

Untitled8

April 17, 2020

1 TAREA # 1 Heiner Romero Leiva

2 Ejercicio #1

3 a) Calcule el resumen numerico, interprete los resultados para dos variables

```
[2]: import pandas as pd
import prince
import os
import pandas as pd
import numpy as np
os.chdir("/Users/heinerleivagmail.com")
print(os.getcwd())
datos = pd.read_csv('SpotifyTop2018_40_V2.csv',delimiter=',',decimal=".")
print(datos.head())
print(datos.shape)
print(datos.dropna().describe())
print(datos.dropna().mean())
print(datos.describe())
print(datos.mean(numeric_only=True))
print(datos.median(numeric_only=True))
print(datos.std(numeric_only=True))
print(datos.max(numeric_only=True))
print(datos.quantile(np.array([0,.25,.50,.75,1])))
```

/Users/heinerleivagmail.com

	danceability	energy	loudness	speechiness	\
Dura	0.791	0.848	-3.456	0.0506	
Plug_Walk	0.876	0.519	-6.531	0.1430	
Silence	0.520	0.761	-3.093	0.0853	
Solo_(feat._Demi_Lovato)	0.737	0.636	-4.546	0.0437	
Sin_Pijama	0.791	0.745	-3.695	0.0464	

	acousticness	instrumentalness	liveness	valence	\
Dura	0.1830	0.000000	0.409	0.828	
Plug_Walk	0.2020	0.000000	0.108	0.158	

Silence	0.2560	0.000005	0.170	0.286
Solo_(feat._Demi_Lovato)	0.0441	0.000067	0.350	0.565
Sin_Pijama	0.3540	0.000029	0.104	0.820

	tempo	duration_ms	time_signature
Dura	95.000	200480	4
Plug_Walk	94.981	175230	4
Silence	141.971	180823	4
Solo_(feat._Demi_Lovato)	105.005	222653	4
Sin_Pijama	94.014	188560	4

(40, 11)

	danceability	energy	loudness	speechiness	acousticness \
count	40.000000	40.000000	40.000000	40.000000	40.000000
mean	0.719200	0.661925	-5.846250	0.121292	0.184621
std	0.151423	0.137874	1.611051	0.094732	0.209052
min	0.258000	0.391000	-9.211000	0.023200	0.000282
25%	0.680500	0.564250	-7.077000	0.049550	0.030800
50%	0.750000	0.659000	-5.930000	0.105500	0.121500
75%	0.817500	0.772500	-4.624000	0.144500	0.227500
max	0.922000	0.909000	-3.093000	0.516000	0.847000

	instrumentalness	liveness	valence	tempo	duration_ms \
count	40.000000	40.000000	40.000000	40.000000	40.000000
mean	0.003695	0.177125	0.467942	122.108850	205645.700000
std	0.021231	0.130476	0.228504	31.178064	34324.034325
min	0.000000	0.021500	0.096700	77.169000	95467.000000
25%	0.000000	0.095275	0.287500	95.037500	189856.000000
50%	0.000003	0.112000	0.437500	122.531500	212904.500000
75%	0.000043	0.294750	0.633250	140.585750	226983.250000
max	0.134000	0.552000	0.931000	191.702000	268867.000000

	time_signature
count	40.000000
mean	3.950000
std	0.220721
min	3.000000
25%	4.000000
50%	4.000000
75%	4.000000
max	4.000000

danceability	0.719200
energy	0.661925
loudness	-5.846250
speechiness	0.121292
acousticness	0.184621
instrumentalness	0.003695
liveness	0.177125
valence	0.467942

```

tempo          122.108850
duration_ms    205645.700000
time_signature 3.950000
dtype: float64

```

```

      danceability  energy  loudness  speechiness  acousticness \
count      40.000000  40.000000  40.000000    40.000000    40.000000
mean        0.719200  0.661925 -5.846250     0.121292     0.184621
std         0.151423  0.137874  1.611051     0.094732     0.209052
min         0.258000  0.391000 -9.211000     0.023200     0.000282
25%         0.680500  0.564250 -7.077000     0.049550     0.030800
50%         0.750000  0.659000 -5.930000     0.105500     0.121500
75%         0.817500  0.772500 -4.624000     0.144500     0.227500
max         0.922000  0.909000 -3.093000     0.516000     0.847000

```

```

      instrumentalness  liveness  valence  tempo  duration_ms \
count      40.000000  40.000000  40.000000  40.000000    40.000000
mean        0.003695  0.177125  0.467942  122.108850  205645.700000
std         0.021231  0.130476  0.228504   31.178064   34324.034325
min         0.000000  0.021500  0.096700   77.169000   95467.000000
25%         0.000000  0.095275  0.287500   95.037500  189856.000000
50%         0.000003  0.112000  0.437500  122.531500  212904.500000
75%         0.000043  0.294750  0.633250  140.585750  226983.250000
max         0.134000  0.552000  0.931000  191.702000  268867.000000

```

```

      time_signature
count      40.000000
mean        3.950000
std         0.220721
min         3.000000
25%         4.000000
50%         4.000000
75%         4.000000
max         4.000000
danceability      0.719200
energy            0.661925
loudness         -5.846250
speechiness       0.121292
acousticness      0.184621
instrumentalness  0.003695
liveness          0.177125
valence           0.467942
tempo            122.108850
duration_ms      205645.700000
time_signature    3.950000
dtype: float64
danceability      0.750000
energy            0.659000
loudness         -5.930000

```

```

speechiness      0.105500
acousticness     0.121500
instrumentalness  0.000003
liveness         0.112000
valence          0.437500
tempo            122.531500
duration_ms      212904.500000
time_signature   4.000000
dtype: float64
danceability     0.151423
energy           0.137874
loudness         1.611051
speechiness      0.094732
acousticness     0.209052
instrumentalness  0.021231
liveness         0.130476
valence          0.228504
tempo            31.178064
duration_ms      34324.034325
time_signature   0.220721
dtype: float64
danceability     0.922
energy           0.909
loudness         -3.093
speechiness      0.516
acousticness     0.847
instrumentalness  0.134
liveness         0.552
valence          0.931
tempo            191.702
duration_ms      268867.000
time_signature   4.000
dtype: float64

```

	danceability	energy	loudness	speechiness	acousticness	\
0.00	0.2580	0.39100	-9.211	0.02320	0.000282	
0.25	0.6805	0.56425	-7.077	0.04955	0.030800	
0.50	0.7500	0.65900	-5.930	0.10550	0.121500	
0.75	0.8175	0.77250	-4.624	0.14450	0.227500	
1.00	0.9220	0.90900	-3.093	0.51600	0.847000	

	instrumentalness	liveness	valence	tempo	duration_ms	\
0.00	0.000000	0.021500	0.09670	77.16900	95467.00	
0.25	0.000000	0.095275	0.28750	95.03750	189856.00	
0.50	0.000003	0.112000	0.43750	122.53150	212904.50	
0.75	0.000043	0.294750	0.63325	140.58575	226983.25	
1.00	0.134000	0.552000	0.93100	191.70200	268867.00	

time_signature

0.00	3.0
0.25	4.0
0.50	4.0
0.75	4.0
1.00	4.0

```
[ ]: # Variables elegidas: energy y duration_ms #####

#Primera variables a interpretar Energy
#count: la cantidad de datos para la variable energy (que mide la intensidad y
↳actividad) es de 40 registros
#mean: el valor de la intensidad y actividad promedio de las 40 canciones mas
↳reproducidas en Spotify
# en el año 2018 es de 0.661925
#std: con respecto a la desviación estándar esta es de 0.137874, lo que indica
↳que la varianza es menor respecto a
# la media (los datos tienen menos variación)
#min: el valor más bajo registrado en cuanto a intensidad y actividad es de 0.
↳391000
#primer cuartil: el 25% de los datos tienen valores inferiores o iguales a 0.
↳564250
#Segundo cuartil: la mitad de los datos son inferiores o iguales a 0.65900
#Tercer cuartil: el 75% de los datos son iguales o menores a 0.77250
#max: el valor más alto registrado de la totalidad de los datos es de 0.909000

#Segunda variable a time_signature
#count: la cantidad de datos para la variable time_signature (beats por cada
↳barra o medida) es de 40 registros
#mean: la cantidad de beats por barra o medida en promedio de las 40 canciones
↳mas reproducidas en Spotify
# en el año 2018 es de 3.950000
#std: con respecto a la desviación estándar esta es de 0.220721, lo que indica
↳que la varianza es menor respecto
# a la media (los datos tienen menos variación)
#min: el dato más bajo registrado en cuanto a la cantidad de beats por barra o
↳medida es de 3
#primer cuartil: el 25% de los datos tienen valores inferiores o iguales a 4
#Segundo cuartil: el 50% de los datos tienen valores inferiores o iguales a 4
#Tercer cuartil: el 75% de los datos tienen valores inferiores o iguales a 4
#max: el dato más alto registrado de la totalidad fue de 4
```

b) Realice el test de normalidad para una variable e interprete el resultado

```
[4]: import scipy.stats
X = datos['loudness'] #En realidad la variable que estoy trayendo es loudness
↳pero
```

```

#en el archivo CSV, las columnas estan mal nombradas y me hala la de la par
print(X)
shapiro_resultados = scipy.stats.shapiro(X)
print(shapiro_resultados)
p_value = shapiro_resultados[1]
print(p_value)
# interpretación
alpha = 0.05
if p_value > alpha:
    print('Sí sigue la curva Normal (No se rechaza H0)')
else:
    print('No sigue la curva Normal (Se rechaza H0)')

# En este caso pvalue (0.53) > 0.05 por lo que, si sigue la curva normal y no
↪ se rechaza H0.

```

Dura	-3.456
Plug_Walk	-6.531
Silence	-3.093
Solo_(feat._Demi_Lovato)	-4.546
Sin_Pijama	-3.695
Call_Out_My_Name	-4.929
Taki_Taki_(with_Selena_Gomez_Ozuna_&_Cardi_B)	-4.206
River_(feat._Ed_Sheeran)	-5.916
Rewrite_The_Stars	-7.005
Him_&_I_(with_Halsey)	-6.343
Look_Alive_(feat._Drake)	-7.495
Rise	-4.650
Perfect	-6.312
Thunder	-4.833
Psycho_(feat._Ty_Dolla_\$ign)	-8.011
no_tears_left_to_cry	-5.507
Everybody_Dies_In_Their_Nightmares	-7.066
Jocelyn_Flores	-9.144
HUMBLE.	-6.638
In_My_Mind	-5.335
Nice_For_What	-6.474
Havana	-4.333
Dusk_Till_Dawn_-_Radio_Edit	-6.593
Candy_Paint	-5.944
This_Is_Me	-7.276
Siguelo_Bailando	-7.110
I_Like_Me_Better	-7.621
Believer	-4.374
Downtown	-4.985
Shape_of_You	-3.183
Gods_Plan	-9.211

Stir_Fry	-5.474
One_Kiss_(with_Dua_Lipa)	-3.240
Finesse_(Remix)_[feat._Cardi_B]	-4.877
Nonstop	-8.074
Mine	-3.874
In_My_Feelings	-5.833
Taste_(feat._Offset)	-7.442
Lucid_Dreams	-7.230
Promises_(with_Sam_Smith)	-5.991

Name: loudness, dtype: float64
(0.9757644534111023, 0.535995364189148)
0.535995364189148
Sí sigue la curva Normal (No se rechaza H0)

4 C) Realice un grafico de dispersion e interprete dos similitudes en el grafico.

```
[6]: #Definicion de grafico
import matplotlib.pyplot as plt
grafico = pd.read_csv('SpotifyTop2018_40_V2.csv',delimiter=',',decimal=".",
    ↳nrows=2, skiprows=[-1])
print(grafico)
print(grafico.head())

x = grafico['tempo']
print(x)
y = grafico['duration_ms']
print(y)

plt.xlabel('tempo')
plt.ylabel('duration_ms')
plt.scatter(x,y, color='red')
```

	danceability	energy	loudness	speechiness	acousticness	\
Dura	0.791	0.848	-3.456	0.0506	0.183	
Plug_Walk	0.876	0.519	-6.531	0.1430	0.202	

	instrumentalness	liveness	valence	tempo	duration_ms	\
Dura	0	0.409	0.828	95.000	200480	
Plug_Walk	0	0.108	0.158	94.981	175230	

	time_signature
Dura	4
Plug_Walk	4

	danceability	energy	loudness	speechiness	acousticness	\
Dura	0.791	0.848	-3.456	0.0506	0.183	

Plug_Walk	0.876	0.519	-6.531	0.1430	0.202
-----------	-------	-------	--------	--------	-------

	instrumentalness	liveness	valence	tempo	duration_ms	\
Dura	0	0.409	0.828	95.000	200480	
Plug_Walk	0	0.108	0.158	94.981	175230	

	time_signature
Dura	4
Plug_Walk	4

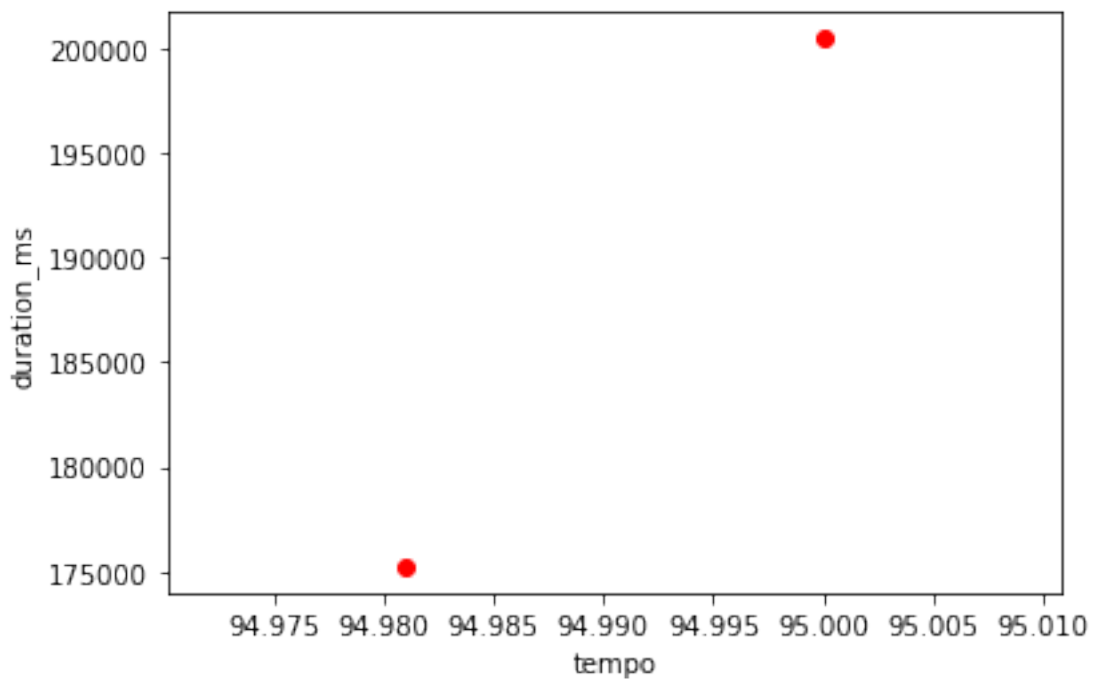
Dura	95.000
Plug_Walk	94.981

Name: tempo, dtype: float64

Dura	200480
Plug_Walk	175230

Name: duration_ms, dtype: int64

[6]: <matplotlib.collections.PathCollection at 0x128fea590>



[10]: ******Interpretacion******
#Se puede ver para estas dos canciones especificos que fueron:
#1. Dura y Plug Walk y tomando como base las variables: tempo y duration_ms
Conforme la duracion en milisegundos (tempo) aumenta se obtienen mas beats
→ por minuto en una pista (duration_ms)

