Proyecto Minería de Datos
Sección 20
Análisis exploratorio y clustering
Matrimonios

Miembros:
JOSE MARIANO REYES HERNANDEZ
ANDREA DE LOURDES LAM PELAEZ
JESSICA PAMELA ORTIZ IXCOT
KENNETH EDUARDO GALVEZ REQUENA

Índice

Situación Problemática:

Las tendencias en el matrimonio y las preferencias de unión han cambiado mucho con el tiempo, lo que hace difícil entender cómo estos cambios afectan a las personas y a la sociedad. Al mismo tiempo, las diferencias entre hombres y mujeres, cómo la edad influye en la elección de un tipo de unión y el papel de factores socioeconómicos y leyes en las tasas de matrimonio se han vuelto más complicados.

Esta falta de entendimiento dificulta la labor de quienes toman decisiones políticas y de organizaciones para solucionar problemas relacionados con la vida familiar y las relaciones de pareja. Al analizar las tendencias matrimoniales, diferencias de género, cambios en las preferencias de unión y otros aspectos, se busca obtener información útil para mejorar la calidad de vida de las parejas y sus comunidades. Comprender estos factores permitirá desarrollar estrategias efectivas para mejorar el bienestar y la estabilidad de las familias en el futuro.

Problema científico:

El problema científico radica en identificar y comprender cómo las tendencias en el matrimonio y las preferencias de unión, las diferencias de género, la relación entre la edad y la clase de unión, y la influencia de factores socioeconómicos y legislativos afectan las tasas de matrimonio y las clases de unión a lo largo del tiempo. Esta comprensión es esencial para establecer relaciones causales y predecir futuras tendencias matrimoniales, lo que permitirá a los responsables de la toma de decisiones desarrollar políticas y programas efectivos para abordar problemas sociales y económicos relacionados con la vida familiar y las relaciones de pareja.

Objetivos:

Objetivo General:

Analizar las tendencias matrimoniales y las preferencias de unión en función del género, la edad, factores socioeconómicos y cambios legislativos para comprender su impacto en la sociedad y desarrollar estrategias efectivas que mejoren la calidad de vida de las parejas y sus comunidades.

Objetivos Específicos:

- ☐ Examinar las diferencias de género en las tasas de matrimonio y preferencias de unión, así como investigar la relación entre la edad y la clase de unión elegida para identificar patrones y cambios significativos en las tendencias matrimoniales a lo largo del tiempo.
- ☐ Evaluar el impacto de los factores socioeconómicos y cambios legislativos en las tasas de matrimonio y las clases de unión, con el fin de determinar qué factores influyen en las decisiones de las parejas y cómo estos pueden ser abordados en políticas y programas efectivos.

Descripción de los datos:

Los datos analizados fueron un total de 751478 filas, con información en 30 columnas. Las columnas variaban en todo tipo de elementos, las cuales se muestran a continuación:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 753177 entries, 0 to 753176
Data columns (total 30 columns):
 #   Column    Non-Null Count   Dtype
---  ------    --------------   -----
 0   DEPREG    753177 non-null  int64
 1   MUPREG    753177 non-null  int64
 2   MESREG    753177 non-null  int64
 3   AÑOREG    753177 non-null  int64
 4   CLAUNI    753177 non-null  int64
 5   EDADHOM   753177 non-null  int64
 6   EDADMUJ   753177 non-null  int64
 7   GETHOM    297767 non-null  float64
 8   GETMUJ    297767 non-null  float64
 9   NACHOM    753177 non-null  int64
 10  NACMUJ    753177 non-null  int64
 11  OCUHOM    297767 non-null  float64
 12  OCUMUJ    297767 non-null  float64
 13  NUPHON    62104 non-null   float64
 14  NUPMUJ    62104 non-null   float64
 15  DEPOCU    753177 non-null  int64
 16  MUPOCU    753177 non-null  object
 17  MESOCU    753177 non-null  int64
 18  AÑOOCU    508678 non-null  float64
 19  AREAG     297767 non-null  float64
 20  ESCHOM    691073 non-null  float64
 21  ESCMUJ    691073 non-null  float64
 22  DIAOCU    691073 non-null  float64
 23  PUEHOM    455410 non-null  float64
 24  PUEMUJ    455410 non-null  float64
 25  CIUOHOM   455410 non-null  object
 26  CIUOMUJ   455410 non-null  object
 27  AREAGOCU  380630 non-null  float64
 28  NUNUHO    295164 non-null  float64
 29  NUNUMU    295164 non-null  float64
dtypes: float64(16), int64(11), object(3)
memory usage: 172.4+ MB
```

Para un mejor entendimiento, el nombre completo de las variables es el siguiente.

1. DEPREG: Departamento de registro (1 a 22 definiendo cada departamento)

2. MUNREG: Municipio de registro (Todos los códigos de municipios del país)
3. MESREG: Mes de registro (1 a 12)
4. AÑOREG: Año de registro
5. CLAUNI: Clase de union (1 Comunidad absoluta, 2 Separación absoluta, 3 Comunidad de gananciales, 9 No especificado)
6. EDADHOM: Edad del hombre
7. EDADMUJ: Edad de la mujer
8. GETHOM: Grupo étnico del hombre (1 Indígena, 2 no indígena, 9 ignorado)
9. GETMUJ: Grupo étnico de la mujer (1 Indígena, 2 no indígena, 9 ignorado)
10. NACHOM: Nacionalidad del hombre (códigos de países definidos en el excel)
11. NACMUJ: Nacionalidad de la mujer (códigos de países definidos en el excel)
12. OCUHOM: Ocupación (Subgrupos, CIUO-08) del hombre, (códigos de ocupaciones definidos en el excel)
13. OCUMUJ: Ocupación (Subgrupos, CIUO-08) de la mujer, (códigos de ocupaciones definidos en el excel)
14. NUPHOM: Número de nupcias del hombre (9 si se ignora)
15. NUPMUJ: Número de nupcias de la mujer (9 si se ignora)
16. DEPOCU: Departamento de ocurrencia (1 a 22 definiendo cada departamento)
17. MUNOCU: Municipio de ocurrencia (Todos los códigos de municipios del país)
18. MESOCU: Mes de ocurrencia
19. AÑOOCU: Año ocurrencia
20. AREAG: Área geográfica de ocurrencia (1 Urbano, 2 Rural, 9 Ignorado)
21. ESCHOM: Escolaridad del hombre
22. ESCMUJ: Escolaridad de la mujer
23. DIAOCU: Día de ocurrencia
24. PUEHOM: Pueblo de pertenencia del hombre (1 Maya, 2 Garífuna, 3 Xinka, 4 Mestizo / Ladino, 5 Otro, 9 Ignorado)
25. PUEMUJ: Pueblo de pertenencia de la mujer (1 Maya, 2 Garífuna, 3 Xinka, 4 Mestizo / Ladino, 5 Otro, 9 Ignorado)
26. CIUOHOM: Ciudad de Ocurrencia Hombre
27. CIUOMUJ: Ciudad de Ocurrencia Mujer
28. AREAGOCU: No hay información de esta columna
29. NUNUHO: No hay información de esta columna
30. NUNUMU: No hay información de esta columna

Limpieza de datos:

Para la limpieza se realizaron los siguientes pasos:

```
from scipy import stats

# reemplazar valores faltantes
df.fillna(9, inplace=True)

# eliminar filas duplicadas
df.drop_duplicates(inplace=True)

# reemplazar los valores que contienen el formato 1-XXX por NaN en todo el dataframe
df = df.applymap(lambda x: np.nan if isinstance(x, str) and re.match('^1-', x) else x)
df = df.applymap(lambda x: np.nan if isinstance(x, str) and re.match('^13-', x) else x)

# eliminar las filas que contienen NaN en todo el dataframe
df.dropna(inplace=True)

# reemplazar todos los valores 'NEOG' por 0
df.replace('NEOG', 0, inplace=True)

# Convertir tipos de datos
df["GETHOM"] = df["GETHOM"].astype(int)
df["GETMUJ"] = df["GETMUJ"].astype(int)
df["OCUHOM"] = df["OCUHOM"].astype(int)
df["OCUMUJ"] = df["OCUMUJ"].astype(int)
df["NUPHON"] = df["NUPHON"].astype(int)
df["NUPMUJ"] = df["NUPMUJ"].astype(int)
df["MUPOCU"] = df["MUPOCU"].astype(int)
df["AÑOOCU"] = df["AÑOOCU"].astype(int)
df["AREAG"] = df["AREAG"].astype(int)
df["ESCHOM"] = df["ESCHOM"].astype(int)
df["ESCMUJ"] = df["ESCMUJ"].astype(int)
df["DIAOCU"] = df["DIAOCU"].astype(int)
```

```
# Calcular el z-score para cada valor de la columna
z_scores = stats.zscore(df['NUPHON'])
z_scores2 = stats.zscore(df['NUPMUJ'])
z_scores3 = stats.zscore(df['NUPMUJ'])

# Identificar los valores que están a más de 3 desviaciones estándar de la media
outliers = (abs(z_scores) > 3)
outliers2 = (abs(z_scores2) > 3)
outliers3 = (abs(z_scores3) > 3)

# Eliminar los valores atípicos
df = df.loc[~outliers]
df = df.loc[~outliers2]
df = df.loc[~outliers3]


# eliminar columnas irrelevantes
df.drop(['PUEHOM', 'PUEMUJ','CIUOHOM', 'CIUOMUJ', 'AREAGOCU','NUNUHO', 'NUNUMU'], axis=1, inplace=True)
```

Y de esta manera queda la data resultado

