Proyecto de Titulación

Segmentación de Municipios de la República Mexicana

Pablo Gómez García

Objetivo

Plantear una solución desde un enfoque Analítico basada en un Modelo no supervisado que logre segmentar a los municipios de la República Mexicana. Se pretende que dicha segmentación resulte de utilidad para estrategias de implementación de Políticas Públicas, de Marketing y para estrategias de Seguridad Pública.

Introducción

En la actualidad existen en México diversos estudios a nivel municipio, en su mayoría muestran características poblacionales de cada uno de ellos, también muestran estados de opinión, preferencias electorales, índices de Pobreza y Marginación etc.

Toda la información que se tiene acerca de los municipios es valiosa para distintos fines, sin embargo a la fecha hay muy pocos estudios que describen a los municipios por medio de varias de sus características, si es que existe tales estudios lo hacen a nivel descriptivo simple.

Dentro de este proyecto planteamos la posibilidad de aplicar una técnica no supervisada para encontrar una segmentación muy característica de los municipios del País.

Planteamiento del Problema

Tomando como punto de partida la información pública que existe de los municipios en diversos temas, plateamos la siguiente Hipótesis:

“Existen variables entre los datos públicos de los municipios que permiten diferenciarlos en segmentos cuyos elementos que los conforman poseen características homogéneas entre sí pero heterogéneas con los elementos de otros segmentos”. Bajo el supuesto anterior se propone recaudar información pública de fuentes como el INEGI, la PGR, la CNBV entre otras más. Con esta información se procedería a construir un Modelo no supervisado para encontrar los segmentos planteados en la hipótesis. Cabe mencionar que este Modelo no tendrá como objetivo clasificar municipios que sean creados en el futuro pues esto no sucede a menudo, sin embargo busca mostrar de manera descriptiva el valor y aporte de cada segmento para los fines planteados en el objetivo del proyecto.

** Importar librerías **

In [1]:
```
#Importamos los paquetes que vamos a utilizar
import json as js
import random as rd
import math as mt
from time import time
from statistics import mean
import pandas as pd
import matplotlib.pyplot as plt
from nltk.corpus import stopwords
import nltk
import unicodedata
import re
from collections import Counter
import numpy as np
import io
import seaborn as sns
```

** Importar el conjunto de datos **

Fuente:

https://www.gob.mx/sesnsp/acciones-y-programas/datos-abiertos-de-incidencia-delictiva

https://drive.google.com/file/d/1caLtpjb1KahDK3dTTBOEwR2syA3FySrD/view?usp=sharing

In [8]:
```
#Cargamos los datos originales
import csv
```

```
with open('IDM_NM_sep22.csv', newline='', encoding='utf-8', errors='ignore') as csvfile:
    csv_reader = list(csv.reader(csvfile, delimiter=','))
#print(csv_reader)
```

```
IOPub data rate exceeded.
The notebook server will temporarily stop sending output
to the client in order to avoid crashing it.
To change this limit, set the config variable
`--NotebookApp.iopub_data_rate_limit`.

Current values:
NotebookApp.iopub_data_rate_limit=1000000.0 (bytes/sec)
NotebookApp.rate_limit_window=3.0 (secs)
```

In [21]:

```
df_natural=pd.DataFrame(csv_reader[1:],columns=csv_reader[0])
df_natural
```

Out[21]:

