

Programa Bienestar - Fertilizantes

En esta parte del proyecto vamos a realizar un análisis exploratorio de los datos obtenidos en los últimos 4 años del programa bienestar del gobierno de México para el apoyo de a productores y productoras de pequeña escala para contribuir en la producción de alimentos. Mas información en: <https://programasparaelbienestar.gob.mx/fertilizantes-para-el-bienestar/>.

El conjunto de datos se obtuvo del sitio web <https://datos.gob.mx/busca/dataset>, donde es posible descargar archivos en diversos formatos para el análisis y uso de la información pública. Este portal de datos abiertos ofrece una amplia gama de conjuntos de datos relacionados con diferentes sectores y programas gubernamentales, incluido el Programa Bienestar - Fertilizantes.

Los datos disponibles en este sitio permiten a los investigadores, analistas y público en general acceder a información detallada sobre la implementación y resultados del programa de fertilizantes. Esto facilita el estudio de su impacto en la producción agrícola de pequeña escala y en el bienestar de las comunidades rurales en México.

Análisis Exploratorio de los datos (EDA)

Objetivo de analisis

El presente proyecto tiene como objetivo, en función de los datos proporcionados, identificar los patrones que nos permitan determinar qué variables o situaciones de las personas solicitantes tienen un mayor impacto en su ingreso al programa Bienestar - Fertilizantes.

Para lograr este objetivo, se llevará a cabo un análisis exhaustivo de las variables socioeconómicas, demográficas y agrícolas de los solicitantes, utilizando técnicas estadísticas y de aprendizaje automático. Este enfoque nos permitirá identificar los factores más influyentes en la selección de beneficiarios y evaluar la eficacia del programa en alcanzar a la población objetivo.

Cargar datos

```
In [1]: # Librerías
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import statsmodels.api as sm
import pylab as py
from scipy import stats
from tabulate import tabulate
import warnings

#ML
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score

# Local modules
from project import Beneficiarios
from project import print_dataframe_info
from project import print_summary
from project import print_describe
from project import distribucion_plot
from project import label_encode
from project import correlacion
from project import bivariado

warnings.filterwarnings('ignore')

In [2]: Beneficiarios_2020 = Beneficiarios(2020)
Beneficiarios_2021 = Beneficiarios(2021)
Beneficiarios_2022 = Beneficiarios(2022)
Beneficiarios_2023 = Beneficiarios(2023)
```

Revision inicial de los datos

Cada columna representa un aspecto clave relacionado con los beneficiarios y la distribución de los fertilizantes permitiendo análisis desde múltiples perspectivas: geográfica, temporal, financiera y socioeconómica. Esto es fundamental para evaluar la efectividad del programa y realizar recomendaciones basadas en datos.

Columnas

- **BENEFICIARIO:** Nombre del beneficiario.

- **ZONA:** Región o zona del país.
- **ENTIDAD:** Estado o entidad federativa.
- **MUNICIPIO:** Municipio del beneficiario.
- **LOCALIDAD:** Localidad específica dentro del municipio.
- **ESTRATIFICACIÓN:** Nivel socioeconómico o clasificación.
- **PROGRAMA:** Nombre del programa, en este caso "FERTILIZANTES".
- **COMPONENTE:** Componente específico del programa.
- **SUBCOMPONENTE:** Subcomponente asociado.
- **PRODUCTO:** Producto agrícola relacionado (por ejemplo, "Cultivo de maíz grano").
- **FECHA:** Fecha en que se realizó el apoyo.
- **MONTO FEDERAL:** Cantidad de dinero otorgada (en pesos).
- **APOYO:** Tipo de apoyo (ej. "Fertilizantes").
- **ACTIVIDAD:** Tipo de actividad (ej. "AGRARIA").
- **ESLABÓN:** Especifica el eslabón dentro de la cadena productiva.
- **CICLO AGRÍCOLA:** Ciclo agrícola correspondiente (ej. "PV2020").

```
In [3]: Beneficiarios_2020.head(5)
```

Out[3]:		BENEFICIARIO	ZONA	ENTIDAD	MUNICIPIO	LOCALIDAD	ESTRATIFICACIÓN	PROGRAMA	COMPONENTE	SUBCOMPONENTE	PRODUCTO	FECHA	MONTO FEDERAL	APOYO	ACTIVIDAD	ESLABÓN	CICLO AGRÍCOLA
	0	ESPINOZA ZAMORA SILVANO	CENTRO PAÍS	GUERRERO	ACATEPEC	ESCALERILLA ZAPATA	Muy alto	FERTILIZANTES	FERTILIZANTES	FERTILIZANTES	Cultivo de maíz grano	2020-06-12	6163.44	Fertilizantes	AGRARIA	AGRARIA	PV2020
	1	CASTRO IGNACIO ESTEBAN	CENTRO PAÍS	GUERRERO	ACATEPEC	ESCALERILLA ZAPATA	Muy alto	FERTILIZANTES	FERTILIZANTES	FERTILIZANTES	Cultivo de maíz grano	2020-06-14	6163.44	Fertilizantes	AGRARIA	AGRARIA	PV2020
	2	ESPINOZA ESPINOZA AMELIA	CENTRO PAÍS	GUERRERO	ACATEPEC	ESCALERILLA ZAPATA	Muy alto	FERTILIZANTES	FERTILIZANTES	FERTILIZANTES	Cultivo de maíz grano	2020-06-15	3081.72	Fertilizantes	AGRARIA	AGRARIA	PV2020
	3	DIRCIO CALIXTO ANUNCIACION	CENTRO PAÍS	GUERRERO	ACATEPEC	ESCALERILLA ZAPATA	Muy alto	FERTILIZANTES	FERTILIZANTES	FERTILIZANTES	Cultivo de maíz grano	2020-06-15	3081.72	Fertilizantes	AGRARIA	AGRARIA	PV2020
	4	CALIXTRO LOPEZ LAURA	CENTRO PAÍS	GUERRERO	ACATEPEC	ESCALERILLA ZAPATA	Muy alto	FERTILIZANTES	FERTILIZANTES	FERTILIZANTES	Cultivo de maíz grano	2020-06-15	3081.72	Fertilizantes	AGRARIA	AGRARIA	PV2020

Tipos de variables

1. Variables Categóricas:

- Son variables que describen atributos o categorías y que no tienen un valor numérico intrínseco. En este conjunto de datos, las variables categóricas son fundamentales para segmentar y clasificar la información.
- Numero de variables categóricas: 14

2. Variables Numericas:

- Son variables que representan cantidades y se pueden medir y ordenar. Las variables numéricas en este dataset permiten realizar análisis estadísticos y financieros.
- Numero de variables categóricas: 2
 - MONTO FEDERAL: Cantidad de dinero otorgada en pesos.
 - FECHA: Fecha en la que se realizó el apoyo, registrada como una variable temporal, pero que puede ser tratada numéricamente para análisis cronológicos.

```
In [4]: # Imprimir información de cada DataFrame
print_dataframe_info(Beneficiarios_2020, 2020)
print_dataframe_info(Beneficiarios_2021, 2021)
print_dataframe_info(Beneficiarios_2022, 2022)
print_dataframe_info(Beneficiarios_2023, 2023)
```

Beneficiarios 2020

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 369121 entries, 0 to 369120
Data columns (total 16 columns):
#   Column                Non-Null Count  Dtype
---  -
0   BENEFICIARIO          369121 non-null object
1   ZONA                  369121 non-null object
2   ENTIDAD               369121 non-null object
3   MUNICIPIO            369121 non-null object
4   LOCALIDAD            369121 non-null object
5   ESTRATIFICACIÓN      369121 non-null object
6   PROGRAMA              369121 non-null object
7   COMPONENTE           369121 non-null object
8   SUBCOMPONENTE        369121 non-null object
9   PRODUCTO             369121 non-null object
10  FECHA                 369121 non-null datetime64[ns]
11  MONTO FEDERAL         369121 non-null float64
12  APOYO                 369121 non-null object
13  ACTIVIDAD             369121 non-null object
14  ESLABÓN              369121 non-null object
15  CICLO AGRÍCOLA        369121 non-null object
dtypes: datetime64[ns](1), float64(1), object(14)
memory usage: 45.1+ MB
```

Beneficiarios 2021

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 394983 entries, 0 to 394982
Data columns (total 16 columns):
#   Column                Non-Null Count  Dtype
---  -
0   BENEFICIARIO          394983 non-null object
1   ZONA                  394983 non-null object
2   ENTIDAD               394983 non-null object
3   MUNICIPIO            394983 non-null object
4   LOCALIDAD            394983 non-null object
5   ESTRATIFICACIÓN      394983 non-null object
6   PROGRAMA              394983 non-null object
7   COMPONENTE           394983 non-null object
8   SUBCOMPONENTE        394983 non-null object
9   PRODUCTO             394983 non-null object
10  FECHA                 394983 non-null datetime64[ns]
11  MONTO FEDERAL         394983 non-null float64
12  APOYO                 394983 non-null object
13  ACTIVIDAD             394983 non-null object
14  ESLABÓN              394983 non-null object
15  CICLO AGRÍCOLA        394983 non-null object
dtypes: datetime64[ns](1), float64(1), object(14)
memory usage: 48.2+ MB
```

Beneficiarios 2022

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 826382 entries, 0 to 826381
Data columns (total 16 columns):
#   Column                Non-Null Count  Dtype
---  -
0   BENEFICIARIO          826382 non-null object
1   ZONA                  826382 non-null object
2   ENTIDAD               826382 non-null object
3   MUNICIPIO            826382 non-null object
4   LOCALIDAD            826382 non-null object
5   ESTRATIFICACIÓN      826382 non-null object
6   PROGRAMA              826382 non-null object
7   COMPONENTE           826382 non-null object
8   SUBCOMPONENTE        826382 non-null object
9   PRODUCTO             826382 non-null object
10  FECHA                 826382 non-null datetime64[ns]
11  MONTO FEDERAL         826382 non-null float64
```

```
12  APOYO          826382 non-null object
13  ACTIVIDAD      826382 non-null object
14  ESLABÓN        826382 non-null object
15  CICLO AGRÍCOLA 826382 non-null object
dtypes: datetime64[ns](1), float64(1), object(14)
memory usage: 100.9+ MB
```

Beneficiarios 2023

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1048575 entries, 0 to 1048574
Data columns (total 16 columns):
#   Column          Non-Null Count  Dtype
---  -
0   BENEFICIARIO    1048575 non-null object
1   ZONA            1048575 non-null object
2   ENTIDAD         1048575 non-null object
3   MUNICIPIO       1048575 non-null object
4   LOCALIDAD       1048575 non-null object
5   ESTRATIFICACIÓN 1048575 non-null object
6   PROGRAMA        1048575 non-null object
7   COMPONENTE      1048575 non-null object
8   SUBCOMPONENTE   1048575 non-null object
9   PRODUCTO        1048575 non-null object
10  FECHA           1048575 non-null datetime64[ns]
11  MONTO FEDERAL   1048575 non-null float64
12  APOYO           1048575 non-null object
13  ACTIVIDAD       1048575 non-null object
14  ESLABÓN         1048575 non-null object
15  CICLO AGRÍCOLA  1048575 non-null object
dtypes: datetime64[ns](1), float64(1), object(14)
memory usage: 128.0+ MB
```

Revision de valores faltantes y duplicados

Al realizar una revisión inicial del conjunto de datos del Programa de Fertilizantes, se llevaron a cabo dos análisis importantes:

Valores Faltantes:

- Resultados: No se encontraron valores faltantes en ninguna de las columnas del conjunto de datos. Esto indica que la base de datos está completa y no hay necesidad de imputar o manejar datos faltantes.
- Importancia: La ausencia de valores faltantes es crucial, ya que garantiza que el análisis será más robusto y que todas las observaciones están completas, lo que permite un análisis más preciso y confiable.