```
 #   Column   Non-Null Count   Dtype
---  ------   --------------   -----
 0   DEPREG   751493 non-null  int64
 1   MUPREG   751493 non-null  int64
 2   MESREG   751493 non-null  int64
 3   AÑOREG   751493 non-null  int64
 4   CLAUNI   751493 non-null  int64
 5   EDADHOM  751493 non-null  int64
 6   EDADMUJ  751493 non-null  int64
 7   GETHOM   751493 non-null  int64
 8   GETMUJ   751493 non-null  int64
 9   NACHOM   751493 non-null  int64
 10  NACMUJ   751493 non-null  int64
 11  OCUHOM   751493 non-null  int64
 12  OCUMUJ   751493 non-null  int64
 13  NUPHON   751493 non-null  int64
 14  NUPMUJ   751493 non-null  int64
 15  DEPOCU   751493 non-null  int64
 16  MUPOCU   751493 non-null  int64
 17  MESOCU   751493 non-null  int64
 18  AÑOOCU   751493 non-null  int64
 19  AREAG    751493 non-null  int64
...
 21  ESCMUJ   751493 non-null  int64
 22  DIAOCU   751493 non-null  int64
dtypes: int64(23)
memory usage: 131.9 MB
```

	DEPREG	MUPREG	MESREG	AÑOREG	CLAUNI	EDADHOM	EDADMUJ	GETHOM	GETMUJ
count	751493.000000	751493.000000	751493.000000	751493.000000	751493.000000	751493.000000	751493.000000	751493.000000	751493.000000
mean	9.499703	957.456383	6.470240	1851.491023	3.006594	32.233748	29.618972	1.299258	1.370263
std	6.401404	640.704754	3.550853	547.249387	1.520520	56.132933	59.578843	2.565399	2.671688
min	1.000000	101.000000	1.000000	9.000000	1.000000	12.000000	10.000000	0.000000	0.000000
25%	4.000000	401.000000	3.000000	2011.000000	3.000000	22.000000	20.000000	0.000000	0.000000
50%	10.000000	1002.000000	6.000000	2014.000000	3.000000	26.000000	23.000000	0.000000	0.000000
75%	14.000000	1413.000000	10.000000	2016.000000	3.000000	32.000000	29.000000	2.000000	2.000000
max	22.000000	2217.000000	12.000000	2019.000000	9.000000	999.000000	999.000000	9.000000	9.000000

Análisis Exploratorio:

o Estudio de las variables

Para este proyecto se posee la base de datos de matrimonios en Guatemala, donde al describir nuestros datos podemos observar las variables de la siguiente manera.

	DEPREG	MUPREG	MESREG	AÑOREG	CLAUNI	EDADHOM	EDADMUJ	GETHOM	GETMUJ	NACHOM	...	NUPHON	NUPMUJ	DEPOCU	MUPOCU	MESOCU	AÑOOCU	AREA
0	13	1302	1	10	1	16	14	9	9	320	...	1	1	13	1302	12	9	
1	14	1412	8	9	1	16	14	1	1	320	...	1	1	14	1412	7	9	
2	14	1412	10	9	1	16	14	1	1	320	...	1	1	14	1412	10	9	
