

Visualización de los datos de las personas fallecidas a causa del Covid-19 en Nuevo León al 20/09/2020

```
In [1]: ▶ import pandas as pd

datos = pd.read_csv("MUERTES COVID NL.csv")
```

```
In [2]: ▶ #Con esto le asignamos un nombre a cada columna
sexo = datos['SEXO']
mingr = datos['FECHA_INGRESO']
muerte = datos['FECHA_DEF']
edad = datos['EDAD']
diasint = datos['T_INT']
embarazo = datos['EMBARAZO']
datos.head()
```

Out[2]:

	SEXO	FECHA_INGRESO	FECHA_DEF	T_INT	EDAD	EMBARAZO
0	M	6	6	6	42	No
1	M	6	6	8	55	No
2	M	4	5	16	74	No
3	M	6	6	2	38	No
4	M	7	7	7	61	No

La función **describe()** sirve para darnos información acerca las columnas, nos da el número de datos, la media, la desviación estándar, los números mínimo y máximo y los 3 cuartiles.

```
In [3]: ▶ edad.describe()
```

```
Out[3]: count      2823.000000
mean         63.253631
std          14.115831
min           0.000000
25%          54.000000
50%          64.000000
75%          73.000000
max         101.000000
Name: EDAD, dtype: float64
```

```
In [4]: ▶ diasint.describe()
```

```
Out[4]: count    2823.000000
mean         8.577754
std          7.800323
min         -1.000000
25%          3.000000
50%          7.000000
75%         12.000000
max          71.000000
Name: T_INT, dtype: float64
```

Con estos ciclos hacemos el conteo de mujeres y hombres fallecidos

```
In [5]: ▶ #número de mujeres fallecidas
mujer = 0
for i in sexo:
    if i == "F":
        mujer = mujer + 1
print("El número de mujeres fallecidas es de: ",mujer)
```

El número de mujeres fallecidas es de: 1094

```
In [6]: ▶ #número de hombres fallecidos
hombre = len(sexo) - mujer
print("El número de hombres fallecidas es de: ",hombre)
```

El número de hombres fallecidas es de: 1729

```
In [7]: ▶ embarazadas = 0
for i in embarazo:
    if i == "Si":
        embarazadas = embarazadas + 1
print("El número de mujeres embarazadas que fallecieron fue de: ",embarazadas)
```

El número de mujeres embarazadas que fallecieron fue de: 1

En el histograma de las edades podemos ver que la mayoría de las personas que fallecieron en Nuevo León por el COVID-19 tenían entre aproximadamente **60 y 75 años**.

```
In [8]: #Histograma de edad
import matplotlib as plt
edad.plot.hist()
```

Out[8]: <matplotlib.axes._subplots.AxesSubplot at 0x28cd4648ac8>

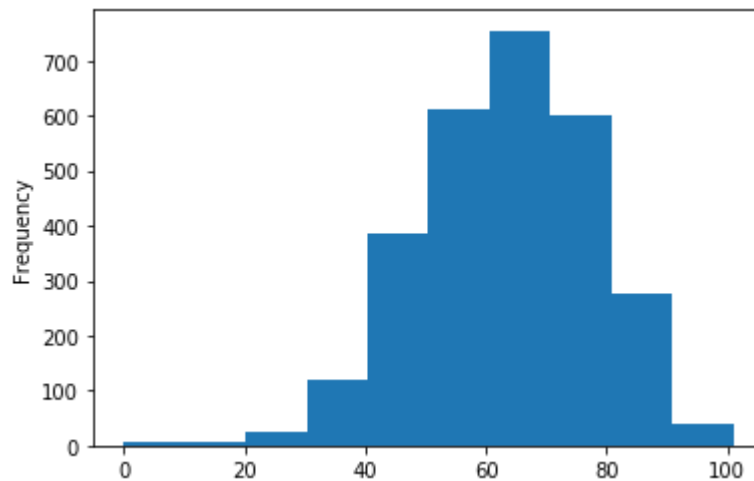


Gráfico de líneas

Debido a que son muchos datos, no es factible una gráfica de barras, así que usamos una de líneas y el resultado que vemos es igual al del histograma, el mayor número de fallecidos está entre **60 y 75** años.

```
In [9]: edad.value_counts().sort_index().plot.area()
```

