



## Tipos de gráficos

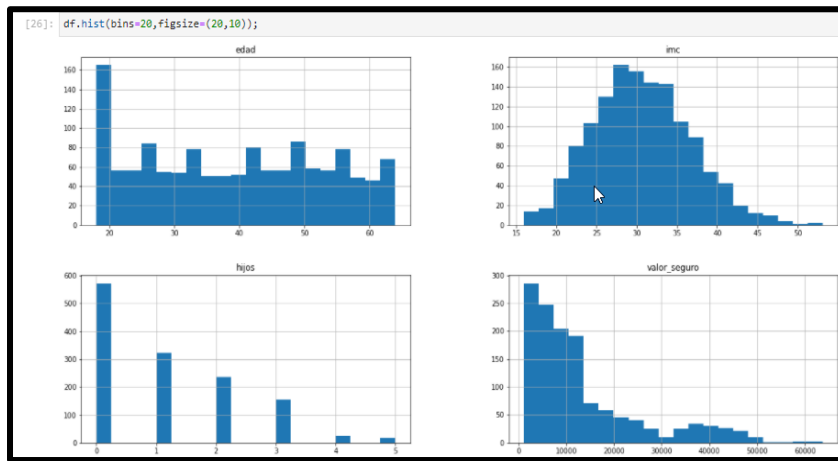
Los tipos de gráficos que pueden utilizarse dentro del análisis exploratorio de datos son:

### 1. Histogramas de frecuencia.

Calculan la cantidad de datos presentes en un elemento o en un rango.

#### Figura 1.

Ejemplos de histogramas de frecuencia



Nota: Captura de pantalla

Para aplicarlo en las variables numéricas del ejemplo trabajado, durante el desarrollo de este espacio de formación, ejecute el comando:

```
df.hist(bins=20,figsize=(20,10))
```

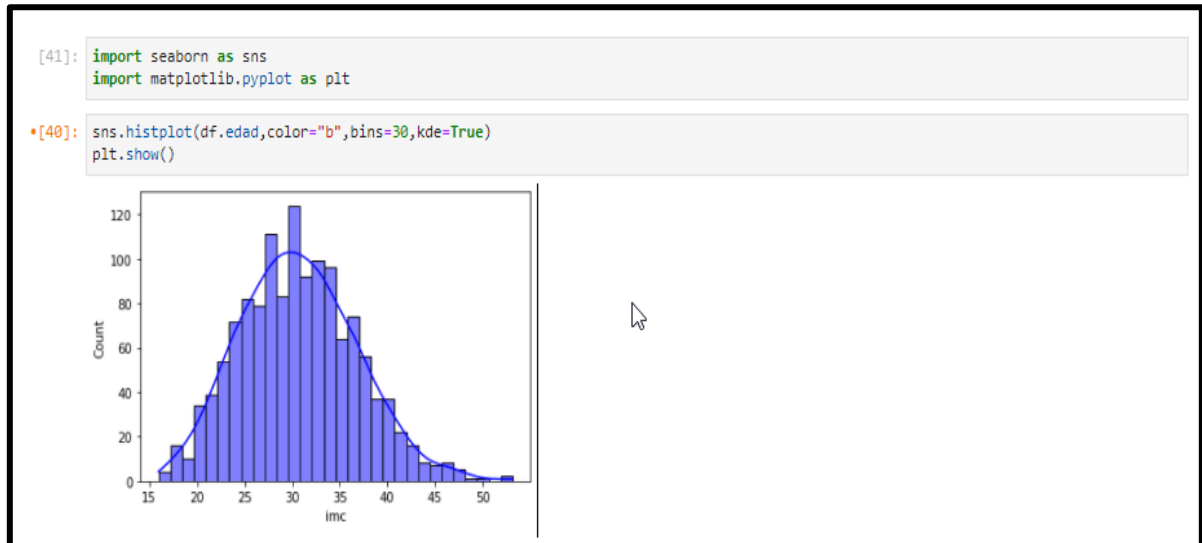
#### Recuerde que:

Para poder realizar mejores análisis gráficos es importante hacer uso de librerías, en el caso del ejemplo trabajado se deben agregar las siguientes: `import seaborn as sns` , `import matplotlib.pyplot as plt` y ejecutar el código: `sns.histplot(df.edad,color="b",bins=30,kde=True)plt.show()` , el cual permite visualizar cómo están distribuidos los datos de la variable o la columna edad, como se ve en la figura:



**Figura 2.**

Uso de librerías



*Nota:* Captura de pantalla

## 2. Gráfico de Barras

Son una muy buena herramienta que permite determinar cómo se distribuyen los datos, para aplicarlo al ejemplo que se ha venido trabajando, se debe usar el siguiente comando:

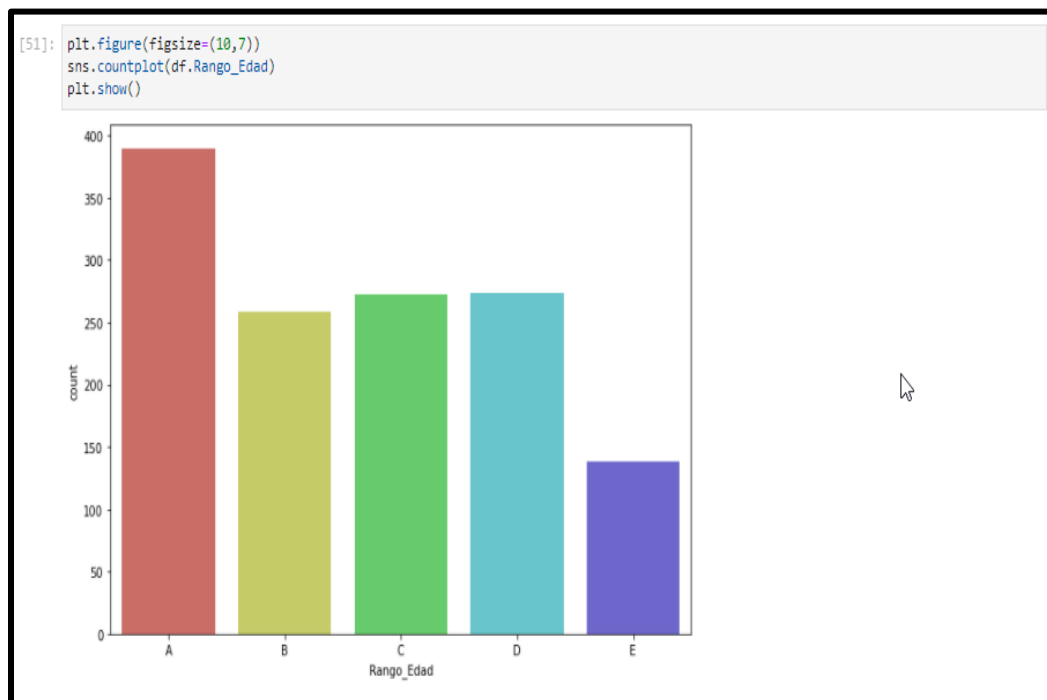
```
plt. Figure(figsize=(10,7))  
sns.countplot(df.Rango_Edad)plt.show()
```

Cómo se puede apreciar en la figura la concentración de los datos están inclinados hacia el rango A, edades entre 18 y 28 años.



**Figura 3.**

Ejemplo gráfico de barras



*Nota: Captura de pantalla*

**Los resultados arrojados en el gráfico demuestran:**

- En el grupo A edades entre 18 y 28 años se presenta una alta concentración de los datos
- La concentración de los datos para los grupos B, C y D es similar.
- El rango con menor número de elementos corresponde al DE edades entre 58 y 68.

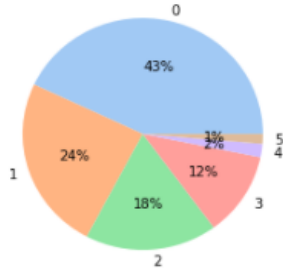
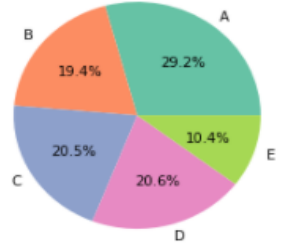
### **3. Gráficos de torta**

Al generar las gráficas propuestas anteriormente, se pudo detectar algunas concentraciones mayores en algunas de las variables, los gráficos de torta ayudan a determinar dichas tendencias, hasta el momento se han obtenidos los análisis descriptivos del comportamiento de las variables, pero es importante cuantificar estos resultados, ejecute los comandos que se muestran en la imagen para calcular y graficar cómo están distribuidos los datos porcentualmente:



**Tabla 1.**

Ejemplos gráficos de torta

Código	Gráfico	Resultado
<p><b>Cálculo de distribución porcentual variable: hijos</b></p> <pre>total_hijos = df['hijos'].groupby(df['hijos']).count() etiquetas = [0,1,2,3,4,5] colors = sns.color_palette('pastel')[0:6] plt.pie(total_hijos, labels = etiquetas, colors = colors, autopct='%0f%%') plt.show()</pre>	<p><b>Figura 4.</b> Distribución porcentual de hijos</p>  <p><i>Nota: Captura de pantalla</i></p>	<ul style="list-style-type: none"> <li>El 43 % de los usuarios no tienen hijos.</li> <li>Mientras que el 54 % al menos tiene un hijo.</li> <li>Solo un tres % tiene más de 3 hijos</li> </ul>
<p><b>Cálculo de distribución porcentual variable: Rango de edad</b></p> <pre>total_rango_edad = df['Rango_Edad'].groupby(df['Rango_Edad']).count() labels = ["A", "B", "C", "D", "E"] colors = sns.color_palette('Set2')[0:5] plt.pie(total_rango_edad, labels = labels, colors = colors, autopct='%1f%%') plt.show()</pre>	<p><b>Figura 5.</b> Distribución porcentual rango de edad</p>  <p><i>Nota: Captura de pantalla</i></p>	<ul style="list-style-type: none"> <li>El rango representado con A, entre 18 y 28 años equivale al 29% de los datos.</li> <li>Para los rangos B, C y D, la equivalencia es del 20% sumando el 60% entre los 3.</li> <li>El rango E equivale al 10%.</li> </ul>



#### 4. Gráficos de Caja y Bigotes

Este tipo de gráfico que permite determinar la distribución de los datos con relación a dos variables indicando los valores medios y cuartiles.