Columnas Duplicadas:

- Resultados: No se encontraron columnas duplicadas en el conjunto de datos.
- Importancia: La inexistencia de columnas duplicadas asegura que no hay redundancia en la información almacenada, lo que podría complicar el análisis o llevar a resultados distorsionados.

```
In [5]: print('Faltantes 2020')
print(Beneficiarios_2020.isna().sum())
print('\nFaltantes 2021')
print(Beneficiarios_2021.isna().sum())
print('\nFaltantes 2022')
print(Beneficiarios_2022.isna().sum())
print('\nFaltantes 2023')
print(Beneficiarios_2023.isna().sum())
```

Faltantes 2020
BENEFICIARIO 0
ZONA 0
ENTIDAD 0
MUNICIPIO 0
LOCALIDAD 0
ESTRATIFICACIÓN 0
PROGRAMA 0
COMPONENTE 0
SUBCOMPONENTE 0
PRODUCTO 0
FECHA 0
MONTO FEDERAL 0
APOYO 0
ACTIVIDAD 0
ESLABÓN 0
CICLO AGRÍCOLA 0
dtype: int64

Faltantes 2021
BENEFICIARIO 0
ZONA 0
ENTIDAD 0
MUNICIPIO 0
LOCALIDAD 0
ESTRATIFICACIÓN 0
PROGRAMA 0
COMPONENTE 0
SUBCOMPONENTE 0
PRODUCTO 0
FECHA 0
MONTO FEDERAL 0
APOYO 0
ACTIVIDAD 0
ESLABÓN 0
CICLO AGRÍCOLA 0
dtype: int64

Faltantes 2022
BENEFICIARIO 0
ZONA 0
ENTIDAD 0
MUNICIPIO 0
LOCALIDAD 0
ESTRATIFICACIÓN 0
PROGRAMA 0
COMPONENTE 0
SUBCOMPONENTE 0
PRODUCTO 0
FECHA 0
MONTO FEDERAL 0
APOYO 0
ACTIVIDAD 0
ESLABÓN 0
CICLO AGRÍCOLA 0
dtype: int64

Faltantes 2023
BENEFICIARIO 0
ZONA 0
ENTIDAD 0
MUNICIPIO 0
LOCALIDAD 0
ESTRATIFICACIÓN 0
PROGRAMA 0
COMPONENTE 0
SUBCOMPONENTE 0
PRODUCTO 0
FECHA 0
MONTO FEDERAL 0
APOYO 0
ACTIVIDAD 0
ESLABÓN 0
CICLO AGRÍCOLA 0
dtype: int64

Filtrado de columnas

En este paso, se procede a la depuración del conjunto de datos, eliminando aquellas columnas categóricas que no aportan valor significativo al análisis. Estas columnas suelen contener valores únicos o repetidos que no contribuyen a la variabilidad del modelo. Como son; programa, Componente, Subcomponente, Apoyo, Actividad, Eslabón.

```
In [6]: print_describe(Beneficiarios_2020, 2020)
print_describe(Beneficiarios_2021, 2021)
print_describe(Beneficiarios_2022, 2022)
print_describe(Beneficiarios_2023, 2023)
```

Beneficiarios 2020

Attribute	BENEFICIARIO	ZONA	ENTIDAD	MUNICIPIO	LOCALIDAD	ESTRATIFICACIÓN	PROGRAMA	COMPONENTE	SUBCOMPONENTE	PRODUCTO	APOYO	ACTIVIDAD	ESLABÓN	CICLO AGRÍCOLA
count	369121	369121	369121	369121	369121	369121	369121	369121	369121	369121	369121	369121	369121	369121
unique	363916	2	5	181	4270	7	1	1	1	1	1	1	1	1
top	XSISTEMA XSISTEMA XSISTEMA	CENTRO PAÍS	GUERRERO	CHILAPA DE ÁLVAREZ	TECOANAPA	Alto	FERTILIZANTES	FERTILIZANTES	FERTILIZANTES	Cultivo de maíz grano	Fertilizantes	AGRARIA	AGRARIA	PV2020
freq	13	369037	340560	16479	3574	237655	369121	369121	369121	369121	369121	369121	369121	369121

Beneficiarios 2021

Attribute	BENEFICIARIO	ZONA	ENTIDAD	MUNICIPIO	LOCALIDAD	ESTRATIFICACIÓN	PROGRAMA	COMPONENTE	SUBCOMPONENTE	PRODUCTO	APOYO	ACTIVIDAD	ESLABÓN	CICLO AGRÍCOLA
count	394983	394983	394983	394983	394983	394983	394983	394983	394983	394983	394983	394983	394983	394983
unique	389218	1	4	246	4714	7	1	1	1	5	1	1	1	1
top	GARCIA GARCIA JUANA	CENTRO PAÍS	GUERRERO	CHILAPA DE ÁLVAREZ	TECOANAPA	Alto	FERTILIZANTES	FERTILIZANTES	FERTILIZANTES	Cultivo de maíz grano blanco	Fertilizantes	AGRARIA	AGRARIA	PV2021
freq	11	394983	334315	16373	3310	251803	394983	394983	394983	368217	394983	394983	394983	394983

Beneficiarios 2022

Attribute	BENEFICIARIO	ZONA	ENTIDAD	MUNICIPIO	LOCALIDAD	ESTRATIFICACIÓN	PROGRAMA	COMPONENTE	SUBCOMPONENTE	PRODUCTO	APOYO	ACTIVIDAD	ESLABÓN	CICLO AGRÍCOLA
count	826382	826382	826382	826382	826382	826382	826382	826382	826382	826382	826382	826382	826382	826382
unique	730262	7	9	1021	9499	7	1	1	1	6	1	1	1	1
top	PEREZ PEREZ JUAN	SUR SURESTE	GUERRERO	LAS MARGARITAS	LAS MARGARITAS	Alto	FERTILIZANTES	FERTILIZANTES	FERTILIZANTES	Cultivo de maíz grano blanco	Fertilizantes	AGRARIA	AGRARIA	PV2022
freq	107	400300	329407	18371	9102	578921	826382	826382	826382	764503	826382	826382	826382	826382

Beneficiarios 2023

Attribute	BENEFICIARIO	ZONA	ENTIDAD	MUNICIPIO	LOCALIDAD	ESTRATIFICACIÓN	PROGRAMA	COMPONENTE	SUBCOMPONENTE	PRODUCTO	APOYO	ACTIVIDAD	ESLABÓN	CICLO AGRÍCOLA
count	1048575	1048575	1048575	1048575	1048575	1048575	1048575	1048575	1048575	1048575	1048575	1048575	1048575	1048575
unique	937866	4	14	862	12644	7	1	1	1	21	1	1	1	1
top	HERNANDEZ HERNANDEZ JUAN	CENTRO PAÍS	GUERRERO	LAS MARGARITAS	LAS MARGARITAS	BAJO	FERTILIZANTES	FERTILIZANTES	FERTILIZANTES	Cultivo de maíz grano	Fertilizantes	AGRARIA	AGRARIA	PV2023
freq	165	548819	331018	18041	9383	332745	1048575	1048575	1048575	876550	1048575	1048575	1048575	1048575

```
In [7]: columns_filter = ['BENEFICIARIO', 'PROGRAMA', 'COMPONENTE', 'SUBCOMPONENTE', 'APOYO', 'ACTIVIDAD', 'ESLABÓN', 'CICLO AGRÍCOLA']

Beneficiarios_2020 = Beneficiarios_2020.drop(columns=columns_filter)
Beneficiarios_2021 = Beneficiarios_2021.drop(columns=columns_filter)
Beneficiarios_2022 = Beneficiarios_2022.drop(columns=columns_filter)
Beneficiarios_2023 = Beneficiarios_2023.drop(columns=columns_filter)

print_describe(Beneficiarios_2020, 2020)
print_describe(Beneficiarios_2021, 2021)
print_describe(Beneficiarios_2022, 2022)
print_describe(Beneficiarios_2023, 2023)
```

Beneficiarios 2020

Attribute	ZONA	ENTIDAD	MUNICIPIO	LOCALIDAD	ESTRATIFICACIÓN	PRODUCTO
count	369121	369121	369121	369121	369121	369121
unique	2	5	181	4270	7	1
top	CENTRO PAÍS	GUERRERO	CHILAPA DE ÁLVAREZ	TECOANAPA	Alto	Cultivo de maíz grano
freq	369037	340560	16479	3574	237655	369121

Beneficiarios 2021

Attribute	ZONA	ENTIDAD	MUNICIPIO	LOCALIDAD	ESTRATIFICACIÓN	PRODUCTO
count	394983	394983	394983	394983	394983	394983
unique	1	4	246	4714	7	5
top	CENTRO PAÍS	GUERRERO	CHILAPA DE ÁLVAREZ	TECOANAPA	Alto	Cultivo de maíz grano blanco
freq	394983	334315	16373	3310	251803	368217

Beneficiarios 2022

Attribute	ZONA	ENTIDAD	MUNICIPIO	LOCALIDAD	ESTRATIFICACIÓN	PRODUCTO
count	826382	826382	826382	826382	826382	826382
unique	7	9	1021	9499	7	6
top	SUR SURESTE	GUERRERO	LAS MARGARITAS	LAS MARGARITAS	Alto	Cultivo de maíz grano blanco
freq	400300	329407	18371	9102	578921	764503

Beneficiarios 2023

Attribute	ZONA	ENTIDAD	MUNICIPIO	LOCALIDAD	ESTRATIFICACIÓN	PRODUCTO
count	1048575	1048575	1048575	1048575	1048575	1048575
unique	4	14	862	12644	7	21
top	CENTRO PAÍS	GUERRERO	LAS MARGARITAS	LAS MARGARITAS	BAJO	Cultivo de maíz grano
freq	548819	331018	18041	9383	332745	876550

Análisis estadístico descriptivo

El análisis estadístico descriptivo proporciona un resumen de las características principales de los datos numéricos en el conjunto de datos del Programa de Fertilizantes. Este resumen incluye información clave como la media y la desviación estándar, que son útiles para comprender la distribución y la variabilidad de los montos federales distribuidos.