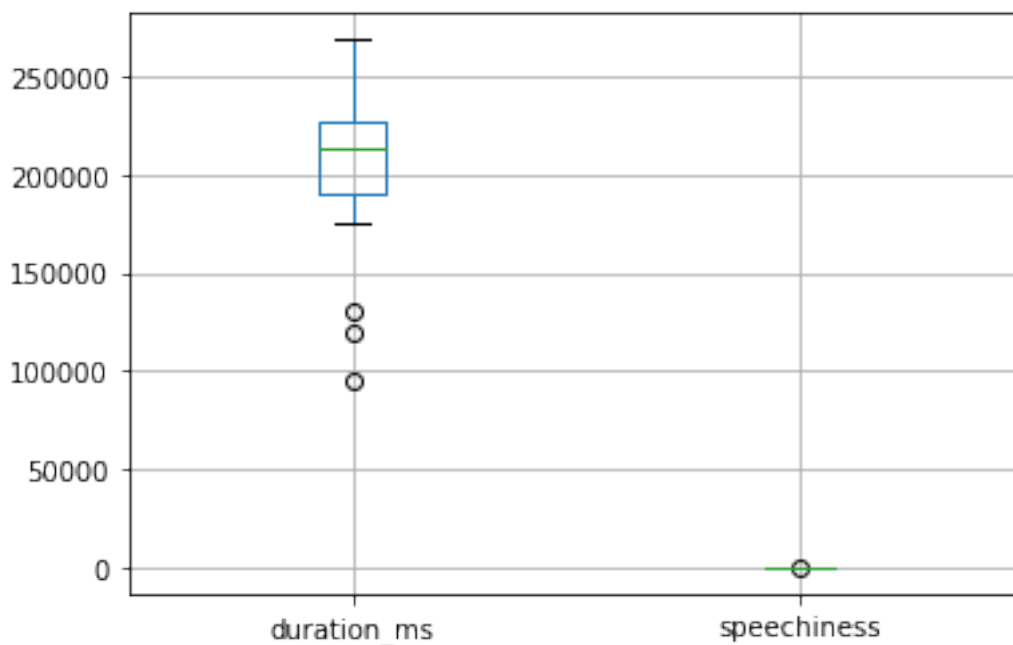

```
# Mientras que si se tienen menos milisegundos, los beats por minuto en una
→ pista disminuyen.
# IMPORTANTE: se trabaja bajo este supuesto usando estos dos datos y la
→ variacion es minima, pero
# se puede decir que hay una correlacion positiva, ya que si duration_ms
→ aumenta,
# tambien aumenta tempo y viceversa, si duration disminuye, tempo tambien lo
→ hara.
```

5 Para dos variables identifique los datos atipicos, si los hay.

```
[11]: datos.head()
boxplots = datos[['duration_ms', 'speechiness']].boxplot(return_type='axes')

# Primera variable: duration_ms, esta nos indica que los rangos estan
→ aproximadamente entre 180000 (es un aproximado) y poco
# mas de 250000 habiendo tres datos atipicos, uno cercano a 100000, y los dos
→ restantes
# cerca de 140000 aproximadamente.

# En cuanto a speechiness los datos no reflejan un rango, ya que estan muy
→ similares (condensados)
# y solo hay un dato atipico que esta apenas por encima de cero.
```



6 E) Calcule la matriz de correlaciones

```
[12]: import seaborn as sns
import matplotlib.pyplot as plt

corr = datos.corr()
print(corr)
f, ax = plt.subplots(figsize=(10, 8))
sns.heatmap(corr, mask=np.zeros_like(corr, dtype=np.bool), cmap=sns.
    ↳diverging_palette(220, 10, as_cmap=True),
    square=True, ax=ax)

# Interpretacion:
# Primera interpretacion: loudness y energy tienen una correlacion positiva,
    ↳porque estan en
# color rosa fuerte, est indica que estan cerca de 0.75, cuando esto pasa,
    ↳indica
# que si loudness crece, energy tambien va a crecer y al contrario, cuando
    ↳loudness baje, energy tambien decrecera
# Segunda interpretacion: valence y duration_ds tienen una correlacion negativa
    ↳casi perfecta
# ya que el cuadro esta de color azul y se ve que es un azul profundo, llegando
    ↳casi a -1
# esto indica que cuando valence aumentan duration_ds disminuye y viceversa.
```

/Users/heinerleivagmail.com/opt/anaconda3/lib/python3.7/site-packages/statsmodels/tools/_testing.py:19: FutureWarning: pandas.util.testing is deprecated. Use the functions in the public API at pandas.testing instead.

```
import pandas.util.testing as tm
```

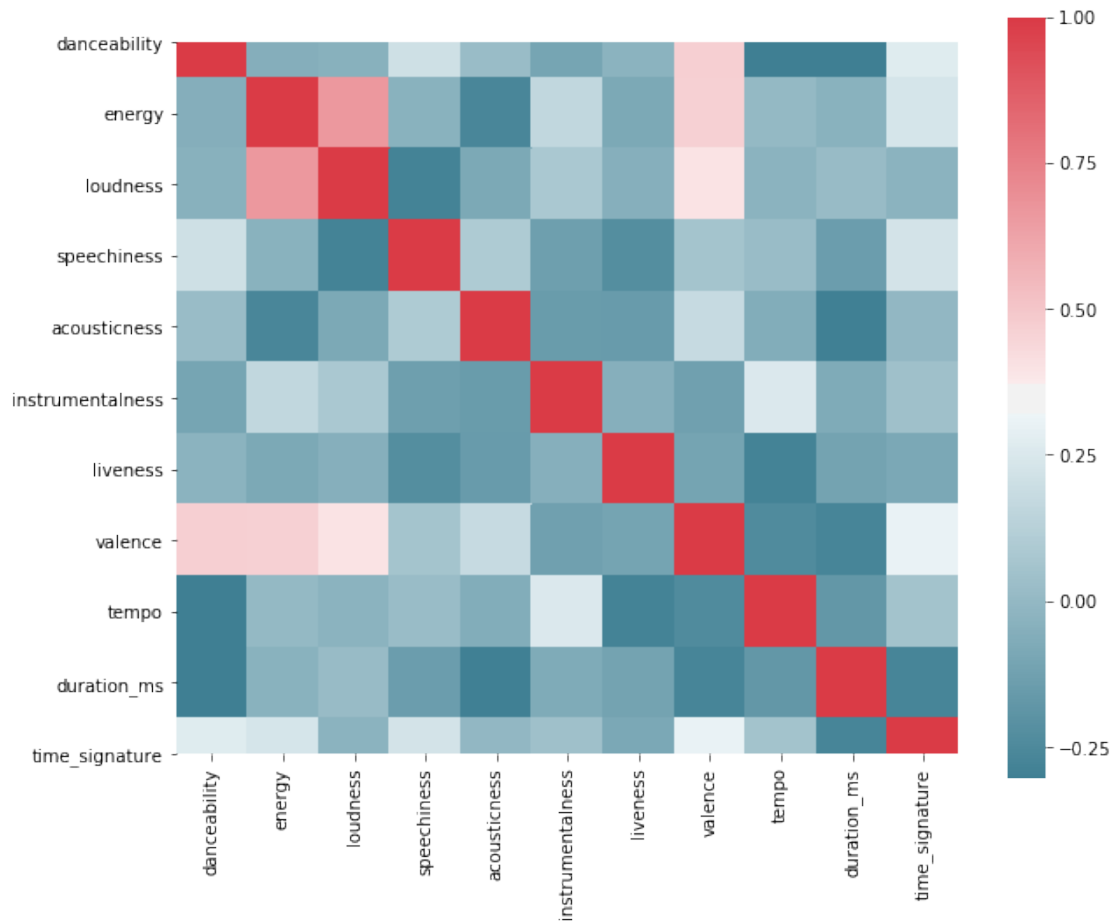
	danceability	energy	loudness	speechiness	acousticness	\
danceability	1.000000	-0.052973	-0.043320	0.207397	0.016291	
energy	-0.052973	1.000000	0.665634	-0.035225	-0.269420	
loudness	-0.043320	0.665634	1.000000	-0.281621	-0.081194	
speechiness	0.207397	-0.035225	-0.281621	1.000000	0.097140	
acousticness	0.016291	-0.269420	-0.081194	0.097140	1.000000	
instrumentalness	-0.102198	0.159487	0.079780	-0.131971	-0.150742	
liveness	-0.031392	-0.082066	-0.050145	-0.225592	-0.155930	
valence	0.470648	0.465742	0.402273	0.059612	0.174967	
tempo	-0.301067	0.005003	-0.034438	0.018453	-0.064332	
duration_ms	-0.305303	-0.035215	0.011911	-0.144740	-0.297226	
time_signature	0.268821	0.234110	-0.032557	0.224884	-0.006534	

	instrumentalness	liveness	valence	tempo	duration_ms	\
danceability	-0.102198	-0.031392	0.470648	-0.301067	-0.305303	
energy	0.159487	-0.082066	0.465742	0.005003	-0.035215	
loudness	0.079780	-0.050145	0.402273	-0.034438	0.011911	

speechiness	-0.131971	-0.225592	0.059612	0.018453	-0.144740
acousticness	-0.150742	-0.155930	0.174967	-0.064332	-0.297226
instrumentalness	1.000000	-0.046094	-0.130534	0.254822	-0.072171
liveness	-0.046094	1.000000	-0.108392	-0.281270	-0.115850
valence	-0.130534	-0.108392	1.000000	-0.240702	-0.273304
tempo	0.254822	-0.281270	-0.240702	1.000000	-0.174676
duration_ms	-0.072171	-0.115850	-0.273304	-0.174676	1.000000
time_signature	0.040438	-0.090593	0.302942	0.056347	-0.272389

	time_signature
danceability	0.268821
energy	0.234110
loudness	-0.032557
speechiness	0.224884
acousticness	-0.006534
instrumentalness	0.040438
liveness	-0.090593
valence	0.302942
tempo	0.056347
duration_ms	-0.272389
time_signature	1.000000

[12]: <matplotlib.axes._subplots.AxesSubplot at 0x129bdcd10>



7 Efectue un ACP y de una interpretaci3n siguiendo los siguientes pasos:

```
[13]: import matplotlib.pyplot as plt
from prince import PCA

class ACP:
    def __init__(self, datos, n_componentes = 5):
        self.__datos = datos
        self.__modelo = PCA(n_components = n_componentes).fit(self.__datos)
        self.__correlacion_var = self.__modelo.column_correlations(datos)
        self.__coordenadas_ind = self.__modelo.row_coordinates(datos)
        self.__contribucion_ind = self.__modelo.row_contributions(datos)
        self.__cos2_ind = self.__modelo.row_cosine_similarities(datos)
        self.__var_explicada = [x * 100 for x in self.__modelo.
→ explained_inertia_]
        @property
```

```

def datos(self):
    return self.__datos
@datos.setter
def datos(self, datos):
    self.__datos = datos
@property
def modelo(self):
    return self.__modelo
@property
def correlacion_var(self):
    return self.__correlacion_var
@property
def coordenadas_ind(self):
    return self.__coordenadas_ind
@property
def contribucion_ind(self):
    return self.__contribucion_ind
@property
def cos2_ind(self):
    return self.__cos2_ind
@property
def var_explicada(self):
    return self.__var_explicada
    self.__var_explicada = var_explicada
def plot_plano_principal(self, ejes = [0, 1], ind_labels = True, titulo =
↳ 'Plano Principal'):
    x = self.coordenadas_ind[ejes[0]].values
    y = self.coordenadas_ind[ejes[1]].values
    plt.style.use('seaborn-whitegrid')
    plt.scatter(x, y, color = 'gray')
    plt.title(titulo)
    plt.axhline(y = 0, color = 'dimgrey', linestyle = '--')
    plt.axvline(x = 0, color = 'dimgrey', linestyle = '--')
    inercia_x = round(self.var_explicada[ejes[0]], 2)
    inercia_y = round(self.var_explicada[ejes[1]], 2)
    plt.xlabel('Componente ' + str(ejes[0]) + ' (' + str(inercia_x) + '%)')
    plt.ylabel('Componente ' + str(ejes[1]) + ' (' + str(inercia_y) + '%)')
    if ind_labels:
        for i, txt in enumerate(self.coordenadas_ind.index):
            plt.annotate(txt, (x[i], y[i]))
def plot_circulo(self, ejes = [0, 1], var_labels = True, titulo = 'Círculo
↳ de Correlación'):
    cor = self.correlacion_var.iloc[:, ejes].values
    plt.style.use('seaborn-whitegrid')
    c = plt.Circle((0, 0), radius = 1, color = 'steelblue', fill = False)
    plt.gca().add_patch(c)
    plt.axis('scaled')

```

```

plt.title(titulo)
plt.axhline(y = 0, color = 'dimgrey', linestyle = '--')
plt.axvline(x = 0, color = 'dimgrey', linestyle = '--')
inercia_x = round(self.var_explicada[ejes[0]], 2)
inercia_y = round(self.var_explicada[ejes[1]], 2)
plt.xlabel('Componente ' + str(ejes[0]) + ' (' + str(inercia_x) + '%)')
plt.ylabel('Componente ' + str(ejes[1]) + ' (' + str(inercia_y) + '%)')
for i in range(cor.shape[0]):
    plt.arrow(0, 0, cor[i, 0] * 0.95, cor[i, 1] * 0.95, color =
↳ 'steelblue',
                alpha = 0.5, head_width = 0.05, head_length = 0.05)
    if var_labels:
        plt.text(cor[i, 0] * 1.05, cor[i, 1] * 1.05, self.
↳ correlacion_var.index[i],
                color = 'steelblue', ha = 'center', va = 'center')
def plot_sobreposicion(self, ejes = [0, 1], ind_labels = True,
    var_labels = True, titulo = 'Sobreposición
↳ Plano-Círculo'):
    x = self.coordenadas_ind[ejes[0]].values
    y = self.coordenadas_ind[ejes[1]].values
    cor = self.correlacion_var.iloc[:, ejes]
    scale = min((max(x) - min(x)/(max(cor[ejes[0]]) - min(cor[ejes[0]]))),
                (max(y) - min(y)/(max(cor[ejes[1]]) - min(cor[ejes[1]]))))
↳ * 0.7
    cor = self.correlacion_var.iloc[:, ejes].values
    plt.style.use('seaborn-whitegrid')
    plt.axhline(y = 0, color = 'dimgrey', linestyle = '--')
    plt.axvline(x = 0, color = 'dimgrey', linestyle = '--')
    inercia_x = round(self.var_explicada[ejes[0]], 2)
    inercia_y = round(self.var_explicada[ejes[1]], 2)
    plt.xlabel('Componente ' + str(ejes[0]) + ' (' + str(inercia_x) + '%)')
    plt.ylabel('Componente ' + str(ejes[1]) + ' (' + str(inercia_y) + '%)')
    plt.scatter(x, y, color = 'gray')
    if ind_labels:
        for i, txt in enumerate(self.coordenadas_ind.index):
            plt.annotate(txt, (x[i], y[i]))
    for i in range(cor.shape[0]):
        plt.arrow(0, 0, cor[i, 0] * scale, cor[i, 1] * scale, color =
↳ 'steelblue',
                alpha = 0.5, head_width = 0.05, head_length = 0.05)
    if var_labels:
        plt.text(cor[i, 0] * scale * 1.15, cor[i, 1] * scale * 1.15,
                self.correlacion_var.index[i],
                color = 'steelblue', ha = 'center', va = 'center')

```

```
[14]: # Modelado
print(datos)
print(datos.shape)

# Declara la instancia de clase
acp = ACP(datos,n_componentes=3)
# Despliega las Componentes Principales
print(acp.coordenadas_ind)
# Despliega los cosenos cuadrados de los individuos
print(acp.cos2_ind)
# Despliega las correlaciones de las variables con respecto a las componentes
print(acp.correlacion_var)

# Plotea el plano principal
acp.plot_plano_principal()
```

	danceability	energy	loudness	\
Dura	0.791	0.848	-3.456	
Plug_Walk	0.876	0.519	-6.531	
Silence	0.520	0.761	-3.093	
Solo_(feat._Demi_Lovato)	0.737	0.636	-4.546	
Sin_Pijama	0.791	0.745	-3.695	
Call_Out_My_Name	0.489	0.598	-4.929	
Taki_Taki_(with_Selena_Gomez_Ozuna_&_Cardi_B)	0.841	0.798	-4.206	
River_(feat._Ed_Sheeran)	0.748	0.749	-5.916	
Rewrite_The_Stars	0.684	0.619	-7.005	
Him_&_I_(with_Halsey)	0.589	0.731	-6.343	
Look_Alive_(feat._Drake)	0.922	0.581	-7.495	
Rise	0.687	0.785	-4.650	
Perfect	0.599	0.448	-6.312	
Thunder	0.605	0.822	-4.833	
Psycho_(feat._Ty_Dolla_\$ign)	0.739	0.559	-8.011	
no_tears_left_to_cry	0.699	0.713	-5.507	
Everybody_Dies_In_Their_Nightmares	0.734	0.570	-7.066	
Jocelyn_Flores	0.872	0.391	-9.144	
HUMBLE.	0.908	0.621	-6.638	
In_My_Mind	0.694	0.770	-5.335	
Nice_For_What	0.586	0.909	-6.474	
Havana	0.765	0.523	-4.333	
Dusk_Till_Dawn_-_Radio_Edit	0.258	0.437	-6.593	
Candy_Paint	0.670	0.654	-5.944	
This_Is_Me	0.284	0.704	-7.276	
Siguelo_Bailando	0.855	0.664	-7.110	
I_Like_Me_Better	0.752	0.505	-7.621	
Believer	0.776	0.780	-4.374	
Downtown	0.775	0.679	-4.985	
Shape_of_You	0.825	0.652	-3.183	

