

REPORTE

El reporte que escogí es acerca del número de muertos en estados a causa de sobredosis de drogas. Los datos que se analizan en este reporte son las edades, género, raza de las víctimas de sobredosis y los tipos de drogas que causaron dicha sobredosis.

Antes de comenzar el análisis del csv, pondré el aproximado de muertes totales en estados unidos en general para comparar un poco los datos.

- Año 2000: alrededor de 2.4 millones de muertes.
- Año 2005: alrededor de 2.5 millones de muertes.
- Año 2010: alrededor de 2.5 millones de muertes.
- Año 2015: alrededor de 2.7 millones de muertes.
- Año 2017: alrededor de 2.8 millones de muertes.

Estas cifras son aproximadas y pueden variar ligeramente según las fuentes de datos y los métodos de recopilación utilizados. Es importante tener en cuenta que estos números incluyen muertes por todas las causas, desde enfermedades hasta accidentes y eventos catastróficos.

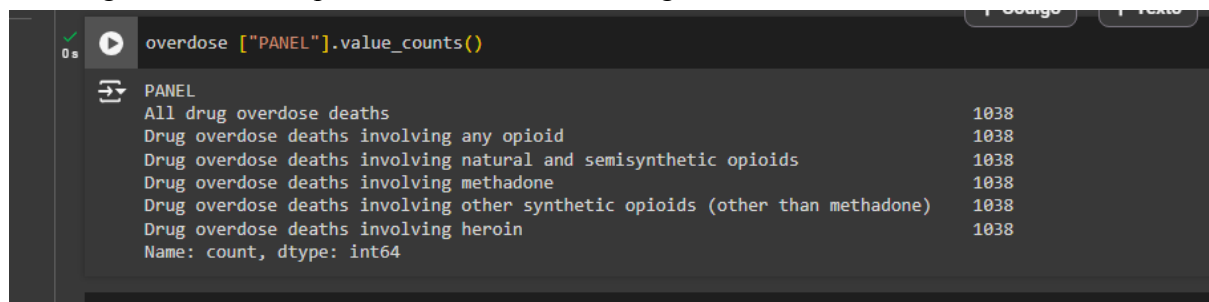
La información que se espera se recolecta en este reporte son:

- Grupo de edades ha muerto más por sobredosis
- Género es el que tiene más casos de sobredosis
- Droga que ha causado más sobredosis

Encontre este reporte en la pagina <https://catalog.data.gov/dataset>

The screenshot shows the Data.gov website interface. The top navigation bar includes 'DATA', 'REPORTS', 'OPEN GOVERNMENT', and 'CONTACT'. The main header is 'DATA CATALOG' with tabs for 'Datasets' and 'Organizations'. The breadcrumb trail indicates the path: 'U.S. Department of Health & Human Services / Centers for Disease Control and Prevention'. The dataset title is 'Drug overdose death rates, by drug type, sex, age, race, and Hispanic origin: United States', with a metadata update date of April 29, 2022. The description states: 'Data on drug overdose death rates, by drug type and selected population characteristics. Please refer to the PDF or Excel version of this table in the HUS 2019 Data Finder (https://www.cdc.gov/nchs/hus/contents2019.htm) for critical information about measures, definitions, and changes over time.' The source is cited as 'SOURCE: INCHS, National Vital Statistics System, numerator data from annual public-use Mortality Files; denominator data from U.S. Census Bureau national population estimates; and Murphy SL, Xu JQ, Kochanek KD, Arias E, Tejada-Vera B. Deaths: Final data for 2018. National Vital Statistics Reports; vol 69 no 13. Hyattsville, MD: National Center for Health Statistics 2021. Available from: https://www.cdc.gov/nchs/products/nvsr.htm. For more information on the National Vital Statistics System, see the corresponding Appendix entry at https://www.cdc.gov/nchs/data/hus19-appendix-508.pdf.' The 'Access & Use Information' section notes that the dataset is public and provides a link to the license. The 'Downloads & Resources' section offers a 'Comma Separated Values File' (161 views) and a 'Download' button, as well as an 'RD File'.