Analisis Anual:

El programa ha mostrado un aumento constante tanto en el número de beneficiarios como en los montos entregados entre 2020 y 2022, seguido de un ajuste a la baja en 2023. El análisis de sesgo y curtosis sugiere que, a lo largo de los años, la distribución de los montos ha sido relativamente simétrica; sin embargo, en 2022 se registró un aumento en la variabilidad, con algunos montos notablemente más altos. Ese mismo año se destacó por una mayor dispersión, posiblemente vinculada a un enfoque en cultivos de mayor valor o en zonas con mayores necesidades.

Beneficiarios 2020

- Media del Monto Federal: \$4,677.48
- Desviación estándar: \$1,531.08
- Sesgo: -0.103 (ligeramente sesgado a la izquierda)
- Curtosis: -1.968 (distribución más plana que la normal)

Beneficiarios 2021

- Media del Monto Federal: \$5,490.92
- Desviación estándar: \$1,828.11

- Sesgo: 0.00152 (casi simétrico)
- Curtosis: -1.744 (ligeramente más plana que la normal)

Beneficiarios 2022

- Media del Monto Federal: \$8,557.01
- Desviación estándar: \$2,526.31
- Sesgo: 0.960 (moderadamente sesgado a la derecha)
- Curtosis: -0.612 (cerca de una distribución normal)

Beneficiarios 2023

- Media del Monto Federal: \$6,626.67
- Desviación estándar: \$2,188.35
- Sesgo: 0.067 (casi simétrico)
- Curtosis: -1.858 (distribución más plana)

```
In [8]: print_summary(Beneficiarios_2020, 2020)
print_summary(Beneficiarios_2021, 2021)
print_summary(Beneficiarios_2022, 2022)
print_summary(Beneficiarios_2023, 2023)
```


Beneficiarios 2020

Attribute	FECHA	MONTO FEDERAL
count	369121	369121
mean	2020-07-06 04:32:40.293779456	4677.48
min	2020-05-07 00:00:00	2701.46
25%	2020-06-04 00:00:00	3081.72
50%	2020-06-17 00:00:00	5577.06
75%	2020-07-04 00:00:00	6163.44
max	2021-03-31 00:00:00	6163.44
std	nan	1531.08

Sesgo (Skewness) de Beneficiarios 2020:
MONTO FEDERAL -0.103783
dtype: float64

Curtosis de Beneficiarios 2020:
MONTO FEDERAL -1.968168
dtype: float64

Beneficiarios 2021

Attribute	FECHA	MONTO FEDERAL
count	394983	394983
mean	2021-05-21 06:05:52.888098816	5490.92
min	2021-03-09 00:00:00	2632.68
25%	2021-04-07 00:00:00	3598.07
50%	2021-04-24 00:00:00	6665.61
75%	2021-05-17 00:00:00	7196.13
max	2022-03-05 00:00:00	9898.42
std	nan	1828.11

Sesgo (Skewness) de Beneficiarios 2021:
MONTO FEDERAL 0.00152
dtype: float64

Curtosis de Beneficiarios 2021:
MONTO FEDERAL -1.744817
dtype: float64

Beneficiarios 2022

Attribute	FECHA	MONTO FEDERAL
count	826382	826382
mean	2022-07-27 09:54:44.897300224	8557.01
min	2022-04-28 00:00:00	3458.89
25%	2022-05-27 00:00:00	7435.6
50%	2022-08-05 00:00:00	7491.34
75%	2022-08-31 00:00:00	12244.2
max	2022-12-31 00:00:00	15414
std	nan	2526.31

```
+-----+-----+-----+
Sesgo (Skewness) de Beneficiarios 2022:
MONTO FEDERAL    0.960458
dtype: float64

Curtosis de Beneficiarios 2022:
MONTO FEDERAL   -0.612583
dtype: float64
```

```
Beneficiarios 2023
+-----+-----+-----+
| Attribute | FECHA                | MONTO FEDERAL |
+-----+-----+-----+
| count     | 1048575              | 1.04858e+06   |
+-----+-----+-----+
| mean      | 2023-06-05 06:34:46.491095296 | 6626.67        |
+-----+-----+-----+
| min       | 2023-01-24 00:00:00   | 2858.06        |
+-----+-----+-----+
| 25%       | 2023-05-05 00:00:00   | 4470.56        |
+-----+-----+-----+
| 50%       | 2023-06-13 00:00:00   | 6399.38        |
+-----+-----+-----+
| 75%       | 2023-07-19 00:00:00   | 8743.39        |
+-----+-----+-----+
| max       | 2023-10-31 00:00:00   | 14195.4        |
+-----+-----+-----+
| std       | nan                  | 2188.35        |
+-----+-----+-----+
Sesgo (Skewness) de Beneficiarios 2023:
MONTO FEDERAL    0.06722
dtype: float64

Curtosis de Beneficiarios 2023:
MONTO FEDERAL   -1.858342
dtype: float64
```