Gods_Plan	0.754	0.449	-9.211
Stir_Fry	0.815	0.816	-5.474
One_Kiss_(with_Dua_Lipa)	0.791	0.862	-3.240
Finesse_(Remix)_[feat._Cardi_B]	0.704	0.859	-4.877
Nonstop	0.912	0.412	-8.074
Mine	0.710	0.789	-3.874
In_My_Feelings	0.835	0.626	-5.833
Taste_(feat._Offset)	0.884	0.559	-7.442
Lucid_Dreams	0.511	0.566	-7.230
Promises_(with_Sam_Smith)	0.781	0.768	-5.991

	speechiness	acousticness	\
Dura	0.0506	0.183000	
Plug_Walk	0.1430	0.202000	
Silence	0.0853	0.256000	
Solo_(feat._Demi_Lovato)	0.0437	0.044100	
Sin_Pijama	0.0464	0.354000	
Call_Out_My_Name	0.0360	0.218000	
Taki_Taki_(with_Selena_Gomez_Ozuna_&_Cardi_B)	0.2290	0.153000	
River_(feat._Ed_Sheeran)	0.5160	0.142000	
Rewrite_The_Stars	0.0386	0.071600	
Him_&_I_(with_Halsey)	0.0868	0.053400	
Look_Alive_(feat._Drake)	0.2700	0.001040	
Rise	0.0333	0.327000	
Perfect	0.0232	0.163000	
Thunder	0.0438	0.006710	
Psycho_(feat._Ty_Dolla_\$ign)	0.1170	0.580000	
no_tears_left_to_cry	0.0594	0.040000	
Everybody_Dies_In_Their_Nightmares	0.1330	0.847000	
Jocelyn_Flores	0.2420	0.469000	
HUMBLE.	0.1020	0.000282	
In_My_Mind	0.1490	0.176000	
Nice_For_What	0.0705	0.089100	
Havana	0.0300	0.184000	
Dusk_Till_Dawn_-_Radio_Edit	0.0390	0.101000	
Candy_Paint	0.1530	0.627000	
This_Is_Me	0.1860	0.005830	
Siguelo_Bailando	0.0607	0.165000	
I_Like_Me_Better	0.2530	0.535000	
Believer	0.1280	0.062200	
Downtown	0.1350	0.180000	
Shape_of_You	0.0802	0.581000	
Gods_Plan	0.1090	0.033200	
Stir_Fry	0.2690	0.002990	
One_Kiss_(with_Dua_Lipa)	0.1100	0.037000	
Finesse_(Remix)_[feat._Cardi_B]	0.0996	0.018500	
Nonstop	0.1240	0.016400	
Mine	0.0722	0.016100	

In_My_Feelings	0.1250	0.058900
Taste_(feat._Offset)	0.1200	0.023600
Lucid_Dreams	0.2000	0.349000
Promises_(with_Sam_Smith)	0.0394	0.011900

	instrumentalness	liveness \
Dura	0.000000	0.4090
Plug_Walk	0.000000	0.1080
Silence	0.000005	0.1700
Solo_(feat._Demi_Lovato)	0.000067	0.3500
Sin_Pijama	0.000029	0.1040
Call_Out_My_Name	0.000000	0.3500
Taki_Taki_(with_Selena_Gomez_Ozuna_&_Cardi_B)	0.000003	0.0618
River_(feat._Ed_Sheeran)	0.000000	0.0713
Rewrite_The_Stars	0.000000	0.1220
Him_&_I_(with_Halsey)	0.000000	0.3080
Look_Alive_(feat._Drake)	0.000059	0.1050
Rise	0.000000	0.2030
Perfect	0.000000	0.1060
Thunder	0.134000	0.1470
Psycho_(feat._Ty_Dolla_\$ign)	0.000000	0.1120
no_tears_left_to_cry	0.000003	0.2940
Everybody_Dies_In_Their_Nightmares	0.000021	0.1120
Jocelyn_Flores	0.000004	0.2970
HUMBLE.	0.000054	0.0958
In_My_Mind	0.000011	0.1180
Nice_For_What	0.000109	0.1190
Havana	0.000036	0.1320
Dusk_Till_Dawn_-_Radio_Edit	0.000001	0.1060
Candy_Paint	0.000001	0.0710
This_Is_Me	0.000115	0.0424
Siguelo_Bailando	0.000040	0.0937
I_Like_Me_Better	0.000003	0.1040
Believer	0.000000	0.0810
Downtown	0.000073	0.0680
Shape_of_You	0.000000	0.0931
Gods_Plan	0.000083	0.5520
Stir_Fry	0.000000	0.1590
One_Kiss_(with_Dua_Lipa)	0.000022	0.0814
Finesse_(Remix)_[feat._Cardi_B]	0.000000	0.0215
Nonstop	0.013000	0.1040
Mine	0.000003	0.4510
In_My_Feelings	0.000060	0.3960
Taste_(feat._Offset)	0.000000	0.1010
Lucid_Dreams	0.000000	0.3400
Promises_(with_Sam_Smith)	0.000005	0.3250

valence	tempo	duration_ms \
---------	-------	---------------

Dura	0.8280	95.000	200480
Plug_Walk	0.1580	94.981	175230
Silence	0.2860	141.971	180823
Solo_(feat._Demi_Lovato)	0.5650	105.005	222653
Sin_Pijama	0.8200	94.014	188560
Call_Out_My_Name	0.1720	134.045	228373
Taki_Taki_(with_Selena_Gomez_Ozuna_&_Cardi_B)	0.5910	95.948	212500
River_(feat._Ed_Sheeran)	0.6590	90.090	221013
Rewrite_The_Stars	0.2840	125.046	217440
Him_&_I_(with_Halsey)	0.1910	87.908	268867
Look_Alive_(feat._Drake)	0.5950	140.022	181263
Rise	0.6550	106.046	194408
Perfect	0.1680	95.050	263400
Thunder	0.2880	167.997	187147
Psycho_(feat._Ty_Dolla_\$ign)	0.4390	140.124	221440
no_tears_left_to_cry	0.3540	121.993	205920
Everybody_Dies_In_Their_Nightmares	0.6890	129.953	95467
Jocelyn_Flores	0.4370	134.021	119133
HUMBLE.	0.4210	150.011	177000
In_My_Mind	0.1630	125.905	184560
Nice_For_What	0.7570	93.394	210747
Havana	0.3940	104.988	217307
Dusk_Till_Dawn_-_Radio_Edit	0.0967	180.043	239000
Candy_Paint	0.4380	180.024	227533
This_Is_Me	0.1000	191.702	234707
Siguelo_Bailando	0.6260	98.015	226800
I_Like_Me_Better	0.4190	91.970	197437
Believer	0.6660	124.949	204347
Downtown	0.6190	166.008	193456
Shape_of_You	0.9310	95.977	233713
Gods_Plan	0.3570	77.169	198973
Stir_Fry	0.4980	181.967	190288
One_Kiss_(with_Dua_Lipa)	0.5920	123.994	214847
Finesse_(Remix)_[feat._Cardi_B]	0.9260	105.115	217289
Nonstop	0.4220	154.983	238614
Mine	0.7170	142.929	131064
In_My_Feelings	0.3500	91.030	217925
Taste_(feat._Offset)	0.3420	97.994	232959
Lucid_Dreams	0.2180	83.903	239836
Promises_(with_Sam_Smith)	0.4860	123.070	213309

time_signature

Dura	4
Plug_Walk	4
Silence	4
Solo_(feat._Demi_Lovato)	4
Sin_Pijama	4
Call_Out_My_Name	3

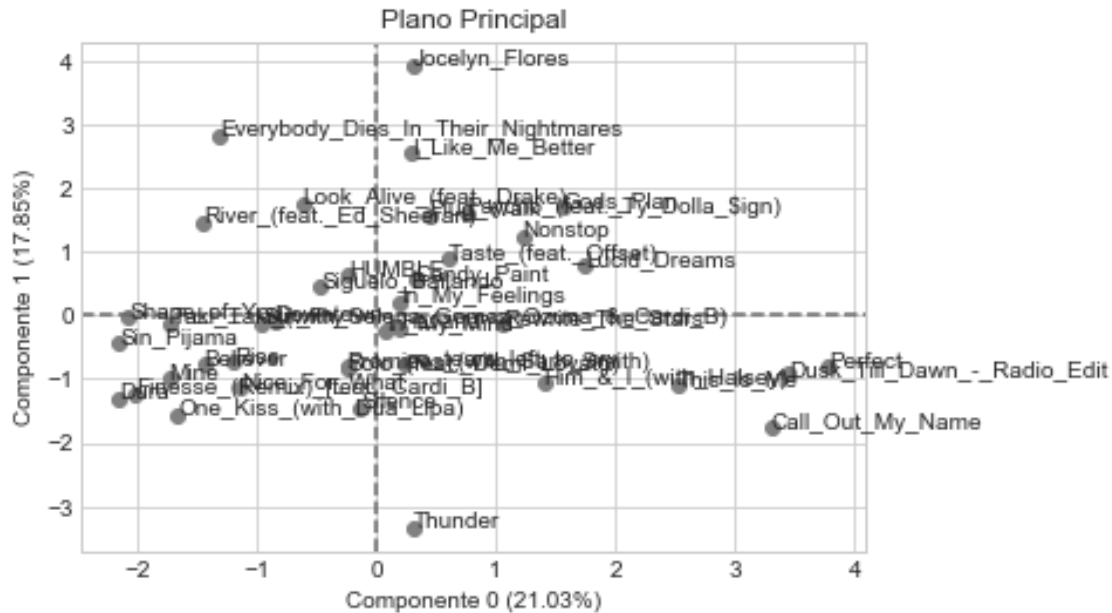
Taki_Taki_(with_Selena_Gomez_Ozuna_&_Cardi_B)	4
River_(feat._Ed_Sheeran)	4
Rewrite_The_Stars	4
Him_&_I_(with_Halsey)	4
Look_Alive_(feat._Drake)	4
Rise	4
Perfect	3
Thunder	4
Psycho_(feat._Ty_Dolla_\$ign)	4
no_tears_left_to_cry	4
Everybody_Dies_In_Their_Nightmares	4
Jocelyn_Flores	4
HUMBLE.	4
In_My_Mind	4
Nice_For_What	4
Havana	4
Dusk_Till_Dawn_-_Radio_Edit	4
Candy_Paint	4
This_Is_Me	4
Siguelo_Bailando	4
I_Like_Me_Better	4
Believer	4
Downtown	4
Shape_of_You	4
Gods_Plan	4
Stir_Fry	4
One_Kiss_(with_Dua_Lipa)	4
Finesse_(Remix)_[feat._Cardi_B]	4
Nonstop	4
Mine	4
In_My_Feelings	4
Taste_(feat._Offset)	4
Lucid_Dreams	4
Promises_(with_Sam_Smith)	4
(40, 11)	

	0	1	2
Dura	-2.158016	-1.298592	-1.826558
Plug_Walk	0.440042	1.566524	-0.112412
Silence	-0.129625	-1.455367	0.746510
Solo_(feat._Demi_Lovato)	-0.238512	-0.854628	-1.454895
Sin_Pijama	-2.146927	-0.437891	-0.756985
Call_Out_My_Name	3.305340	-1.771822	-1.588707
Taki_Taki_(with_Selena_Gomez_Ozuna_&_Cardi_B)	-1.719642	-0.137449	-0.158241
River_(feat._Ed_Sheeran)	-1.442549	1.459012	0.507475
Rewrite_The_Stars	1.055307	-0.131063	0.108222
Him_&_I_(with_Halsey)	1.400136	-1.070508	-1.347914
Look_Alive_(feat._Drake)	-0.614537	1.758094	0.892328
Rise	-1.193979	-0.735019	-0.634045

Perfect	3.783530	-0.817431	-1.993337
Thunder	0.308155	-3.343676	3.672982
Psycho_(feat._Ty_Dolla_\$ign)	0.753171	1.609924	0.601211
no_tears_left_to_cry	0.233400	-0.776135	-0.496387
Everybody_Dies_In_Their_Nightmares	-1.317104	2.825336	1.146942
Jocelyn_Flores	0.308408	3.911444	0.678860
HUMBLE.	-0.237480	0.627157	0.812685
In_My_Mind	0.076500	-0.250685	0.721275
Nice_For_What	-1.137052	-1.118727	-0.451627
Havana	0.194813	-0.215735	-0.701307
Dusk_Till_Dawn_-_Radio_Edit	3.443868	-0.925762	1.629353
Candy_Paint	0.329813	0.549212	1.656991
This_Is_Me	2.522739	-1.102386	2.644651
Siguelo_Bailando	-0.462747	0.437508	-0.707156
I_Like_Me_Better	0.288081	2.543927	0.069641
Believer	-1.437497	-0.780162	0.190421
Downtown	-0.851143	-0.095884	1.221652
Shape_of_You	-2.071086	-0.033630	-1.037857
Gods_Plan	1.559168	1.708317	-2.138288
Stir_Fry	-0.956002	-0.128360	1.720920
One_Kiss_(with_Dua_Lipa)	-1.662786	-1.568682	0.055450
Finesse_(Remix)_[feat._Cardi_B]	-2.021838	-1.235200	-0.198275
Nonstop	1.226078	1.229698	0.728505
Mine	-1.732012	-1.000946	-0.315014
In_My_Feelings	0.198550	0.211286	-1.574303
Taste_(feat._Offset)	0.607186	0.888489	-0.551418
Lucid_Dreams	1.741865	0.777630	-0.967698
Promises_(with_Sam_Smith)	-0.245616	-0.817816	-0.793645
	0	1	2
Dura	0.481114	0.174214	0.344672
Plug_Walk	0.072789	0.922461	0.004750
Silence	0.006241	0.786760	0.206999
Solo_(feat._Demi_Lovato)	0.019590	0.251512	0.728899
Sin_Pijama	0.857692	0.035680	0.106628
Call_Out_My_Name	0.658601	0.189247	0.152152
Taki_Taki_(with_Selena_Gomez_Ozuna_&_Cardi_B)	0.985361	0.006295	0.008344
River_(feat._Ed_Sheeran)	0.465829	0.476522	0.057649
Rewrite_The_Stars	0.974715	0.015034	0.010251
Him_&_I_(with_Halsey)	0.398189	0.232771	0.369040
Look_Alive_(feat._Drake)	0.088552	0.724745	0.186703
Rise	0.602058	0.228162	0.169780
Perfect	0.755148	0.035248	0.209604
Thunder	0.003834	0.451434	0.544732
Psycho_(feat._Ty_Dolla_\$ign)	0.161129	0.736202	0.102669
no_tears_left_to_cry	0.060310	0.666900	0.272790
Everybody_Dies_In_Their_Nightmares	0.157237	0.723529	0.119234
Jocelyn_Flores	0.005999	0.964935	0.029066
HUMBLE.	0.050800	0.354290	0.594910

In_My_Mind	0.009937	0.106707	0.883356
Nice_For_What	0.470414	0.455373	0.074213
Havana	0.065852	0.080756	0.853393
Dusk_Till_Dawn_-_Radio_Edit	0.771545	0.055753	0.172702
Candy_Paint	0.034466	0.095574	0.869960
This_Is_Me	0.436693	0.083387	0.479920
Siguelo_Bailando	0.236451	0.211362	0.552186
I_Like_Me_Better	0.012652	0.986609	0.000739
Believer	0.762140	0.224486	0.013374
Downtown	0.325436	0.004130	0.670434
Shape_of_You	0.799116	0.000211	0.200673
Gods_Plan	0.245021	0.294140	0.460839
Stir_Fry	0.234826	0.004233	0.760940
One_Kiss_(with_Dua_Lipa)	0.528785	0.470627	0.000588
Finesse_(Remix)__[feat._Cardi_B]	0.723143	0.269902	0.006955
Nonstop	0.423916	0.426423	0.149661
Mine	0.731498	0.244305	0.024197
In_My_Feelings	0.015384	0.017421	0.967194
Taste_(feat._Offset)	0.252146	0.539899	0.207956
Lucid_Dreams	0.663155	0.132170	0.204676
Promises_(with_Sam_Smith)	0.044390	0.492135	0.463475

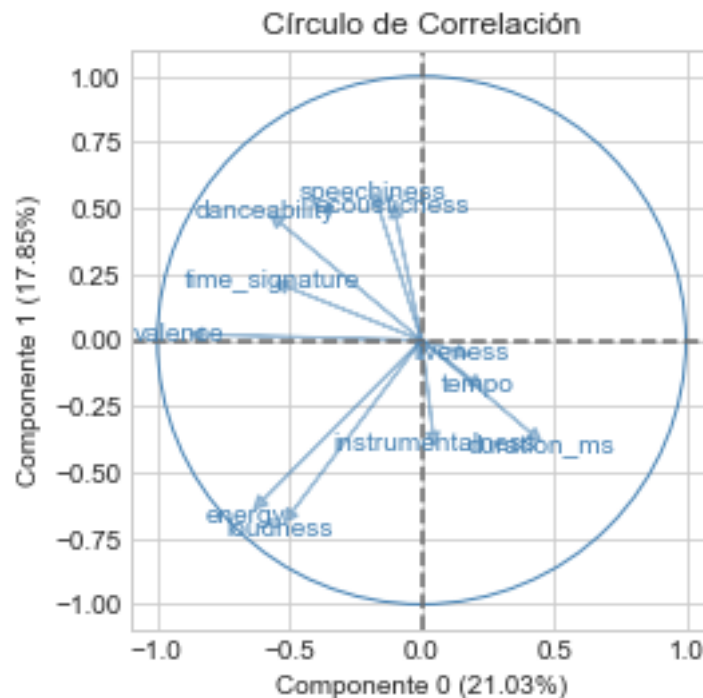
	0	1	2
acousticness	-0.102040	0.489755	0.045054
danceability	-0.561408	0.466725	-0.201905
duration_ms	0.435534	-0.369733	-0.287193
energy	-0.630813	-0.636606	0.093463
instrumentalness	0.044901	-0.367711	0.479720
liveness	0.140395	-0.044909	-0.542022
loudness	-0.514836	-0.683866	-0.109206
speechiness	-0.173728	0.537291	0.305905
tempo	0.199634	-0.157487	0.810632
time_signature	-0.534599	0.211960	0.329464
valence	-0.870490	0.022651	-0.162053