	Anio	Clave_Ent	Entidad	Cve_Municipio	Municipio	Bien_juridico_afectado	Tipo_delito	Subtipo_delito	Modalidad	Enero
0	2015	1	Aguascalientes	1001	Aguascalientes	La vida y la Integridad corporal	Homicidio	Homicidio doloso	Con arma de fuego	2
1	2015	1	Aguascalientes	1001	Aguascalientes	La vida y la Integridad corporal	Homicidio	Homicidio doloso	Con arma blanca	1
2	2015	1	Aguascalientes	1001	Aguascalientes	La vida y la Integridad corporal	Homicidio	Homicidio doloso	Con otro elemento	0
3	2015	1	Aguascalientes	1001	Aguascalientes	La vida y la Integridad corporal	Homicidio	Homicidio doloso	No especificado	1
4	2015	1	Aguascalientes	1001	Aguascalientes	La vida y la Integridad corporal	Homicidio	Homicidio culposo	Con arma de fuego	0
...	...	...	...	...	...	...	...	...	...	...
1832399	2022	32	Zacatecas	32058	Santa Mara de la Paz	Otros bienes jurdicos afectados (del fuero comn)	Falsificacin	Falsificacin	Falsificacin	0
1832400	2022	32	Zacatecas	32058	Santa Mara de la Paz	Otros bienes jurdicos afectados (del fuero comn)	Contra el medio ambiente	Contra el medio ambiente	Contra el medio ambiente	0
1832401	2022	32	Zacatecas	32058	Santa Mara de la Paz	Otros bienes jurdicos afectados (del fuero comn)	Delitos cometidos por servidores pblicos	Delitos cometidos por servidores pblicos	Delitos cometidos por servidores pblicos	0
1832402	2022	32	Zacatecas	32058	Santa Mara de la Paz	Otros bienes jurdicos afectados (del fuero comn)	Electorales	Electorales	Electorales	0
1832403	2022	32	Zacatecas	32058	Santa Mara de la Paz	Otros bienes jurdicos afectados (del fuero comn)	Otros delitos del Fuero Comn	Otros delitos del Fuero Comn	Otros delitos del Fuero Comn	0

1832404 rows × 21 columns

In [51]:

```
#Cargamos los datos que previamente fueron Manipulados con el software R para obtner la siguiente tabla
# Se importa  el archivo con los nombres de municipios y Estados
df_ini=pd.read_csv('Delitos_totales_2022_final.csv')
df = df_ini.set_index('id')
drop_col=['Entidad','Municipio','XCOORD','YCOORD']
df.drop(drop_col, inplace=True, axis=1)
df
```

Out[51]:

id	Aborto	Abuso_de_confianza	Abuso_sexual	Acoso_sexual	Allanamiento_de_morada	Amenazas	Contra_el_medio_ambiente	Corrupc
1010001	0.000000	0.291667	0.177083	0.031250	0.010417	0.479167	0.000000	
1010002	0.000000	0.000000	0.010417	0.000000	0.000000	0.062500	0.000000	
1010003	0.000000	0.000000	0.000000	0.000000	0.000000	0.083333	0.000000	
1010004	0.000000	0.468750	0.333333	0.041667	0.062500	0.947917	0.000000	
1010005	0.125000	34.020833	17.354167	1.895833	8.583333	66.093750	0.041667	
...	...	...	...	...	...	...	...	
99014	0.281250	34.770833	10.406250	3.895833	4.781250	61.072917	3.989583	
99015	0.552083	66.395833	34.760417	11.427083	6.125000	131.354167	6.343750	

99016	0.364583	22.166667	12.062500	4.020833	3.500000	52.333333	3.312500
99017	0.854167	19.531250	14.541667	3.239583	3.187500	63.093750	2.291667
99998	0.072917	0.187500	1.447917	0.604167	0.000000	0.656250	0.156250

2479 rows × 40 columns

** Análisis exploratorio de datos **

In [17]:
```
# Visualizamos las columnas que contiene el dataset
df.columns
```

Out[17]:
```
Index(['Aborto', 'Abuso_de_confianza', 'Abuso_sexual', 'Acoso_sexual',
       'Allanamiento_de_morada', 'Amenazas', 'Contra_el_medio_ambiente',
       'Corrupcion_de_menores', 'Dano_a_la_propiedad',
       'por_servidores_publicos', 'Despojo', 'Electorales',
       'Evasion_de_presos', 'Extorsion', 'Falsedad', 'Falsificacion',
       'Feminicidio', 'Fraude', 'Homicidio', 'Hostigamiento_sexual', 'Incesto',
       'Incump_obligaciones_asistencia_fam', 'Lesiones', 'Narcomenudeo',
       'contra_el_patrimonio', 'Otros_contra_la_familia',
       'Otros_contra_la_sociedad', 'Otros_Fuero_Comun', 'libertad_personal',
       'libertad_seguridad_sexual', 'vida_integridad_corporal', 'Rapto',
       'Robo', 'Secuestro', 'Trafico_de_menores', 'Trata_de_personas',
       'Violacion_equiparada', 'Violacion_simple', 'Violencia_de_genero',
       'Violencia_familiar'],
      dtype='object')
```