Visualización de la distribucion de los datos

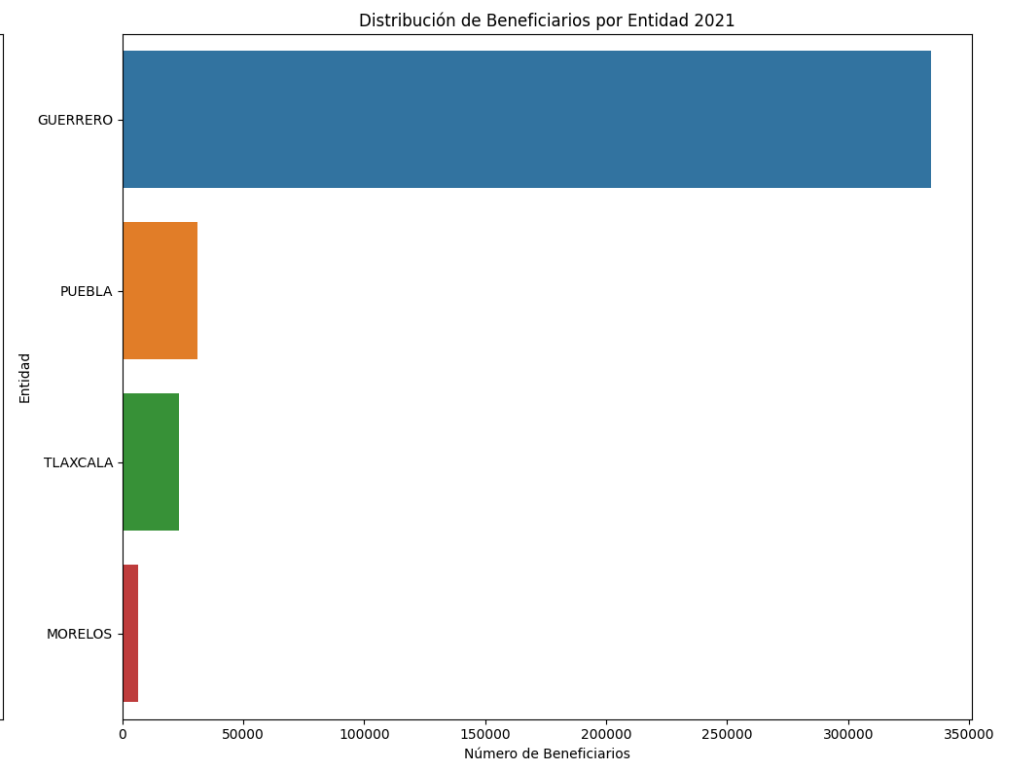
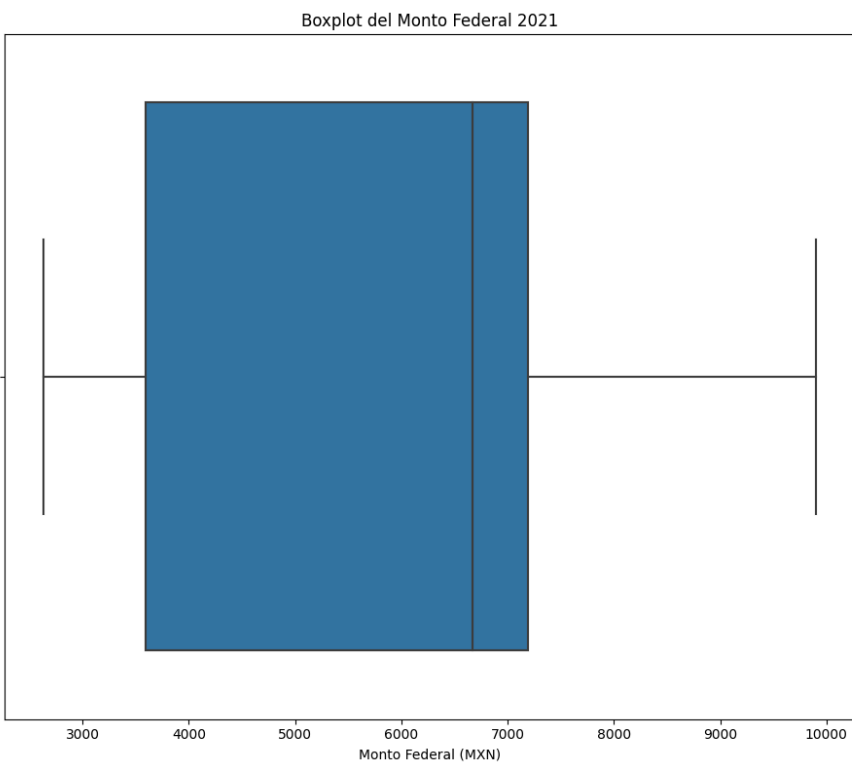
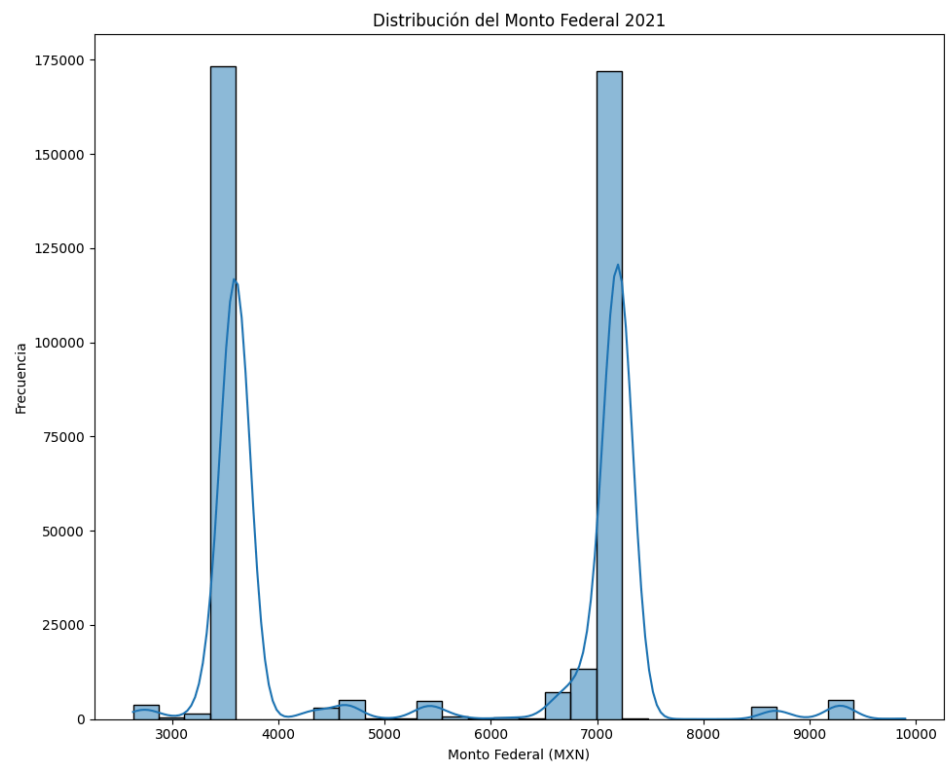
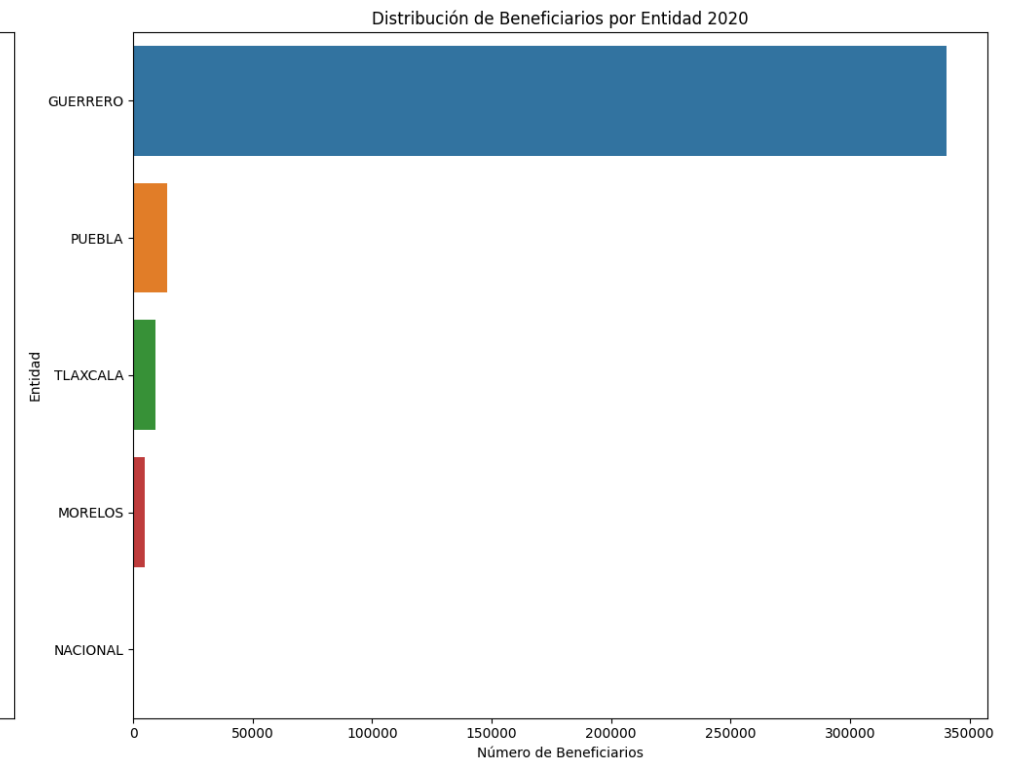
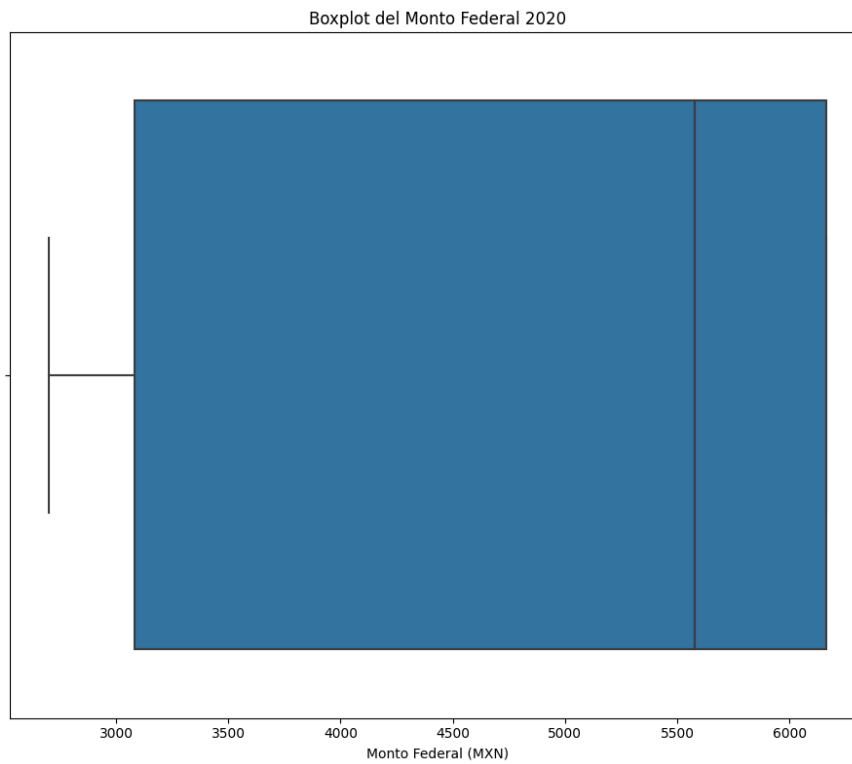
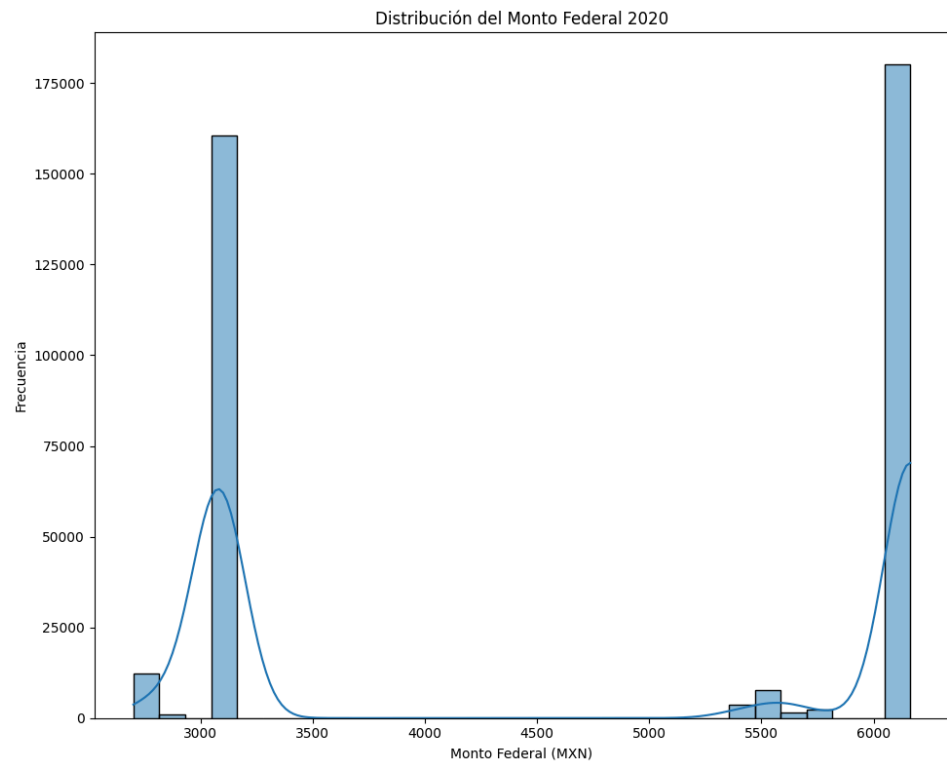
1. Distribución del Monto Federal
- El **histograma** muestra la distribución de los montos federales otorgados a los beneficiarios. La mayoría de los beneficiarios recibieron montos dentro de un rango específico, con una curva que podría estar sesgada hacia la izquierda o derecha dependiendo de la forma observada.
2. Identificación de Valores Atípicos
- El **boxplot** revela la presencia de posibles valores atípicos (outliers) en los montos federales. Estos valores atípicos pueden ser casos donde se otorgaron montos significativamente mayores o menores comparados con la mayoría de los beneficiarios.
3. Distribución de Beneficiarios por Entidad
- El gráfico de barras muestra cómo se distribuyen los beneficiarios a lo largo de las diferentes entidades federativas. Algunas entidades pueden tener un mayor número de beneficiarios, lo que podría reflejar necesidades más altas en esas áreas o una mayor focalización del programa.