8 En el círculo de correlacion determine la correlacion entre las variables

```
[16]: # Plotea el círculo de correlación
acp.plot_circulo()
# Inercia explicada = 38.88%
# Antes de iniciar se debe recalcar, que hay ciertas variables que no tienen
↳ tan buenas representaciones en el plano,
# como es el caso de las variables: danceability, tempo, liveness, duration_ms,
↳ time_signature, entre otras, ya que estas
# (flechas) no alcanzan la frontera del círculo. La mas mal representada de las
↳ variables es liveness y tempo.
# En este caso con este círculo se puede ver que por ejemplo hay variables que
↳ siguen esta relacion:
#  $q = 180^\circ$  implica que  $\cos(q) = R(X,Y) = -1$  como por ejemplo: duration_ms y
↳ danceability, que al estar
# en 180 grados indica que son correlaciones fuertes y perfectamente negativas
↳ y que entre mas dure
# la cancion en milisegundos menos apta para bailar sera (danceability)
# lo mismo pasa para aunque ligeramente menor para time signature con
↳ duration_ms
# otras como loudness y danceability no tienen relacion lineal y lo que pase en
↳ una, no afectara a la otra
# lo mismo pasa con energy y danceability; en este caso no existe correlacion
↳ entre estas variables.
```

```
# Mientras que por ejemplo energy y loudness se ven que tiene relaciones
→ fuertes y positivamente correlacionadas
# porque tienden a esta forma  $q = 0$  implica que  $\cos(q) = R(X,Y) = 1$  (aunque no
→ es 1, pero estan muy cerca) provoca
# que si por ejemplo energy aumenta loudness tambien aumentara y lo mismo
→ sucede (aunque en otro plano) con speechiness y
# acousticness si una aumenta, la otra tambien aumentara
```



9 Explique la formacion de los clusteres basado en la sobreposicion del circulo y el plano

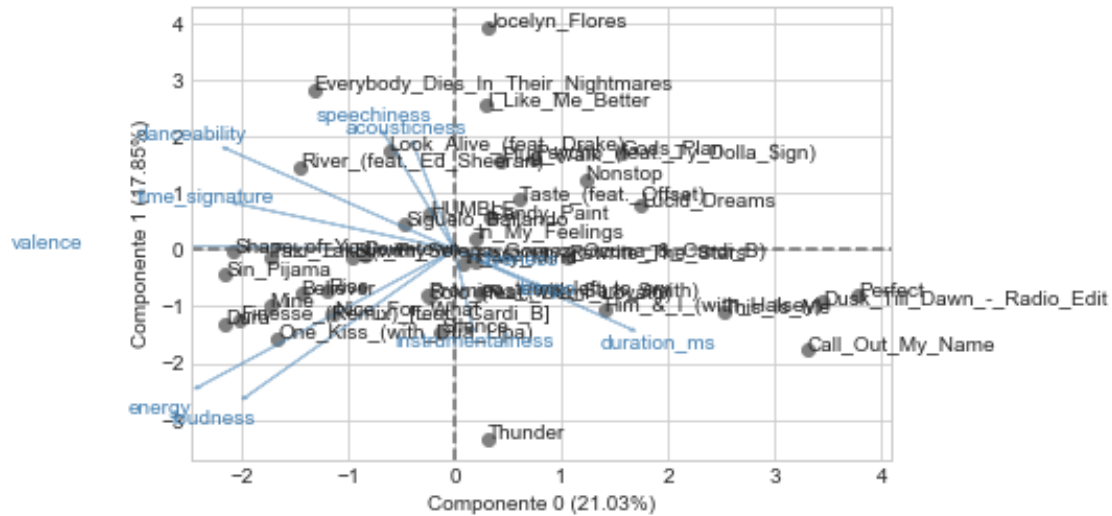
```
[17]: # Plotea la sobreposición plano-correlación
acp.plot_sobreposicion()

#El cluster 1 (Everybody_Dies, River, etc) se caracteriza por agrupar variables
→ como danceability, time_signature, speechiness,
#acousticness, es decir variables que tienen mas que ver como lo apta que es la
→ cancion para bailar, palabras, que tan acustica es, etc.,
# y se ve como a medida que esta aumenta otras como las del cluster 3
→ descienden porque tienen correlaciones negativas
# mientras que las del cluster 2 tienden a ser mas relaciones mas
→ perpendiculares (aunque no totalmente) y pueden guardar cierta relacion
```

entre sus variables. En este caso la canción síguelo, se puede ver como esta,
→ mal representada

El cluster 2 (Sin pijama, One kiss, etc.) Se ve fuertemente impactado - de
→ manera positiva- por las variables de: energy,
loudness, valence, etc., es decir, entre más intensidad y actividad habrá
→ mayor sonoridad general de la pista
en decibiles y habrá más positividad musical transmitida por la canción. Pero
→ estas variables no tendrán relación
alguna con speechiness o acousticness y solo un poco con danceability, sin
→ embargo con duration sí tendrán correlaciones negativas
y entre más duren las canciones en milisegundos, menos intensidad y
→ actividad, así como sonoridad y positividad musical
tendrán. También hay que agregar que hay unos datos mal representados porque
→ se encuentran cerca del origen.

El cluster 3 (call out my name, Dusk Till Dawn - Radio Edit, etc.) parece ser
→ el opuesto al cluster 1, en la medida en la que
las canciones duren más en milisegundos, y sean más instrumentalness, menos
→ aptas para bailar serán y menos beats habrán en cada
barra medida. Con respecto a el cluster 2 que tiene las variables de energy y
→ loudness, entre otras,
por estar de forma perpendicular o correlación nula, no es correcto asumir
→ nada sobre su comportamiento. Además,
se presentan outliers para este cluster, como es el caso de call out my name,
→ thunder, perfect, Dusk till dawn
de Radio Edit y se podría evaluar eliminarlos. También hay datos mal
→ representados como las canciones: In My Feelings,
In My Mind, Havana, Candy Paint y HUMBLE. Por último, hay unos datos que se
→ encuentran en el cuadrante 1, que no comparten
ninguna variable, esto se debe a que estos datos no siguen la tendencia
→ general de los demás y por ende se salen de los parámetros.

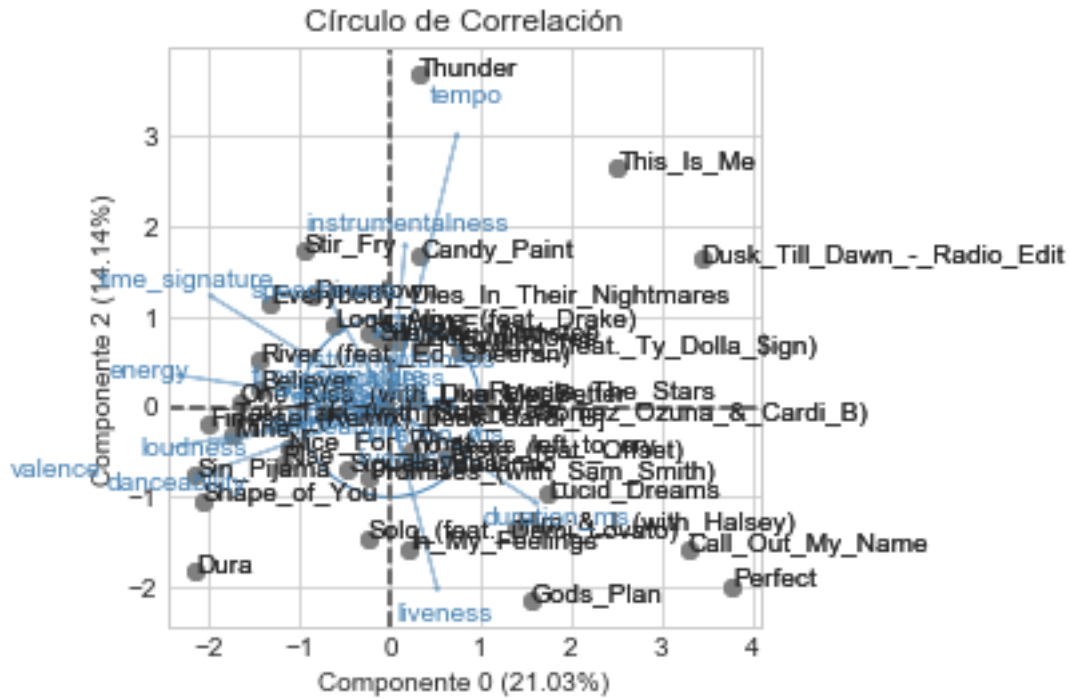


10 En el plano de los componentes 1 y 3 interprete las canciones In My Feelings, In My Mind, Havana, Candy Paint y HUMBLE, que son mal representadas en los componentes 1 y 2.

```
[18]: # Plotea el plano principal
acp.plot_plano_principal(ejes = [0, 2])

# Plotea el círculo de correlación
acp.plot_circulo(ejes = [0, 2])

# Plotea la sobreposición plano-correlación
acp.plot_sobreposicion(ejes = [0, 2])
# En este caso estas canciones en el componente 1 y 2 son mal representadas
→ porque se encuentran muy cerca del origen
# en el caso del componente 1 y 3, se ve que ahora mas bien se reacomodan estos
→ datos
# y estas canciones estan quedando ahora por tendencia general muy cerca de
→ donde se encuentran la mayoria de datos (en el centro),
# en el caso de In my feelings quedo algo mas alejado del centro y como cambio
→ de estar en el cuadrante 1 al cuadrante 4,
# Humble quedo algo mas cercana al centro de la distribucion y asi con las
→ restantes.
```



11 Ejercicio #2

```
[19]: import pandas as pd
import prince
import os
import pandas as pd
import numpy as np
os.chdir("/Users/heinerleivagmail.com")
print(os.getcwd())
casados = pd.read_csv('TablaAffairs.csv',delimiter=';',decimal=".")
print(casados.head())
print(casados.shape)
```

```
/Users/heinerleivagmail.com
   Unnamed: 0  TiempoInfiel  Genero  Edad  AnnosCasado  Hijos  Religioso  \
0           1             0   male  37.0           10.00    no         2.98
1           2             0  female  27.0            4.00    no         4.16
2           3             0  female  32.0           15.00   yes         0.97
3           4             0   male  57.0           15.00   yes         4.53
4           5             0   male  22.0            0.75    no         1.66

   Educacion  Ocupacion  Valoracion
0        17.84         7        feliz
```

```

1      13.38      6      feliz
2      12.69      1      feliz
3      17.79      6      muy feliz
4      16.87      6      neural
(601, 10)

```

12 a) Calcule el resumen numerico, interprete los resultados para una variable

```

[20]: print(casados.dropna().describe())
print(casados.dropna().mean())
print(casados.describe())
print(casados.mean(numeric_only=True))
print(casados.median(numeric_only=True))
print(casados.std(numeric_only=True))
print(casados.max(numeric_only=True))
print(casados.quantile(np.array([0,.25,.50,.75,1])))

##### Variable elegida: Edad #####

#count: la cantidad de datos para la variable edad es de 601 registros
#mean: de las 601 parejas estudiadas la edad promedio de estas personas es de
→32.487521 annos.
#std: con respecto a la desviacion estandar esta es de 9.288762, lo que indica
→que la varianza es
# menor respecto a la media es decir, los datos tienen baja variacion.
#min: la edad mas baja registrada es de 17.5 annos.
#primer cuartil: el 25% de los datos tienen edades inferiores o iguales a 27
→annos.
#Segundo cuartil: la mitad de los datos son inferiores o iguales a 32 annos.
#Tecer cuartil: el 75% de los datos son iguales o menores a 37 annos.
#max: el valor mas alto registrado de la totalidad de los datos es de 57 annos.

```

	Unnamed: 0	TiempoInfiel	Edad	AnnosCasado	Religioso \
count	601.000000	601.000000	601.000000	601.000000	601.000000
mean	301.000000	1.455907	32.487521	8.177696	3.128286
std	173.638033	3.298758	9.288762	5.571303	1.206544
min	1.000000	0.000000	17.500000	0.125000	0.510000
25%	151.000000	0.000000	27.000000	4.000000	2.110000
50%	301.000000	0.000000	32.000000	7.000000	3.200000
75%	451.000000	0.000000	37.000000	15.000000	4.140000
max	601.000000	12.000000	57.000000	15.000000	5.490000

	Educacion	Ocupacion
count	601.000000	601.000000
mean	16.185042	4.194676

```

std      2.427380    1.819443
min      8.890000    1.000000
25%     14.270000    3.000000
50%     16.390000    5.000000
75%     17.860000    6.000000
max     20.500000    7.000000

```

```

Unnamed: 0      301.000000
TiempoInfiel    1.455907
Edad            32.487521
AnnosCasado     8.177696
Religioso      3.128286
Educacion       16.185042
Ocupacion       4.194676

```

dtype: float64

```

      Unnamed: 0  TiempoInfiel      Edad  AnnosCasado  Religioso \
count  601.000000    601.000000  601.000000    601.000000  601.000000
mean   301.000000      1.455907   32.487521      8.177696    3.128286
std    173.638033      3.298758    9.288762     5.571303    1.206544
min      1.000000      0.000000   17.500000     0.125000    0.510000
25%    151.000000      0.000000   27.000000     4.000000    2.110000
50%    301.000000      0.000000   32.000000     7.000000    3.200000
75%    451.000000      0.000000   37.000000    15.000000    4.140000
max    601.000000     12.000000   57.000000    15.000000    5.490000

```

```

      Educacion  Ocupacion
count  601.000000  601.000000
mean   16.185042   4.194676
std     2.427380   1.819443
min      8.890000   1.000000
25%    14.270000   3.000000
50%    16.390000   5.000000
75%    17.860000   6.000000
max    20.500000   7.000000

```

```

Unnamed: 0      301.000000
TiempoInfiel    1.455907
Edad            32.487521
AnnosCasado     8.177696
Religioso      3.128286
Educacion       16.185042
Ocupacion       4.194676

```

dtype: float64

```

Unnamed: 0      301.00
TiempoInfiel     0.00
Edad             32.00
AnnosCasado       7.00
Religioso        3.20
Educacion        16.39
Ocupacion         5.00

```

```

dtype: float64
Unnamed: 0      173.638033
TiempoInfie1    3.298758
Edad            9.288762
AnnosCasado     5.571303
Religioso      1.206544
Educacion       2.427380
Ocupacion       1.819443
dtype: float64
Unnamed: 0      601.00
TiempoInfie1    12.00
Edad            57.00
AnnosCasado     15.00
Religioso      5.49
Educacion       20.50
Ocupacion       7.00
dtype: float64

```

	Unnamed: 0	TiempoInfie1	Edad	AnnosCasado	Religioso	Educacion	\
0.00	1.0	0.0	17.5	0.125	0.51	8.89	
0.25	151.0	0.0	27.0	4.000	2.11	14.27	
0.50	301.0	0.0	32.0	7.000	3.20	16.39	
0.75	451.0	0.0	37.0	15.000	4.14	17.86	
1.00	601.0	12.0	57.0	15.000	5.49	20.50	

	Ocupacion
0.00	1.0
0.25	3.0
0.50	5.0
0.75	6.0
1.00	7.0

13 b) Calcule la matriz de correlaciones

```

[21]: import seaborn as sns
import matplotlib.pyplot as plt

corr = casados.corr()
print(corr)
f, ax = plt.subplots(figsize=(10, 8))
sns.heatmap(corr, mask=np.zeros_like(corr, dtype=np.bool), cmap=sns.
    ↪diverging_palette(220, 10, as_cmap=True),
            square=True, ax=ax)

# Interpretacion

# Primera correlacion:

```

```

# Se puede observar como la variable religioso y la variable tiempo infiel son
↳ fuertemente y negativamente opuestas
# es decir, tienen una correlacion negativa, porque entre mas religiosa la
↳ persona sea menos tiempo va a ser infiel.
# Esto se puede observar, ya que la coordenada que los entrecruza es de un
↳ color muy azul, acercandose casi al -1.

# Segunda correlacion:

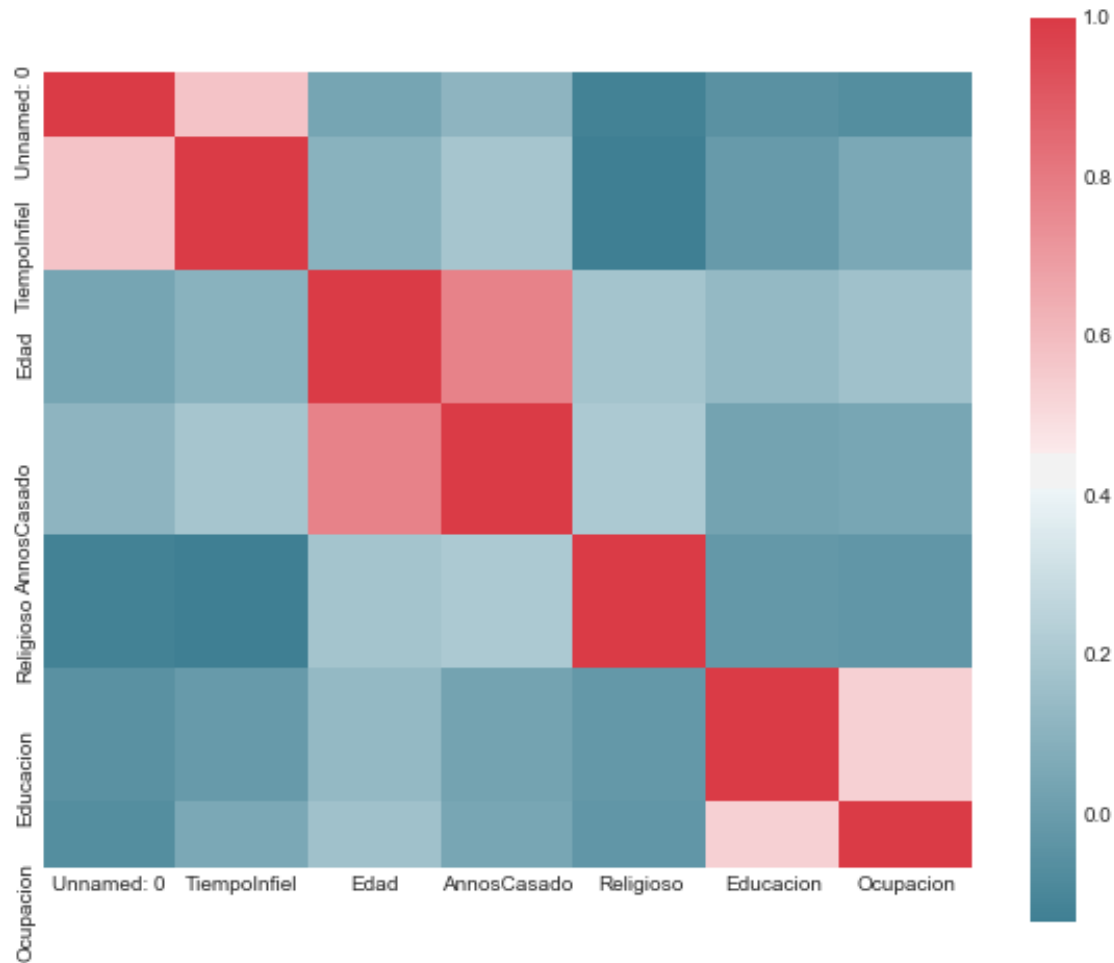
# Se puede observar como la variable annos casado y edad son positivas y
↳ fuermente correlacionadas, es decir, entre mas annos
# tenga una persona mas anos casado tendra y viceversa, en este caso se puede
↳ ver esa correlacion porque el cuadrante es de color
# rojo fuerte y ese color esta muy cerca del 1, lo que indica una correlacion
↳ positiva.

```

	Unnamed: 0	TiempoInfiel	Edad	AnnosCasado	Religioso	\
Unnamed: 0	1.000000	0.576919	0.036228	0.107761	-0.120416	
TiempoInfiel	0.576919	1.000000	0.095237	0.186842	-0.137501	
Edad	0.036228	0.095237	1.000000	0.777546	0.178952	
AnnosCasado	0.107761	0.186842	0.777546	1.000000	0.200261	
Religioso	-0.120416	-0.137501	0.178952	0.200261	1.000000	
Educacion	-0.049671	-0.010330	0.130390	0.030205	-0.021423	
Ocupacion	-0.069130	0.049612	0.166413	0.044592	-0.028387	

	Educacion	Ocupacion
Unnamed: 0	-0.049671	-0.069130
TiempoInfiel	-0.010330	0.049612
Edad	0.130390	0.166413
AnnosCasado	0.030205	0.044592
Religioso	-0.021423	-0.028387
Educacion	1.000000	0.534744
Ocupacion	0.534744	1.000000

```
[21]: <matplotlib.axes._subplots.AxesSubplot at 0x12c678590>
```



14 C) Usando solo las variables numericas efectue un ACP y de una interpretacion siguiendo los siguientes pasos:

15 1 en el plano principal encuentre 4 cluseres

```
[22]: print(casados.head())
casados2= pd.DataFrame(data=casados, columns=['TiempoInfiel', 'Edad',
↳ 'AnnosCasado',
    'Religioso', 'Educacion'])
acp = ACP(casados2, n_componentes=4)

# en el plano principal encuentre 4 cluseres
# Plotea el plano principal
acp.plot_plano_principal()
```

Unnamed: 0 TiempoInfiel Genero Edad AnnosCasado Hijos Religioso \

	Educacion	Ocupacion	Valoracion
0	17.84	7	feliz
1	13.38	6	feliz
2	12.69	1	feliz
3	17.79	6	muy feliz
4	16.87	6	neural



tiempo infiel, otro por Educacion, annos casados y edad y el ultimo por
 ↳ religioso y un cuarto que en realidad no contiene ninguna variable.

se puede ver como la variable tiempo infiel tiene cierto grado de relacion
 ↳ positiva con las variables educacion, annos casados y edad, ya que forman
 ↳ angulos

agudos, es decir si una crece las demas creceran, aunque con la que guarda
 ↳ mas relacion es con educacion, ya que su angulo agudo es mas pequenno y hay
 ↳ mayor correlacion, aunque

de todas las variables educacion es la peor representada porque se encuentra
 ↳ muy cerca del origen. Ademas, tiempo infiel presenta correlacion negativa
 ↳ con religioso, es decir si una de las dos aumenta, la demas disminuira.

El segundo cluster, annos casados, edad y educacion, presentan altas
 ↳ correlaciones entrre ellas y si una crece la otra crecera, presentan tambien
 ↳ correlacion positiva con tiempo

infiel y tambien con reloigioso, porque se forman angulos agudos y existe
 ↳ correlacion. Por ejemplo: annos casados, y edad, entre mas edad tenga las
 ↳ personas

mas tiempo habran estado casados, asi tambien si estas aumentan, la educacion
 ↳ tambien aumentaran e indudablemente el tiempo infiel tambien aumentara (en
 ↳ menos medida) pero aumentara este vector.

Con respecto al tercer cluster, religioso, se ve como la varible religioso y
 ↳ tiempo infiel tienen una correlacion negativa

y entre mas religiosa sea la persona menos tiempo infiel sera, ya que el
 ↳ angulo que se forma entre ambas es obtuso, lo mismo pasa (aunque en menor
 ↳ medida) por ejemplo

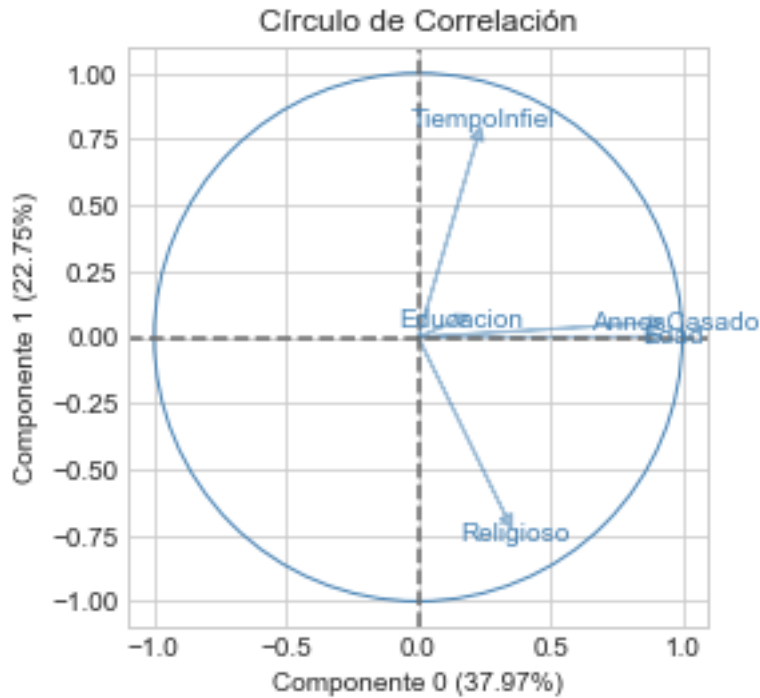
con educacion si elongamos el angulo se ve como este tiende a ser un angulo
 ↳ ligeramente obtuso y entonces indica que la educacion tiene una correlacion
 ↳ negativa con la

religiosidad de una persona, lo cual indica que entre mas religiosa sea una
 ↳ persona menos estudiada sera y viceversa (aunque en realidad no es tan
 ↳ fuerte como tiempo infiel)

Otras como annos casados, edad y religioso (aunque estan en un cluster
 ↳ diferente) se ve que tienen angulos agudos con respecto a religioso,
 ↳ entonces guardan

relacion y se puede decir que entre mas edad tenga una persona mas religiosa
 ↳ sera.

Y se cuenta con un cuarto cluster que en realidad carece de variables
 ↳ asociadas.



17 3 explique la formación de los clusteres basado en la sobreposición del círculo y el plano.

[25]: `acp.plot_sobreposicion()`

(ver análisis de arriba para mejor apoyo)

En este caso se ve que se forman clusteres con variables muy separadas, aunque
 → hay bastantes datos que están dispersos por todo

el plano cartesiano y se pierde interpretabilidad hay una cantidad gruesa de
 → datos que se agrupan en el cuadrante III y unos pocos

en el cuadrante II, aunque estos no tienen variables asociadas y solo están
 → dispersos, se ve como hay muchos datos

formando un cluster en el cuarto cuadrante y en donde se aglutinan datos que
 → tienen que ver con religiosidad, años casados y

edad (por la correlación que se menciona arriba) y solo otros pocos más
 → dispersos con las variables como tiempo infiel y

educación que tienen correlaciones positivas y fuertes.

También se ve que hay muchísimos datos mal representados, porque se
 → encuentran muy cerca del origen y no en el centro.

Por otro lado, se pueden ver outliers, como el 558, el 525, 587, etc., que
 → son datos que no se tienden a agrupar en el centro.


```

acp.plot_plano_principal()

# Plotea el círculo de correlación
acp.plot_circulo()

# Plotea la sobreposición plano-correlación
acp.plot_sobreposicion()

# ¿Se gana interpretabilidad al convetir Genero e Hijos enCodigo Disyuntivo
→Completo?

# Por supuesto, ahora se puede ver que en cada uno de los 4 cuadrantes del
→plano cartesiano existen clusters y se llega
# al objetivo de tener cuatro clusters bien definidos. Se ve como en los
→clusters ya los datos se encuentran mejor definidos y se pueden ver
# correlaciones como religion, tiempo infiel, que si tienen hijos y edad, y
→tiempo casados que hay una gran cantidad de datos (Cluster 2)
# mientras que el cluster 3 va a estar definido por la educacion, personas de
→genero masculino, ya que se ve que tiene correlacin positiva y fuerte,
→mientras
# que estos mismos tienen correlaciones negativas perfectas con genero
→femenino, se aprecia como la educacion y el genero masculino si una aumenta
→la otra
# aumenta, mientras que si la educacion aumenta el genero femenino disminuye.
# Asimismo, se puede ver una correlacion negativa interesante, ya que se ve que
→si la pareja tiene hijos, el tiempo que es infiel aumenta,
# mientras que si no tienen hijos el tiempo de infidelidad va a disminuir.
# Por ultimo se puede ver que ahora hay una minomeria bastante reducida de
→datos mal representados, quiza solo el 416, 404 y 596
# sean los mas notables, se puede ver como los datos tienden al centro y no se
→tienen casi outliers.

```

	TiempoInfiel	Edad	AnnosCasado	Religioso	Educacion	Genero_female \
0	0	37.0	10.00	2.98	17.84	0
1	0	27.0	4.00	4.16	13.38	1
2	0	32.0	15.00	0.97	12.69	1
3	0	57.0	15.00	4.53	17.79	0
4	0	22.0	0.75	1.66	16.87	0

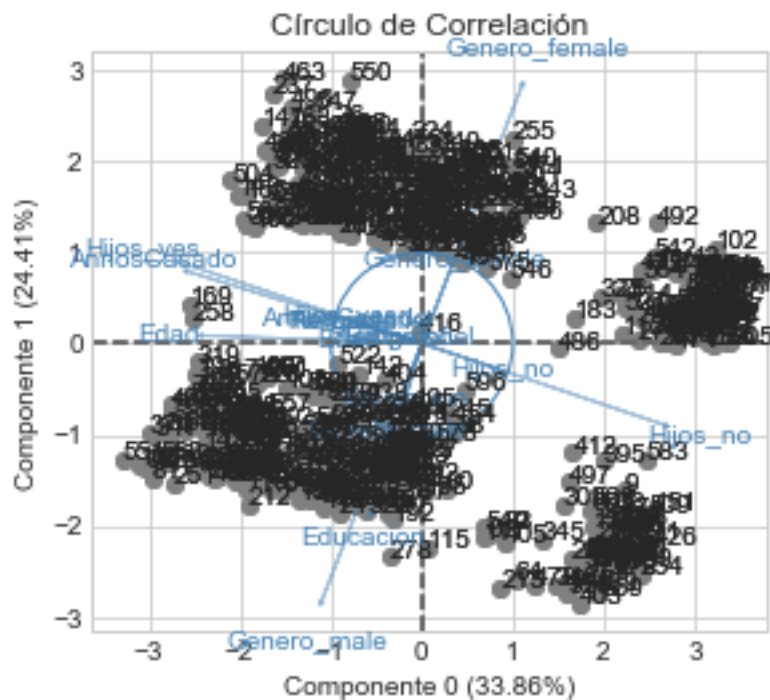
	Genero_male	Hijos_no	Hijos_yes
0	1	1	0
1	0	1	0
2	0	0	1
3	1	0	1
4	1	1	0

TiempoInfiel	int64
Edad	float64

```

AnnosCasado      float64
Religioso        float64
Educacion         float64
Genero_female     uint8
Genero_male       uint8
Hijos_no          uint8
Hijos_yes         uint8
dtype: object

```



19 Ejercicio #3

```

[27]: import pandas as pd
import prince
import os
import pandas as pd
import numpy as np
os.chdir("/Users/heinerleivagmail.com")
print(os.getcwd())
corazon = pd.read_csv('SAheart.csv',delimiter=';',decimal=".")
print(corazon.head())
print(corazon.shape)

```

```

/Users/heinerleivagmail.com
  sbp  tobacco  ldl  adiposity  famhist  typea  obesity  alcohol  age  chd

```

0	160	12.00	5.73	23.11	Present	49	25.30	97.20	52	Si
1	144	0.01	4.41	28.61	Absent	55	28.87	2.06	63	Si
2	118	0.08	3.48	32.28	Present	52	29.14	3.81	46	No
3	170	7.50	6.41	38.03	Present	51	31.99	24.26	58	Si
4	134	13.60	3.50	27.78	Present	60	25.99	57.34	49	Si

(462, 10)

20 a) Efectue un ACP usando solo las variables numericas y de una interpretacion siguiendo los siguientes pasos:

21 En el plano principal encuentre los clusteres

```
[28]: coraz2 = pd.DataFrame(data=corazon, columns=['sbp', 'tobacco', 'ldl',
        'adiposity', 'typea', 'obesity', 'alcohol', 'age'])

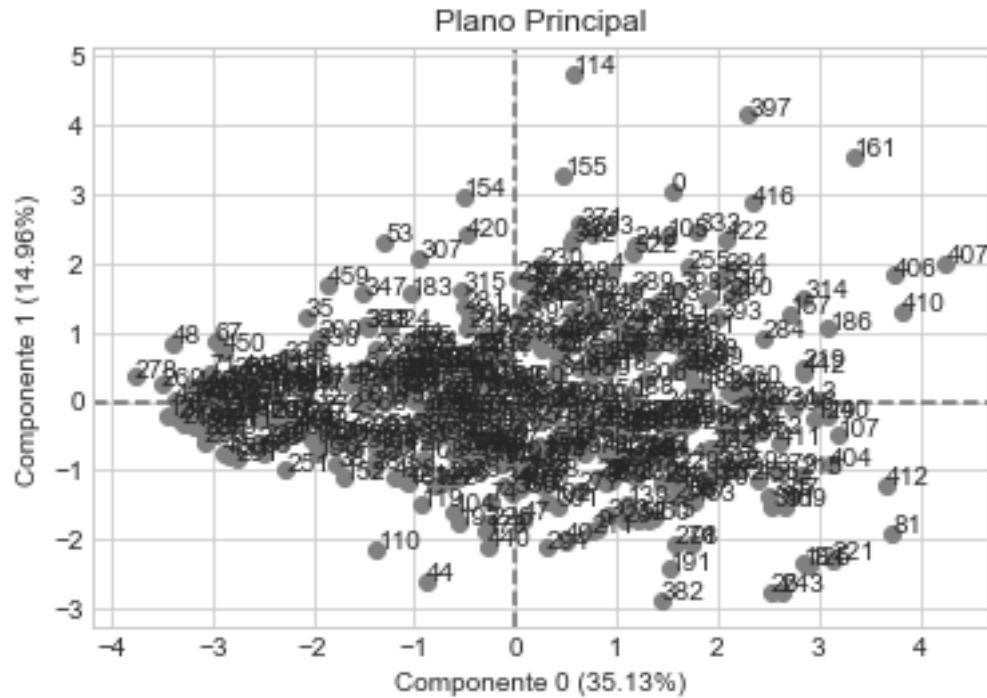
print(corazon2.head())
print(corazon2.describe())

acp = ACP(corazon2, n_componentes=3)
# Plotea el plano principal
acp.plot_plano_principal()
```

	sbp	tobacco	ldl	adiposity	typea	obesity	alcohol	age
0	160	12.00	5.73	23.11	49	25.30	97.20	52
1	144	0.01	4.41	28.61	55	28.87	2.06	63
2	118	0.08	3.48	32.28	52	29.14	3.81	46
3	170	7.50	6.41	38.03	51	31.99	24.26	58
4	134	13.60	3.50	27.78	60	25.99	57.34	49

	sbp	tobacco	ldl	adiposity	typea	obesity
count	462.000000	462.000000	462.000000	462.000000	462.000000	462.000000
mean	138.326840	3.635649	4.740325	25.406732	53.103896	26.044113
std	20.496317	4.593024	2.070909	7.780699	9.817534	4.213680
min	101.000000	0.000000	0.980000	6.740000	13.000000	14.700000
25%	124.000000	0.052500	3.282500	19.775000	47.000000	22.985000
50%	134.000000	2.000000	4.340000	26.115000	53.000000	25.805000
75%	148.000000	5.500000	5.790000	31.227500	60.000000	28.497500
max	218.000000	31.200000	15.330000	42.490000	78.000000	46.580000

	alcohol	age
count	462.000000	462.000000
mean	17.044394	42.816017
std	24.481059	14.608956
min	0.000000	15.000000
25%	0.510000	31.000000
50%	7.510000	45.000000
75%	23.892500	55.000000
max	147.190000	64.000000



22 En el círculo de correlacion determine la correlacion entre las variables

[30]: `# Plotea el círculo de correlación`
`acp.plot_circulo()`
Inercia explicada = 50.09%
En este caso se ve como hay mayor preponderancia de las variables numericas
→ al quedar del lado derecho del plano cartesiano, específicamente en el
→ cuadrante 1 y 4 y una en
el cuadrante 3. Además se observa como la variable mas mal representada es
→ typea, ya que se encuentra muy cerca del origen.

En este caso, analizando las variables del cuadrante 1, se puede ver como
→ alcohol, tobacco, sbp, y age guardan correlaciones lineales positivas, en
→ este caso

si una persona consume alcohol y acumula tabaco, su presión arterial
→ también va a aumentar, y en la medida en que las primeras dos aumenten su
→ presión aumentará signi-

ficativamente y lo mismo con la edad, ya que si la edad aumenta, la ingesta
→ de alcohol, y el tabaco así como la presión también lo harán.

En el caso de adiposity, obesity y ldl; estas tres guardan correlaciones
 ↳ positivas y muy fuertes, por lo que, si una persona aumenta en su porcentaje
 ↳ de obesidad, también

aumentara la adiposidad así como el ldl, ahora bien, aunque estas últimas
 ↳ mencionadas se encuentren en el cuadrante 4, guardan relación con sbp, age y
 ↳ tobacco, es decir,

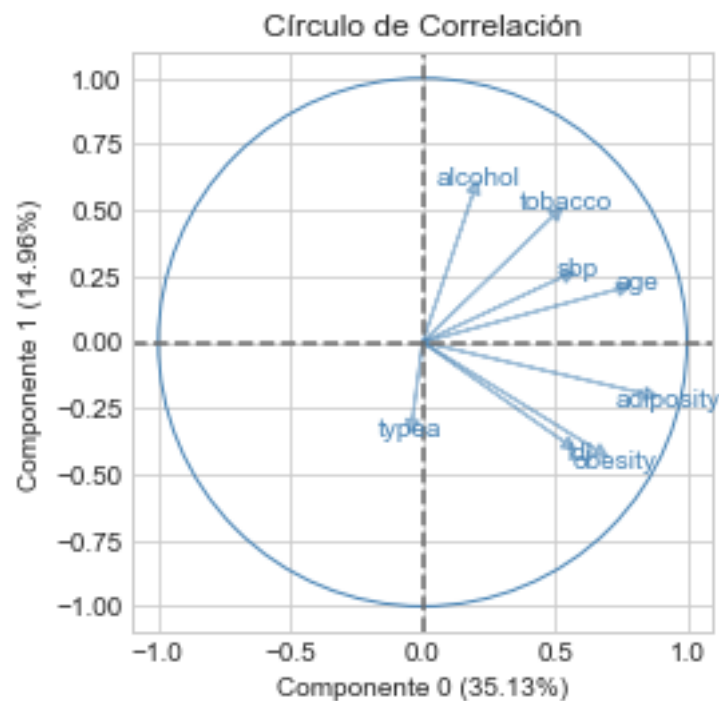
si estas aumentan, las demás también, ya que en su interior se forman ángulos
 ↳ agudos, excepto por ejemplo, ldl y alcohol, ya que más bien se puede
 ↳ apreciar un ángulo obtuso,

y por lo tanto una correlación negativa, es decir, si alcohol aumenta, ldl
 ↳ bajara. Otras como por ejemplo alcohol y adiposity tienden a no tener
 ↳ relación lineal, porque se

se ve un ángulo más o menos recto, entonces una no tendrá impacto en la otra.

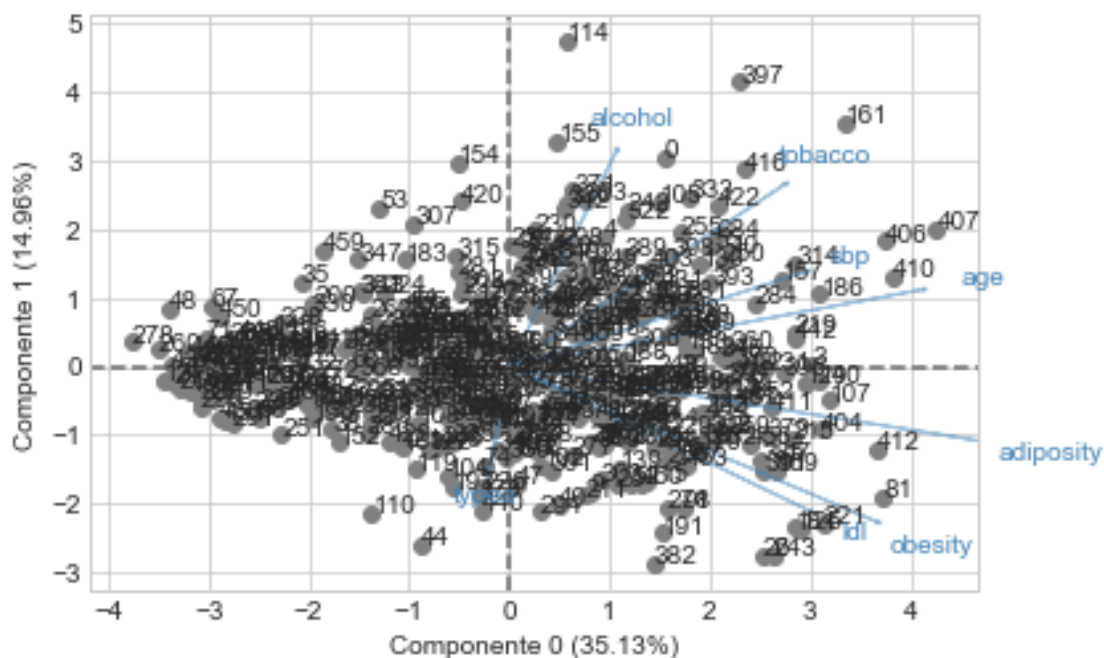
Finalmente, el consumo de alcohol y la variable typea tienen correlaciones
 ↳ negativas perfectas, es decir si el consumo de alcohol aumenta, el
 ↳ comportamiento del tipo A va a

descender y viceversa.



23 Explique la formacion de los clusteres basado en la sobreposicion del circulo y el plano.

```
[31]: # Plotea la sobreposición plano-correlación
acp.plot_sobreposicion()
# (ver analisis de arriba para un mejor apoyo)
# En este caso se ve con la sobreposicion que no hay como tal un cluster facil
  ↳ de diferenciar, ya que hay datos muy agrupados en el centro de la grafica, y
  ↳ hay una cantidad
# muy grande los graficos mal representados, ya que estan muy cerca o casi en
  ↳ el origen y otros estan muy aglutinados
# en el centro, lo que hace dificil poder separar los cluster para poder
  ↳ analizar, tambien se ve que en el cuadrante 2 hay datos (con menos
  ↳ preponderancia)
# pero existen datos sin variables asociadas. La mayoría de los datos se
  ↳ sobreponen en el cluster de obesity, adiposity y en el cluster de alcohol,
  ↳ tobacco, y age tambien hay datos,
# hay tambien en el cluster de type, y se ve quye hay una alta concentracion de
  ↳ los datos en este cuadrante.
# tambien se observa que hay valores que se alejan de los centros (outliers)
  ↳ como por ejemplo 114, 397, 161, 407, etc.
```



24 b) Efectue un ACP usando las variables numericas y las variables categoricas (recuerde recodificar las categoricas usando codigo disyuntivo completo).

25 En el plano principal encuentre los clusters

```
[32]: # Convirtiendo la variables en Dummy
datos_dummy = pd.get_dummies(corazon)
print(datos_dummy.head())
print(datos_dummy.dtypes)

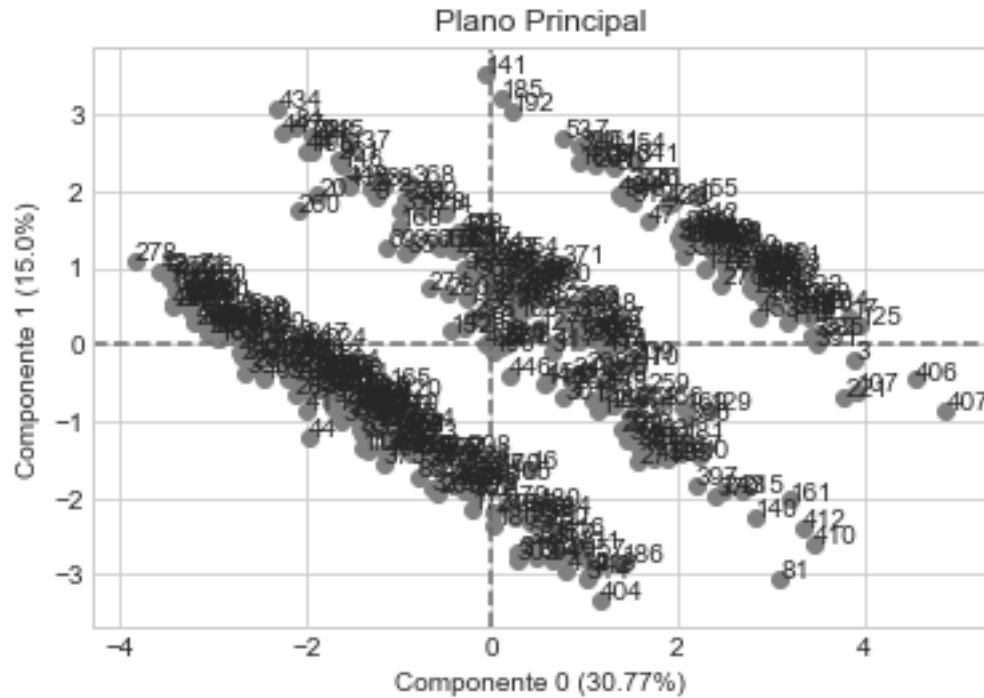
acp = ACP(datos_dummy,n_componentes=3)

# Plotea el Plano Principal
acp.plot_plano_principal()
```

	sbp	tobacco	ldl	adiposity	typea	obesity	alcohol	age	\
0	160	12.00	5.73	23.11	49	25.30	97.20	52	
1	144	0.01	4.41	28.61	55	28.87	2.06	63	
2	118	0.08	3.48	32.28	52	29.14	3.81	46	
3	170	7.50	6.41	38.03	51	31.99	24.26	58	
4	134	13.60	3.50	27.78	60	25.99	57.34	49	

	famhist_Absent	famhist_Present	chd_No	chd_Si
0	0	1	0	1
1	1	0	0	1
2	0	1	1	0
3	0	1	0	1
4	0	1	0	1

```
sbp          int64
tobacco      float64
ldl          float64
adiposity    float64
typea        int64
obesity      float64
alcohol      float64
age          int64
famhist_Absent  uint8
famhist_Present  uint8
chd_No        uint8
chd_Si        uint8
dtype: object
```



26 En el círculo de correlacion determine la correlacion entre las variables

[33]: *# Plotea el círculo de correlación*

```
acp.plot_circulo()
```

#Inercia explicada: 45.77%

En este caso se ve como se han formado ahora tres clusters, en el cluster 1

→representando por las variables typea, chd_si, famhist_present

para este cluster en particular la variable que tiene peor representacion es

→la de typea, ya que no se encuentra cerca de la frontera y se encuentra

→cerca del origen

sin embargo todas estas variables presentan correlaciones positivas y

→lineales, es decir, si hay presencia de ataque al corazon en la familia

→aumenta la enfermedad del

corazon y el tipo de comportamiento A.

Con respecto al segundo cluster, representado por tobacco, age, sbp, obesity,

→adiposity, y alcohol, la variaoable que tiene la peor representacion es

→alcohol, ya que se encuentra cerca

del origen, sin embargo todas estas variables presentan correlaciones
 ↳positivas y si una aumenta las demas tambien aumentaran y si una disminuye,
 ↳las demas tambien lo haran

se ve como obesity no tiene relacion lineal con famhist_present, ya que hay
 ↳un angulo de 90 grados, entonces no es correcto asumir nada sobre estas
 ↳variables, otras como obesity y chd_si

tienen una correlacion positiva tambien, ya que el angulo es agudo y guardan
 ↳relacion, mientras que por ejemplo age y chd_no

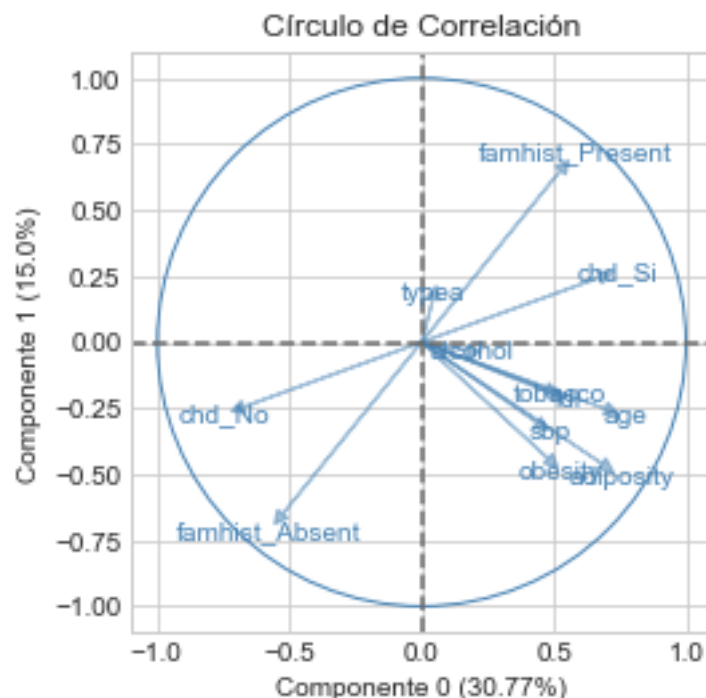
presentan correlaciones negativas y opuestas, es decir, su la edad aumenta,
 ↳el diagnostico tendra a disminuir.