3	21	2102	5	9	1	16	14	2	2	320	...	1	1	21	2102	4	9	
4	12	1202	7	9	1	16	14	9	9	320	...	1	1	12	1202	6	9	

Out[26]:

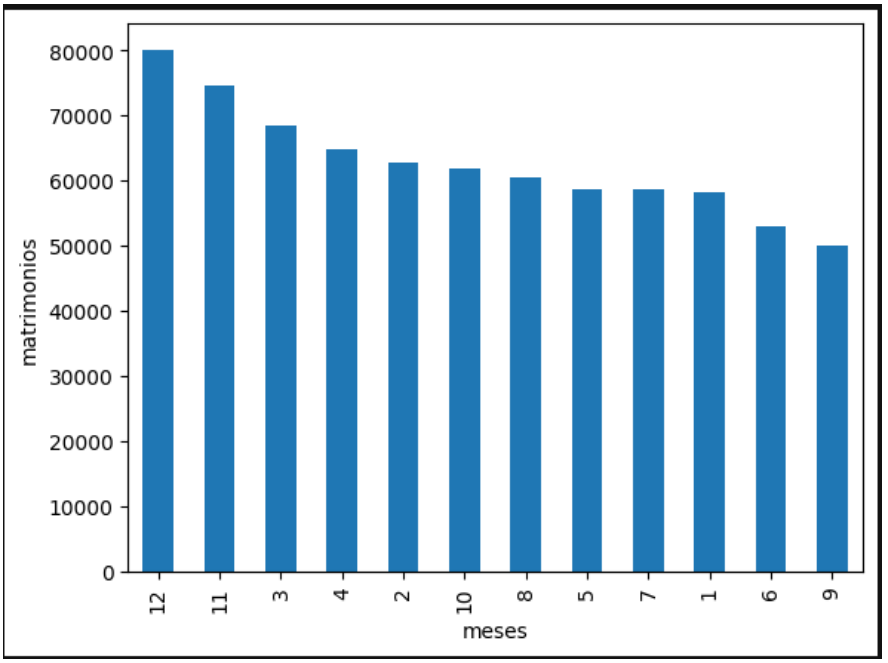
	DEPREG	MUPREG	MESREG	AÑOREG	CLAUNI	EDADHOM	EDADMUJ	GETHOM	GETMUJ	NACHOM	...	NUPHON
count	718045.000000	718045.000000	718045.000000	718045.000000	718045.000000	718045.000000	718045.000000	718045.000000	718045.000000	718045.000000	...	718045.000000
mean	9.532187	960.754008	6.482173	1937.265550	3.100023	31.117912	28.363635	1.140156	1.212049	321.990665	...	0.038627
std	6.388962	639.474049	3.550496	384.770918	1.491131	45.191211	47.973962	2.385953	2.505738	35.622448	...	0.194811
min	1.000000	101.000000	1.000000	9.000000	1.000000	12.000000	10.000000	0.000000	0.000000	8.000000	...	0.000000
25%	4.000000	401.000000	3.000000	2012.000000	3.000000	22.000000	20.000000	0.000000	0.000000	320.000000	...	0.000000
50%	10.000000	1002.000000	6.000000	2014.000000	3.000000	26.000000	23.000000	0.000000	0.000000	320.000000	...	0.000000
75%	14.000000	1414.000000	10.000000	2016.000000	3.000000	32.000000	29.000000	1.000000	1.000000	320.000000	...	0.000000
max	22.000000	2217.000000	12.000000	2019.000000	9.000000	999.000000	999.000000	9.000000	9.000000	9999.000000	...	6.000000

8 rows × 23 columns

Este tipo de variables ya fueron definidas anteriormente, por si se poseen dudas de sus significados.
Para una parte de las variables que solo contaban con cierta cantidad de respuestas, se decidió de mejor manera cambiarles de tipo escrito a números (por ejemplo el (área demográfica que ocurrió el casamiento 1-Urbana, 2-Rural), para un mejor manejo de datos.

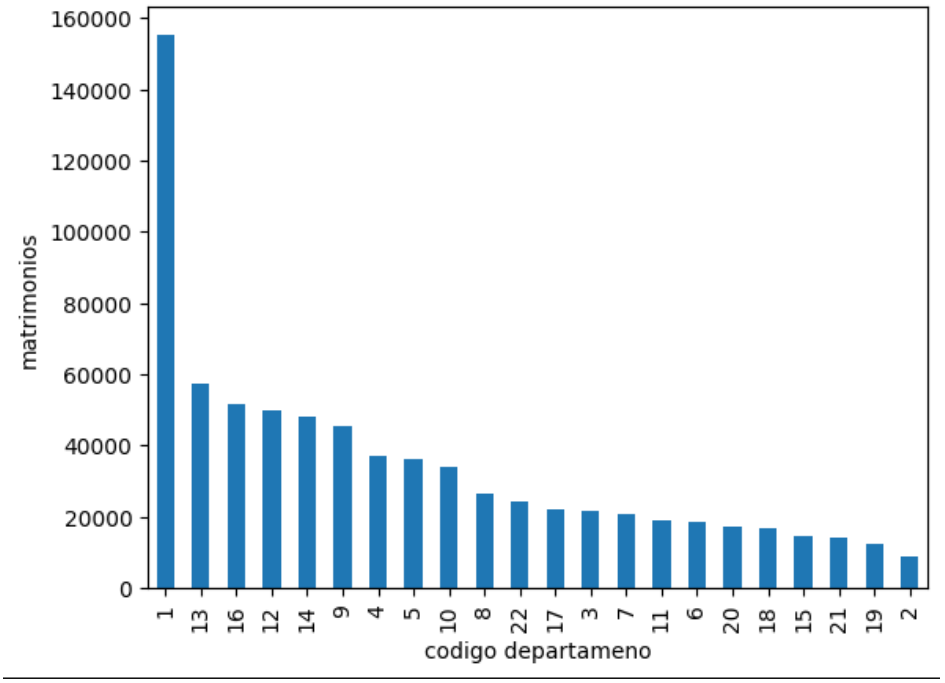
o Gráficos

a. Cantidad de matrimonios por mes

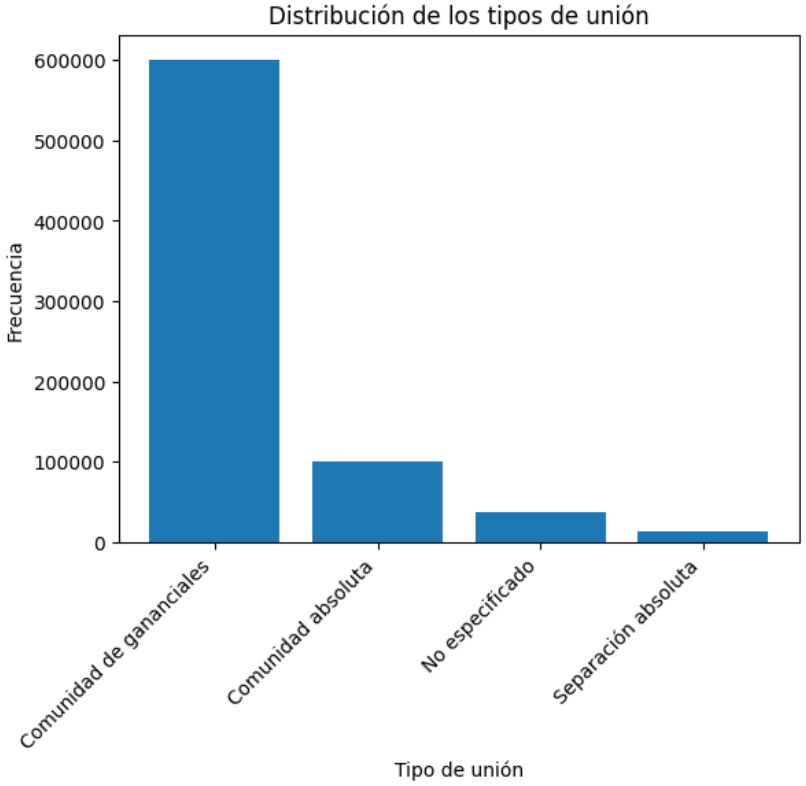


Se puede observar que el mes de Diciembre es el que más casamientos ha registrado a lo largo de 10 años, seguidamente de Noviembre y Marzo.

b. Cantidad de matrimonios por departamento

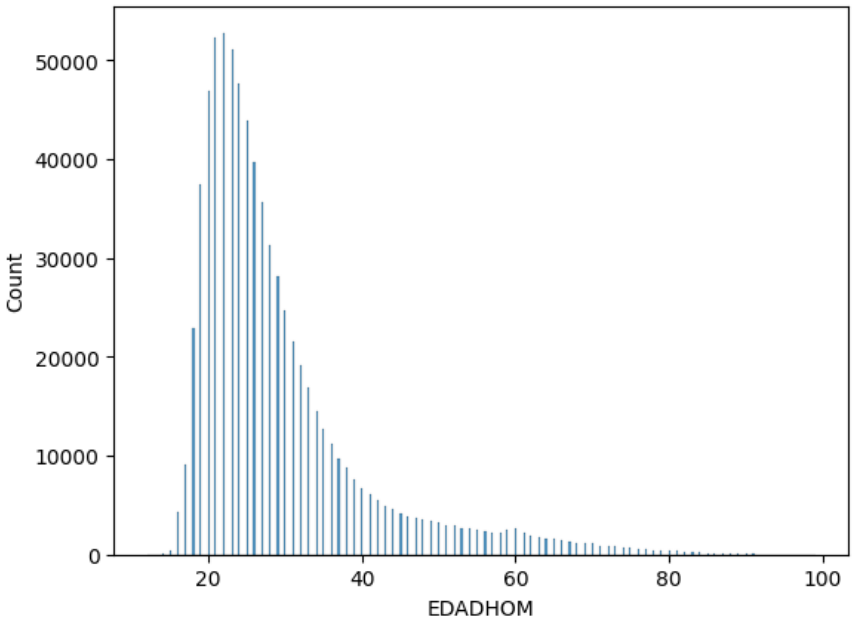


c. Frecuencia con la que se unen por bienes y sus tipos



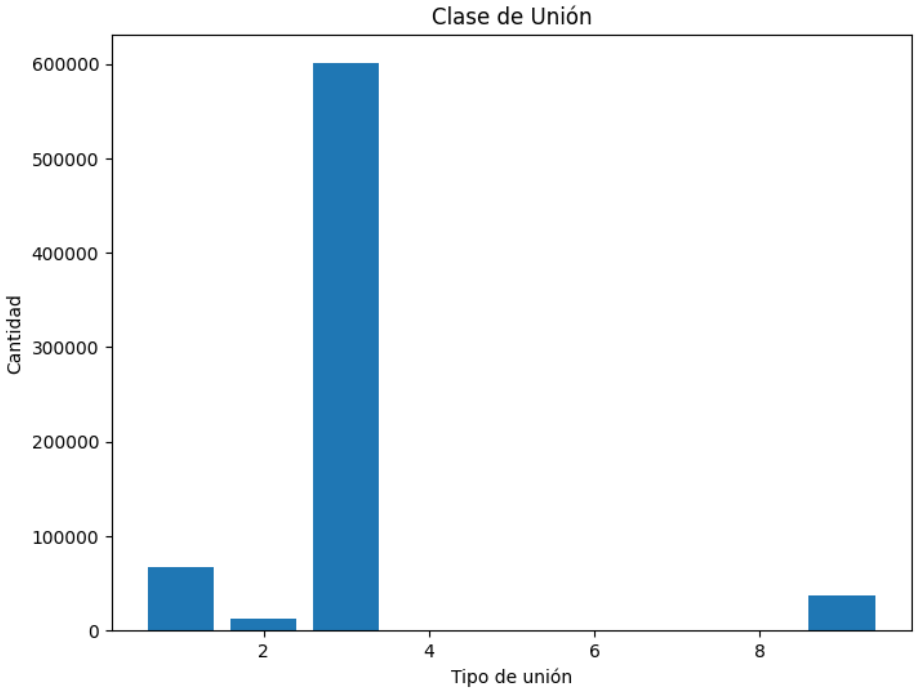
El tipo de unión matrimonial más popular y practicada es la ganancial con un aproximado de 600000 matrimonios registrados, seguida de la absoluta, con un aproximado de 130000.

d. Las edades en las que se casan las personas



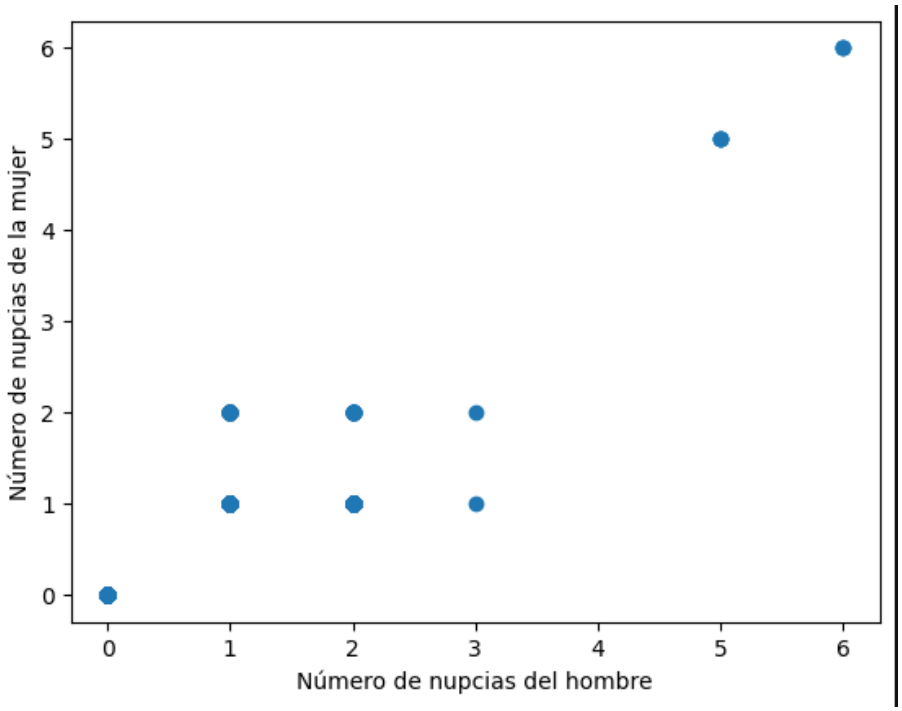