Out[9]: <matplotlib.axes._subplots.AxesSubplot at 0x28cd4eecd8>

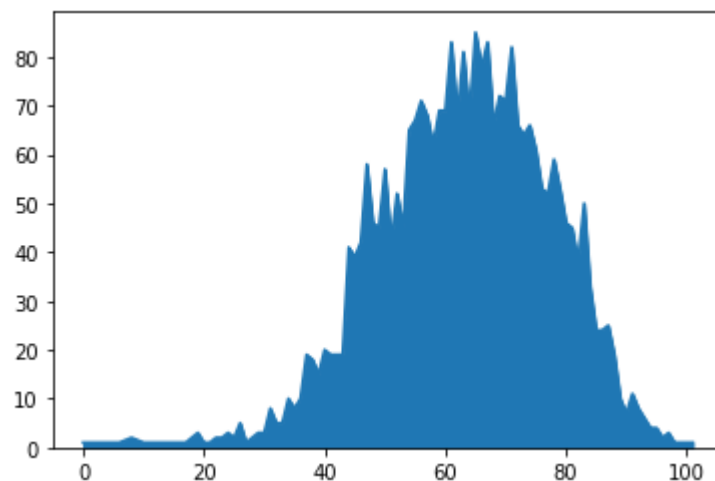


Gráfico de barras

Con este gráfico podemos comparar la cantidad de mujeres y hombres que fallecieron a causa de COVID en NL, podemos ver que el número de hombres que han fallecido es mayor.

```
In [10]: ▶ #Grafico de barras de sexo
import seaborn as sns
sns.countplot(x=sexo, palette = "Set2")
```

```
Out[10]: <matplotlib.axes._subplots.AxesSubplot at 0x28cd716c8c8>
```

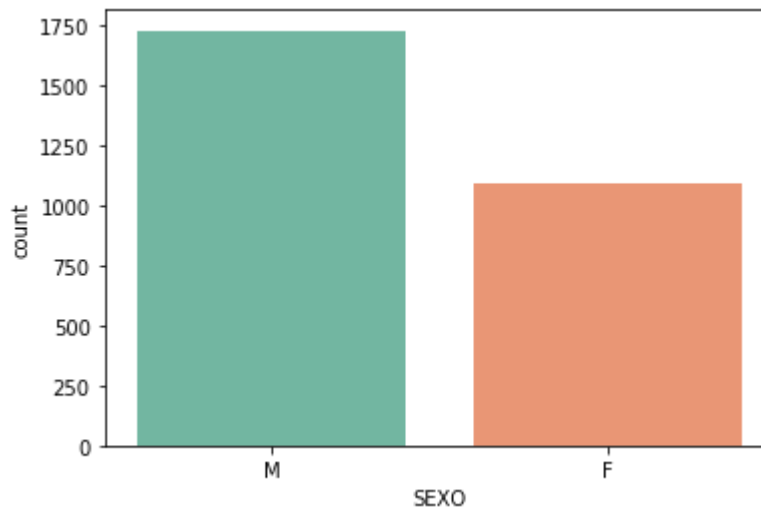
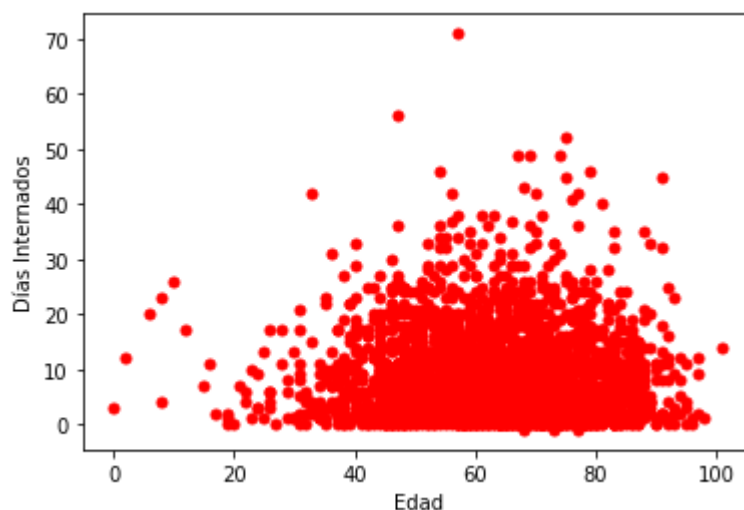


Gráfico de dispersión

Se realiza para identificar de manera gráfica si existe una relación entre las variables, en este caso la edad de la persona y los días que estuvieron internados. Como podemos observar en la gráfica, no existe una relación lineal.

```
In [11]: ▶ import matplotlib.pyplot as plt
plt.scatter(x=edad , y=diasint, marker='o', c='red', s=25)
plt.xlabel("Edad")
plt.ylabel("Días Internados")
plt.show()
```