In [18]:
```
df.dtypes
```

Out[18]:
```
Aborto                              float64
Abuso_de_confianza                  float64
Abuso_sexual                        float64
Acoso_sexual                        float64
Allanamiento_de_morada              float64
Amenazas                            float64
Contra_el_medio_ambiente            float64
Corrupcion_de_menores               float64
Dano_a_la_propiedad                 float64
por_servidores_publicos             float64
Despojo                             float64
Electorales                         float64
Evasion_de_presos                   float64
Extorsion                           float64
Falsedad                            float64
Falsificacion                       float64
Feminicidio                         float64
Fraude                              float64
Homicidio                           float64
Hostigamiento_sexual                float64
Incesto                             float64
Incump_obligaciones_asistencia_fam  float64
Lesiones                            float64
Narcomenudeo                        float64
contra_el_patrimonio                float64
Otros_contra_la_familia             float64
Otros_contra_la_sociedad            float64
Otros_Fuero_Comun                   float64
libertad_personal                   float64
libertad_seguridad_sexual           float64
vida_integridad_corporal            float64
Rapto                               float64
Robo                                float64
Secuestro                           float64
Trafico_de_menores                  float64
Trata_de_personas                   float64
Violacion_equiparada                float64
Violacion_simple                    float64
Violencia_de_genero                 float64
Violencia_familiar                  float64
dtype: object
```

In [19]:
```
#Visualizamos los Estadigrafos de la variables
df_describe=df.describe().T
```

```
df_describe
```

Out[19]:

	count	mean	std	min	25%	50%	75%	max
Aborto	2479.0	0.020641	0.103366	0.0	0.00	0.00	0.010	1.97
Abuso_de_confianza	2479.0	0.824712	3.707413	0.0	0.00	0.04	0.240	66.40
Abuso_sexual	2479.0	0.676958	2.973642	0.0	0.01	0.05	0.220	50.75
Acoso_sexual	2479.0	0.134433	0.677515	0.0	0.00	0.00	0.030	15.64
Allanamiento_de_morada	2479.0	0.434885	3.573997	0.0	0.00	0.03	0.150	154.16
Amenazas	2479.0	3.179000	14.026048	0.0	0.04	0.21	0.930	217.26
Contra_el_medio_ambiente	2479.0	0.060803	0.398351	0.0	0.00	0.00	0.010	9.18
Corrupcion_de_menores	2479.0	0.071154	0.512703	0.0	0.00	0.00	0.020	16.99
Dano_a_la_propiedad	2479.0	4.244078	18.110651	0.0	0.06	0.29	1.445	308.27
por_servidores_publicos	2479.0	0.588495	3.766842	0.0	0.01	0.04	0.130	104.36
Despojo	2479.0	0.874296	3.068362	0.0	0.03	0.14	0.440	45.21
Electorales	2479.0	0.043667	0.163427	0.0	0.00	0.01	0.030	2.70
Evasion_de_presos	2479.0	0.003780	0.020601	0.0	0.00	0.00	0.000	0.41
Extorsion	2479.0	0.243852	1.074441	0.0	0.00	0.01	0.080	21.84
Falsedad	2479.0	0.104558	0.753174	0.0	0.00	0.00	0.020	27.36
Falsificacion	2479.0	0.543647	3.503101	0.0	0.00	0.01	0.070	83.49
Feminicidio	2479.0	0.006523	0.020495	0.0	0.00	0.00	0.010	0.36
Fraude	2479.0	2.399137	12.807845	0.0	0.02	0.10	0.570	294.99
Homicidio	2479.0	0.148471	0.548769	0.0	0.01	0.03	0.090	15.85
Hostigamiento_sexual	2479.0	0.049484	0.304511	0.0	0.00	0.00	0.020	6.95
Incesto	2479.0	0.000710	0.010123	0.0	0.00	0.00	0.000	0.45
Incump_obligaciones_asistencia_fam	2479.0	0.764712	3.491686	0.0	0.00	0.03	0.230	73.43
Lesiones	2479.0	0.732283	2.920760	0.0	0.01	0.06	0.280	55.86
Narcomenudeo	2479.0	1.937608	19.184218	0.0	0.00	0.04	0.250	731.50
contra_el_patrimonio	2479.0	0.348616	2.120363	0.0	0.00	0.01	0.070	43.01
Otros_contra_la_familia	2479.0	0.386023	2.685532	0.0	0.00	0.01	0.060	80.17
Otros_contra_la_sociedad	2479.0	0.186837	1.187383	0.0	0.00	0.00	0.020	32.25
Otros_Fuero_Comun	2479.0	5.899855	28.703586	0.0	0.06	0.25	1.355	523.40
libertad_personal	2479.0	0.582711	2.463326	0.0	0.00	0.04	0.180	40.28
libertad_seguridad_sexual	2479.0	0.228439	1.106819	0.0	0.00	0.02	0.080	26.36
vida_integridad_corporal	2479.0	0.258766	1.633317	0.0	0.00	0.01	0.070	46.20
Rapto	2479.0	0.005389	0.054069	0.0	0.00	0.00	0.000	1.66
Robo	2479.0	0.629685	3.041599	0.0	0.00	0.03	0.150	54.31
Secuestro	2479.0	0.006196	0.022908	0.0	0.00	0.00	0.000	0.44
Trafico_de_menores	2479.0	0.002481	0.045548	0.0	0.00	0.00	0.000	2.09
Trata_de_personas	2479.0	0.016002	0.112320	0.0	0.00	0.00	0.000	3.26
Violacion_equiparada	2479.0	0.126370	0.640366	0.0	0.00	0.01	0.040	16.88
Violacion_simple	2479.0	0.412291	1.588285	0.0	0.01	0.05	0.190	33.27
Violencia_de_genero	2479.0	0.098362	0.739735	0.0	0.00	0.00	0.000	19.48
Violencia_familiar	2479.0	6.418080	27.651832	0.0	0.08	0.38	1.830	514.15

In [20]:
```
# Obtenemos el % total de valores perdidos que aún quedan en la base
df.isnull().values.mean() * 100
```

Out[20]: 0.0

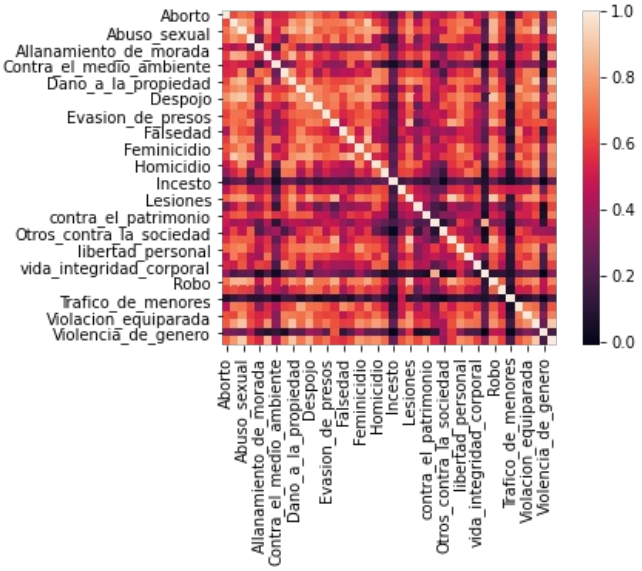
In [21]:
```
df.shape
```

Out[21]: (2479, 40)

Approach 1

In [23]:
```
# Análisis de correlación entre las variables
sns.heatmap(df.corr(method='pearson'),square=True)
```