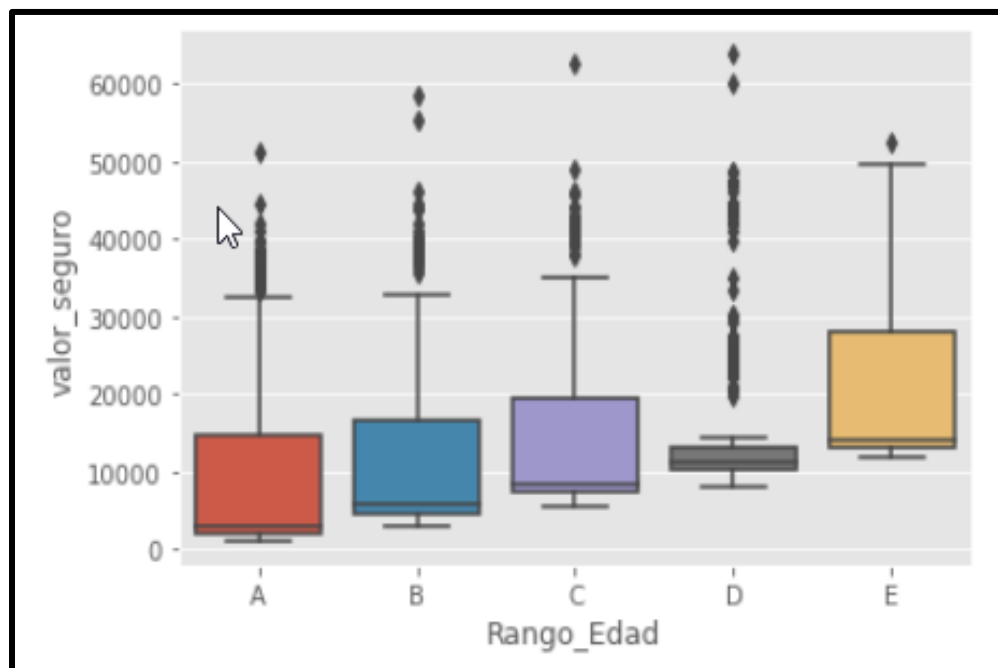
Siguiendo el ejemplo y para establecer la relación entre el valor del seguro y el rango de edades es necesario ejecutar el siguiente comando:

```
Redad_valor =  
sns.boxplot(x=df["Rango_Edad"],  
y=df["valor_seguro"])
```

Al ejecutar el comando se desprende la siguiente gráfica:

**Figura 6.**

Ejemplo para el rango de edad



Nota: Captura de pantalla



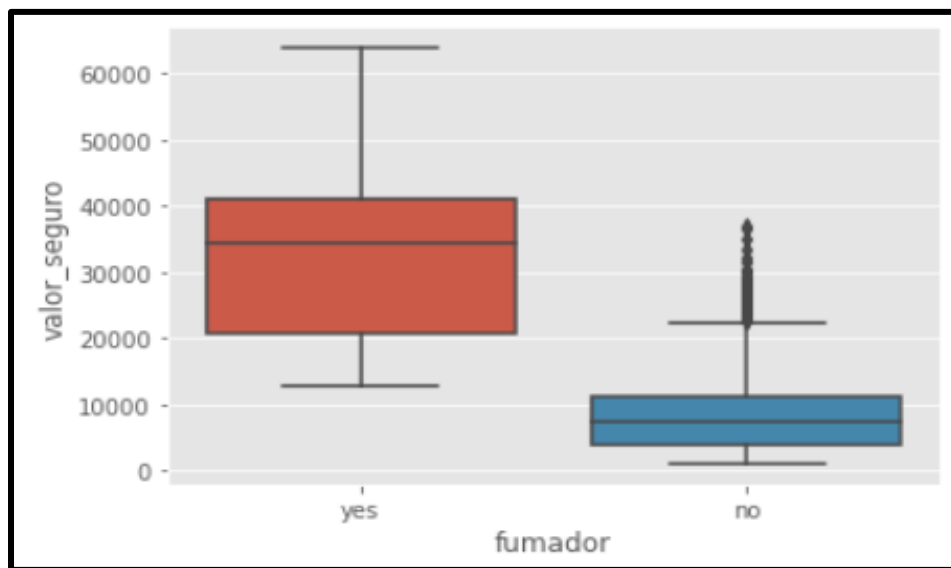
Para establecer la relación entre el valor del seguro y el estado fumador ejecute el siguiente comando:

```
fumador_valor =  
sns.boxplot(x=df["fumador"],  
y=df["valor_seguro"])
```

Al ejecutar el comando se desprende la siguiente gráfica:

**Figura 7.**

Ejemplo para fumadores



*Nota:* Captura de pantalla