Las drogas en cuestión que se mencionan son las siguientes:

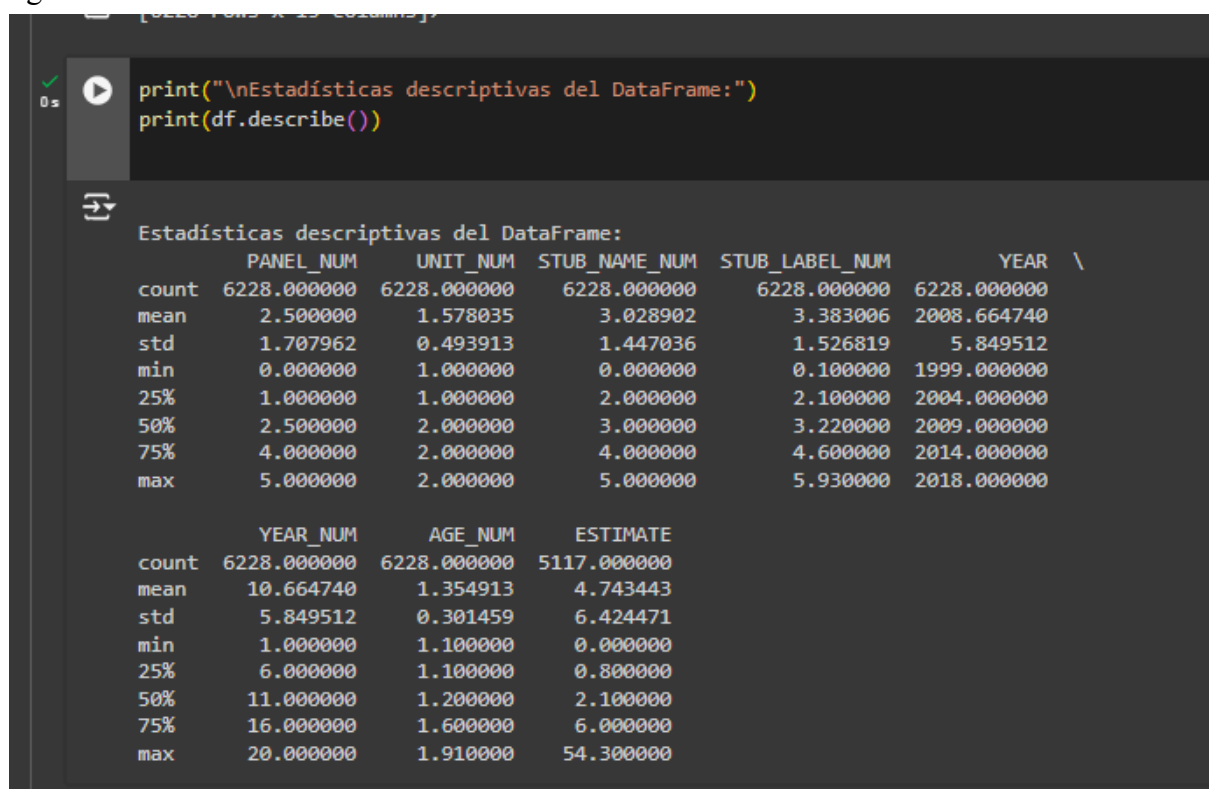


```
overdose ["PANEL"].value_counts()
```

PANEL	count
All drug overdose deaths	1038
Drug overdose deaths involving any opioid	1038
Drug overdose deaths involving natural and semisynthetic opioids	1038
Drug overdose deaths involving methadone	1038
Drug overdose deaths involving other synthetic opioids (other than methadone)	1038
Drug overdose deaths involving heroin	1038

Name: count, dtype: int64

Lo primero que intente hacer para poder analizar el csv fue checar los valores numéricos con los que contará el csv, intentar ver si a primera vista se nota cierta tendencia, me salio lo siguiente:



```
print("\nEstadísticas descriptivas del DataFrame:")
print(df.describe())
```

Estadísticas descriptivas del DataFrame:

	PANEL_NUM	UNIT_NUM	STUB_NAME_NUM	STUB_LABEL_NUM	YEAR
count	6228.000000	6228.000000	6228.000000	6228.000000	6228.000000
mean	2.500000	1.578035	3.028902	3.383006	2008.664740
std	1.707962	0.493913	1.447036	1.526819	5.849512
min	0.000000	1.000000	0.000000	0.100000	1999.000000
25%	1.000000	1.000000	2.000000	2.100000	2004.000000
50%	2.500000	2.000000	3.000000	3.220000	2009.000000
75%	4.000000	2.000000	4.000000	4.600000	2014.000000
max	5.000000	2.000000	5.000000	5.930000	2018.000000

	YEAR_NUM	AGE_NUM	ESTIMATE
count	6228.000000	6228.000000	5117.000000
mean	10.664740	1.354913	4.743443
std	5.849512	0.301459	6.424471
min	1.000000	1.100000	0.000000
25%	6.000000	1.100000	0.800000
50%	11.000000	1.200000	2.100000
75%	16.000000	1.600000	6.000000
max	20.000000	1.910000	54.300000

Como se puede observar, son números bastante altos. Entonces, decidí checar por separado cada columna para poder ver qué representan estos números...

