DATA

	age	anemia	creatinine phosphokinase	diabetes	ejection fraction	high_blood_pressure	platelets	serum_creatinine	serum_sodium	sex	smoking	time	death_event
0	75	1	582	0	20	1	265000.00	1.9	130	1	1	4	1
1	55	1	7861	0	38	0	263358.03	1.1	136	1	1	6	1
2	65	1	146	0	20	0	162000.00	1.3	129	1	1	7	1
3	50	1	111	0	20	0	210000.00	1.9	137	1	1	7	1
4	65	1	160	1	20	0	327000.00	2.7	116	0	1	8	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...
294	62	1	61	1	38	1	155000.00	1.1	143	1	1	270	0
295	55	1	1820	0	38	0	270000.00	1.2	139	0	0	271	0
296	45	1	2060	1	60	0	742000.00	0.8	138	0	0	278	0
297	45	1	2413	0	38	0	140000.00	1.4	140	1	1	280	0
298	50	1	196	0	45	0	395000.00	1.6	136	1	1	285	0

299 rows × 13 columns

Se determina cuales son los factores de riesgo asociado mortalidad en los paciente de falla cardiaca donde existe 13 variables .

```
# Cumulative proportion of variance (from PC1 to PC6)
np.cumsum(pca_out.explained_variance_ratio_)
```

array([0.13858049, 0.27017178, 0.37585032, 0.46884517, 0.55466877,
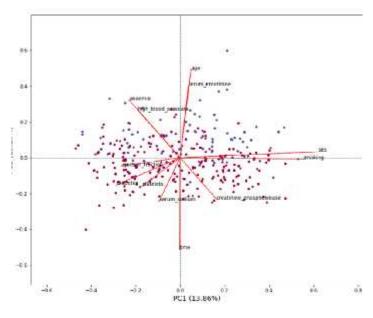   0.63707509, 0.71068826, 0.78112898, 0.84479713, 0.90376138])

Se toman 5 variables donde evidencia varianza acumulada de del 55%

variable	PC1	PC2	PC3	PC4	PC5
age	0.049101	0.490902	-0.102108	0.052074	0.197514
anaemia	-0.228183	0.370464	-0.160436	-0.414443	-0.205258
creatinine_phosphokinase	0.162765	-0.230922	0.083985	0.481709	0.510700
diabetes	-0.204097	-0.161573	0.309240	0.151390	-0.280351
ejection_fraction	-0.261512	-0.046662	-0.443315	-0.027672	0.354845
high_blood_pressure	-0.190258	0.265189	-0.229776	0.413401	-0.312432
platelets	-0.173213	-0.157015	-0.124657	0.503584	-0.270750
serum_creatinine	0.037199	0.403600	0.295003	0.047074	0.482013
serum_sodium	-0.093530	-0.240871	-0.025028	-0.010290	0.102830
sex	0.622375	0.033969	-0.092557	-0.035026	-0.115356
smoking	0.663570	0.008669	0.216302	0.077003	2.300657
time	-0.001082	-0.510580	0.041592	-0.353603	2.066587

Se elimina la variable dependiente de con variable de estudio(muerte por evento)

SdI1   Salas de Informática; 4/06/2022

PC1 (13.86%)

Se escojen las siguientes variable
X=data[["sex","smoking","ejection_fracti
on","platelets","diabetes"]]
Y=data["DEATH_EVENT"]

```
print_model=model.summary()
print(print_model)
```


CONCLUSION: DEACUERDO AL VALOR
DEL R CUADRADO QUE FUE DEL 7.6% NO
PUDO ESTIMAR LA FUERZA DE RELACION
ENTRE MODELO Y LAS VARIABLES, EN
CUAL LAS VARIABLES INDIPENDIENTE NO
SE RELACIONA CON LA VARIABLE
DEPENDIENTE