Con el ultimo cluster, formado por chd_no y famhist_absent, ambas estan bien
 ↳representadas, la segunda mejor que la primera, pero guardan correlaciones
 ↳positivas entre ellas y si una aumenta

la otra aumentara de igual forma, existen correlaciones negativas perfectas,
 ↳como chd_no, ya que si el diagnostico en una persona es ausente, la
 ↳enfermedad coronaria sera no tambien, mientras

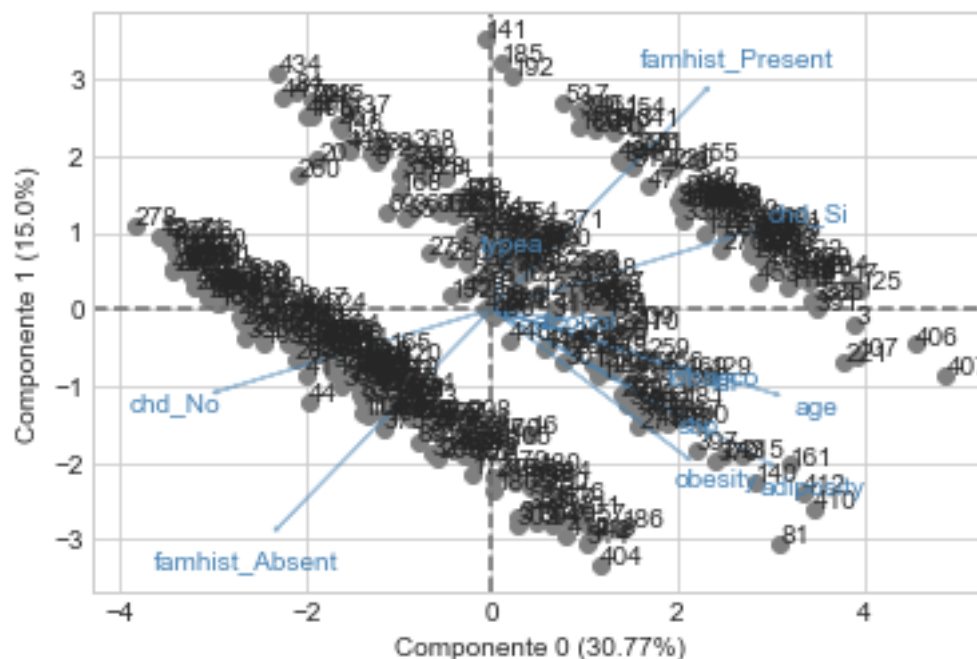
que si en una esta presente la enfermedad coronaria, su diagnostico sera
 ↳presente obviamente porque son variables opuestas. Otras como edad y
 ↳enfermedad del corazon tienen una correlacion

negativa, es decir entre mas aumente la edad, menos enfermedad coronaria de
 ↳corazon tendra.



27 Explique la formacion de los clusteres basado en la sobreposicion del circulo y el plano.

```
[35]: # Plotea la sobreposición plano-correlación
acp.plot_sobreposicion()
# Habiendo incluido las variables categoricas se puede observar como se forman
↳ 3 cluster bastante marcados,
# aunque muchos de los datos siguen estando en el origen y no estan tan bien
↳ representados
# se ve como el cluster de famhist_present, es el que tiene la mejor
↳ representacion, el de obesity, age,
# tiene datos bien representados como el caso del 81, el 161, 410, etc, pero se
↳ siguen teniendo muchos
# datos en el centro y ademas se ve como se traslapan con el cuadrante 2 que no
↳ tiene asociadas variables,
# de igual forma se mantienen las correlaciones positivas para los datos que se
↳ encuentran en el cuadrante 4
# finalmente el cluster formado por las variables famhist_absent y chd_no
↳ presenta datos muy aglutinados y
# hay algunos que se traslapan con el cuadrante 4 y presentan las correlaciones
↳ que se mencionaron en el
# apartado anterior, aunque los datos que se presnetan en el cuadrante 2 no
↳ presentan variables asociadas.
```



28 Explique las diferencias de este ACP respecto al anterior (usando solo las variables numericas. ¿Cual le parece mas interesante? ¿Por que?

```
[36]: # La principal diferencia es que en este nuevo ACP hay mejor representacion de
      ↪ los datos en los clusters, apartado que los datos ya
# se logran agrupar en tres cluster y no solo en un cluster enorme que era mas
      ↪ difuso y dificil de entender porque todos los datos se
# tendieron a apilar en el centro sin embargo en este ultimo hay mas evidencia
      ↪ de datos en el cuadrante 2 y en el, no hay variables asociadas,
# sin embargo en este ultimo se logra ver mejor la delimitacion correcta de
      ↪ cada cluster y es mas facil ver las relaciones.
# al centro aunque no esten bien representados.
# Por otro lado, se logra ver que justo en el cluster de la mitad es donde mas
      ↪ datos cerca al origen hay y ese cluster es el que
# tiene mas outliers, pero se ve como los datos tienden a irse al centro, que
      ↪ en este caso es el origen.
# El que me parece mas interesante es este ultimo (grafico) ya que solo
      ↪ analizando las variables numericas todas eran bastante similares,
# de ahi que se dio un amontonamiento de los datos
# en cambio en este, con la introduccion de las variables categoricas, ayudo a
      ↪ segmentar mejor los cluster con otras variables cualitativas
# y es lo que me llama la atencion ya que como en este caso, las variables
      ↪ cualitativas pueden jugar un papel decisivo en la representacion
# y nos ayudan a ver mejor y entender los datos y las correlaciones que tienen
```

29 Ejercicio #4

30 Programe una clase derivada (que herede) de la clase class ACP, que fue presentada durante la leccion, que incluya adicionalmente lo siguiente:

```
[37]: # Parte A
      # Que sobrecargue el constructor de la clase init para seleccionar variables,
      ↪ es decir,
# que reciba adicionalmente un vector con los números de una columna o nombres
      ↪ de la
# variables respectivas de manera que el atributo datos sea modificado para
      ↪ eliminar esas
# columnas. Y asi todos los calculos sean realizados eliminando estas columnas.
```

```
[38]: # Parte B
      # Que sobrecargue los metodos plot plano principal y plot sobreposicion de
      ↪ manera
```

```

# tal que en estos graficos se puedan eliminar individuos mal representados,
→esto basado
# en en el cos2 (x) mediante el atributo cos2 ind. Es decir, que estos metodos
→reciban un
# parametros adicional que el es porcentaje minimo en el coseno cuadrado
→aceptable para
# que los individuos aparezcan en estos graficos.

```

```

[39]: import matplotlib.pyplot as plt
from prince import PCA
import pandas as pd
import prince
import os
import pandas as pd
import numpy as np
datos = pd.read_csv('SpotifyTop2018_40_V2.csv',delimiter=',',decimal=".")

class ACP_new(ACP):
    def __init__(self, datos, n_componentes = 5,delete_column=[]):
        super().__init__(datos, n_componentes)
        self.__datos = datos
# PARTE A, sobrecarga el constructor de la clase init
        self.__delete_column = delete_column
        for col in self.__delete_column:
            del self.__datos[col]
# PARTE B, Sobrecarga de los metodos plot plano principal y plot sobreposicion
        def plot_plano_principal(self, ejes = [0, 1], ind_labels = True, titulo =
→'Plano Principal', valor_aceptable = 50):
            coordenadas = self.coordenadas_ind.drop(self.cos2_ind[(self.cos2_ind.
→iloc[:,0] +self.cos2_ind.iloc[:,1]) * 100 < valor_aceptable].index)
            x = coordenadas[ejes[0]].values
            y = coordenadas[ejes[1]].values
            plt.style.use('seaborn-whitegrid')
            plt.scatter(x, y, color = 'gray')
            plt.title(titulo)
            plt.axhline(y = 0, color = 'dimgrey', linestyle = '--')
            plt.axvline(x = 0, color = 'dimgrey', linestyle = '--')
            inercia_x = round(self.var_explicada[ejes[0]], 2)
            inercia_y = round(self.var_explicada[ejes[1]], 2)
            plt.xlabel('Componente ' + str(ejes[0]) + ' (' + str(inercia_x) + '%)')
            plt.ylabel('Componente ' + str(ejes[1]) + ' (' + str(inercia_y) + '%)')
            if ind_labels:
                for i, txt in enumerate(coordenadas.index):
                    plt.annotate(txt, (x[i], y[i]))
        def plot_sobreposicion(self, ejes = [0, 1], ind_labels = True, var_labels =
→True, titulo = 'Sobreposición Plano-Círculo',valor_aceptable = 50):

```

```

    coordenadas = self.coordenadas_ind.drop(self.cos2_ind[((self.cos2_ind.
↪iloc[:,0] +self.cos2_ind.iloc[:,1]) * 100) < valor_aceptable].index)
    x = coordenadas[ejes[0]].values
    y = coordenadas[ejes[1]].values
    cor = self.correlacion_var.iloc[:, ejes]
    scale = min((max(x) - min(x)/(max(cor[ejes[0]]) - min(cor[ejes[0]]))),
    (max(y) - min(y)/(max(cor[ejes[1]]) - min(cor[ejes[1]])))) * 0.7
    cor = self.correlacion_var.iloc[:, ejes].values
    plt.style.use('seaborn-whitegrid')
    plt.axhline(y = 0, color = 'dimgrey', linestyle = '--')
    plt.axvline(x = 0, color = 'dimgrey', linestyle = '--')
    inercia_x = round(self.var_explicada[ejes[0]], 2)
    inercia_y = round(self.var_explicada[ejes[1]], 2)
    plt.xlabel('Componente ' + str(ejes[0]) + ' (' + str(inercia_x) + '%)')
    plt.ylabel('Componente ' + str(ejes[1]) + ' (' + str(inercia_y) + '%)')
    plt.scatter(x, y, color = 'gray')
    if ind_labels:
        for i, txt in enumerate(coordenadas.index):
            plt.annotate(txt, (x[i], y[i]))
    for i in range(cor.shape[0]):
        plt.arrow(0, 0, cor[i, 0] * scale, cor[i, 1] * scale, color = ↪
↪'steelblue',alpha = 0.5, head_width = 0.05, head_length = 0.05)
    if var_labels:
        plt.text(cor[i, 0] * scale * 1.15, cor[i, 1] * scale * 1.15, self.
↪correlacion_var.index[i],color = 'steelblue', ha = 'center', va = 'center')

```

[40]: # PRUEBAS PARTE A

```

nuevo = ACP_new(datos, n_componentes = 5, delete_column = ↪
↪['duration_ms', 'tempo'])
print(datos) #esta es la prueba
nuevo.datos
print(nuevo.datos.head())
print(nuevo.datos.shape)
print(nuevo.datos.describe())

```

	danceability	energy	loudness \
Dura	0.791	0.848	-3.456
Plug_Walk	0.876	0.519	-6.531
Silence	0.520	0.761	-3.093
Solo_(feat._Demi_Lovato)	0.737	0.636	-4.546
Sin_Pijama	0.791	0.745	-3.695
Call_Out_My_Name	0.489	0.598	-4.929
Taki_Taki_(with_Selena_Gomez_Ozuna_&_Cardi_B)	0.841	0.798	-4.206
River_(feat._Ed_Sheeran)	0.748	0.749	-5.916
Rewrite_The_Stars	0.684	0.619	-7.005

Him_&_I_(with_Halsey)	0.589	0.731	-6.343
Look_Alive_(feat._Drake)	0.922	0.581	-7.495
Rise	0.687	0.785	-4.650
Perfect	0.599	0.448	-6.312
Thunder	0.605	0.822	-4.833
Psycho_(feat._Ty_Dolla_\$ign)	0.739	0.559	-8.011
no_tears_left_to_cry	0.699	0.713	-5.507
Everybody_Dies_In_Their_Nightmares	0.734	0.570	-7.066
Jocelyn_Flores	0.872	0.391	-9.144
HUMBLE.	0.908	0.621	-6.638
In_My_Mind	0.694	0.770	-5.335
Nice_For_What	0.586	0.909	-6.474
Havana	0.765	0.523	-4.333
Dusk_Till_Dawn_-_Radio_Edit	0.258	0.437	-6.593
Candy_Paint	0.670	0.654	-5.944
This_Is_Me	0.284	0.704	-7.276
Siguelo_Bailando	0.855	0.664	-7.110
I_Like_Me_Better	0.752	0.505	-7.621
Believer	0.776	0.780	-4.374
Downtown	0.775	0.679	-4.985
Shape_of_You	0.825	0.652	-3.183
Gods_Plan	0.754	0.449	-9.211
Stir_Fry	0.815	0.816	-5.474
One_Kiss_(with_Dua_Lipa)	0.791	0.862	-3.240
Finesse_(Remix)_[feat._Cardi_B]	0.704	0.859	-4.877
Nonstop	0.912	0.412	-8.074
Mine	0.710	0.789	-3.874
In_My_Feelings	0.835	0.626	-5.833
Taste_(feat._Offset)	0.884	0.559	-7.442
Lucid_Dreams	0.511	0.566	-7.230
Promises_(with_Sam_Smith)	0.781	0.768	-5.991

	speechiness	acousticness \
Dura	0.0506	0.183000
Plug_Walk	0.1430	0.202000
Silence	0.0853	0.256000
Solo_(feat._Demi_Lovato)	0.0437	0.044100
Sin_Pijama	0.0464	0.354000
Call_Out_My_Name	0.0360	0.218000
Taki_Taki_(with_Selena_Gomez_Ozuna_&_Cardi_B)	0.2290	0.153000
River_(feat._Ed_Sheeran)	0.5160	0.142000
Rewrite_The_Stars	0.0386	0.071600
Him_&_I_(with_Halsey)	0.0868	0.053400
Look_Alive_(feat._Drake)	0.2700	0.001040
Rise	0.0333	0.327000
Perfect	0.0232	0.163000
Thunder	0.0438	0.006710
Psycho_(feat._Ty_Dolla_\$ign)	0.1170	0.580000

no_tears_left_to_cry	0.0594	0.040000
Everybody_Dies_In_Their_Nightmares	0.1330	0.847000
Jocelyn_Flores	0.2420	0.469000
HUMBLE.	0.1020	0.000282
In_My_Mind	0.1490	0.176000
Nice_For_What	0.0705	0.089100
Havana	0.0300	0.184000
Dusk_Till_Dawn_-_Radio_Edit	0.0390	0.101000
Candy_Paint	0.1530	0.627000
This_Is_Me	0.1860	0.005830
Siguelo_Bailando	0.0607	0.165000
I_Like_Me_Better	0.2530	0.535000
Believer	0.1280	0.062200
Downtown	0.1350	0.180000
Shape_of_You	0.0802	0.581000
Gods_Plan	0.1090	0.033200
Stir_Fry	0.2690	0.002990
One_Kiss_(with_Dua_Lipa)	0.1100	0.037000
Finesse_(Remix)_[feat._Cardi_B]	0.0996	0.018500
Nonstop	0.1240	0.016400
Mine	0.0722	0.016100
In_My_Feelings	0.1250	0.058900
Taste_(feat._Offset)	0.1200	0.023600
Lucid_Dreams	0.2000	0.349000
Promises_(with_Sam_Smith)	0.0394	0.011900

	instrumentalness	liveness \
Dura	0.000000	0.4090
Plug_Walk	0.000000	0.1080
Silence	0.000005	0.1700
Solo_(feat._Demi_Lovato)	0.000067	0.3500
Sin_Pijama	0.000029	0.1040
Call_Out_My_Name	0.000000	0.3500
Taki_Taki_(with_Selena_Gomez_Ozuna_&_Cardi_B)	0.000003	0.0618
River_(feat._Ed_Sheeran)	0.000000	0.0713
Rewrite_The_Stars	0.000000	0.1220
Him_&_I_(with_Halsey)	0.000000	0.3080
Look_Alive_(feat._Drake)	0.000059	0.1050
Rise	0.000000	0.2030
Perfect	0.000000	0.1060
Thunder	0.134000	0.1470
Psycho_(feat._Ty_Dolla_\$ign)	0.000000	0.1120
no_tears_left_to_cry	0.000003	0.2940
Everybody_Dies_In_Their_Nightmares	0.000021	0.1120
Jocelyn_Flores	0.000004	0.2970
HUMBLE.	0.000054	0.0958
In_My_Mind	0.000011	0.1180
Nice_For_What	0.000109	0.1190

Havana	0.000036	0.1320
Dusk_Till_Dawn_-_Radio_Edit	0.000001	0.1060
Candy_Paint	0.000001	0.0710
This_Is_Me	0.000115	0.0424
Siguelo_Bailando	0.000040	0.0937
I_Like_Me_Better	0.000003	0.1040
Believer	0.000000	0.0810
Downtown	0.000073	0.0680
Shape_of_You	0.000000	0.0931
Gods_Plan	0.000083	0.5520
Stir_Fry	0.000000	0.1590
One_Kiss_(with_Dua_Lipa)	0.000022	0.0814
Finesse_(Remix)_[feat._Cardi_B]	0.000000	0.0215
Nonstop	0.013000	0.1040
Mine	0.000003	0.4510
In_My_Feelings	0.000060	0.3960
Taste_(feat._Offset)	0.000000	0.1010
Lucid_Dreams	0.000000	0.3400
Promises_(with_Sam_Smith)	0.000005	0.3250

	valence	time_signature
Dura	0.8280	4
Plug_Walk	0.1580	4
Silence	0.2860	4
Solo_(feat._Demi_Lovato)	0.5650	4
Sin_Pijama	0.8200	4
Call_Out_My_Name	0.1720	3
Taki_Taki_(with_Selena_Gomez_Ozuna_&_Cardi_B)	0.5910	4
River_(feat._Ed_Sheeran)	0.6590	4
Rewrite_The_Stars	0.2840	4
Him_&_I_(with_Halsey)	0.1910	4
Look_Alive_(feat._Drake)	0.5950	4
Rise	0.6550	4
Perfect	0.1680	3
Thunder	0.2880	4
Psycho_(feat._Ty_Dolla_\$ign)	0.4390	4
no_tears_left_to_cry	0.3540	4
Everybody_Dies_In_Their_Nightmares	0.6890	4
Jocelyn_Flores	0.4370	4
HUMBLE.	0.4210	4
In_My_Mind	0.1630	4
Nice_For_What	0.7570	4
Havana	0.3940	4
Dusk_Till_Dawn_-_Radio_Edit	0.0967	4
Candy_Paint	0.4380	4
This_Is_Me	0.1000	4
Siguelo_Bailando	0.6260	4
I_Like_Me_Better	0.4190	4