Out[23]: <AxesSubplot:>



Clustering de variables

In [ ]:
```
#!pip install varclushi
```

In [25]:
```
#Clusterin de variables
from varclushi import VarClusHi
#Realizamos el clustering
var_clust_model=VarClusHi(df,maxeigval2=0.7, maxclus=None)
var_clust_model.varclus()
```

Out[25]: <varclushi.varclushi.VarClusHi at 0x225c15f9d30>

In [26]:
```
#Visualizamos los clusters formados
var_clust_model.rsquare
```

Out[26]:

	Cluster	Variable	RS_Own	RS_NC	RS_Ratio
0	0	Abuso_de_confianza	0.830808	0.769288	0.733349
1	0	por_servidores_publicos	0.824154	0.621512	0.464602
2	0	Falsedad	0.760237	0.409843	0.406269
3	0	Falsificacion	0.822116	0.567534	0.411324
4	0	Fraude	0.909022	0.681342	0.285505
5	0	Trata_de_personas	0.770087	0.522953	0.481950
6	1	Abuso_sexual	0.872288	0.757028	0.525624
7	1	Amenazas	0.838310	0.567483	0.373835
8	1	Dano_a_la_propiedad	0.895934	0.750435	0.416989
9	1	contra_el_patrimonio	0.586850	0.414141	0.705204
10	1	libertad_seguridad_sexual	0.599838	0.477581	0.765978
11	1	Violacion_simple	0.874484	0.742551	0.487539
12	1	Violencia_familiar	0.934205	0.729318	0.243070
13	2	Otros_contra_la_familia	0.919739	0.276414	0.110921
14	2	Rapto	0.919739	0.137501	0.093056
15	3	Feminicidio	0.775939	0.627296	0.601178
16	3	libertad_personal	0.798321	0.570432	0.469494
17	3	Secuestro	0.636053	0.453582	0.666061

18	3	Aborto	0.690397	0.459194	0.572485
19	3	Despojo	0.843609	0.798878	0.777591
20	3	Evasion_de_presos	0.668282	0.515286	0.684358
21	4	Trafico_de_menores	1.000000	0.105126	0.000000
22	5	Otros_Fuero_Comun	0.889798	0.500419	0.220589
23	5	Violencia_de_genero	0.492870	0.190336	0.626346
24	5	Extorsion	0.877798	0.651777	0.350929
25	5	Lesiones	0.920972	0.700007	0.263434
26	5	Robo	0.847986	0.753226	0.616005
27	6	Incesto	1.000000	0.074333	0.000000
28	7	Homicidio	0.811694	0.607992	0.480364
29	7	Narcomenudeo	0.682930	0.382159	0.513191
30	7	Incump_obligaciones_asistencia_fam	0.667902	0.432368	0.585058
31	8	Contra_el_medio_ambiente	1.000000	0.335585	0.000000
32	9	Allanamiento_de_morada	0.627176	0.341939	0.566550
33	9	Corrupcion_de_menores	0.853143	0.592095	0.360026
34	9	vida_integridad_corporal	0.760559	0.437726	0.425843
35	9	Hostigamiento_sexual	0.678358	0.521357	0.671986
36	9	Violacion_equiparada	0.705156	0.568175	0.682785
37	10	Acoso_sexual	0.811100	0.601113	0.473568
38	10	Electorales	0.804232	0.648085	0.556294
39	10	Otros_contra_la_sociedad	0.752623	0.377773	0.397567

In [27]:
```
#Verificamos el proceso anterior
var_clus=var_clust_model.rsquare
var_clus.to_csv('var_clust.csv')
```

Observación

Tras el Análisis de Clustering de variables podemos tomar la variable con la distancia menor al centroide dentro de cada cluster, esto es con el valor del RS_Ratio y así elegiriamos una variable representante de cada grupo de acuerdo a su nivel de correlación. Por otro lado aquellos cluster cuyo numero de variables es uno significa que son variables que no se correlaconan con ninguna otra.