La etapa en la que se realizan más matrimonios es en la etapa del inicio de los 20´s, mayormente entre la edad de 22 a 25 años. Pero este tipo de números sigue siendo alto hasta llegar a los 30´s.

e. La cantidad de matrimonios que se unen por los diferentes tipos de uniones



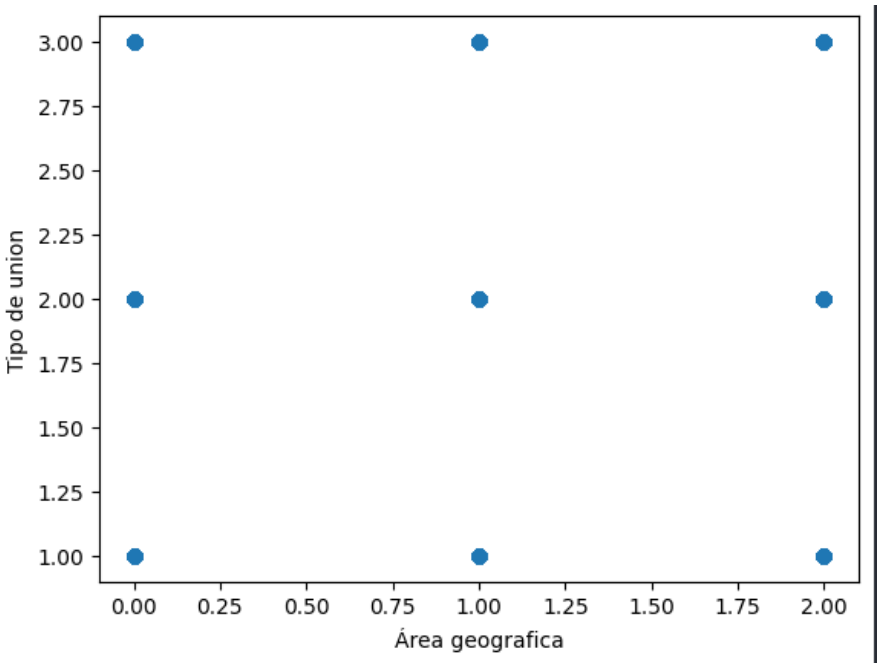
Aqui esta otra forma de ver la cantidad de matrimonios registrados con su tipo, donde como se explico anteriormente el mayormente practicado es el matrimonio por comunidad de gananciales. Donde de igual forma se puede visualizar que gran parte de los matrimonios no tienen un tipo de unión especificado.

f. Cantidad de nupcias por género



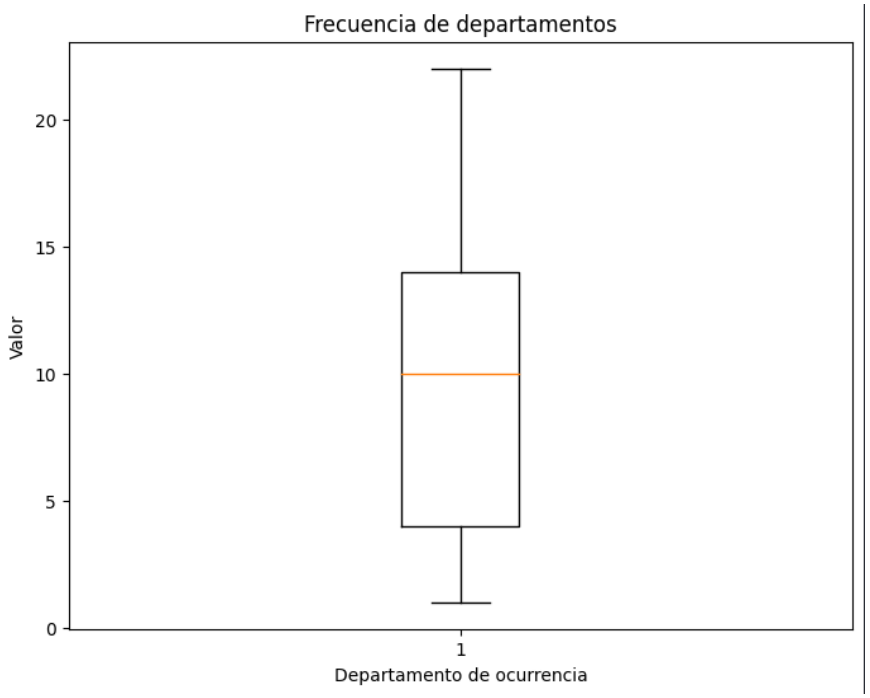
En este tipo de gráfico se puede observar que entre un hombre más cantidad de casamientos posea, la pareja tiende a tener menos cantidad de casamientos. Se puede ver que los hombres tienden a tener más matrimonios que las mujeres.
En lo cual se puede observar que se dio una o varias épocas donde personas con varios casamientos tanto hombres como mujeres, consumieron el matrimonio, explicando los datos atípicos.

g. Tipo de unión dependiendo el área geográfica



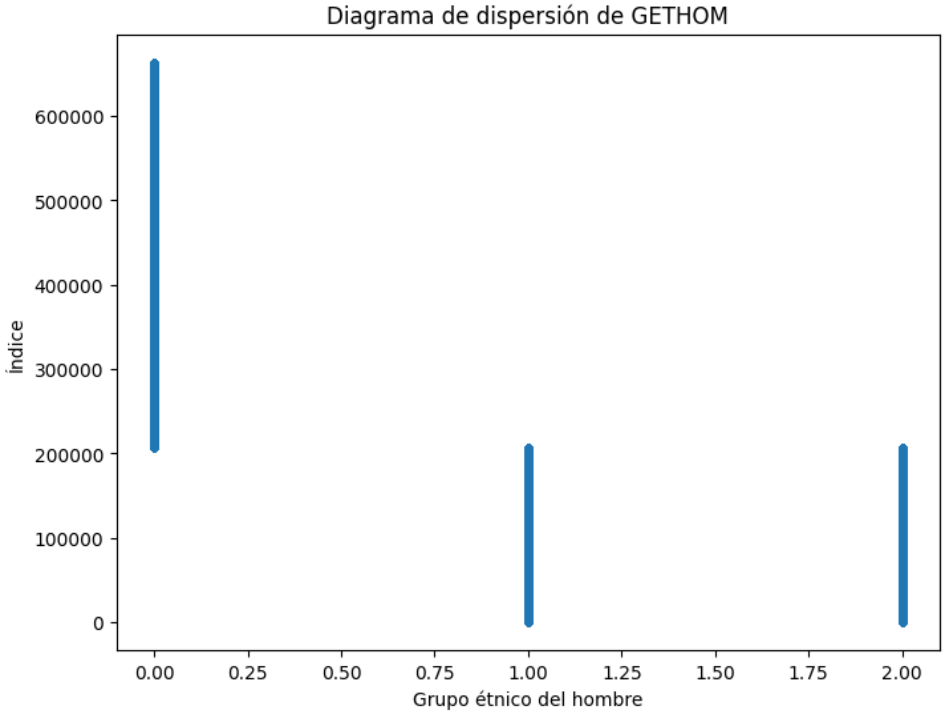
Las áreas geográficas se basan en 1 Urbano, 2 Rural y el tipo de unión en 1 Comunidad absoluta, 2 Separación absoluta, 3 Comunidad de gananciales. Donde puede observarse que las dos variables no tienen una mayor correlación entre sí ; demostrando que el área del casamiento no afecta directamente en el tipo de unión que se lleva a cabo.

h. Frecuencia de matrimonios por departamento



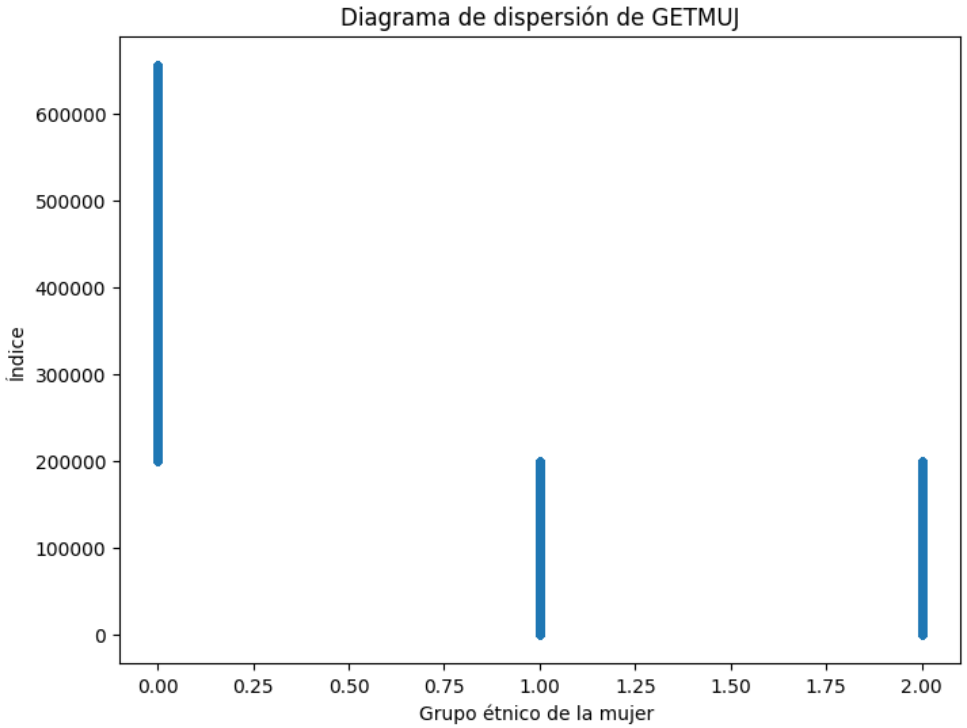
Lo que nos presenta este diagrama es que la mediana de matrimonios está enfocada en el departamento de Suchitepéquez, el tercer cuartil se encuentran los departamentos de Guatemala, El Progreso, Sacatepéquez, Chimaltenango y Escuintla: siendo estos los departamentos con mayores registros de matrimonios. Seguidamente los de departamento que poseen una menor cantidad que se encuentran en el primer cuartil son los departamentos de Retalhuleu, San Marcos, Huehuetenango, Quiché, Baja Verapaz, Alta Verapaz, Petén, Izabal, Zacapa, Chiquimula, Jalapa y Jutiapa.

i. Grupo etnico del hombre



