

```
In [1]: import pandas as pd
import numpy as np

import matplotlib.pyplot as plt
from matplotlib import style
import seaborn as sns

import statsmodels.api as sm
from scipy import stats

import warnings
warnings.filterwarnings('ignore')

from matplotlib.pyplot import figure
figure(figsize=(8, 6), dpi=80)
```

Out[1]: <Figure size 640x480 with 0 Axes>
<Figure size 640x480 with 0 Axes>

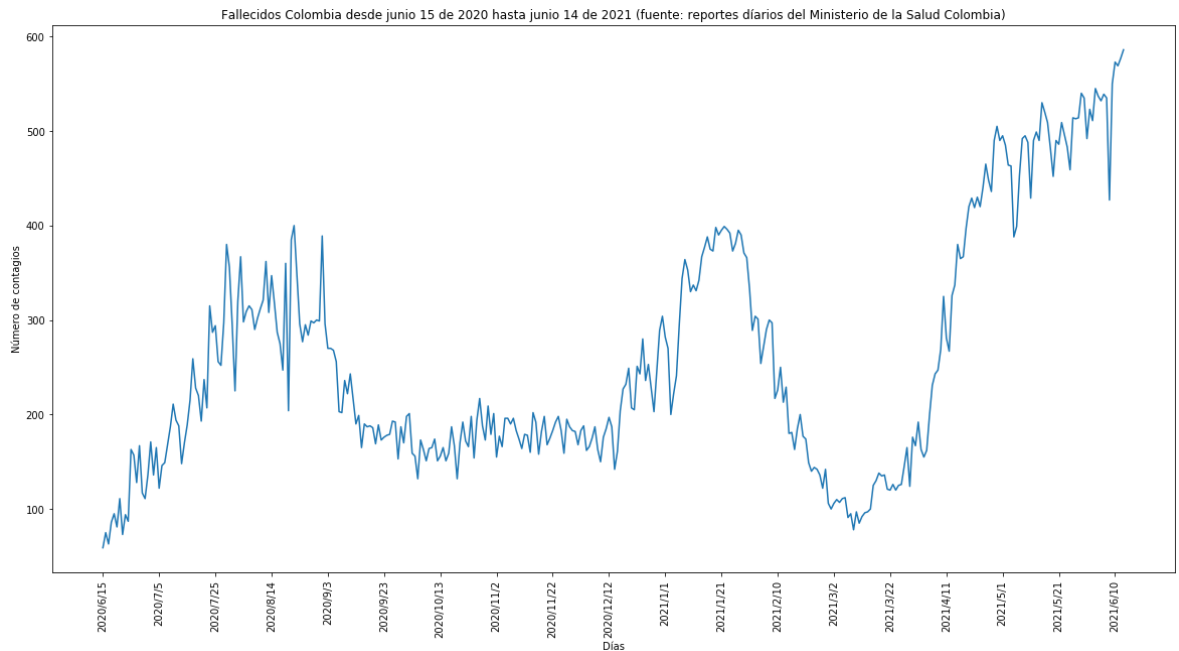
```
In [2]: data1 = pd.read_csv("../datasets/reportes_diarios.csv")
days = pd.read_csv("../datasets/days.csv")
#days["days"] = pd.to_datetime(days['days'])
days
```

Out[2]:

	days
0	2020/6/15
1	2020/6/16
2	2020/6/17
3	2020/6/18
4	2020/6/19
...	...
359	2021/6/9
360	2021/6/10
361	2021/6/11
362	2021/6/12
363	2021/6/13