El criterio tomado es que la iteración del algoritmo pare cuando se tenga el 70% de la Varianza explicada.

En este caso de 40 variables que entraron al análisis podemos xplicar el 70% de la varianza del fenómeno con solo con solo 11 variables.

In [53]:
```
#Eliminamos la variables correlacionada de nuestro Data set original
col_drop2 = ['Falsedad','Falsificacion','por_servidores_publicos','Trata_de_personas','Abuso_de_confianza','Amena
             'Dano_a_la_propiedad','Violacion_simple','Abuso_sexual','contra_el_patrimonio','libertad_seguridad_se
             'Otros_contra_la_familia','Aborto','Feminicidio','Secuestro','Evasion_de_presos','Despojo','Lesiones'
             'Extorsion','Robo','Violencia_de_genero','Narcomenudeo','Incump_obligaciones_asistencia_fam',
              'vida_integridad_corporal','Allanamiento_de_morada','Hostigamiento_sexual','Violacion_equiparada',
              'Acoso_sexual','Electorales']
df.drop(col_drop2, inplace=True, axis=1)
```

In [29]:
```
df
```

Out[29]:

id	Contra_el_medio_ambiente	Corrupcion_de_menores	Fraude	Homicidio	Incesto	Otros_contra_la_sociedad	Otros_Fuero_Comun	liberta
1010001	0.00	0.00	1.32	0.08	0.00	0.00	2.35	
1010002	0.00	0.00	0.00	0.02	0.00	0.00	0.07	
1010003	0.00	0.00	0.00	0.01	0.00	0.00	0.06	
1010004	0.00	0.00	0.74	0.10	0.00	0.02	1.83	
1010005	0.04	0.99	104.42	0.87	0.01	1.74	34.98	
...	...	...	...	...	...	...	...	
99014	3.99	0.76	154.69	0.53	0.00	5.32	58.61	
99015	6.34	2.58	294.99	1.43	0.00	18.76	122.91	

99016	3.31	0.74	114.60	0.67	0.00	5.34	45.83
99017	2.29	0.83	46.89	1.51	0.00	4.08	29.92
99998	0.16	0.24	1.59	0.01	0.00	0.32	5.84

2479 rows × 11 columns

** Modelación **

En esta parte vamos a probar un modelo simple de kmeans para tener un primer acercamiento de si los datos propuestos arrojan resultados orientados a probar nuestra hipótesis.

** Análisis de Clustering: K-Means **

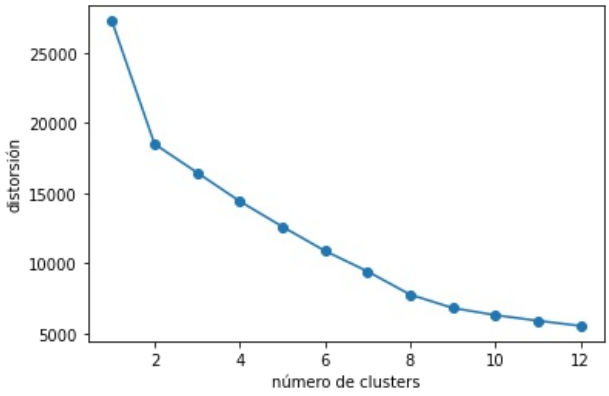
In [56]:
```
# Realizamos un proceso de estandarización de las variables
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
X = pd.DataFrame(scaler.fit_transform(df), index=df.index, columns=df.columns)
```

In [32]:
```
# cluster es la clase para implementar agrupamiento con sklearn.
from sklearn import cluster
```

Método del codo

In [57]:
```
# Ejecución de K-means con 12 valores de clusters.
distorsion = []
for i in range(1,13):
  modeloK = cluster.KMeans(n_clusters = i)
  modeloK.fit(X)
  distorsion.append(modeloK.inertia_)
```