0 s



```
df["STUB_LABEL"].value_counts()
```



```
STUB_LABEL
All persons                240
Female                    240
Male                      240
Male: Hispanic or Latino: All races    126
Female: Not Hispanic or Latino: American Indian or Alaska Native    126
Female: Not Hispanic or Latino: Black    126
Female: Not Hispanic or Latino: White    126
Male: Not Hispanic or Latino: American Indian or Alaska Native    126
Male: Not Hispanic or Latino: Black    126
Male: Not Hispanic or Latino: White    126
Female: Hispanic or Latino: All races    126
Female: Black or African American    126
Female: White    126
Male: American Indian or Alaska Native    126
Male: Black or African American    126
Female: American Indian or Alaska Native    126
Male: White    126
Female: 75-84 years    120
Female: 15-24 years    120
Male: 55-64 years    120
Male: 65-74 years    120
Male: 75-84 years    120
Male: 85 years and over    120
Female: Under 15 years    120
Female: 45-54 years    120
Female: 25-34 years    120
Female: 35-44 years    120
Female: 55-64 years    120
Female: 85 years and over    120
Male: 35-44 years    120
Female: 65-74 years    120
Male: 45-54 years    120
55-64 years    120
Male: 25-34 years    120
25-34 years    120
Male: Asian or Pacific Islander    120
Female: Asian or Pacific Islander    120
Male: Not Hispanic or Latino: Asian or Pacific Islander    120
Female: Not Hispanic or Latino: Asian or Pacific Islander    120
Under 15 years    120
Male: 15-24 years    120
15-24 years    120
35-44 years    120
45-54 years    120
65-74 years    120
75-84 years    120
85 years and over    120
Male: Under 15 years    120
Male: Not Hispanic or Latino: Asian    6
Male: Not Hispanic or Latino: Native Hawaiian or Other Pacific Islander    6
Female: Not Hispanic or Latino: Asian    6
Female: Not Hispanic or Latino: Native Hawaiian or Other Pacific Islander    6
Name: count, dtype: int64
```