Los grupos étnicos se separan en 1 Indígena, 2 no indígena, mostrando que los hombres indígenas contraen en mayor cantidad matrimonio a los hombres que no se identifican como indígenas. El dato atípico de esta gráfica es que en los matrimonios registrados mayormente no se indicaron si se identificaban como indígenas o no.

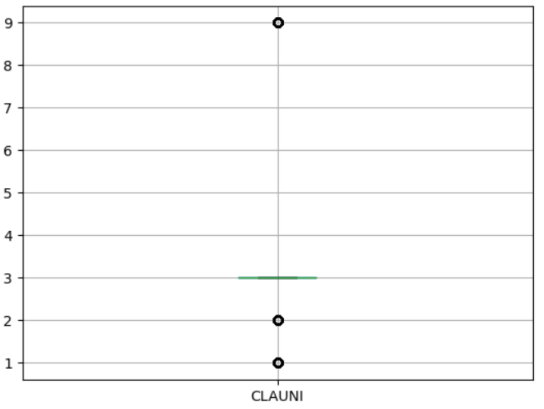
j. Grupo etnico de la mujer

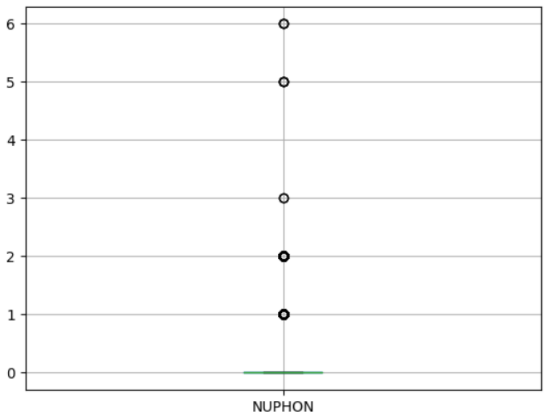


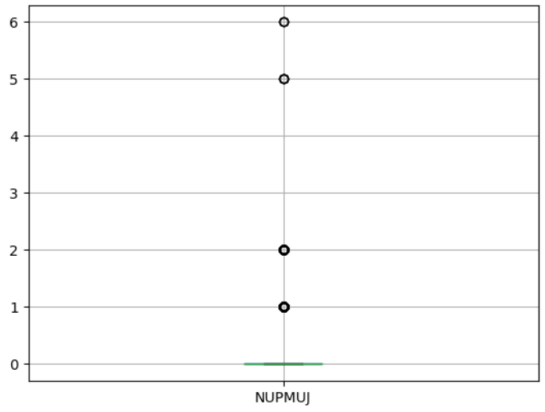
Por otro lado en el grupo etnico de las mujeres de igual manera se puede observar que tanto mujeres indígenas como no indígenas contraen de igual manera matrimonio, sin ser un grupo mayor al otro. El dato atípico de esta gráfica es que en los matrimonios registrados mayormente no se indicaron si se identificaban como indígenas o no.

Clustering

	DEPREG	MUPREG	MESREG	AÑOREG	CLAUNI	EDADHOM	EDADMUJ
DEPREG	1.000000	0.999948	0.008328	-0.041457	0.118861	0.015716	0.009233
MUPREG	0.999948	1.000000	0.008366	-0.041899	0.119584	0.015684	0.009198
MESREG	0.008328	0.008366	1.000000	0.009482	-0.001220	0.002602	0.003098
AÑOREG	-0.041457	-0.041899	0.009482	1.000000	0.279948	-0.124359	-0.145461
CLAUNI	0.118861	0.119584	-0.001220	0.279948	1.000000	-0.040829	-0.046095
EDADHOM	0.015716	0.015684	0.002602	-0.124359	-0.040829	1.000000	0.436270
EDADMUJ	0.009233	0.009198	0.003098	-0.145461	-0.046095	0.436270	1.000000
GETHOM	-0.067941	-0.068397	-0.006675	-0.141830	0.005191	0.068974	0.052704
GETMUJ	-0.061897	-0.062331	-0.004404	-0.134990	0.012143	0.047841	0.068508
NACHOM	-0.019089	-0.019211	-0.001908	0.019624	-0.007546	0.010497	0.001573
NACMUJ	-0.020463	-0.020562	-0.001606	0.021196	-0.006007	-0.004153	0.005954
OCUHOM	-0.001061	-0.000791	-0.007939	-0.258502	0.038534	0.041066	0.043612
OCUMUJ	0.038694	0.039166	-0.006607	-0.268716	0.058749	0.043869	0.044427
NUPHON	0.041531	0.041971	-0.010115	-0.993592	-0.279247	0.124623	0.144442
NUPMUJ	0.041455	0.041893	-0.009922	-0.995401	-0.279755	0.124084	0.144746
DEPOCU	0.989770	0.989689	0.007548	-0.040823	0.118355	0.015581	0.009046
MUPOCU	0.989730	0.989747	0.007587	-0.041212	0.119048	0.015541	0.009004
MESOCU	-0.021495	-0.021601	0.569895	0.015785	-0.016783	0.013857	0.013703
AÑOOCU	-0.010174	-0.010303	0.034189	0.255696	0.011840	-0.022817	-0.028803
AREAG	0.059894	0.060365	-0.006288	-0.787787	-0.176099	0.101706	0.116523
ESCHOM	-0.141254	-0.142038	-0.007045	0.288180	0.028842	-0.048786	-0.054243