In [58]:
```
# Grafica de distorsión para identificar el codo de la curva con el que se
# obtiene el número óptimo de clusters.
clusters = np.linspace(1,12,12)
plt.plot(clusters, distorsion, marker = 'o')
plt.xlabel('número de clusters')
plt.ylabel('distorsión')
plt.show()
```



In [59]:
```
# Construcción del modelo para obtener etiquetas mediante kmeans.
# Debemos indicar el numero de clusters (grupos) que el algorítmo ajustará.
modeloK = cluster.KMeans(n_clusters = 7)
```

In [60]:
```
# Ajuste de los clusters.
modeloK.fit(X)
```

Out[60]:
```
KMeans(n_clusters=7)
```

In [63]:
```
prediccion = modeloK.predict(X)
```

```
tabla={'id':df_ini['id'],
    'cluster':prediccion}
clusters=pd.DataFrame(tabla)
#Cruzamos los data frames de frecuencias de bigramas y de unigramas
df_final=pd.merge(df_ini,clusters, left_on='id', right_on='id')
df_final
```

Out[63]:

	id	Entidad	Municipio	Aborto	Abuso_de_confianza	Abuso_sexual	Acoso_sexual	Allanamiento_de_morada	Amenazas	Contra
0	1010001	Durango	Canatlan	0.000000	0.291667	0.177083	0.031250	0.010417	0.479167	
1	1010002	Durango	Canelas	0.000000	0.000000	0.010417	0.000000	0.000000	0.062500	
2	1010003	Durango	Coneto de Comonfort	0.000000	0.000000	0.000000	0.000000	0.000000	0.083333	
3	1010004	Durango	Cuencame	0.000000	0.468750	0.333333	0.041667	0.062500	0.947917	
4	1010005	Durango	Durango	0.125000	34.020833	17.354167	1.895833	8.583333	66.093750	
...	...	...	...	...	...	...	...	...	...	
2474	99014	Ciudad de Mexico	Benito Juarez	0.281250	34.770833	10.406250	3.895833	4.781250	61.072917	
2475	99015	Ciudad de Mexico	Cuauhtemoc	0.552083	66.395833	34.760417	11.427083	6.125000	131.354167	
2476	99016	Ciudad de Mexico	Miguel Hidalgo	0.364583	22.166667	12.062500	4.020833	3.500000	52.333333	
2477	99017	Ciudad de Mexico	Venustiano Carranza	0.854167	19.531250	14.541667	3.239583	3.187500	63.093750	
2478	99998	Ciudad de Mexico	No Especificado	0.072917	0.187500	1.447917	0.604167	0.000000	0.656250	

2479 rows × 46 columns

Resultados del Análisis

In [6]:

```
from IPython.display import Image
Image(filename = "resultados.png", width=600, height=5)
```

Out[6]:

In [64]:

```
df_final.to_csv('clusters.csv')
```

Conclusiones

Como primer acercamiento a los resultados buscados de nuestro proyecto, concluimos lo siguiente:

1. 7 segmentos no son suficientes para discriminar la totalidad de los municipios
2. Se logra diferenciar el 5% de los municipios del resto como aquellos con mayo incidencia delictiva en fraudes, secuestros, trata de

personas y homicidios.
3. Es necesario incorporar más variables que ayuden a discriminar mejor a los municipios.
4. Detectamos un sesgo en cuanto a la cultura de la denuncia, es decir municipios que se perciben con altos índices de delitos (por ejemplo los mencionados constantemente en las noticias azotados por el narcotráfico) no reflejan dicha característica en los datos oficiales.

Referencias

1. Mitchell, Tom, “Machine Learning”, Ed. McGraw-Hill (1997), cap 6 pp 154-199.
2. Everitt, B.S. (2011). Cluster analysis, 5th Edition. Wiley.
3. Peña Sánchez de Rivera, D. “Estadística. Modelos y Métodos. Volumen 2” Ed. Alianza. Madrid, 1987
4. Introduction to machine learning, Third Edition. Ethem Alpaydin. MIT Press
5. Understanding Machine Learning: From Theory to Algorithms. Shai Shalev-Shwartz and Shai Ben-David. Cambridge University Press.