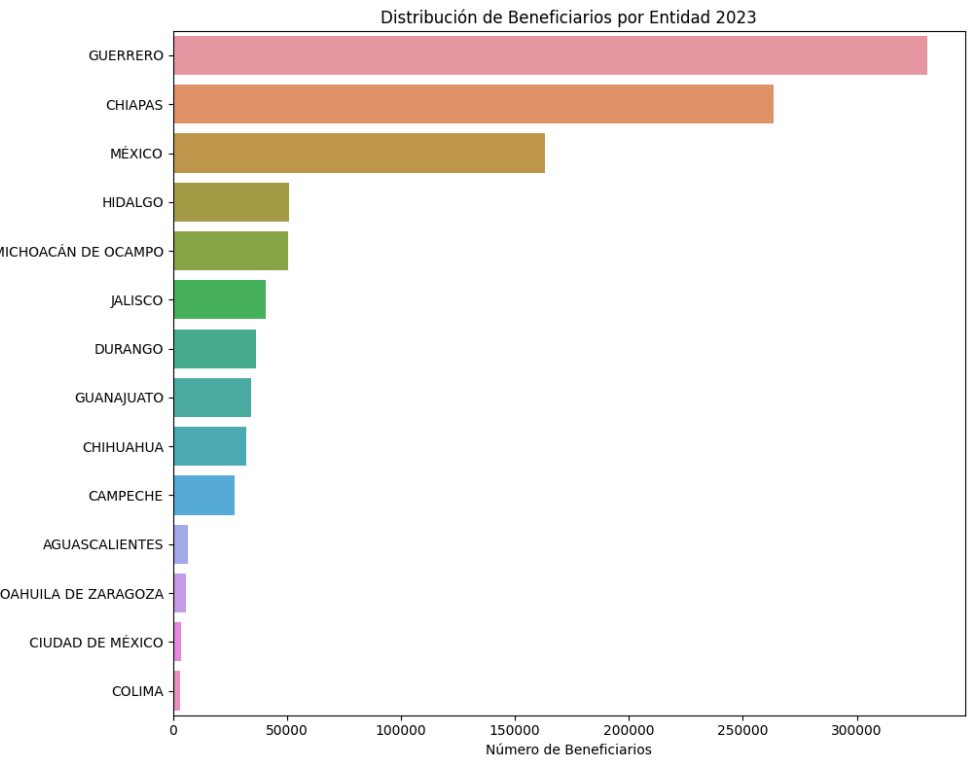
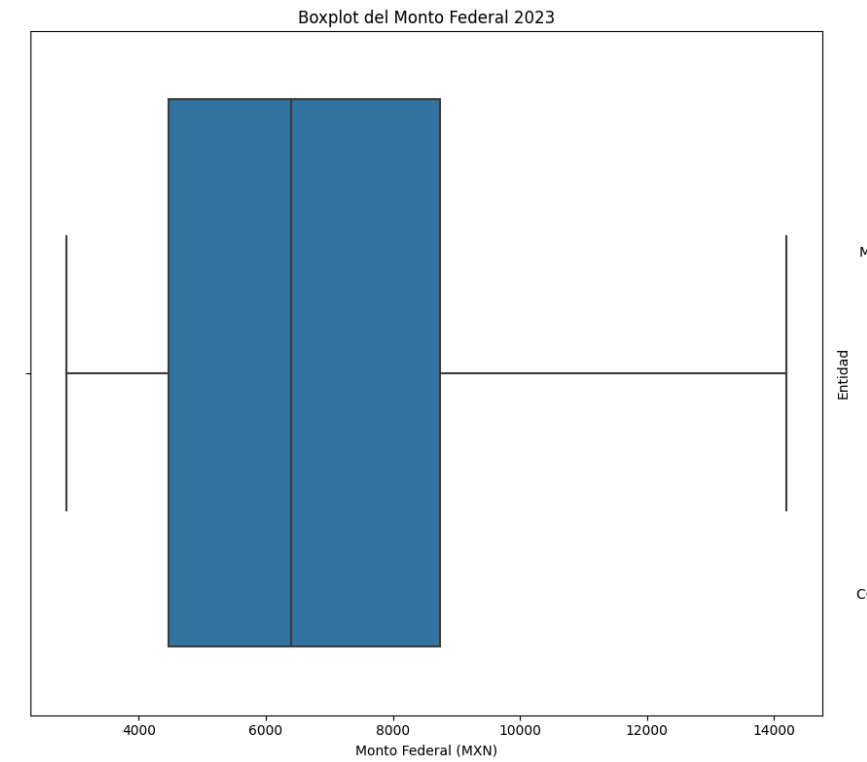
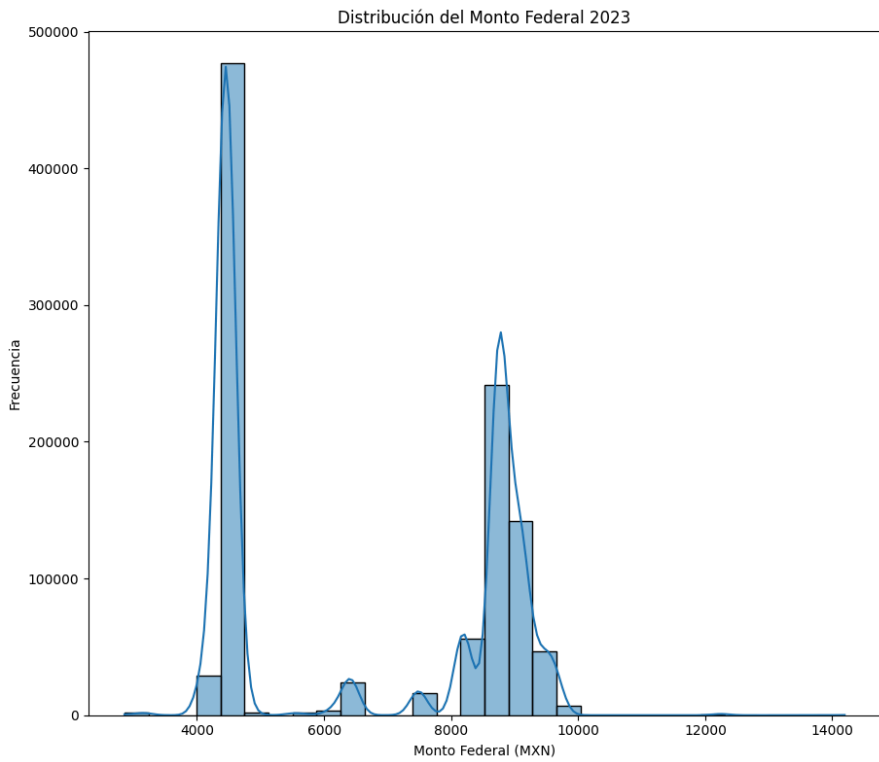
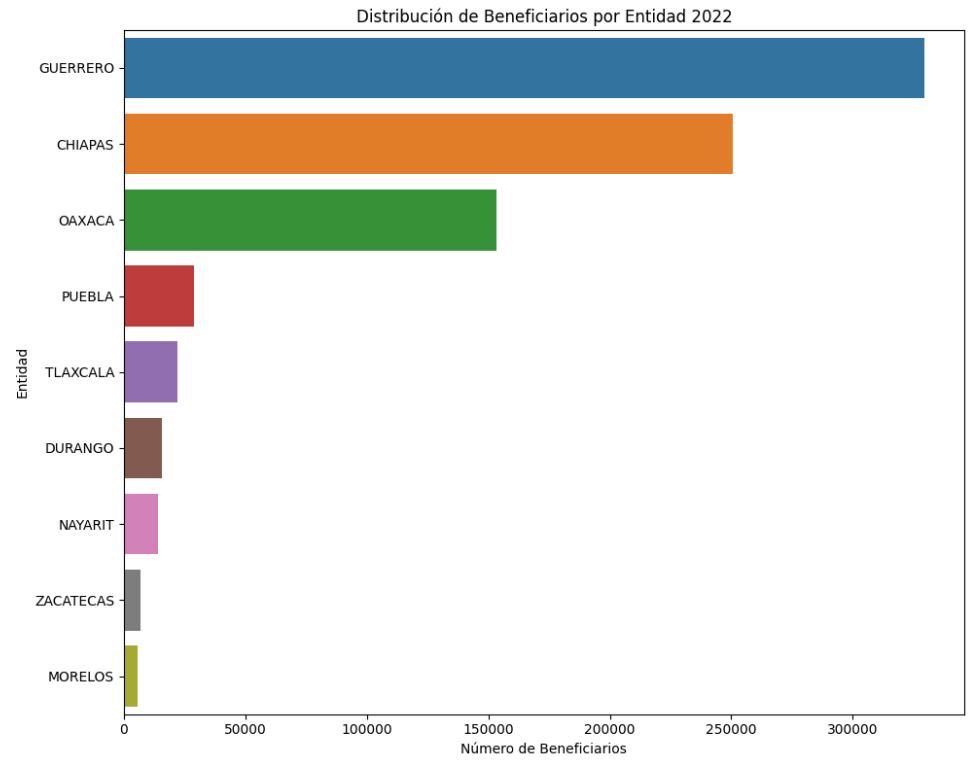
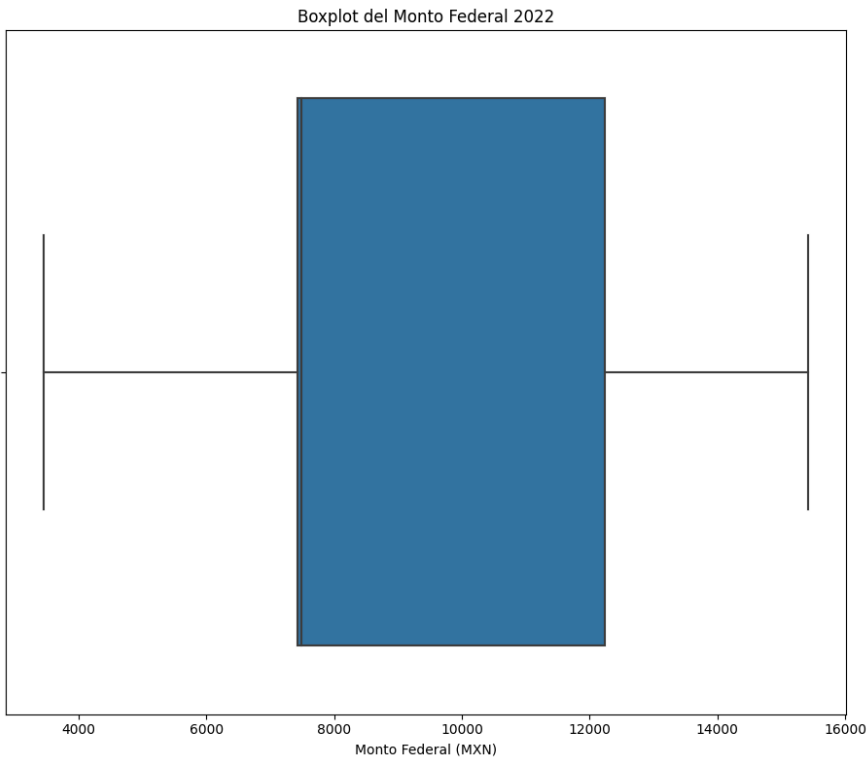
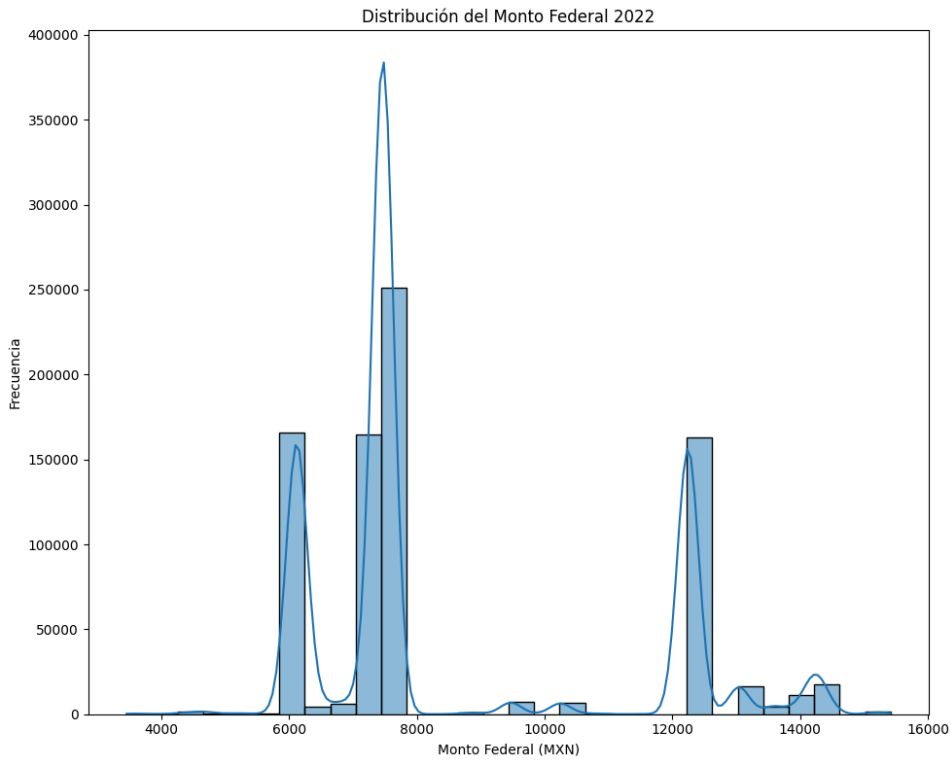
```
In [9]: distribucion_plot(Beneficiarios_2020,year=2020)

distribucion_plot(Beneficiarios_2021,year=2021)

distribucion_plot(Beneficiarios_2022,year=2022)

distribucion_plot(Beneficiarios_2023,year=2023)
```