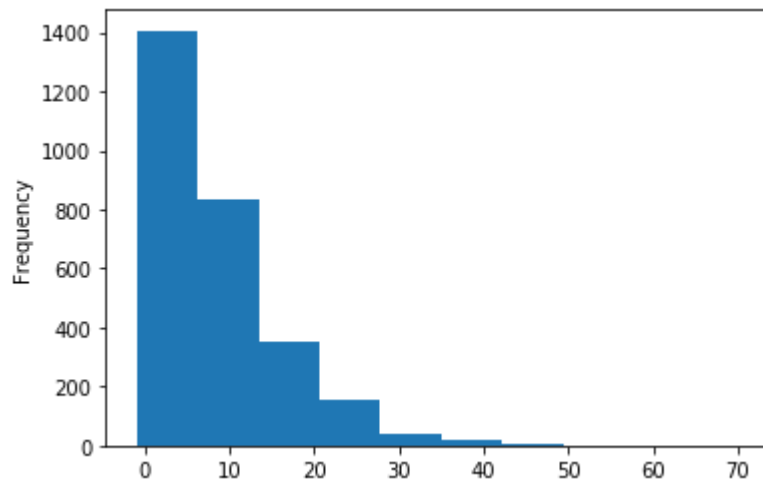


Histograma

Los histogramas nos ayudan a observar la frecuencia de los valores representados, en el siguiente histograma podemos ver la frecuencia de los días que las personas que fallecieron por COVID 19 en Nuevo León estuvieron internadas, como lo muestra el gráfico, la mayor parte de las personas estuvieron internadas entre 0 y 12 días aproximadamente.

```
In [12]: #Histograma de Los días internados  
diasint.plot.hist()
```

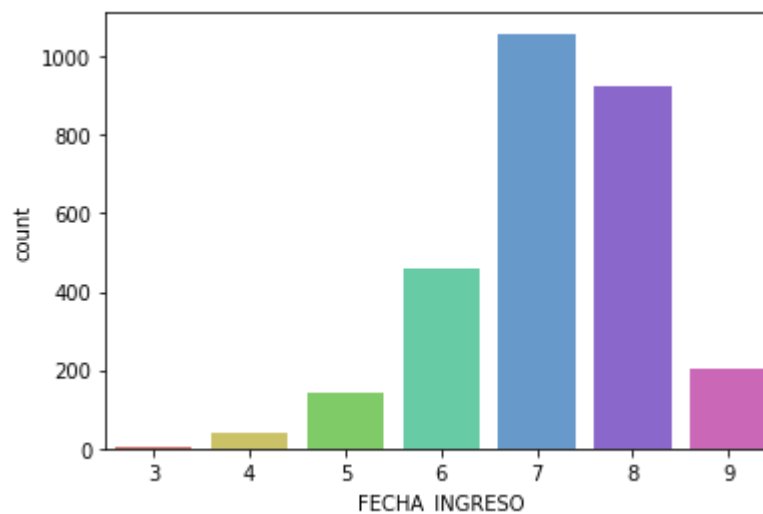
```
Out[12]: <matplotlib.axes._subplots.AxesSubplot at 0x28cd7260908>
```



Al comparar los meses en los que ingresaron al hospital con una gráfica de barras, nos damos cuenta que Julio y Agosto son los meses con más ingresos al hospital.

```
In [13]: #gráfica de barras del mes de ingreso  
sns.countplot(x=mingr, palette = "hls")
```

```
Out[13]: <matplotlib.axes._subplots.AxesSubplot at 0x28cd72cd648>
```

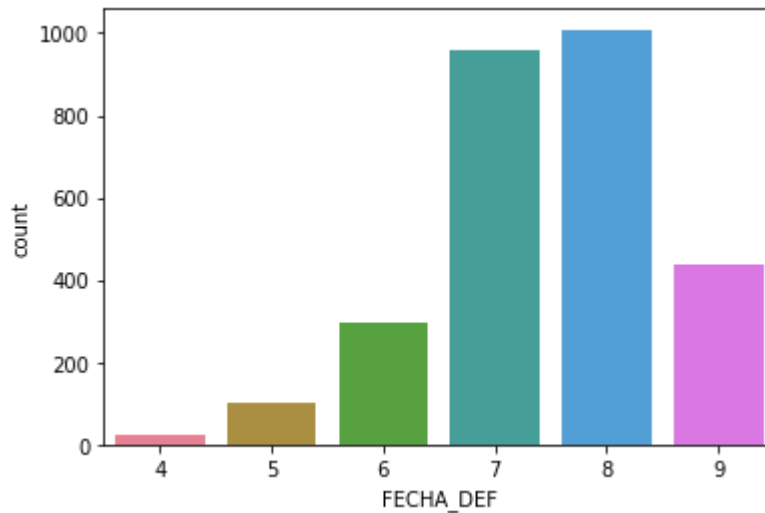


Después comparamos las fechas de muerte y tenemos que Julio y Agosto fueron los meses con más muertes registradas. También podemos observar que en el mes de Septiembre se han

reducido la cantidad de muertes, aunque queda esperar a que cierre el mes para ver si efectivamente bajaran las defunciones.

```
In [14]: #gráfico de barras del mes de muerte  
sns.countplot(x=muerte, palette = "husl")
```