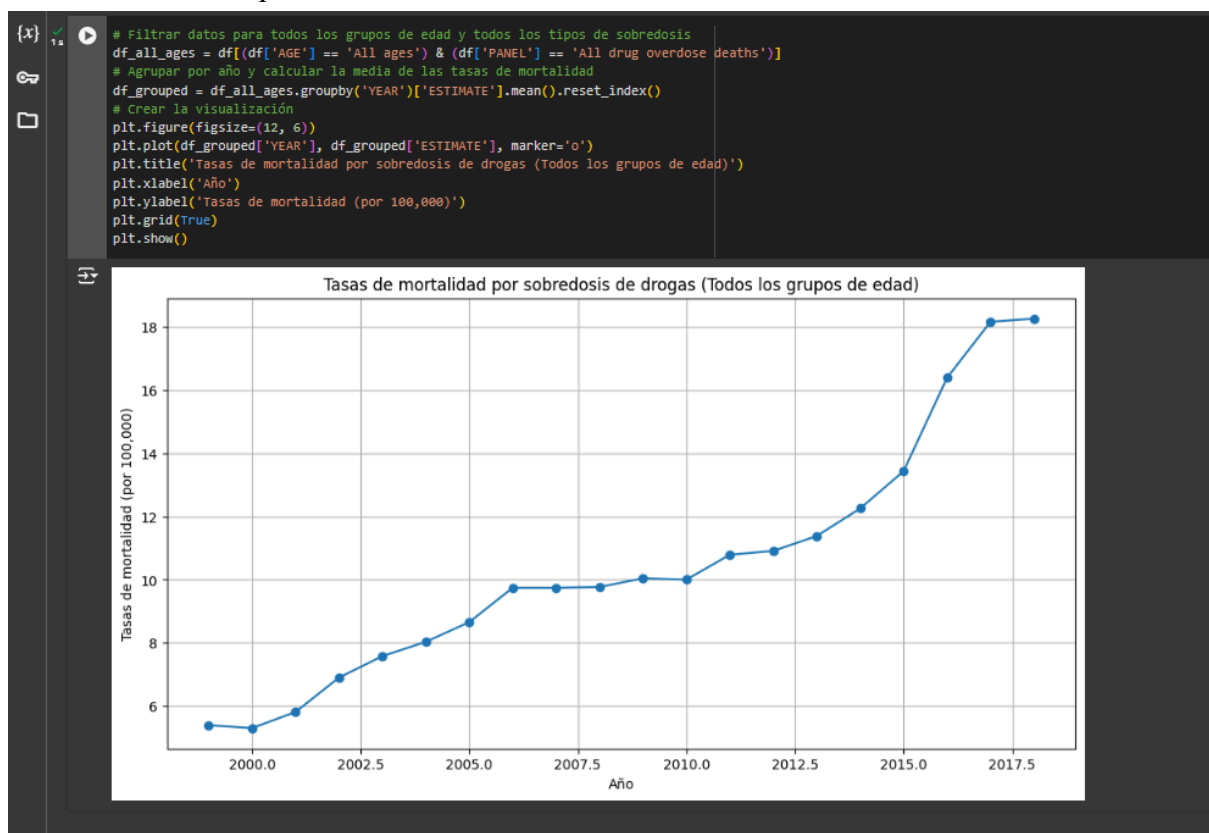
no es disponible

```
[34] df["AGE"].value_counts()

AGE
All ages          2988
Under 15 years     360
15-24 years        360
25-34 years        360
35-44 years        360
45-54 years        360
55-64 years        360
65-74 years        360
75-84 years        360
85 years and over  360
Name: count, dtype: int64
```

Como se puede observar, la columna "STUB_LABEL" y "AGE" y "PANEL" indican el tipo de drogas que causaron la sobredosis y las características de personas que entran en este csv, su raza, intervalo de edades y su sexo.

Después de entender un poco más lo que representaban esos números, considere primero una tabla de mortalidad por sobredosis.