Distribución del Monto Federal (2020 y 2021)

2020

- Media: El promedio del monto federal distribuido en 2020 fue de 4,677.48 MXN.
- Sesgo: Un valor de sesgo de -0.103 implica que la distribución es levemente asimétrica hacia la izquierda, con una mayor concentración de montos cercanos al valor máximo.
- Curtosis: Con un valor de -1.968, la curtosis es negativa, indicando una distribución más plana o menos concentrada en los extremos que una distribución normal.

En los gráficos:

- El histograma muestra una concentración significativa de beneficiarios en montos cercanos a 3,000 y 6,000 MXN.
- El boxplot destaca que la mayoría de los montos están agrupados en este rango, con pocos valores atípicos.

- Guerrero aparece como la entidad con el mayor número de beneficiarios en 2020, seguido por Puebla, Tlaxcala y Morelos.

2021

- Media: En 2021, el promedio del monto federal subió a 5,490.92 MXN, lo que representa un aumento considerable respecto al año anterior.
- Sesgo: El sesgo es cercano a 0 (0.00152), lo que sugiere que la distribución es bastante simétrica, aunque con dos picos significativos.
- Curtosis: La curtosis de -1.744 aún sugiere una distribución más plana que la normal.

En los gráficos:

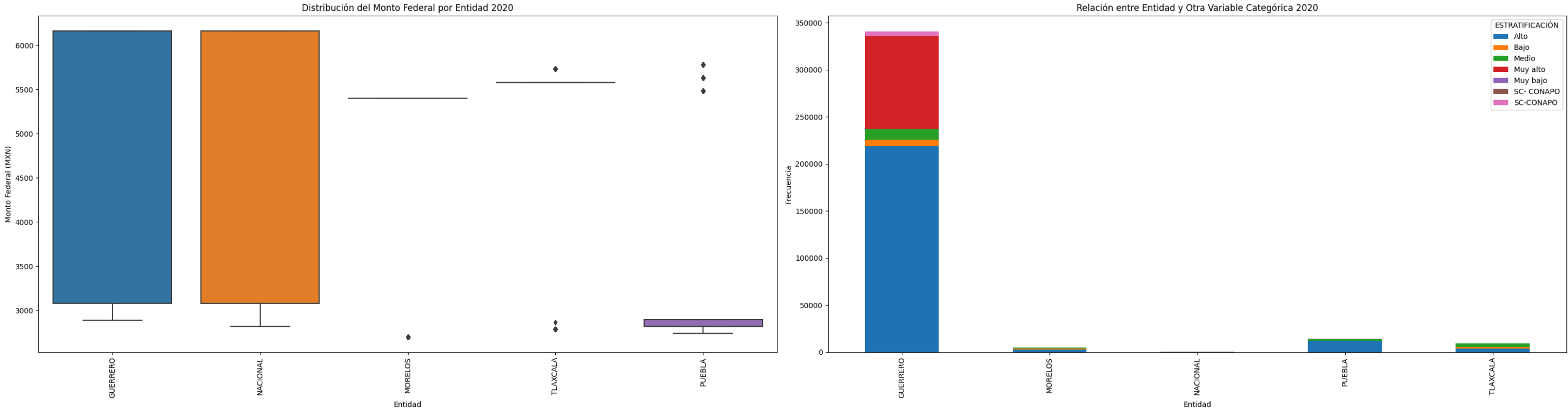
- En el gráfico de distribución, se observan dos picos claros alrededor de los 3,000 MXN y 7,000 MXN, lo que sugiere la existencia de dos grupos de beneficiarios en términos de los montos recibidos.