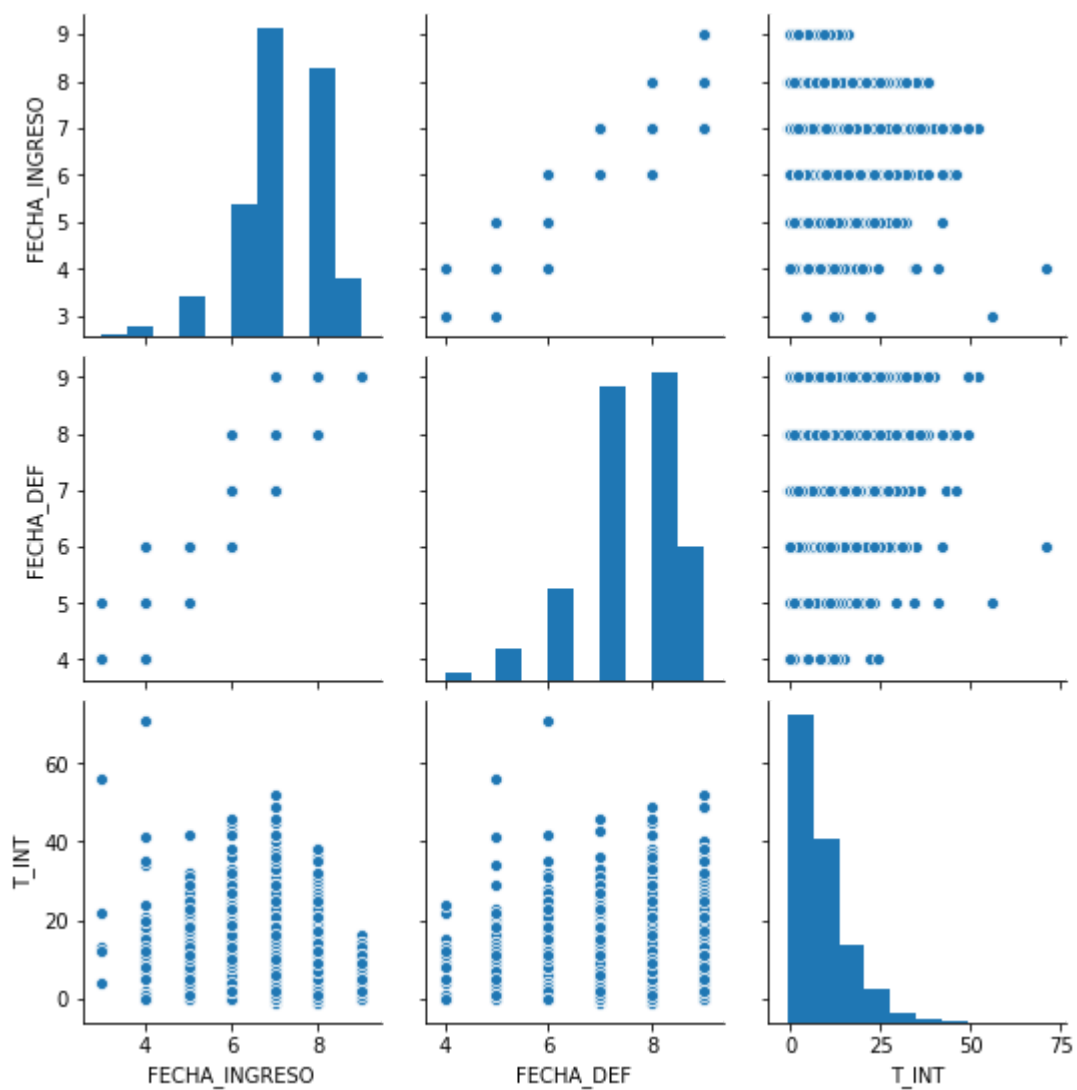
```
Out[14]: <matplotlib.axes._subplots.AxesSubplot at 0x28cd7346288>
```