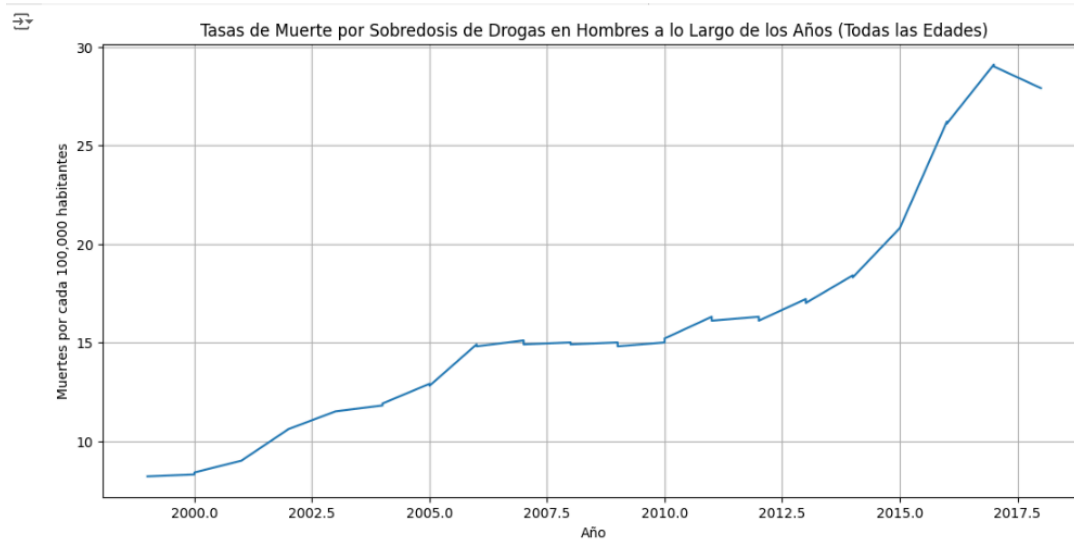
La siguiente gráfica de la misma manera expresa la tasa de mortalidad por sobredosis, pero ahora solo aplicada a los hombres.

```
# Filtrar los datos para incluir solo 'All drug overdose deaths', 'Total' y 'Male' para todas las edades
men_filtered_data = df[(df['AGE'] == 'All ages') &
                      (df['PANEL'] == 'All drug overdose deaths') &
                      (df['STUB_LABEL'] == 'Male')]

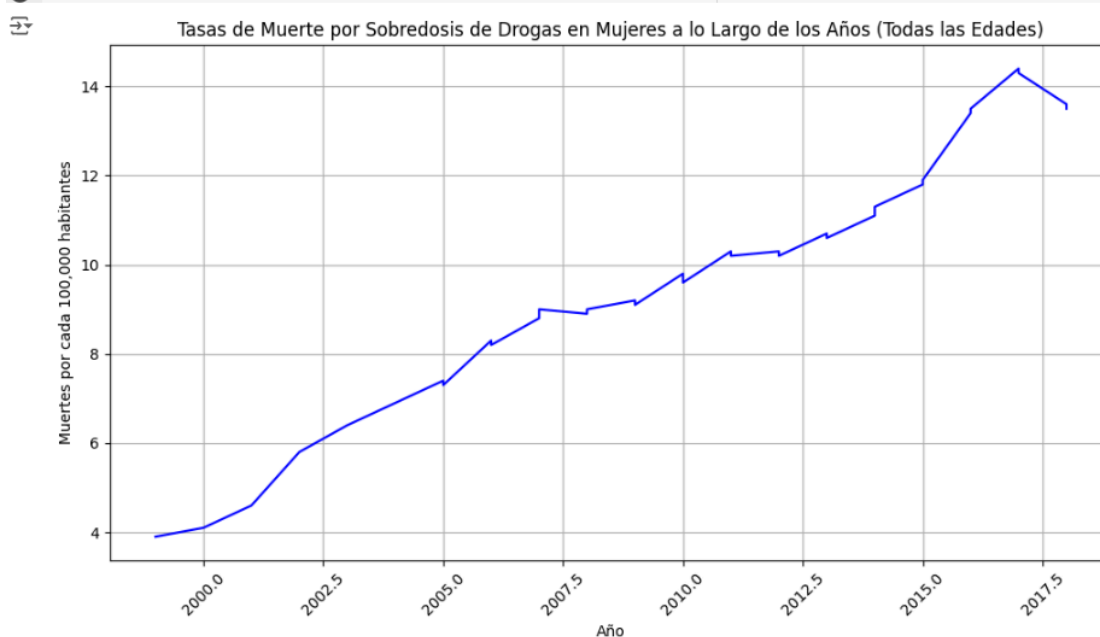
# Ordenar los datos por año en orden ascendente
men_filtered_data_sorted = men_filtered_data.sort_values(by='YEAR')

# Extraer las columnas relevantes para la visualización
men_years = men_filtered_data_sorted['YEAR']
men_death_rates = men_filtered_data_sorted['ESTIMATE']

# Generar la gráfica
plt.figure(figsize=(13, 6))
plt.plot(men_years, men_death_rates)
plt.title('Tasas de Muerte por Sobredosis de Drogas en Hombres a lo Largo de los Años (Todas las Edades)')
plt.xlabel('Año')
plt.ylabel('Muertes por cada 100,000 habitantes')
plt.grid(True)
plt.show()
```



La siguiente tabla ahora es aplicada para mujeres.