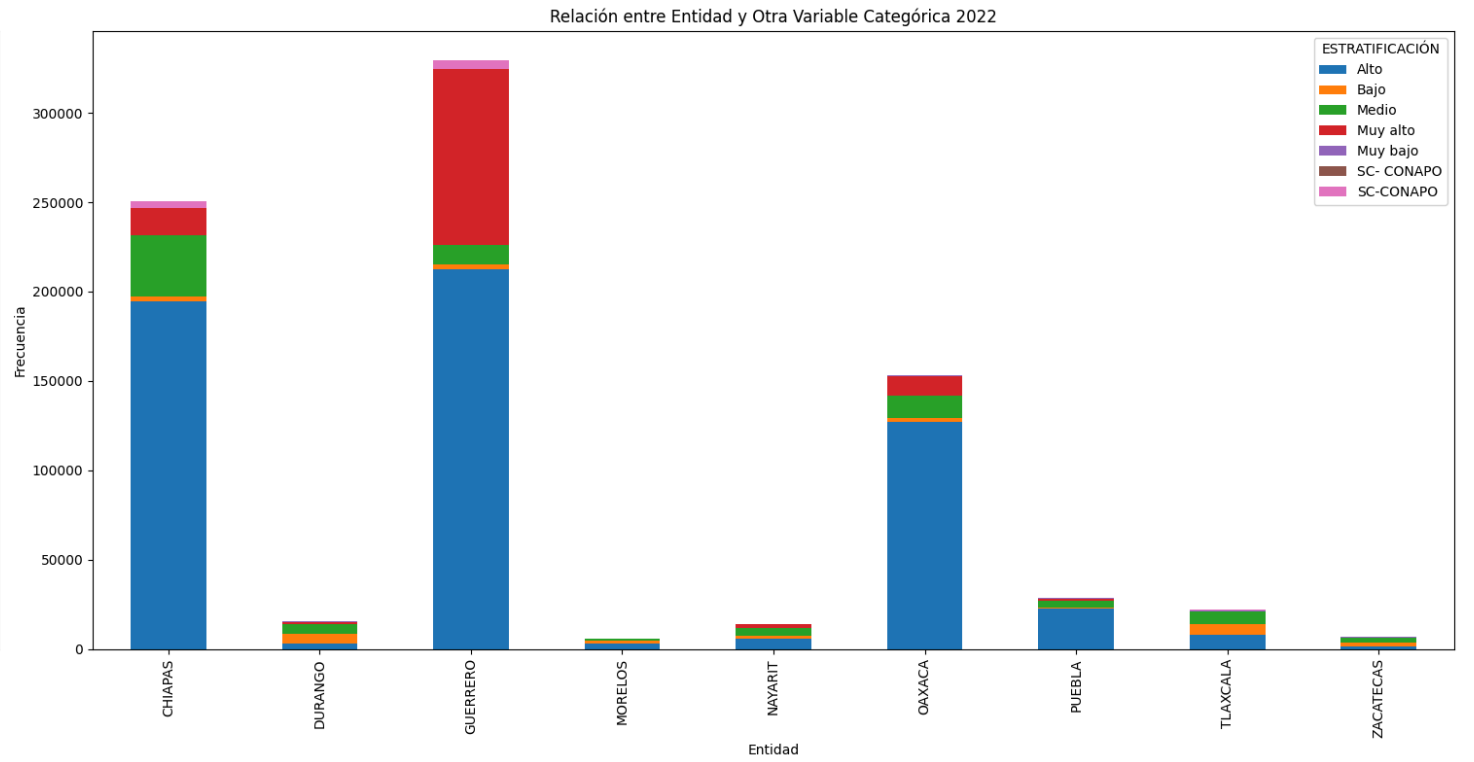
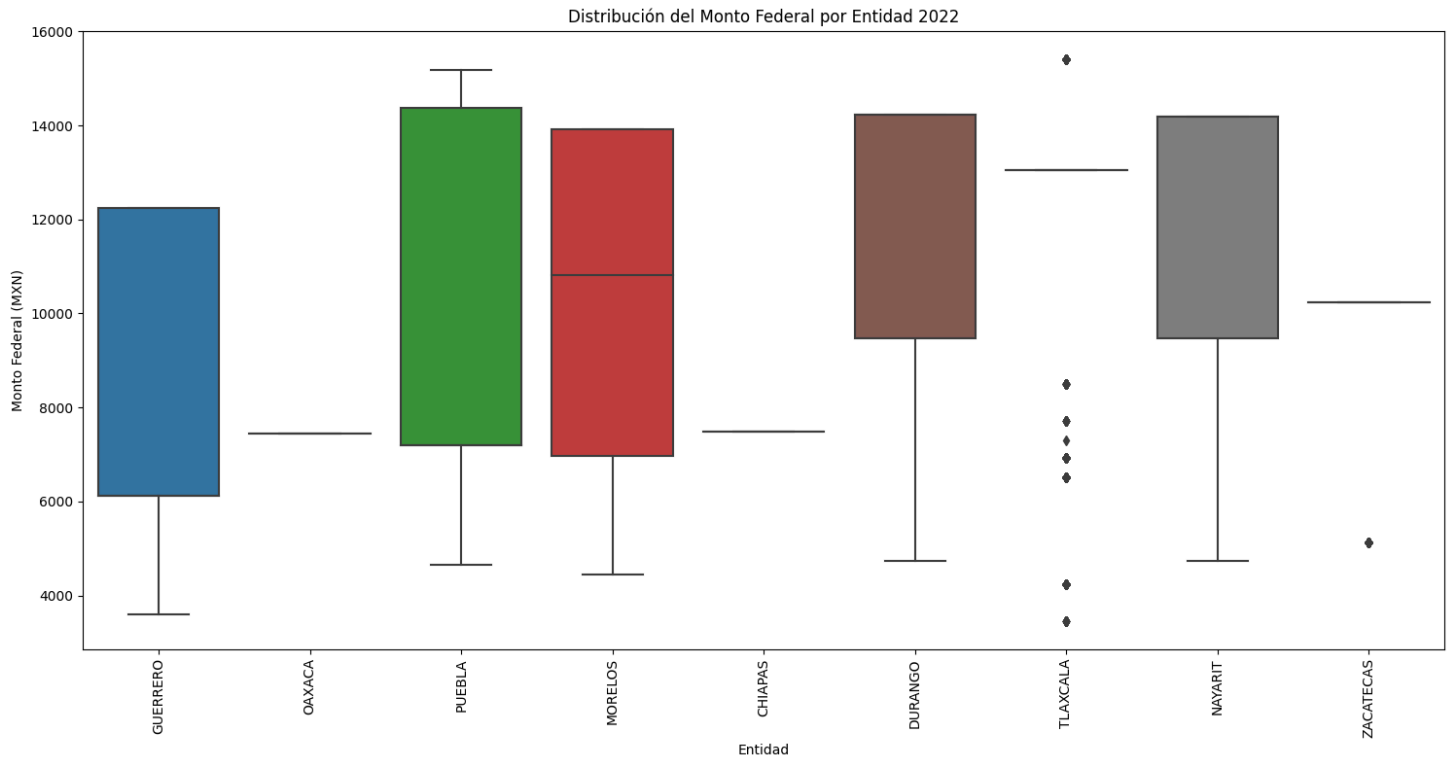
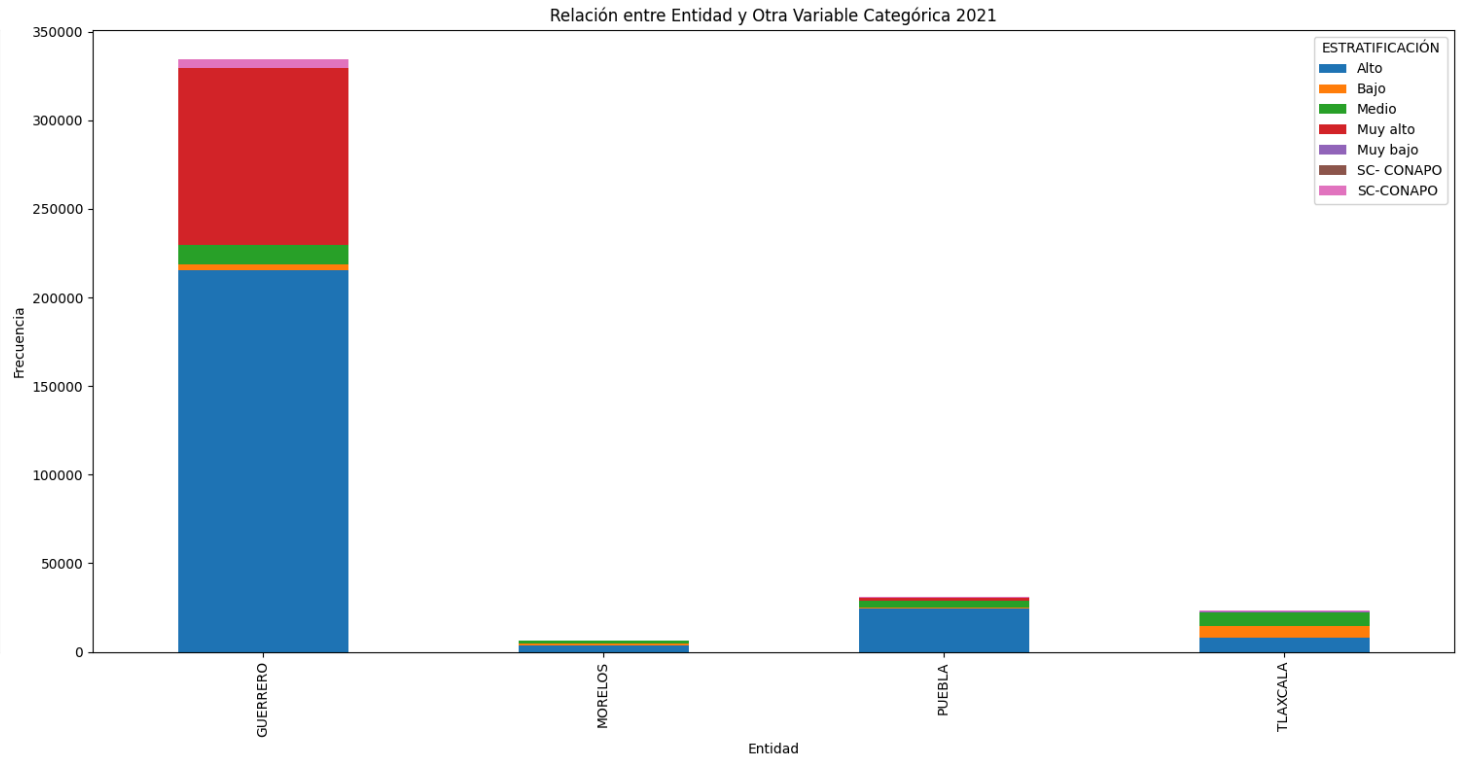
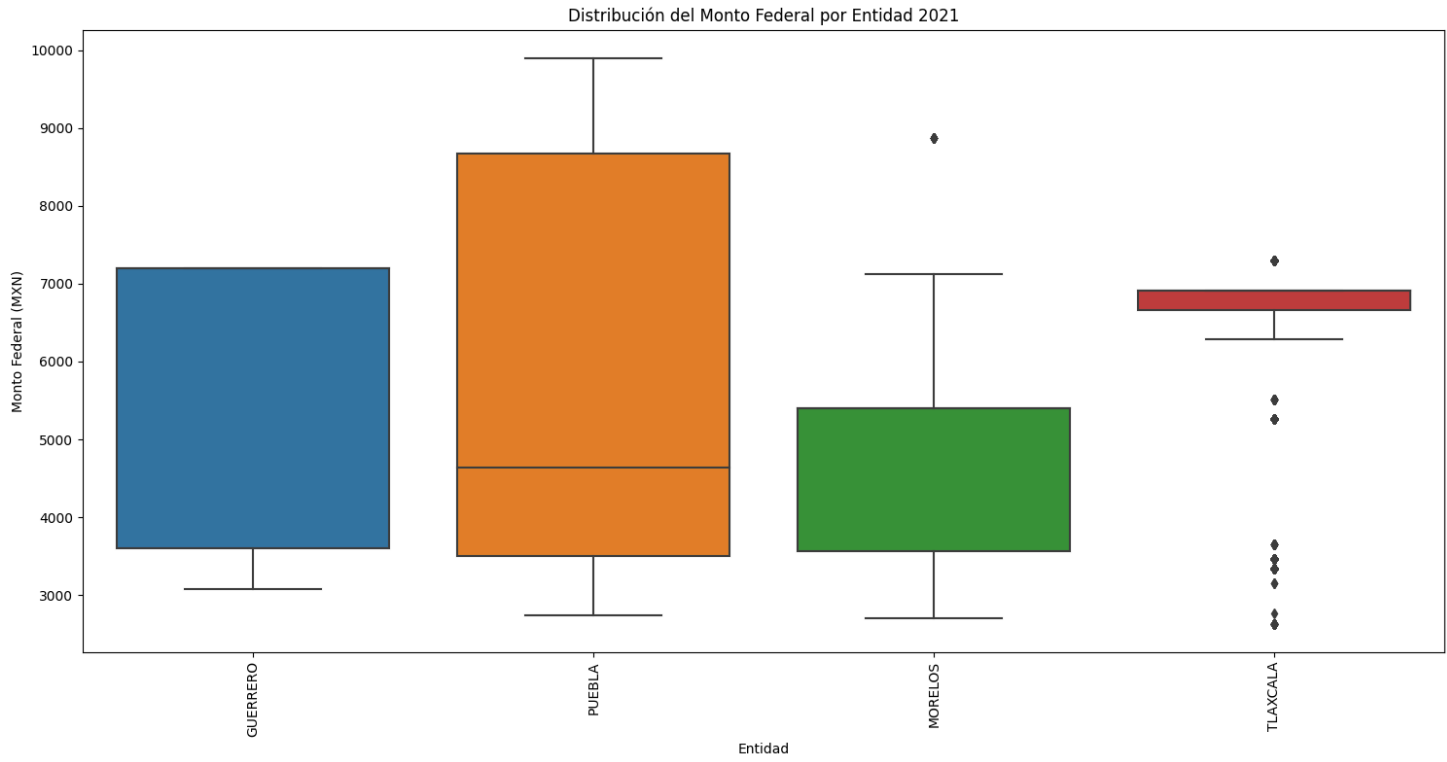
Conclusiones Preliminares