ESCMUJ	-0.128536	-0.129264	-0.007799	0.262734	0.027358	-0.063025	-0.057014
DIAOCU	-0.016921	-0.017037	0.008695	0.341679	0.094952	-0.043847	-0.051590
...							
ESCMUJ	-0.128548	0.000446	0.083425	-0.186249	0.553424	1.000000	0.094057
DIAOCU	-0.016840	-0.024139	0.086854	-0.268776	0.100106	0.094057	1.000000



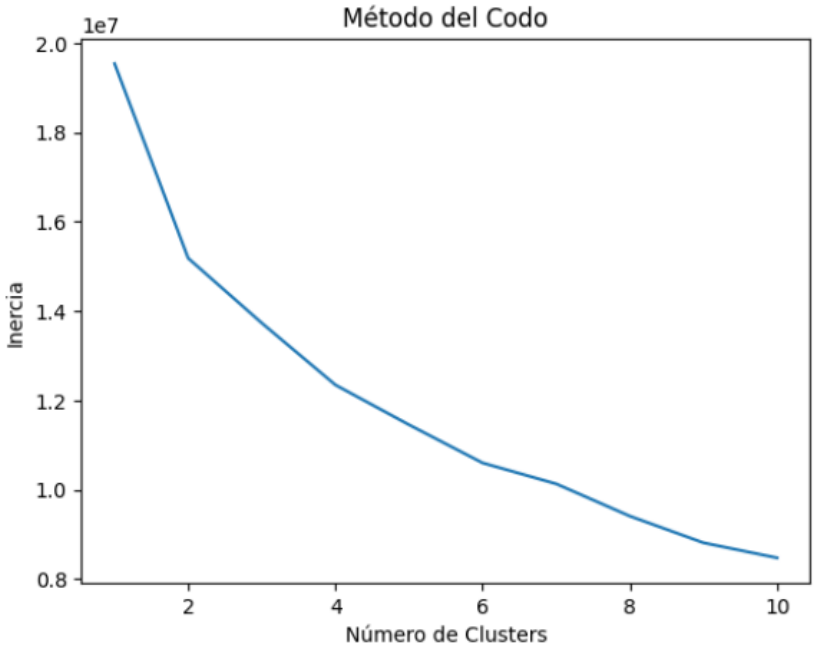




Para analizar las correlaciones entre las variables, utilizamos la matriz de correlación. Las variables usadas rondan una correlación cercana a 1 lo que indica una correlación positiva fuerte. Para los puntos atípicos, en los diagramas de caja apenas hay valor que rondan por el 1 por lo que son muy pocos en los datos usados. Para valores faltantes consideramos obviarlos utilizando el número 9 como referencia al momento de agruparlos.

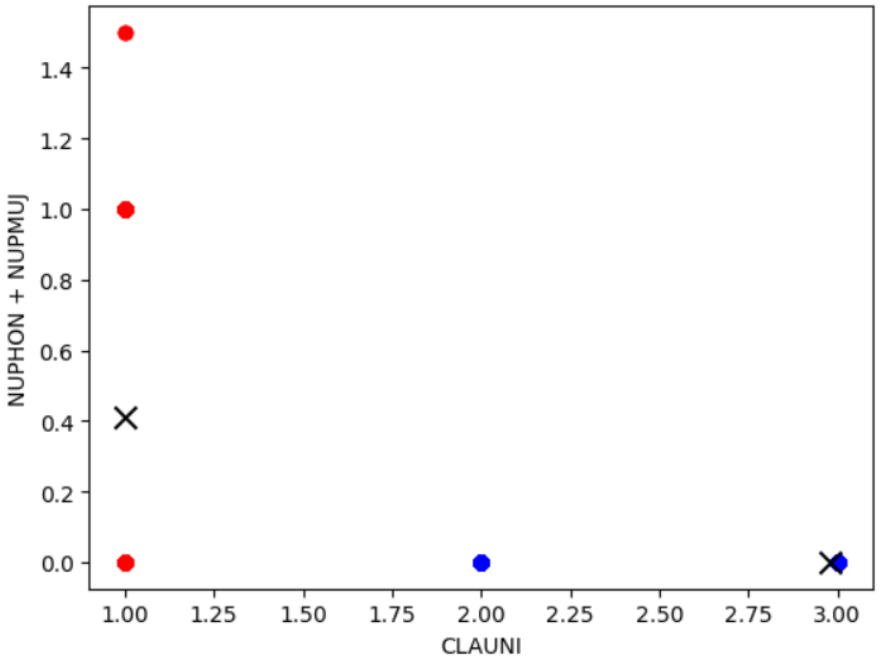
- Tendencias de agrupamientos

Utilizando la suma de las distancias entre los puntos y sus respectivos centroides frente al número de clusters determinamos que el número de clusters ideal sería 2. Si observamos la gráfica el punto en el que la reducción de la variación total ya no es significativa empieza en el 2.



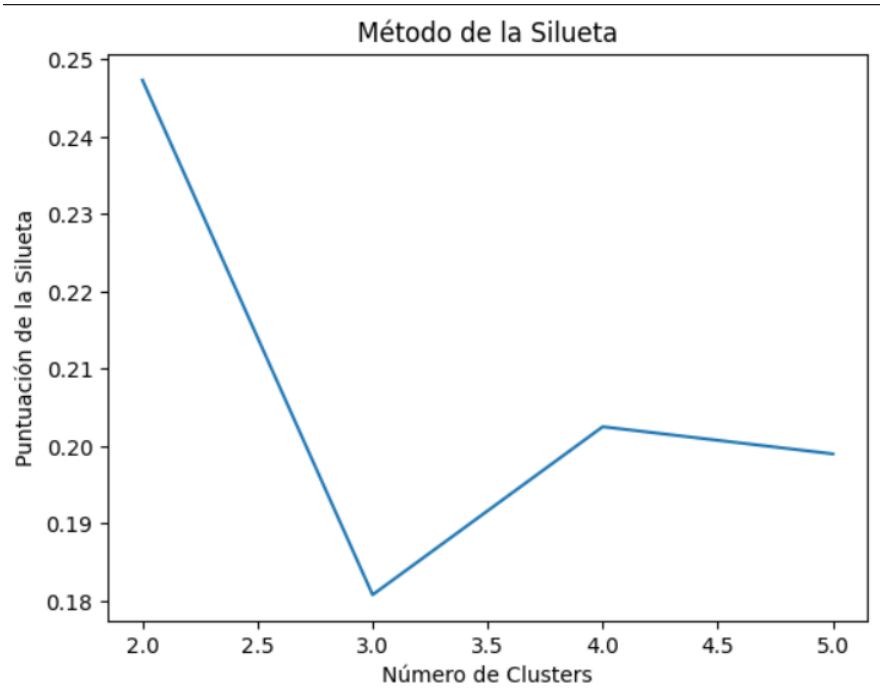
- Algoritmo de clustering

Utilizamos el algoritmo de clustering K-means para agrupar los datos en dos clusters y visualizarlos en un gráfico de dispersión. Las dos columnas que decidimos utilizar fue “CLAUNI”(Clase de union (1 Comunidad absoluta, 2 Separación absoluta, 3 Comunidad de gananciales, 9 No especificado)) y “NUPHON + NUPMUJ”(NUPHOM: Número de nupcias del hombre (9 si se ignora) NUPMUJ: Número de nupcias de la mujer (9 si se ignora)).



o Verificación calidad de agrupamiento

El método de silueta lo utilizamos para medir qué tan bien se ajustan los puntos de un cluster con los demás puntos de ese cluster en comparación con los puntos de los clusters vecinos lo que también nos ayudó a colaborar que el número de clusters que escogimos es el indicado.



o Interpretación de los grupos

La variable 'CLAUNI' describe la clase de unión de las personas, Comunidad absoluta, Separación absoluta, Comunidad de gananciales. Las personas que tienen una clase de unión similar podría serán agrupadas juntas en un cluster.