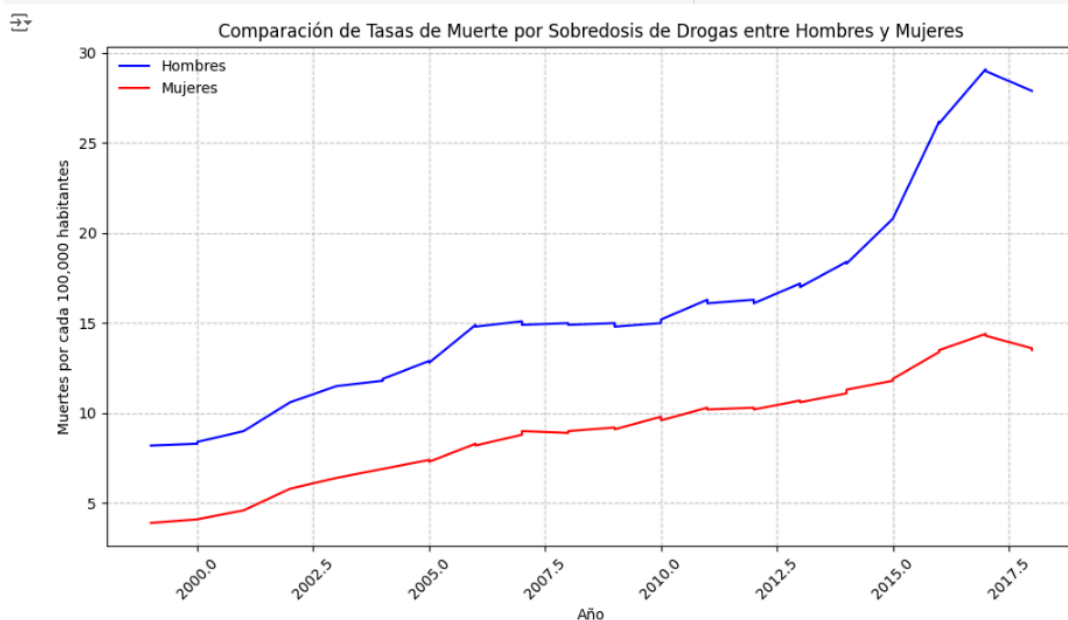
La siguiente gráfica sirve para comparar las gráficas de los dos sexos por separado.

```
[ ] # Generar la comparación con líneas para hombres y mujeres sin relieves
plt.figure(figsize=(10, 6))

# Línea para hombres
plt.plot(men_years, men_death_rates, linestyle='-', color='b', label='Hombres')

# Línea para mujeres
plt.plot(female_years, female_death_rates, linestyle='-', color='r', label='Mujeres')

plt.title('Comparación de Tasas de Muerte por Sobredosis de Drogas entre Hombres y Mujeres')
plt.xlabel('Año')
plt.ylabel('Muertes por cada 100,000 habitantes')
plt.legend(frameon=False) # Quitar el marco de la leyenda para un aspecto más limpio
plt.grid(True, linestyle='--', alpha=0.7) # Ajustar la apariencia de la cuadrícula
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



El siguiente histograma representa las víctimas de sobredosis separadas por grupos de edades.

```

# Filtrar los datos para incluir solo 'All drug overdose deaths' y todos los grupos de edad
age_filtered_data = df[(df['PANEL'] == 'All drug overdose deaths') &
                      (df['STUB_NAME'] == 'Total') &
                      (df['AGE'] != 'All ages')]

# Inspeccionar las categorías únicas en la columna 'AGE'
unique_ages_in_data = df['AGE'].unique()
unique_ages_in_data

# Ver algunas filas del conjunto de datos filtrado
age_filtered_data.head()

# Filtrar los datos para incluir solo los grupos de edad sin considerar otras condiciones
age_only_filtered_data = df[df['AGE'].isin(unique_ages_in_data) & (df['AGE'] != 'All ages')]

# Ver algunas filas del conjunto de datos filtrado
age_only_filtered_data.head() #Moficar numero para visualizar los datos

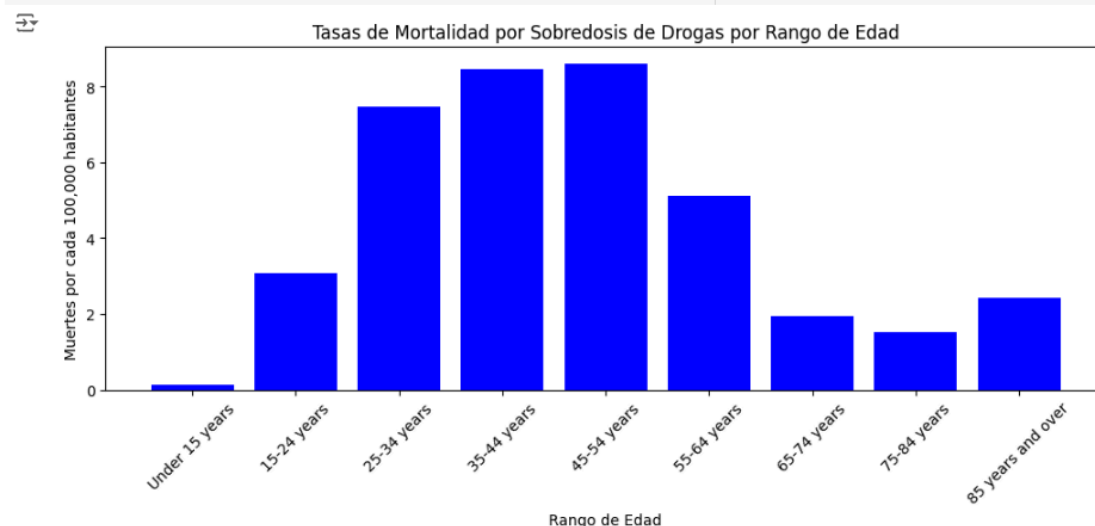
# Agrupar por edad y calcular la media de las tasas de muerte
age_grouped_data = age_only_filtered_data.groupby('AGE')['ESTIMATE'].mean().reset_index()

# Ordenar por las categorías de edad
age_order = ['Under 15 years', '15-24 years', '25-34 years', '35-44 years', '45-54 years', '55-64 years', '65-74 years', '75-84 years', '85 years and over']
age_grouped_data['AGE'] = pd.Categorical(age_grouped_data['AGE'], categories=age_order, ordered=True)
age_grouped_data = age_grouped_data.sort_values('AGE')

# Generar la gráfica de barras para las tasas de mortalidad por rango de edad
plt.figure(figsize=(10, 5))
plt.bar(age_grouped_data['AGE'], age_grouped_data['ESTIMATE'], color='b')
plt.title('Tasas de Mortalidad por Sobredosis de Drogas por Rango de Edad')
plt.xlabel('Rango de Edad')
plt.ylabel('Muertes por cada 100,000 habitantes')
plt.xticks(rotation=45)
plt.tight_layout()

plt.show()

```



A continuación, la próxima gráfica representa los tipos de droga que causan las sobredosis

```

# Filtrar los datos para incluir solo los tipos de drogas analizadas
drug_types_filtered_data = df[(df['PANEL'] != 'All drug overdose deaths') & (df['AGE'] == 'All ages')]

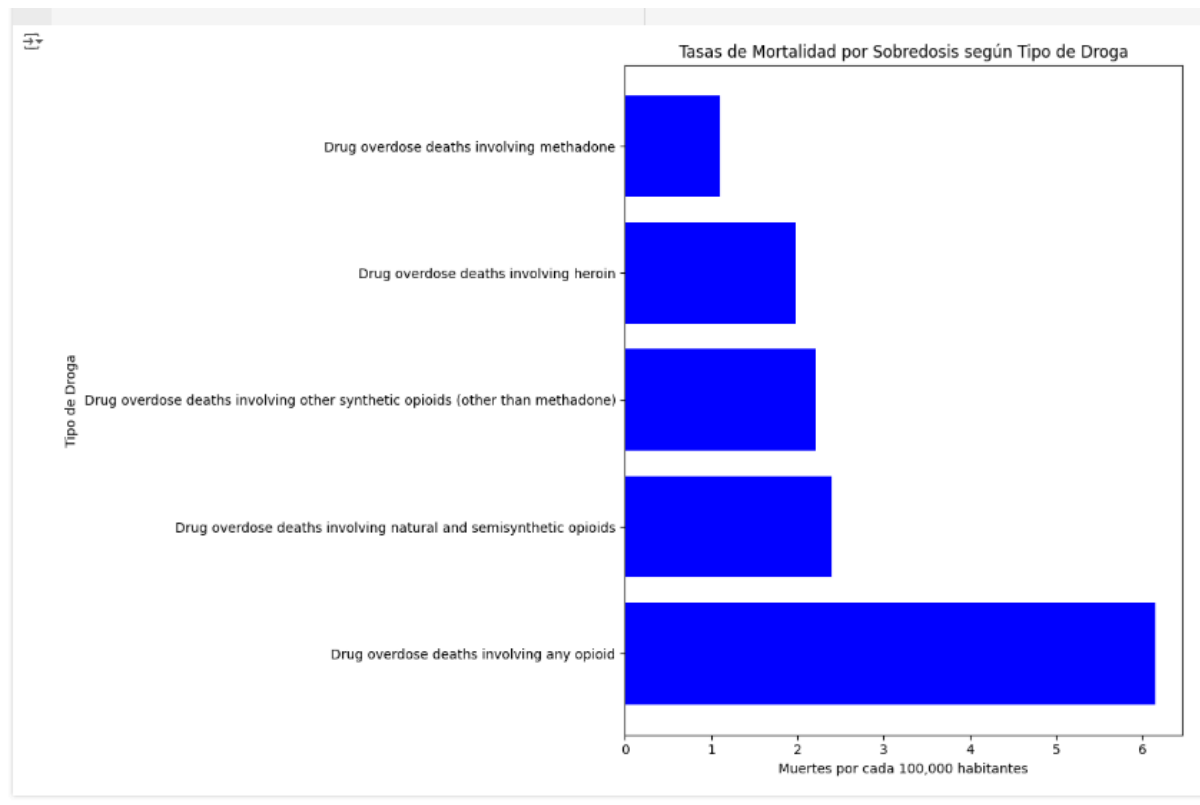
# Agrupar por tipo de droga y calcular la media de las tasas de muerte
drug_types_grouped_data = drug_types_filtered_data.groupby('PANEL')['ESTIMATE'].mean().reset_index()

# Ordenar por las tasas de muerte de forma descendente
drug_types_grouped_data = drug_types_grouped_data.sort_values(by='ESTIMATE', ascending=False)

# Generar la gráfica de barras para las tasas de mortalidad por tipo de droga
plt.figure(figsize=(12, 8))
plt.barh(drug_types_grouped_data['PANEL'], drug_types_grouped_data['ESTIMATE'], color='b')
plt.title('Tasas de Mortalidad por Sobredosis según Tipo de Droga')
plt.xlabel('Muertes por cada 100,000 habitantes')
plt.ylabel('Tipo de Droga')
plt.tight_layout()

# Mostrar la gráfica
plt.show()

```



Conclusiones particulares

Gracias a las gráficas de las librerías pandas, numpy y matplotlib podemos conseguir la información que planteamos al inicio:

i. Grupo de edades ha muerto más por sobredosis

La gráfica indica que la mortalidad por sobredosis de drogas es significativamente mayor en los adultos de mediana edad, especialmente en los grupos de 35 a 54 años. Los jóvenes menores de 15 años y los adultos mayores (especialmente aquellos de 75 a 84 años) tienen las tasas más bajas de mortalidad por sobredosis. El grupo de 85 años y más muestra un leve incremento en comparación con los grupos de edad inmediatamente anteriores, pero sigue siendo menor que en los adultos de mediana edad.

ii. Género es el que tiene más casos de sobredosis

La gráfica indica que las tasas de mortalidad por sobredosis de drogas han aumentado significativamente tanto para hombres como para mujeres desde el año 2000. Sin embargo, los hombres han experimentado tasas de mortalidad consistentemente más altas que las mujeres a lo largo de este período. El aumento más pronunciado ocurrió alrededor de 2015, y aunque ambas tasas muestran una ligera disminución después de 2017, las tasas actuales siguen siendo mucho más altas que las observadas al inicio del período.

iii. Droga que ha causado más sobredosis

La gráfica sugiere que los opioides, en general, son responsables de la mayor parte de las muertes por sobredosis, con los opioides sintéticos distintos a la metadona y la heroína

también representando una parte significativa de estas muertes. Las muertes por sobredosis relacionadas con la metadona son comparativamente menores.

Conclusiones generales

Teniendo en cuenta la información del inicio acerca la tasa de mortalidad en estados unidos en un intervalo de años (2000-2017)

- Año 2000: alrededor de 2.4 millones de muertes.
- Año 2005: alrededor de 2.5 millones de muertes.
- Año 2010: alrededor de 2.5 millones de muertes.
- Año 2015: alrededor de 2.7 millones de muertes.
- Año 2017: alrededor de 2.8 millones de muertes.

Estas cifras son aproximadas y pueden variar ligeramente según las fuentes de datos y los métodos de recopilación utilizados. Es importante tener en cuenta que estos números incluyen muertes por todas las causas, desde enfermedades hasta accidentes y eventos catastróficos.

Podemos concluir gracias a las gráficas que las tasas de mortalidad por sobredosis han incrementado de manera preocupante y contribuyen en mayor medida a la tasa de mortalidad total, aunque la tasa general no ha variado de manera tan significativa. Esta comparación resalta la importancia de abordar la crisis de opioides como un problema crítico de salud pública dentro del contexto más amplio de la mortalidad en EE.UU.