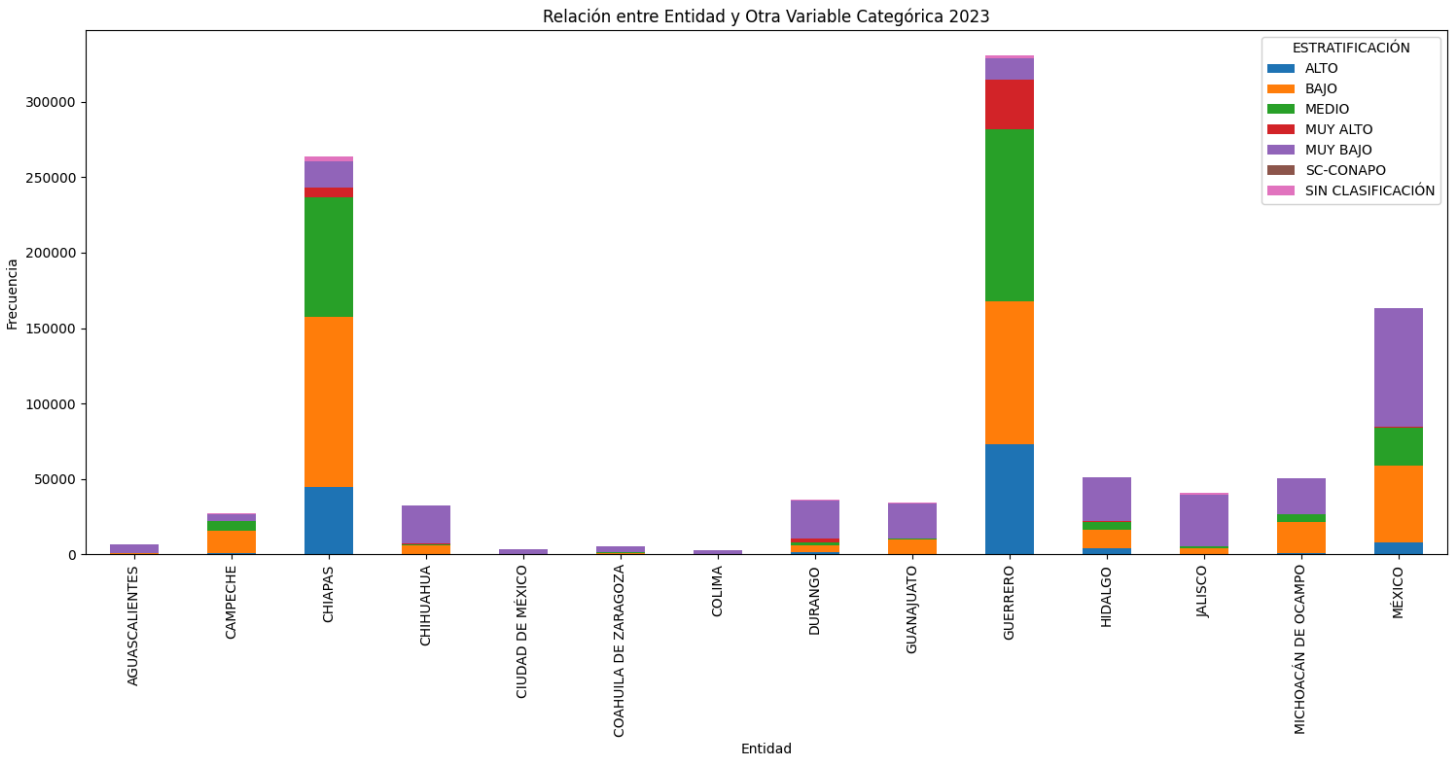
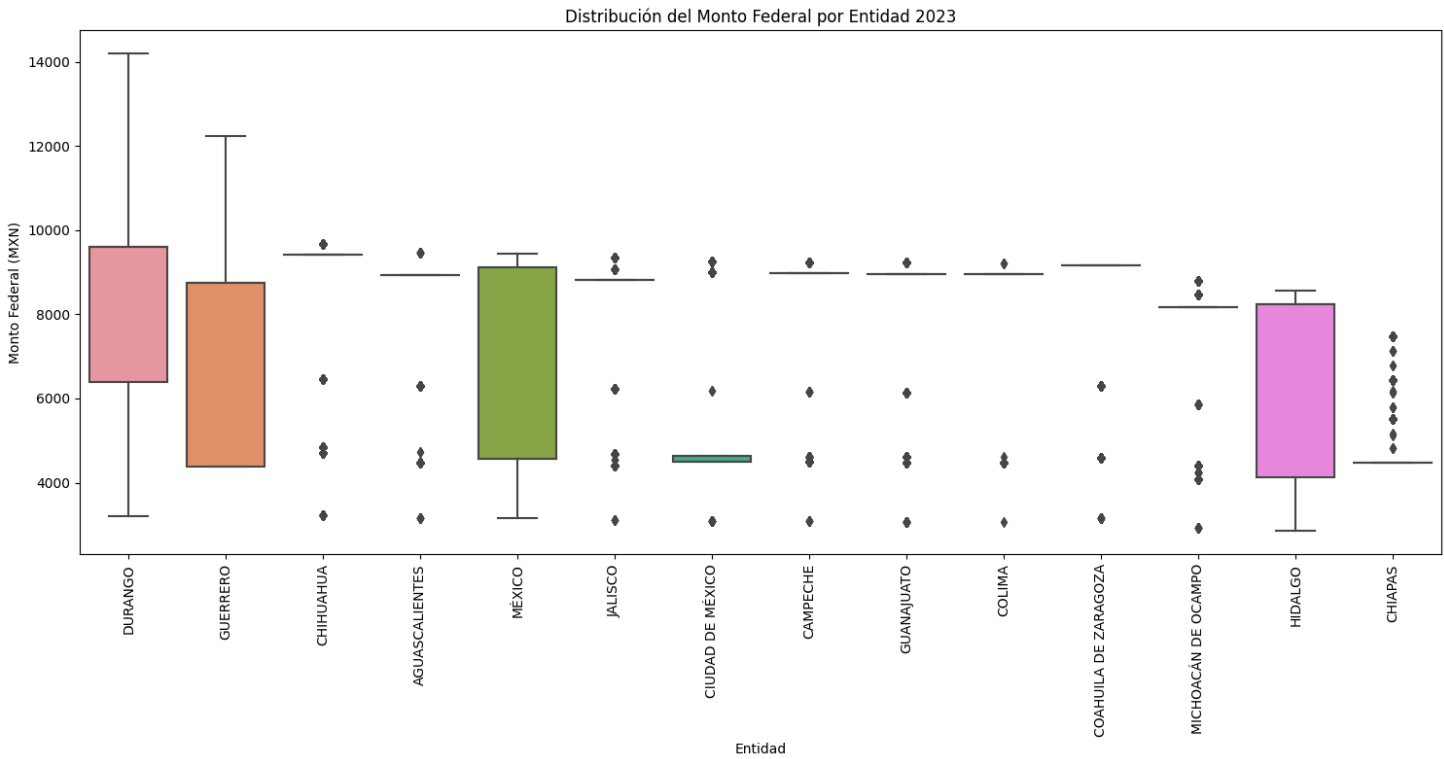
El incremento en el monto promedio de 2020 a 2021 muestra un cambio significativo en la cantidad de apoyos. Este aumento puede estar relacionado con una mayor necesidad de insumos agrícolas o una mayor cobertura del programa. La distribución por entidades revela que Guerrero sigue siendo la entidad con el mayor número de beneficiarios, lo que puede estar relacionado con la relevancia agrícola de esta región en términos de producción de fertilizantes. El sesgo negativo en 2020 y el sesgo neutral en 2021 reflejan una mejor distribución de los montos en 2021, sugiriendo una posible mejora en la equidad de los apoyos otorgados. La presencia de dos picos en la distribución de 2021 podría indicar cambios en la política de asignación, con diferentes montos asignados según ciertos criterios.

Análisis bivariado

```
In [10]: biviariado(Beneficiarios_2020, 2020)
biviariado(Beneficiarios_2021, 2021)
biviariado(Beneficiarios_2022, 2022)
biviariado(Beneficiarios_2023, 2023)
```





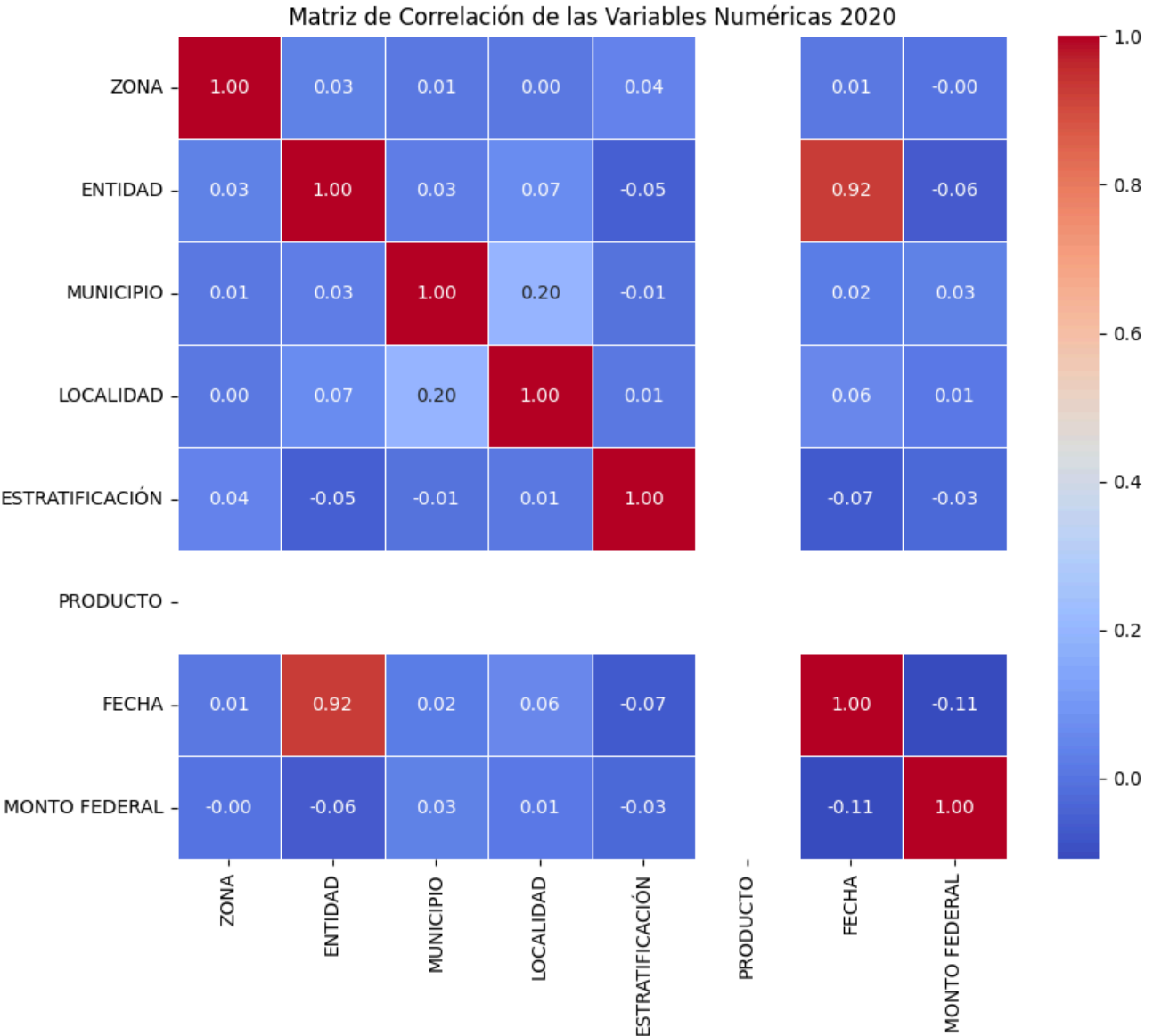


Análisis de Correlación

El análisis de correlación se realizó para identificar la relación entre las variables numéricas en el dataset del Programa de Fertilizantes. Dado que el dataset principalmente contiene datos categóricos, vamos a realizar una transformacion de los datos

```
In [11]: data_2020_encode = label_encode(Beneficiarios_2020)
data_2021_encode = label_encode(Beneficiarios_2021)
data_2022_encode = label_encode(Beneficiarios_2022)
data_2023_encode = label_encode(Beneficiarios_2023)
```

```
In [12]: correlacion(data_2020_encode,2020)
correlacion(data_2021_encode,2021)
correlacion(data_2022_encode,2022)
correlacion(data_2023_encode,2023)
```



Matriz de Correlación de las Variables Numéricas 2021

