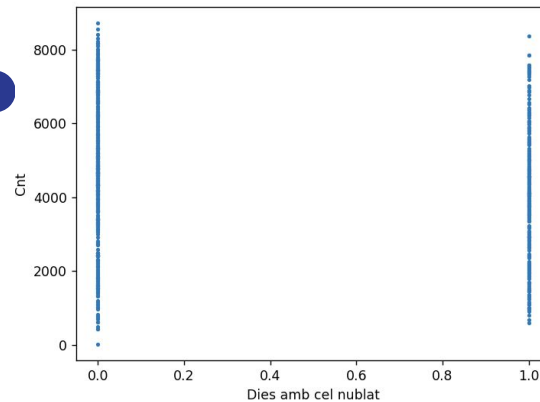
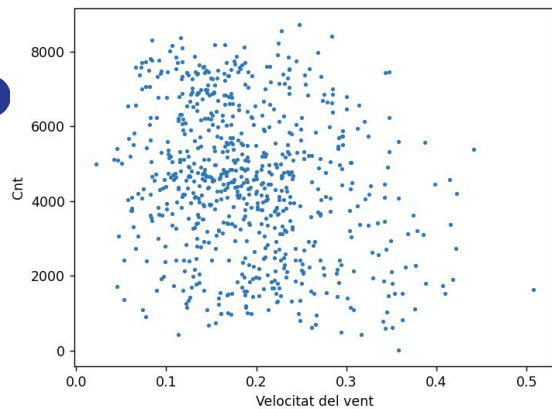
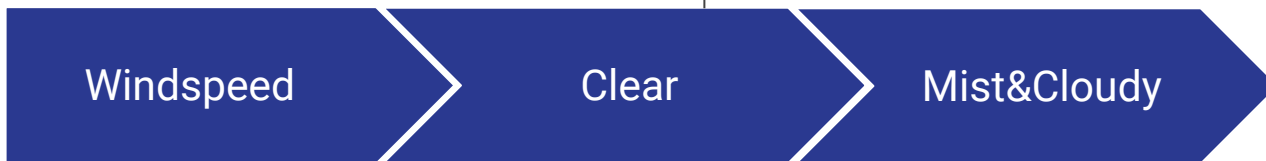
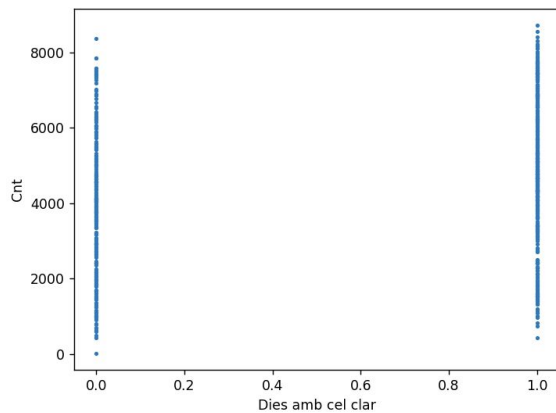
Workingday

