

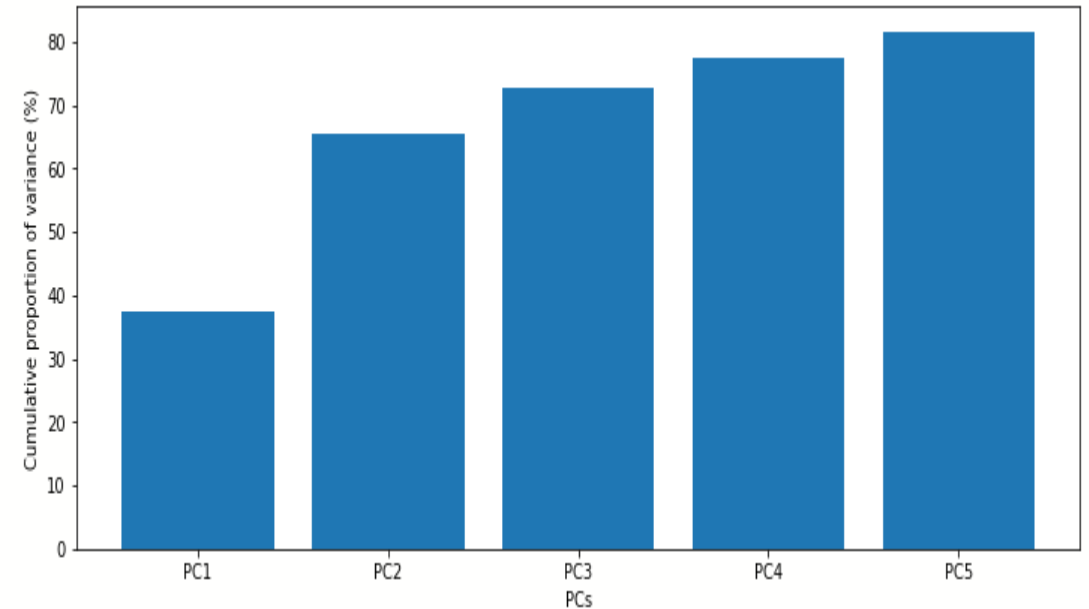
# Análisis de Componentes Principales y Predicción

Mildred Caro Álvarez

Esta base de datos contiene información con el fin de establecer el consumo energético de los electrodomésticos en una vivienda. Los datos contienen información de temperatura, y humedad relativa de los diferentes espacios de la casa, y condiciones atmosféricas de su alrededor, tomados en intervalos de 10 minutos. Esta información fue recogida en 137 días. De acuerdo a la descripción de la data, se observa que hay un total de 19.735 observaciones y un total de 28 variables. Para estandarizar las variables, se suprimieron 4 variables.

Una vez estandarizados los datos, a continuación se procede a validar si las primeras 5 componentes recogen la información suficiente para explicar el conjunto original de datos.

Al validar las primeras 5 componentes, se observa que las tres primeras de ellas explican la mayor parte de la variación de los datos originales, aproximadamente en un 72.82%.



```
np.cumsum(pca_out.explained_variance_ratio_)
```

```
array([0.37291678, 0.65542713, 0.72821719, 0.77366811, 0.81543227])
```

```
loadings_df[(loadings_df["PC1"]>0.28)][["PC1"]]
variable
T1      0.300893
T3      0.303848
T4      0.297670
T5      0.299199
T7      0.295028
T9      0.304527
Name: PC1, dtype: float64
```

```
loadings_df[(loadings_df["PC2"]<-0.31)][["PC2"]]
variable
RH_1    -0.331781
RH_3    -0.348843
RH_4    -0.358000
RH_7    -0.347312
RH_8    -0.345572
RH_9    -0.337168
Name: PC2, dtype: float64
```

```
loadings_df[(loadings_df["PC3"]>0.25)][["PC3"]]
variable
T6      0.318593
T_out   0.310488
Windspeed 0.496594
Name: PC3, dtype: float64
```

OLS Regression Results			
<b>Dep. Variable:</b>	Appliances	<b>R-squared (uncentered):</b>	0.537
<b>Model:</b>	OLS	<b>Adj. R-squared (uncentered):</b>	0.536
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	1632.
<b>Date:</b>	Sat, 04 Jun 2022	<b>Prob (F-statistic):</b>	0.00
<b>Time:</b>	20:33:32	<b>Log-Likelihood:</b>	-1.1816e+05
<b>No. Observations:</b>	19735	<b>AIC:</b>	2.363e+05
<b>Df Residuals:</b>	19721	<b>BIC:</b>	2.365e+05
<b>Df Model:</b>	14		
<b>Covariance Type:</b> nonrobust			

## R Cuadrado sin Intercepto

OLS Regression Results			
<b>Dep. Variable:</b>	Appliances	<b>R-squared:</b>	0.116
<b>Model:</b>	OLS	<b>Adj. R-squared:</b>	0.116
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	185.3
<b>Date:</b>	Sat, 04 Jun 2022	<b>Prob (F-statistic):</b>	0.00
<b>Time:</b>	20:50:35	<b>Log-Likelihood:</b>	-1.1816e+05
<b>No. Observations:</b>	19735	<b>AIC:</b>	2.363e+05
<b>Df Residuals:</b>	19720	<b>BIC:</b>	2.365e+05
<b>Df Model:</b>	14		
<b>Covariance Type:</b> nonrobust			

## R Cuadrado con Intercepto

18 variables sugeridas por los tres primeros componentes, pueden ser utilizadas para la predicción, arrojando un R cuadrado del 53,7%. Este R Cuadrado fue logrado mediante la eliminación del intercepto.