Las variables 'NUPHON' y 'NUPMUJ' describen el número de veces que un hombre y una mujer se han casado. Estas variables nos ayudarán a identificar patrones de comportamiento o características de grupos de personas que han tenido múltiples matrimonios o relaciones.

Hallazgos y conclusiones:

Como se puede observar en la sección de gráficos el apartado a, Diciembre es el mayor mes con registros de matrimonios ingresados con un aproximado de 80000 matrimonios, seguidamente de Noviembre con aproximadamente 75000 matrimonios y el tercero es Marzo con aproximadamente 68000. También se puede observar en el apartado b de gráficos que el departamento con mayor cantidad de registros matrimoniales es la Ciudad de Guatemala con aproximadamente 150000 matrimonios registrados. Seguidamente de los demás departamentos que no pasan de la cantidad de 60000 matrimonios registrados.

En el apartado c de las gráficas se puede observar que la distribución de bienes mayormente elegida por los matrimonios es el de la comunidad de bienen gananciales, el cual posee una cantidad aproximada de 600000 matrimonios registrados. Seguidamente va el de la comunidad absoluta por 100000 matrimonios registrados.

Para las edades en las que mayoritariamente se casan las personas de Guatemala, se encuentra en el apartado d de gráficos. Donde se puede contemplar que los guatemaltecos se casan mayormente en sus 20´s, mayormente alrededor de sus 22 a 25 años. Además en los casamientos registrados se puede ver que los hombres han tenido una mayor cantidad de nupcias que las mujeres, tendiendo así los hombres a casarse hasta 3 veces a comparación de las mujeres que tienden a solo tener 1 o 2 nupcias más, como se puede observar en el apartado f de gráficos.

Con la visualización del gráfico de dispersión, se puede observar que la relación es un poco tenue en comparación al valor que nosotros le encontramos a estos datos. Además, podemos notar que los dos clusters están separados por una línea vertical en el centro del gráfico. Esto sugiere que la variable CLAUNI tiene un gran impacto en la separación de los datos en dos grupos distintos. El cluster rojo tiene valores de "CLAUNI" y "NUPHON + NUPMUJ" más bajos, mientras que el cluster azul tiener valores más altos. Las cruces indican los valores medios de "CLAUNI" y "NUPHON + NUPMUJ" para cada cluster.