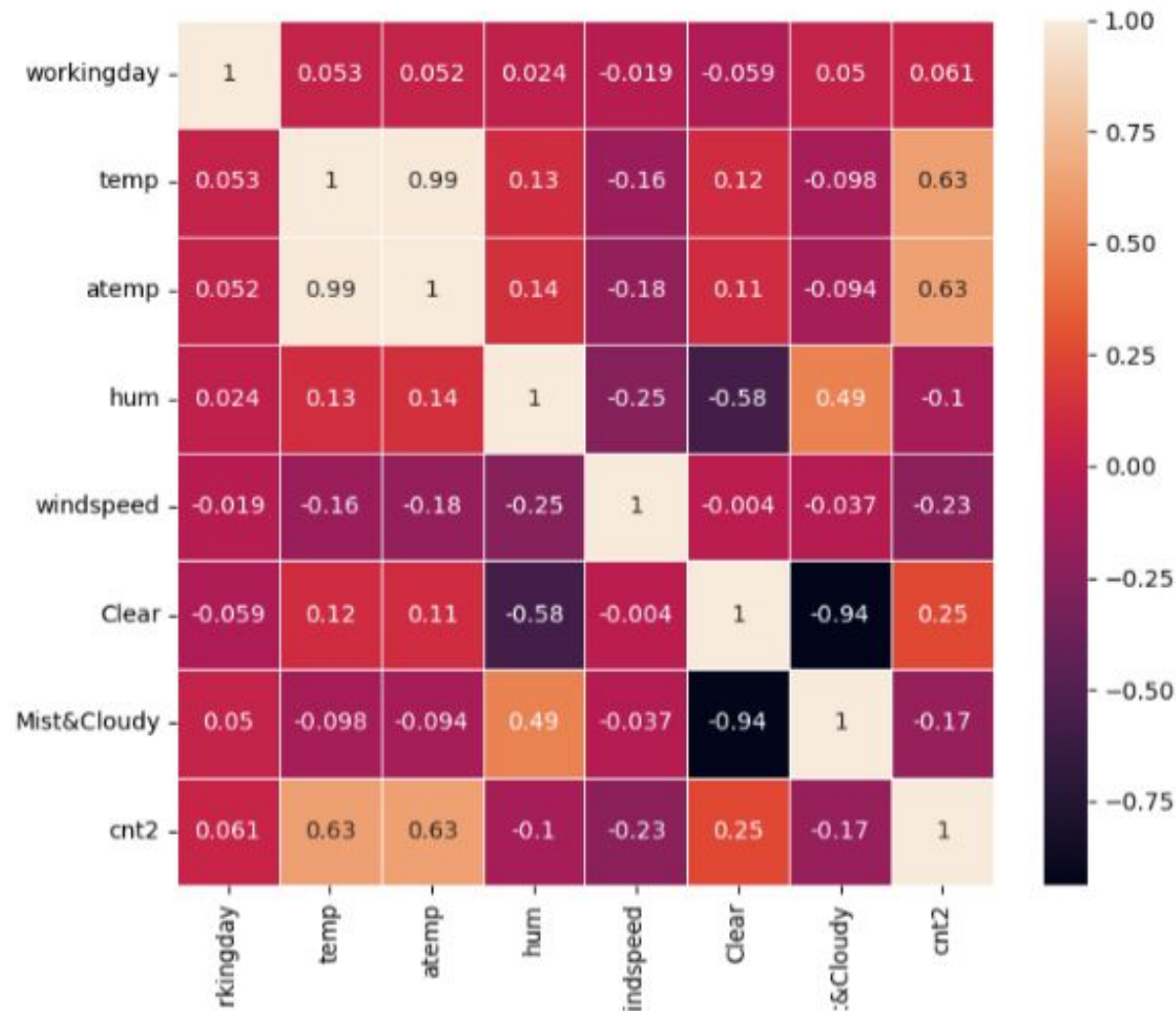
temp

atemp

hum







Mapa de calor  
de la correlació  
de les variables

## Test de correlació de Pearson

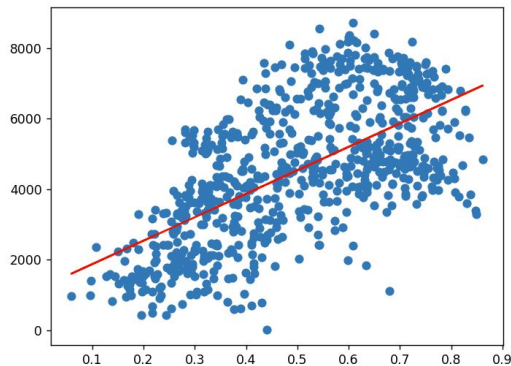
Dades	Resultats	Conclusió
TEMP	stat=0.627, p=0.000	Dependent
ATEMP	stat=0.631, p=0.000	Dependent
HUM	stat=-0.101, p=0.006	Dependent
WINDSPEED	stat=-0.235, p=0.000	Dependent

Estudi numèric de la  
correlació

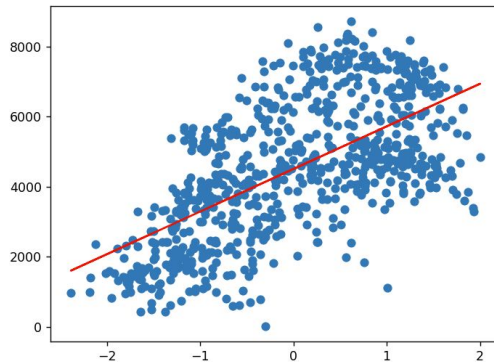
# Regressió

Correlació

## Regressió inicial



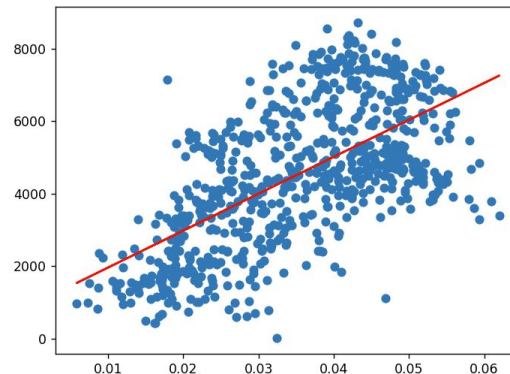
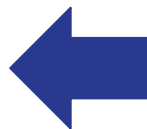
## Regressió x estandarizada