364 rows × 1 columns

```
In [3]: x = days["days"]#np.arange(0, 364, 1)#
y = data1["fallecidos"]
plt.figure(figsize=(20, 10))
plt.plot(x, y)
xt = np.arange(0, 364, 20)
plt.xticks(xt, rotation="vertical")
plt.title("Fallecidos Colombia desde junio 15 de 2020 hasta junio 14 de 2021 (fuente: reportes diarios del Ministerio de la Salud Colombi
a)")
plt.xlabel("Días")
plt.ylabel("Número de contagios")
plt.show()
```



```

In [4]: premierDay = 200 #200 es el índice del día primero de enero de 2021
b1 = data1[:premierDay] #Conjunto de datos hasta el 31 de diciembre d
e 2020
a1 = data1[premierDay:] #Conjunto de datos desde el 1 de enero de 202
1
b1["date"] = days[:premierDay]
a1["date"] = days[premierDay:]
#Guardando en forma de csv los reportes diarios para cada año
b1.to_csv("./out/reportes_covid_minsalud_2020.csv")
a1.to_csv("./out/reportes_covid_minsalud_2021.csv")
#La vacunación en Colombia empezó a finales de febrero del 2021, por
lo que asumimos, para más seguridad
#que desde marzo el plan de vacunación en Colombia, contra Covid-19,
ha estado activo
inicioVacunacion = 259 #El día 259 de la recolección de reportes corr
esponde al 1 de marzo de 2021
bv = data1[inicioVacunacion]
av = data1[inicioVacunacion:]
bv["date"] = days[:inicioVacunacion]
av["date"] = days[inicioVacunacion:]
bv.to_csv("./out/reportes_antes_vacunacion_minsalud.csv")
av.to_csv("./out/reportes_despues_vacunacion_minsalud.csv")

```

```

In [5]: #Seperación en los grupos deseados:
#Grupo1: 20 días antes del 28 de abril
#Grupo2: 20 días despues del 28 de abril
#Nota: El 28 de abril iniciaron las protestas del Paro Nacional 2021
d28a = 317 #El día 328 de los datos corresponde al 28 de abril
b28a = data1[d28a-20:d28a] #Grupo1
a28a = data1[d28a:d28a+20] #Grupo2
b28a["date"] = days[d28a-20:d28a]
a28a["date"] = days[d28a:d28a+20]

```

```

In [21]: #Guardando los datos de los casos reportados 20 días antes del 28 de
abril y 20 días despues del 28 de abril

b28a.to_csv("./out/reportes_antes_28a_minsalud.csv")
a28a.to_csv("./out/reportes_despues_28a_minsalud.csv")