La siguiente gráfica produce una matriz de relaciones entre cada variable en sus datos para un examen instantáneo de nuestros datos. También puede ser un gran punto de partida para determinar los tipos de análisis de regresión a utilizar.

```
In [15]: import seaborn as sns
sns.pairplot(datos[['FECHA_INGRESO', 'FECHA_DEF', 'T_INT']])
```

Out[15]: <seaborn.axisgrid.PairGrid at 0x28cd73b5208>



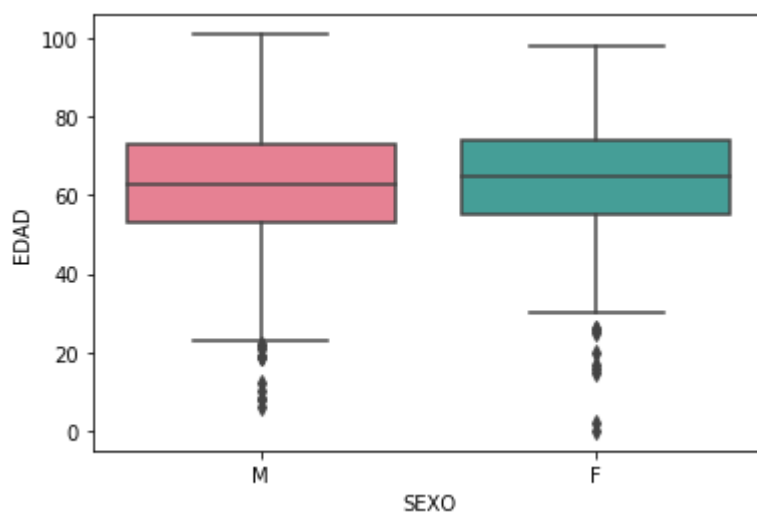
Boxplot

Muestra la distribución de datos basada en un resumen de cinco números ("mínimo", primer cuartil (Q1), mediana, tercer cuartil (Q3) y "máximo").

En la siguiente gráfica podemos ver que los dos boxplots presentan datos que se salen del valor mínimo, estos datos podrían ser outliers. Al comparar la edad de los pacientes en base al sexo, podemos ver que no hay mucha diferencia entre ambos.

```
In [16]: sns.boxplot(x=datos["SEXO"],y=datos["EDAD"],palette="hus1")
```

```
Out[16]: <matplotlib.axes._subplots.AxesSubplot at 0x28cd78e1488>
```



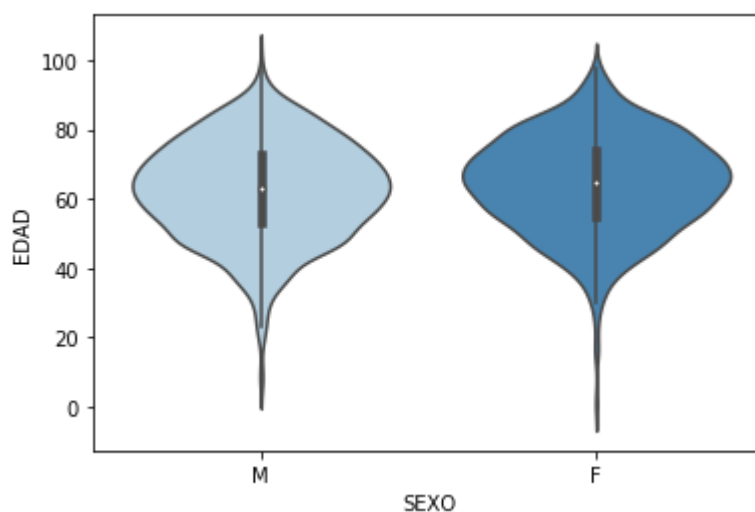
Gráfica de violín

Sirve para visualizar la distribución de datos numéricos de diferentes variables. Es similar al diagrama de caja pero con un diagrama rotado en cada lado, que brinda más información sobre la estimación de densidad en el eje y.

Podemos ver que al igual que en el boxplot no existe mucha diferencia entre hombres y mujeres.


```
In [17]: sns.violinplot(x=datos["SEXO"],y=datos["EDAD"],palette="Blues")
```

```
Out[17]: <matplotlib.axes._subplots.AxesSubplot at 0x28cd7a80888>
```



Histograma de probabilidad

Representa la variación en la distribución de datos.

```
In [18]: sns.distplot(edad)
```

```
Out[18]: <matplotlib.axes._subplots.AxesSubplot at 0x1e500ba3948>
```