MSE: 2272020.25  
 $R^2$ : 0.39

## Regressió amb les dades normalitzades

MSE: 2255173.85  
 $R^2$ : 0.40



# Regressió polinomial

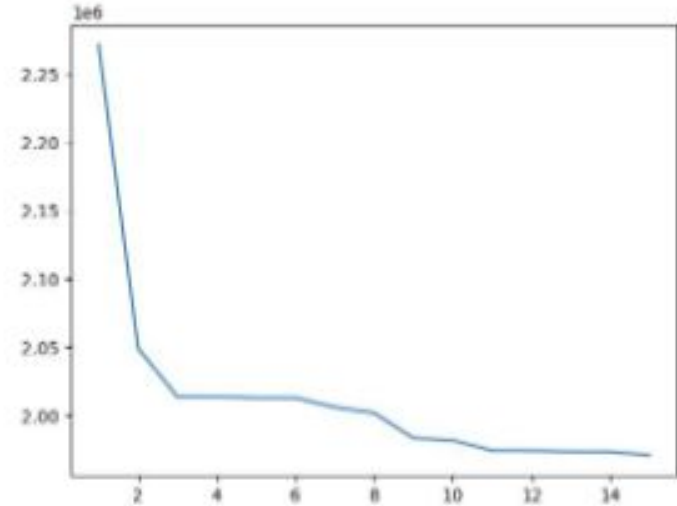
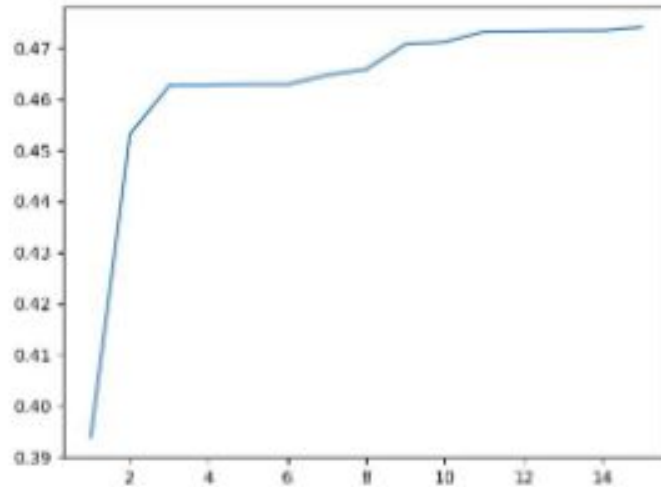


Figura 18: Evolució de l' $R^2$  i de l'MSE, respectivament, en funció del grau del polinomi

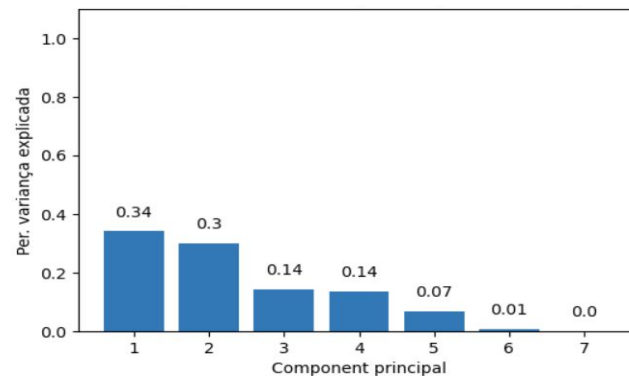


# Principal Component Analysis (PCA)

## Resultats PCA

	workingday	temp	atemp	hum	windspeed	Clear	Mist&Cloudy
PC1	0.046	-0.142	-0.138	0.455	-0.062	-0.620	0.603
PC2	-0.076	-0.652	-0.657	-0.256	0.262	0.035	-0.042
PC3	-0.880	-0.069	-0.056	0.163	-0.428	0.068	-0.058
PC4	-0.463	0.214	0.192	-0.125	0.801	-0.157	0.145
PC5	-0.053	0.082	0.080	-0.819	-0.315	-0.210	0.411
PC6	-0.010	0.032	-0.004	-0.120	-0.056	-0.735	-0.664
PC7	-0.000	0.704	-0.709	0.010	-0.016	0.020	0.016

## Percentatge de variància explicada per cada component



## Conclusions

No eliminem cap variable



Conclusions

Preguntas?