```

```
In [6]: #Mostrando graficamente los nuevos casos de Covid, días antes del 28
de abril y días despues del 28
x = np.arange(0, 20, 1)

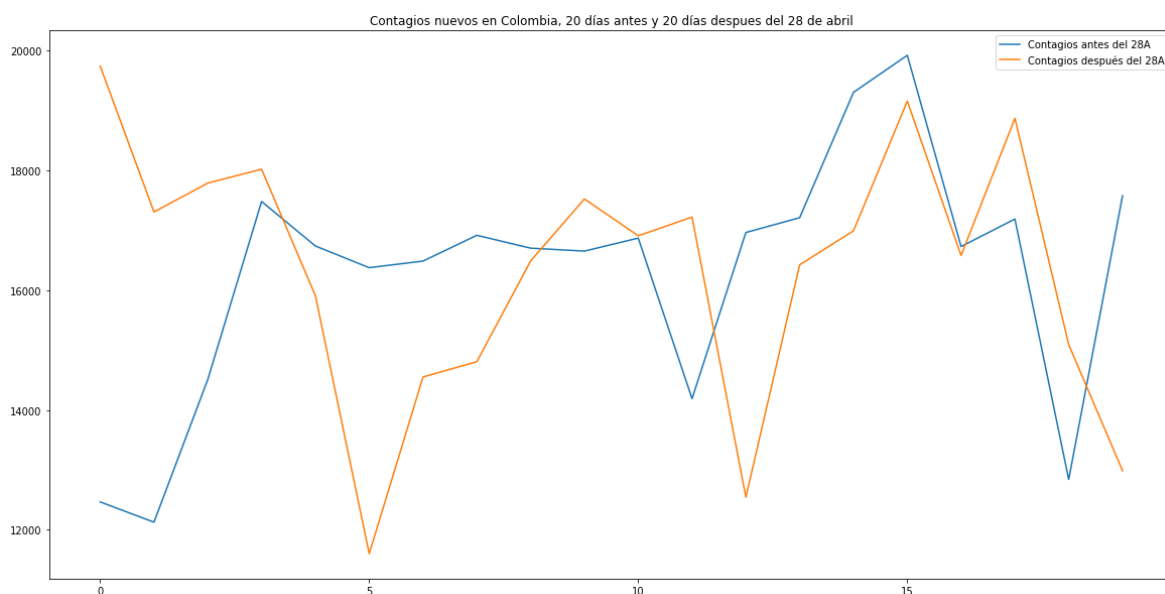
plt.figure(figsize=(20, 10))

plt.plot(x, b28a["nuevos"])
#plt.xticks(xt, rotation="vertical")

plt.plot(x, a28a["nuevos"])
xt = np.arange(0, 20, 5)
plt.legend(["Contagios antes del 28A", "Contagios después del 28A"])
plt.xticks(xt)

plt.title("Contagios nuevos en Colombia, 20 días antes y 20 días desp
ues del 28 de abril")

plt.show()
```



```

In [7]: #Mostrando graficamente los fallecimientos por Covid, días antes del
        28 de abril y días despues del 28
x = np.arange(0, 20, 1)

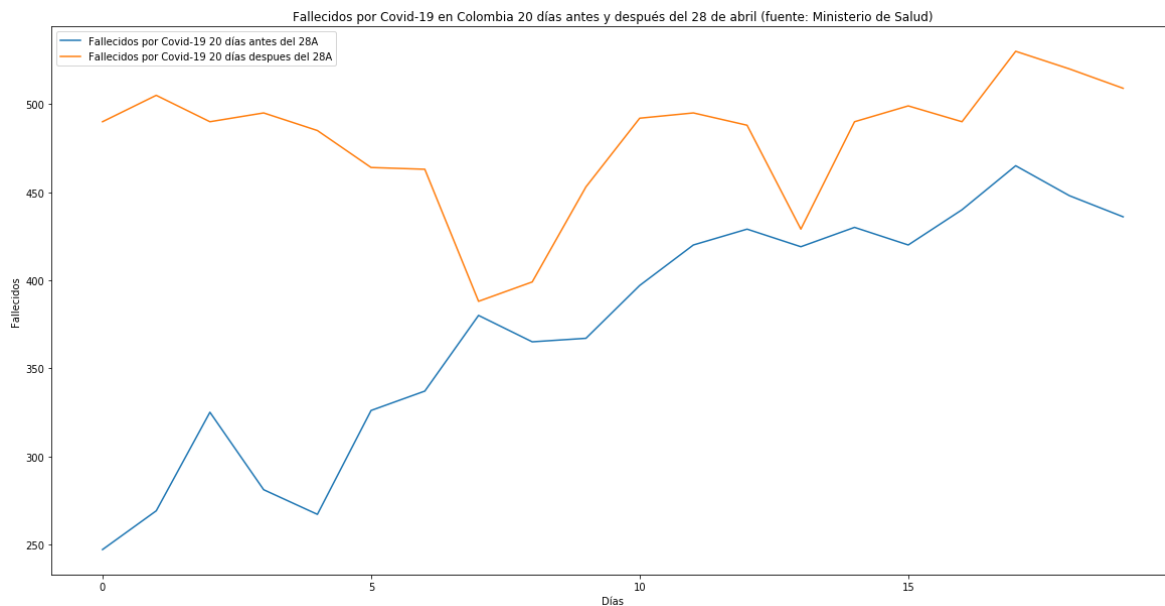
plt.figure(figsize=(20, 10))

plt.plot(x, b28a["fallecidos"])
#plt.xticks(xt, rotation="vertical")

plt.plot(x, a28a["fallecidos"])
xt = np.arange(0, 20, 5)
plt.xlabel("Días")
plt.ylabel("Fallecidos")
plt.legend(["Fallecidos por Covid-19 20 días antes del 28A", "Falleci
dos por Covid-19 20 días despues del 28A"])
plt.xticks(xt)
plt.title("Fallecidos por Covid-19 en Colombia 20 días antes y después
s del 28 de abril (fuente: Ministerio de Salud)")

plt.show()

```



```
In [8]: data2 = pd.read_csv("./datasets/Casos_positivos_de_COVID-19_en_Colomb
ia.csv")
data2
```

Out[8]:

	fecha reporte web	ID de caso	Fecha de notificación	Código DIVIPOLA departamento	Nombre departamento	Código DIVIPOLA municipio	Nombre municipio
0	6/3/2020 0:00:00	1	2/3/2020 0:00:00	11	BOGOTA	11001	BOGOTA
1	9/3/2020 0:00:00	2	6/3/2020 0:00:00	76	VALLE	76111	BUGA
2	9/3/2020 0:00:00	3	7/3/2020 0:00:00	5	ANTIOQUIA	5001	MEDELLIN
3	11/3/2020 0:00:00	4	9/3/2020 0:00:00	5	ANTIOQUIA	5001	MEDELLIN
4	11/3/2020 0:00:00	5	9/3/2020 0:00:00	5	ANTIOQUIA	5001	MEDELLIN
...
3777595	14/6/2021 0:00:00	3777636	2/6/2021 0:00:00	13	BOLIVAR	13549	PINILLOS
3777596	14/6/2021 0:00:00	3777637	2/6/2021 0:00:00	13	BOLIVAR	13549	PINILLOS
3777597	14/6/2021 0:00:00	3777638	2/6/2021 0:00:00	13	BOLIVAR	13549	PINILLOS
3777598	14/6/2021 0:00:00	3777639	2/6/2021 0:00:00	13	BOLIVAR	13549	PINILLOS
3777599	14/6/2021 0:00:00	3777640	1/6/2021 0:00:00	13	BOLIVAR	13042	ARENAL

3777600 rows × 23 columns

```
In [9]: data2.columns
```

```
Out[9]: Index(['fecha reporte web', 'ID de caso', 'Fecha de notificación',
              'Código DIVIPOLA departamento', 'Nombre departamento',
              'Código DIVIPOLA municipio', 'Nombre municipio', 'Edad',
              'Unidad de medida de edad', 'Sexo', 'Tipo de contagio',
              'Ubicación del caso', 'Estado', 'Código ISO del país',
              'Nombre del país', 'Recuperado', 'Fecha de inicio de síntoma
s',
              'Fecha de muerte', 'Fecha de diagnóstico', 'Fecha de recuperac
ión',
              'Tipo de recuperación', 'Pertenencia étnica',
              'Nombre del grupo étnico'],
              dtype='object')
```

```
In [10]: #Cambiando los campos de texto por campos de tipo datetime, de esta forma podemos separar por intervalos de tiempo este
#DataFrame
data2["Recuperado"] = data2["Recuperado"].replace("Fallecido", "fallecido")
data2["Recuperado"].unique()
data2["fecha reporte web"] = pd.to_datetime(data2["fecha reporte web"], format="%d/%m/%Y %H:%M:%S")
data2["fecha reporte web"]
```

```
Out[10]: 0          2020-03-06
1          2020-03-09
2          2020-03-09
3          2020-03-11
4          2020-03-11
...
3777595    2021-06-14
3777596    2021-06-14
3777597    2021-06-14
3777598    2021-06-14
3777599    2021-06-14
Name: fecha reporte web, Length: 3777600, dtype: datetime64[ns]
```

```
In [11]: from datetime import datetime

year_starts = datetime(2021, 1, 1) #Este día empieza el año 2021

cond1 = data2["fecha reporte web"] < year_starts #Casos positivos de 2020
cond2 = data2["fecha reporte web"] >= year_starts #Casos positivos de 2021

datos2020 = data2[cond1] #DATASET Casos positivos 2020
datos2021 = data2[cond2] #DATASET Casos positivos 2021
```

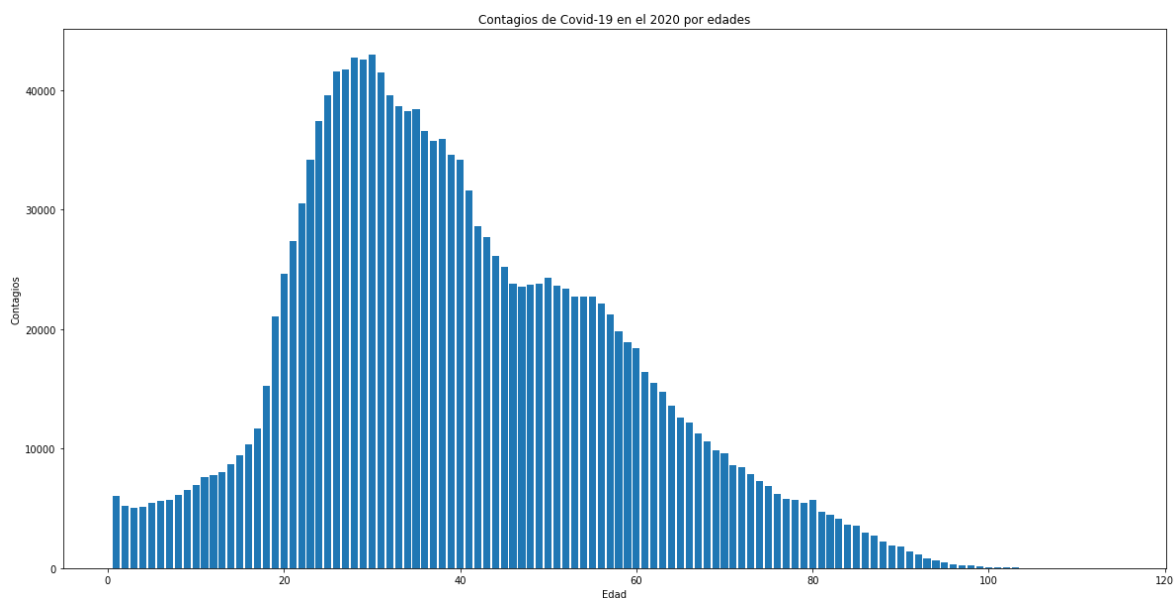
```
In [12]: #Reparando el dataset, dado que existe ambigüedad en las minúsculas y mayúsculas de los sexos
datos2021["Sexo"] = datos2021["Sexo"].replace(["f", "m"], ["F", "M"])
datos2021["Sexo"].unique()
```

```
Out[12]: array(['F', 'M'], dtype=object)
```

```
In [ ]:
```

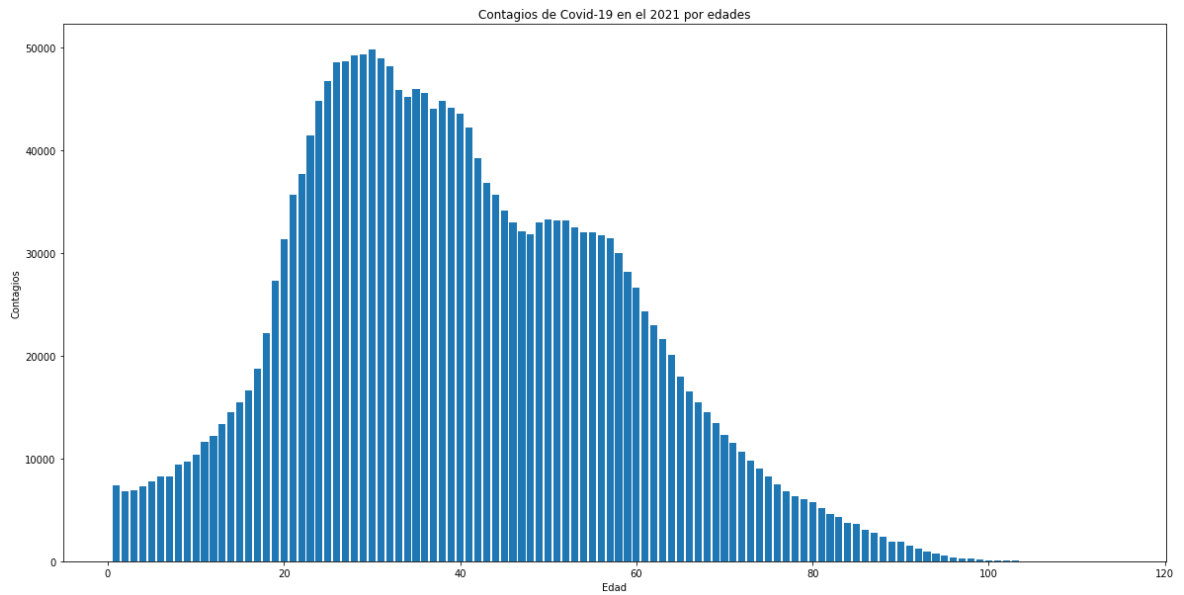
```
In [13]: #Sacando los contagios por edades para cada año
#Frecuencia para edades de 2020
freq_edad_2020=datos2020.groupby(["Edad"]).count()
edad = freq_edad_2020.index
contagios = freq_edad_2020["fecha reporte web"]

plt.figure(figsize=(20, 10))
plt.bar(edad, contagios)
plt.title("Contagios de Covid-19 en el 2020 por edades")
plt.xlabel("Edad")
plt.ylabel("Contagios")
plt.show()
```




```
In [14]: #Sacando los contagios por edades para cada año
#Frecuencia para edades de 2021
freq_edad_2021=datos2021.groupby(["Edad"]).count()
edad = freq_edad_2021.index
contagios = freq_edad_2021["fecha reporte web"]

plt.figure(figsize=(20, 10))
plt.bar(edad, contagios)
plt.title("Contagios de Covid-19 en el 2021 por edades")
plt.xlabel("Edad")
plt.ylabel("Contagios")
plt.show()
```

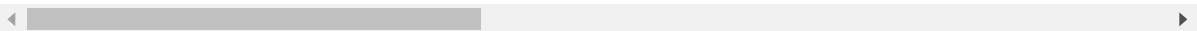


```
In [15]: cond1 = datos2020["Recuperado"]=="fallecido"
datos2020_fallecidos = datos2020[cond1]
datos2020_fallecidos
```

Out[15]:

	fecha reporte web	ID de caso	Fecha de notificación	Código DIVIPOLA departamento	Nombre departamento	Código DIVIPOLA municipio	Nombre municipio
151	2020-03-20	152	18/3/2020 0:00:00	11	BOGOTA	11001	BOGOTA
152	2020-03-20	153	18/3/2020 0:00:00	11	BOGOTA	11001	BOGOTA
156	2020-03-20	157	20/3/2020 0:00:00	47001	STA MARTA D.E.	47001	SANTA MARTA
187	2020-03-21	188	17/3/2020 0:00:00	76	VALLE	76892	YUMBO
196	2020-03-16	197	13/3/2020 0:00:00	13001	CARTAGENA	13001	CARTAGENA
...
1642691	2020-12-31	1642732	21/12/2020 0:00:00	76	VALLE	76520	PALMIRA
1642726	2020-12-31	1642767	26/12/2020 0:00:00	76	VALLE	76520	PALMIRA
1642757	2020-12-31	1642798	19/12/2020 0:00:00	76	VALLE	76520	PALMIRA
1642767	2020-12-31	1642808	22/12/2020 0:00:00	76	VALLE	76147	CARTAGO
1701113	2020-12-22	1701154	22/12/2020 0:00:00	20	CESAR	20011	AGUACHICA

46809 rows × 23 columns

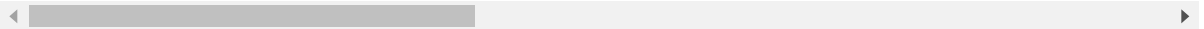


```
In [16]: cond1 = datos2021["Recuperado"]=="fallecido"
datos2021_fallecidos = datos2021[cond1]
datos2021_fallecidos
```

Out[16]:

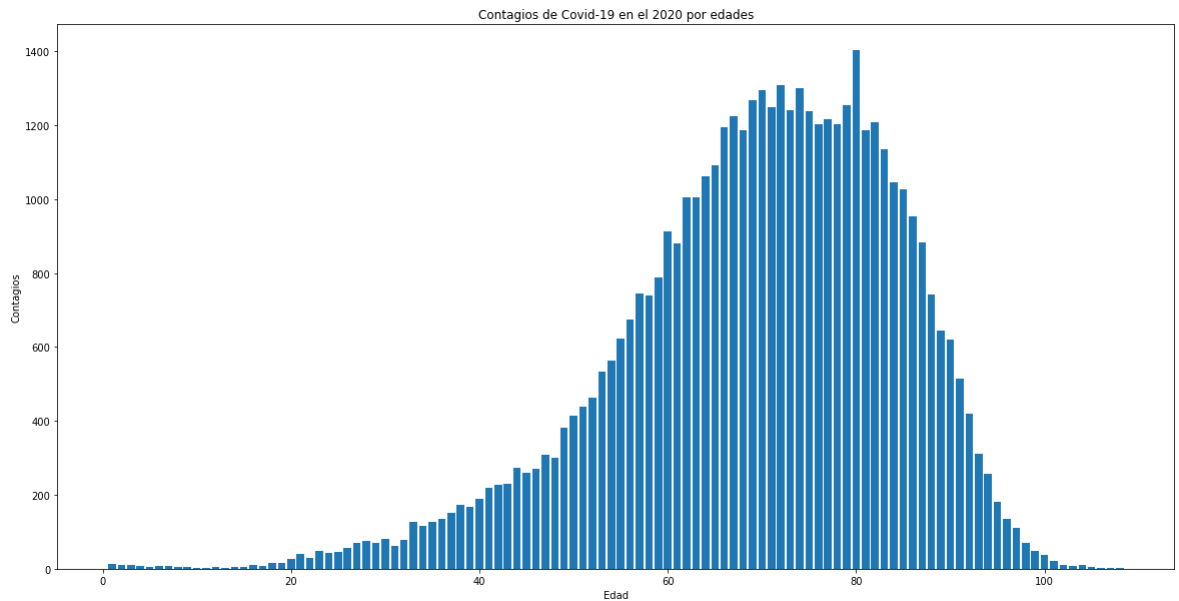
	fecha reporte web	ID de caso	Fecha de notificación	Código DIVIPOLA departamento	Nombre departamento	Código DIVIPOLA municipio	No muni
1642781	2021-01-01	1642822	29/12/2020 0:00:00	11	BOGOTA	11001	BOC
1642801	2021-01-01	1642842	29/12/2020 0:00:00	63	QUINDIO	63001	ARM
1642869	2021-01-01	1642910	29/12/2020 0:00:00	52	NARIÑO	52215	CORC
1642919	2021-01-01	1642960	30/12/2020 0:00:00	41	HUILA	41001	N
1642923	2021-01-01	1642964	29/12/2020 0:00:00	8001	BARRANQUILLA	8001	BARRANQL
...	
3752312	2021-06-13	3752353	30/5/2021 0:00:00	52	NARIÑO	52835	TUN
3752766	2021-06-13	3752807	31/5/2021 0:00:00	13001	CARTAGENA	13001	CARTAG
3752988	2021-06-13	3753029	11/5/2021 0:00:00	25	CUNDINAMARCA	25899	ZIPAQ
3753005	2021-06-13	3753046	23/5/2021 0:00:00	13001	CARTAGENA	13001	CARTAG
3753080	2021-06-13	3753121	17/5/2021 0:00:00	50	META	50006	ACA

49557 rows × 23 columns



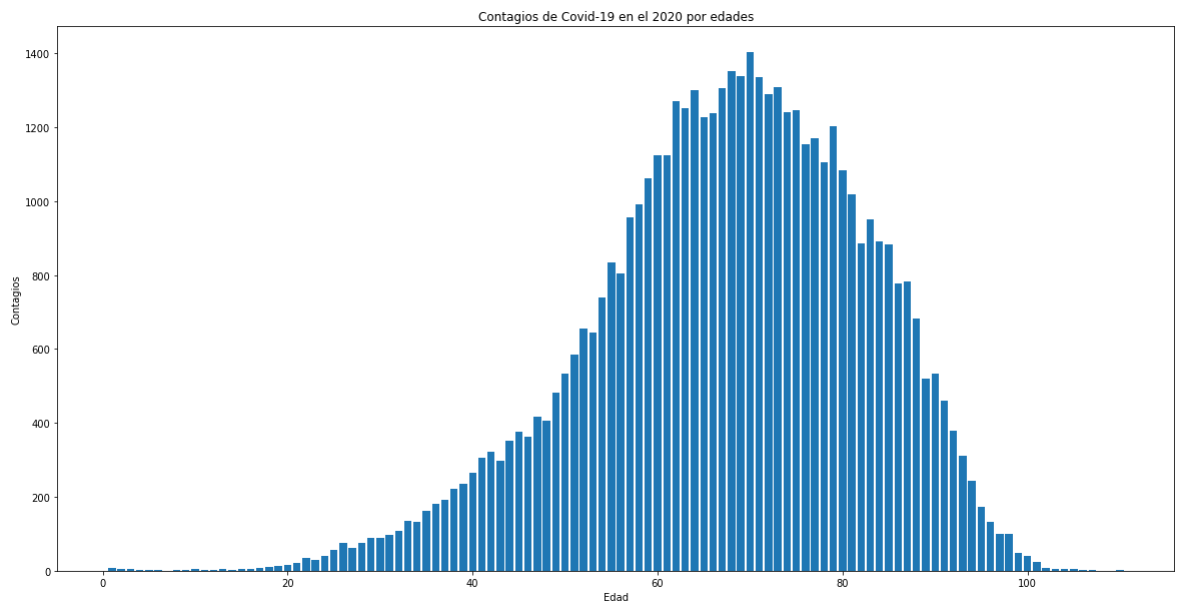
```
In [17]: #Sacando los muertos por edades para cada año
#Frecuencia para edades de 2020
freq_edad_2020_f=datos2020_fallecidos.groupby(["Edad"]).count()
edad = freq_edad_2020_f.index
contagios = freq_edad_2020_f["fecha reporte web"]

plt.figure(figsize=(20, 10))
plt.bar(edad, contagios)
plt.title("Contagios de Covid-19 en el 2020 por edades")
plt.xlabel("Edad")
plt.ylabel("Contagios")
plt.show()
```



```
In [18]: freq_edad_2021_f=datos2021_fallecidos.groupby(["Edad"]).count()
edad = freq_edad_2021_f.index
contagios = freq_edad_2021_f["fecha reporte web"]

plt.figure(figsize=(20, 10))
plt.bar(edad, contagios)
plt.title("Contagios de Covid-19 en el 2020 por edades")
plt.xlabel("Edad")
plt.ylabel("Contagios")
plt.show()
```



```
In [19]: datos2020_fallecidos.to_csv("./out/fallecidos2020.csv")
```

```
In [20]: datos2021_fallecidos.to_csv("./out/fallecidos2021.csv")
```

```
In [ ]:
```