

Análisis Exploratorio de Datos

1. Descripción general del conjunto de datos

El conjunto de datos analizado proviene del Instituto Nacional de Estadística (INE) de Guatemala y corresponde a información consolidada de divorcios registrada durante más de diez años. Los datos originales se encontraban en formato .sav y fueron unificados y normalizados para facilitar su análisis exploratorio.

El conjunto de datos contiene 71,576 observaciones, donde cada observación representa un registro individual, y 23 variables, que describen características demográficas, temporales y geográficas relacionadas con los divorcios registrados.

2. Significado y tipo de las variables

A continuación, se describen las variables incluidas en el conjunto de datos junto con su tipo:

Variable	Tipo de dato	Tipo de variable
DEPREG	object	Categórica
MUPREG	object	Categórica
MESREG	object	Categórica
AÑOREG	float64	Categórica
DIAOCU	float64	Numérica
MESOCU	object	Categórica
AÑOOCU	float64	Categórica
DEPOCU	object	Categórica
MUPOCU	object	Categórica
EDADHOM	float64	Numérica
EDADMUJ	float64	Numérica
PUEHOM	object	Categórica
PUEMUJ	object	Categórica
NACHOM	object	Categórica
NACMUJ	object	Categórica
ESCHOM	object	Categórica
ESCMUJ	object	Categórica
CIUOHOM	object	Categórica
CIUOMUJ	object	Categórica

GETHOM	object	Categórica
GETMUJ	object	Categórica
OCUHOM	object	Categórica
OCUMUJ	object	Categórica

3. Cantidad de variables y observaciones

Cantidad de variables: 23

Cantidad de observaciones: 71,576

Esta cantidad de datos permite realizar un análisis exploratorio estadísticamente representativo.

4. Exploración de variables numéricas

Las variables numéricas analizadas fueron EDADHOM, EDADMUJ y DIAOCU.

4.1. Medidas de tendencia central

Variable	Media	Mediana	Moda
EDADHOM	36.02	34	31
EDADMUJ	32.80	31	27
DIAOCU	15.73	16	22

Los resultados indican que, en promedio, los hombres presentan una edad mayor que las mujeres al momento del divorcio.

4.2. Medidas de dispersión, orden y distribución

Rango, varianza y desviación estándar:

Variable	Mín	Máx	Rango	Varianza	Desv. Std	Coef. Variación
EDADHOM	15	98	83	100.56	10.03	0.28
EDADMUJ	15	81	66	83.12	9.12	0.28
DIAOCU	1	31	30	75.40	8.68	0.55

Cuartiles:

Variable	Q1 (25%)	Mediana	Q3 (75%)
EDADHOM	29	34	41
EDADMUJ	26	31	38
DIAOCU	8	16	23

Distribución:

- EDADHOM y EDADMUJ presentan asimetría positiva, lo que indica una mayor concentración de valores en edades menores.
- DIAOCU presenta una distribución aproximadamente simétrica.

5. Exploración de variables categóricas – Tablas de frecuencia

5.1. Departamento de registro (DEPREG)

Los departamentos con mayor frecuencia de registros son:

- Guatemala (38.52%)
- Quetzaltenango (8.06%)
- Escuintla (4.45%)

5.2. Departamento de ocurrencia (DEPOCU)

Se observa un patrón similar, donde el departamento de Guatemala concentra el mayor porcentaje de casos (37.07%).

5.3. Mes de ocurrencia (MESOCU)

Los meses con mayor número de divorcios registrados son:

- Octubre (9.78%)
- Marzo (9.76%)
- Agosto (9.25%)

Diciembre presenta el menor porcentaje (5.38%).

5.4. Escolaridad

En cuanto a la escolaridad:

La categoría “Ignorado” es la más frecuente tanto para hombres (46.91%) como para mujeres (46.28%).

El nivel Diversificado es el segundo más frecuente en ambos casos.

6. Relaciones entre las variables

6.1. Relación entre edades

La correlación entre EDADHOM y EDADMUJ es de 0.79, lo que indica una relación positiva fuerte: generalmente, cuando aumenta la edad de uno de los cónyuges, también aumenta la del otro.

6.2. Diferencia de edad

La diferencia de edad promedio entre los cónyuges es de 3.16 años, con una mediana de 2 años, lo que sugiere que en la mayoría de los casos el hombre es ligeramente mayor que la mujer.

Link del repo: <https://github.com/Ninaswiftie09/Mineria-de-Datos>