Believer	0.6660	4
Downtown	0.6190	4
Shape_of_You	0.9310	4
Gods_Plan	0.3570	4
Stir_Fry	0.4980	4
One_Kiss_(with_Dua_Lipa)	0.5920	4
Finesse_(Remix)_[feat._Cardi_B]	0.9260	4
Nonstop	0.4220	4
Mine	0.7170	4
In_My_Feelings	0.3500	4
Taste_(feat._Offset)	0.3420	4
Lucid_Dreams	0.2180	4
Promises_(with_Sam_Smith)	0.4860	4

	danceability	energy	loudness	speechiness	\
Dura	0.791	0.848	-3.456	0.0506	
Plug_Walk	0.876	0.519	-6.531	0.1430	
Silence	0.520	0.761	-3.093	0.0853	
Solo_(feat._Demi_Lovato)	0.737	0.636	-4.546	0.0437	
Sin_Pijama	0.791	0.745	-3.695	0.0464	

	acousticness	instrumentalness	liveness	valence	\
Dura	0.1830	0.000000	0.409	0.828	
Plug_Walk	0.2020	0.000000	0.108	0.158	
Silence	0.2560	0.000005	0.170	0.286	
Solo_(feat._Demi_Lovato)	0.0441	0.000067	0.350	0.565	
Sin_Pijama	0.3540	0.000029	0.104	0.820	

time_signature	
Dura	4
Plug_Walk	4
Silence	4
Solo_(feat._Demi_Lovato)	4
Sin_Pijama	4

(40, 9)

	danceability	energy	loudness	speechiness	acousticness	\
count	40.000000	40.000000	40.000000	40.000000	40.000000	
mean	0.719200	0.661925	-5.846250	0.121292	0.184621	
std	0.151423	0.137874	1.611051	0.094732	0.209052	
min	0.258000	0.391000	-9.211000	0.023200	0.000282	
25%	0.680500	0.564250	-7.077000	0.049550	0.030800	
50%	0.750000	0.659000	-5.930000	0.105500	0.121500	
75%	0.817500	0.772500	-4.624000	0.144500	0.227500	
max	0.922000	0.909000	-3.093000	0.516000	0.847000	

	instrumentalness	liveness	valence	time_signature
count	40.000000	40.000000	40.000000	40.000000
mean	0.003695	0.177125	0.467942	3.950000
std	0.021231	0.130476	0.228504	0.220721

min	0.000000	0.021500	0.096700	3.000000
25%	0.000000	0.095275	0.287500	4.000000
50%	0.000003	0.112000	0.437500	4.000000
75%	0.000043	0.294750	0.633250	4.000000
max	0.134000	0.552000	0.931000	4.000000

[41]: *# PRUEBAS PARTE B*

```
# Declara la instancia de clase
nuevo = ACP_new(datos, n_componentes = 5)
# Despliega las Componentes Principales
print(nuevo.coordenadas_ind)
# Despliega los cosenos cuadrados de los individuos
print(nuevo.cos2_ind)
# Despliega las correlaciones de las variables con respecto a las componentes
print(nuevo.correlacion_var)

# Plotea el Plano Principal
nuevo.plot_plano_principal()

# Plotea el círculo de correlación
nuevo.plot_circulo()

# Plotea la sobreposición plano-correlación
nuevo.plot_sobreposicion()
```

	0	1	2 \
Dura	-2.278720	0.945061	-1.193561
Plug_Walk	1.159023	-0.951205	0.465000
Silence	-0.528992	1.505003	-0.419113
Solo_(feat._Demi_Lovato)	-0.472680	0.879671	-0.503637
Sin_Pijama	-2.021807	-0.048405	-1.292254
Call_Out_My_Name	2.551000	3.020351	-2.307211
Taki_Taki_(with_Selena_Gomez_Ozuna_&_Cardi_B)	-1.733442	-0.682530	0.317330
River_(feat._Ed_Sheeran)	-1.048800	-2.409509	1.528666
Rewrite_The_Stars	0.878545	0.346397	0.413156
Him_&_I_(with_Halsey)	0.773103	1.163493	0.514450
Look_Alive_(feat._Drake)	0.002071	-1.965323	1.105894
Rise	-1.242542	0.561984	-1.003659
Perfect	3.154030	1.859633	-1.831960
Thunder	-0.800442	3.389671	3.809890
Psycho_(feat._Ty_Dolla_\$ign)	1.113364	-1.533193	-0.613648
no_tears_left_to_cry	0.012229	0.968221	0.112182
Everybody_Dies_In_Their_Nightmares	0.273465	-2.042978	-1.649178
Jocelyn_Flores	2.077180	-2.662302	-0.094821
HUMBLE.	0.022720	-0.624251	0.617425
In_My_Mind	0.051550	0.347837	0.579414
Nice_For_What	-1.442100	0.635676	0.109055

Havana	0.113318	0.288816	-0.557940
Dusk_Till_Dawn_-_Radio_Edit	2.608916	1.453729	0.379019
Candy_Paint	0.181564	-1.021980	-0.786794
This_Is_Me	1.573995	1.013842	1.576090
Siguelo_Bailando	-0.363539	-0.743445	-0.061197
I_Like_Me_Better	1.208829	-2.219185	-0.088125
Believer	-1.687346	0.063340	0.071913
Downtown	-0.953459	-0.457103	-0.109839
Shape_of_You	-2.060495	-0.876378	-2.013576
Gods_Plan	2.370392	-0.336401	0.337225
Stir_Fry	-1.110471	-0.569362	1.041473
One_Kiss_(with_Dua_Lipa)	-2.217469	0.634064	0.049028
Finesse_(Remix)_[feat._Cardi_B]	-2.437364	0.173799	-0.006313
Nonstop	1.281419	-1.301899	1.015962
Mine	-1.529846	1.309448	-0.608460
In_My_Feelings	0.363528	0.132647	0.149148
Taste_(feat._Offset)	0.757438	-0.893124	0.786920
Lucid_Dreams	1.888410	-0.147748	0.127735
Promises_(with_Sam_Smith)	-0.486575	0.793637	0.034310

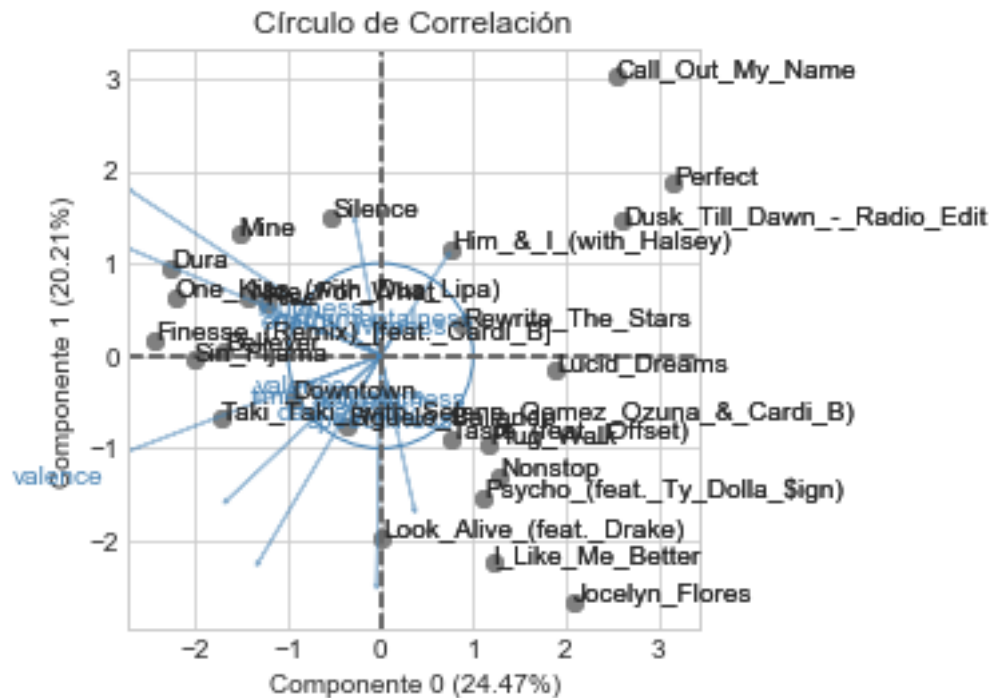
	3	4
Dura	1.535690	0.275984
Plug_Walk	0.056174	0.208525
Silence	-0.970490	-0.561157
Solo_(feat._Demi_Lovato)	1.456574	0.170764
Sin_Pijama	-0.472633	0.643775
Call_Out_My_Name	-0.341498	-0.695993
Taki_Taki_(with_Selena_Gomez_Ozuna_&_Cardi_B)	-0.538827	-0.737109
River_(feat._Ed_Sheeran)	-0.984258	-2.040711
Rewrite_The_Stars	0.070687	-0.010201
Him_&_I_(with_Halsey)	0.675089	-0.614978
Look_Alive_(feat._Drake)	0.579536	-0.511142
Rise	-0.121732	0.393035
Perfect	-1.099012	-0.363990
Thunder	-1.370757	3.558295
Psycho_(feat._Ty_Dolla_\$ign)	-0.813539	0.995005
no_tears_left_to_cry	0.971458	-0.231947
Everybody_Dies_In_Their_Nightmares	-1.386261	1.532217
Jocelyn_Flores	0.826979	0.910950
HUMBLE.	0.599884	0.016571
In_My_Mind	-0.590747	-0.706130
Nice_For_What	-0.337744	-0.575180
Havana	0.007523	0.527307
Dusk_Till_Dawn_-_Radio_Edit	-1.257134	-0.558619
Candy_Paint	-1.623082	0.505033
This_Is_Me	-1.690890	-1.857878
Siguelo_Bailando	0.268044	0.500261
I_Like_Me_Better	-0.962132	0.366113

Believer	-0.207148	-0.483219
Downtown	-0.442665	-0.103241
Shape_of_You	-0.903393	1.209270
Gods_Plan	3.000638	0.435119
Stir_Fry	0.246824	-1.239781
One_Kiss_(with_Dua_Lipa)	-0.255419	-0.663567
Finesse_(Remix)_[feat._Cardi_B]	-0.515468	-0.581613
Nonstop	0.717989	0.818026
Mine	1.883349	-0.224333
In_My_Feelings	1.861556	-0.007915
Taste_(feat._Offset)	0.561560	0.040158
Lucid_Dreams	0.041334	-0.312630
Promises_(with_Sam_Smith)	1.523943	-0.025076
	0	1 \
Dura	5.221386e-01	0.089810
Plug_Walk	5.349825e-01	0.360332
Silence	7.035785e-02	0.569495
Solo_(feat._Demi_Lovato)	6.568146e-02	0.227483
Sin_Pijama	6.389238e-01	0.000366
Call_Out_My_Name	3.019156e-01	0.423233
Taki_Taki_(with_Selena_Gomez_Ozuna_&_Cardi_B)	6.821338e-01	0.105753
River_(feat._Ed_Sheeran)	7.651624e-02	0.403855
Rewrite_The_Stars	7.229476e-01	0.112390
Him_&_I_(with_Halsey)	1.959629e-01	0.443840
Look_Alive_(feat._Drake)	7.550440e-07	0.679702
Rise	5.084734e-01	0.104015
Perfect	5.495311e-01	0.191036
Thunder	1.555633e-02	0.278973
Psycho_(feat._Ty_Dolla_\$ign)	2.206167e-01	0.418367
no_tears_left_to_cry	7.677800e-05	0.481308
Everybody_Dies_In_Their_Nightmares	6.654648e-03	0.371406
Jocelyn_Flores	3.338178e-01	0.548373
HUMBLE.	4.561878e-04	0.344384
In_My_Mind	2.033279e-03	0.092573
Nice_For_What	7.072367e-01	0.137419
Havana	1.872800e-02	0.121656
Dusk_Till_Dawn_-_Radio_Edit	6.212601e-01	0.192895
Candy_Paint	7.188396e-03	0.227751
This_Is_Me	2.014161e-01	0.083566
Siguelo_Bailando	1.307580e-01	0.546846
I_Like_Me_Better	1.960498e-01	0.660729
Believer	9.088352e-01	0.001281
Downtown	6.800948e-01	0.156312
Shape_of_You	3.741766e-01	0.067689
Gods_Plan	3.736174e-01	0.007525
Stir_Fry	2.908389e-01	0.076457
One_Kiss_(with_Dua_Lipa)	8.438347e-01	0.068994
Finesse_(Remix)_[feat._Cardi_B]	9.035391e-01	0.004594

Nonstop	2.956580e-01	0.305184
Mine	2.917282e-01	0.213727
In_My_Feelings	3.633122e-02	0.004837
Taste_(feat._Offset)	2.486197e-01	0.345673
Lucid_Dreams	9.628501e-01	0.005894
Promises_(with_Sam_Smith)	7.419888e-02	0.197397

	2	3	4
Dura	0.143249	0.237143	0.007659
Plug_Walk	0.086112	0.001257	0.017317
Silence	0.044165	0.236808	0.079174
Solo_(feat._Demi_Lovato)	0.074566	0.623697	0.008572
Sin_Pijama	0.261015	0.034915	0.064780
Call_Out_My_Name	0.246967	0.005411	0.022474
Taki_Taki_(with_Selena_Gomez_Ozuna_&_Cardi_B)	0.022860	0.065910	0.123343
River_(feat._Ed_Sheeran)	0.162552	0.067388	0.289688
Rewrite_The_Stars	0.159885	0.004680	0.000097
Him_&_I_(with_Halsey)	0.086773	0.149425	0.123999
Look_Alive_(feat._Drake)	0.215218	0.059103	0.045976
Rise	0.331756	0.004880	0.050876
Perfect	0.185393	0.066721	0.007319
Thunder	0.352430	0.045621	0.307419
Psycho_(feat._Ty_Dolla_\$ign)	0.067020	0.117793	0.176203
no_tears_left_to_cry	0.006461	0.484532	0.027622
Everybody_Dies_In_Their_Nightmares	0.242023	0.171006	0.208911
Jocelyn_Flores	0.000696	0.052912	0.064202
HUMBLE.	0.336894	0.318023	0.000243
In_My_Mind	0.256869	0.267016	0.381508
Nice_For_What	0.004044	0.038793	0.112507
Havana	0.454010	0.000083	0.405524
Dusk_Till_Dawn_-_Radio_Edit	0.013112	0.144250	0.028483
Candy_Paint	0.134989	0.574455	0.055618
This_Is_Me	0.201952	0.232444	0.280622
Siguelo_Bailando	0.003705	0.071085	0.247605
I_Like_Me_Better	0.001042	0.124196	0.017983
Believer	0.001651	0.013697	0.074536
Downtown	0.009026	0.146594	0.007974
Shape_of_You	0.357330	0.071926	0.128878
Gods_Plan	0.007562	0.598707	0.012589
Stir_Fry	0.255820	0.014369	0.362516
One_Kiss_(with_Dua_Lipa)	0.000413	0.011196	0.075564
Finesse_(Remix)_[feat._Cardi_B]	0.000006	0.040412	0.051449
Nonstop	0.185850	0.092820	0.120488
Mine	0.046147	0.442124	0.006273
In_My_Feelings	0.006116	0.952699	0.000017
Taste_(feat._Offset)	0.268351	0.136658	0.000699
Lucid_Dreams	0.004405	0.000461	0.026389
Promises_(with_Sam_Smith)	0.000369	0.727838	0.000197

	0	1	2	3	4
acousticness	0.096628	-0.439251	-0.545175	-0.442216	0.386642
danceability	-0.350337	-0.592663	-0.008921	0.435702	0.259060
energy	-0.822127	0.350493	0.134573	-0.083278	-0.226046
instrumentalness	-0.072786	0.386625	0.577708	-0.196551	0.626378
liveness	0.193149	0.287537	-0.173383	0.778742	0.073842
loudness	-0.716053	0.472953	-0.276422	-0.142776	-0.067611
speechiness	-0.011348	-0.657215	0.367686	-0.160167	-0.398630
time_signature	-0.441006	-0.415012	0.439005	0.156404	0.131071
valence	-0.802538	-0.303717	-0.293533	0.090824	0.119925



[]: