



# Servicio nacional de Aprendizaje -SENA

## **CURSO:**

ANÁLISIS EXPLORATORIO DE DATOS EN PYTHON

## **FICHA:**

2769316

## **APRENDIZ:**

JAMITH ALEXANDER GARCIA ARRIETA

# Evidencia AA3-EV03.

---

Presentación en la que dé a conocer los tipos de gráficos y su función en el análisis de datos. AA3-EV03.





# Introducción

---

- En esta presentación, exploraremos la importancia de diferentes gráficos en el análisis de datos. Los histogramas de frecuencia nos permiten comprender la distribución de una variable y detectar valores atípicos. Los gráficos de torta nos muestran la proporción de diferentes categorías en nuestros datos. Y las tablas nos brindan una visualización estructurada de la información. A través de estos gráficos y tablas, obtendremos insights valiosos sobre los factores clave que influyen en el precio de las viviendas y locales para la venta en A&A Ltda. ¡Comencemos a explorar y descubrir patrones en nuestros datos!

# Histogramas de frecuencia

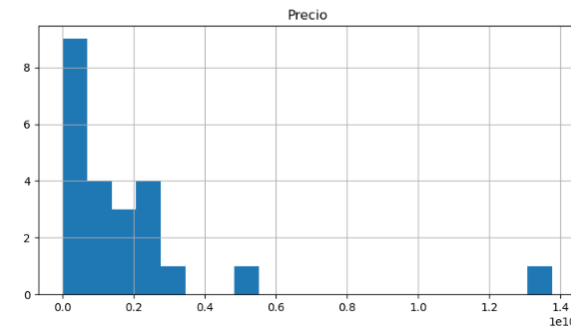
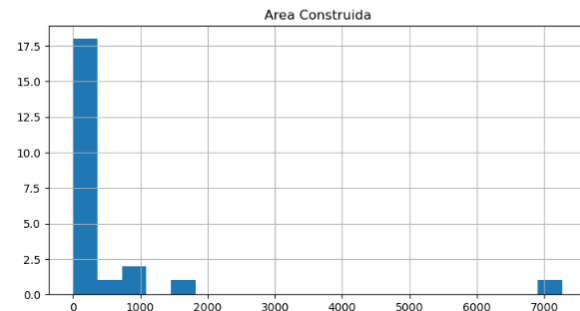
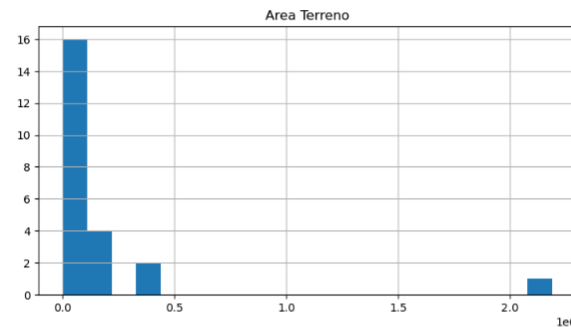
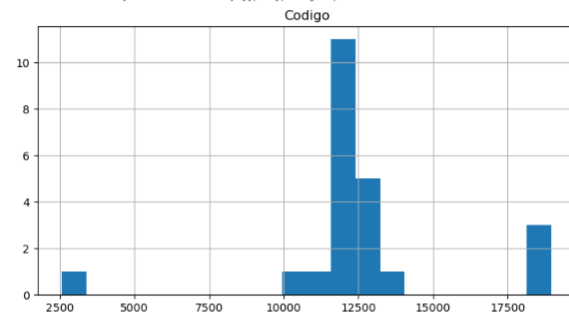
- El histograma de frecuencia es una herramienta importante en el análisis de datos debido a su capacidad para mostrar la distribución de una variable de manera visual y comprensible. Nos ayuda a identificar la forma de la distribución (si es simétrica, sesgada a la derecha o sesgada a la izquierda) y nos proporciona información sobre la concentración de valores en diferentes rangos. Esto nos permite detectar valores atípicos, identificar patrones o tendencias y realizar comparaciones entre diferentes grupos o subconjuntos de datos. En resumen, el histograma de frecuencia nos ayuda a comprender mejor los datos y a tomar decisiones informadas basadas en su distribución.

- Histogramas de frecuencia

```
23]: # Calcular la cantidad de datos presentes en un elemento o en un rango.
```

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

```
23]: array([[<Axes: title='center': 'Codigo'>],  
          [<Axes: title='center': 'Area Terreno'>]],  
        [[<Axes: title='center': 'Area Construida'>],  
          [<Axes: title='center': 'Precio'>]]], dtype=object)
```



#Importar librerías para generación de gráficos a partir del analisis

```
import pandas as pd  
import seaborn as sns  
import matplotlib.pyplot as plt
```

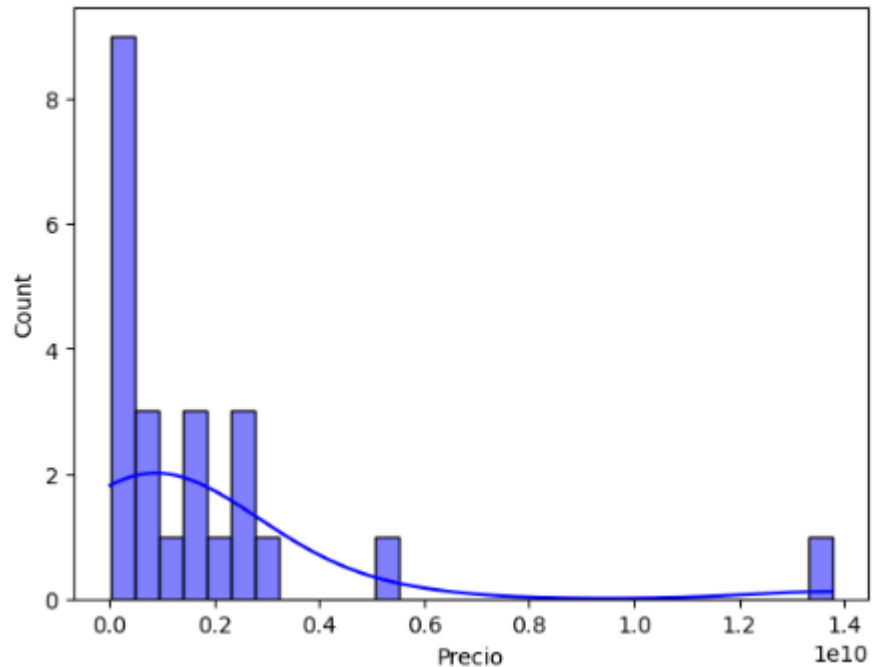
# Calcular la cantidad de datos presentes en un elemento o en un rango.

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

## Gráficos de Barras

- El gráfico de barras es una herramienta importante en el análisis de datos debido a su capacidad para comparar categorías o variables de manera visual y clara. Permite representar la frecuencia o el valor de cada categoría mediante barras rectangulares de longitud proporcional a los datos que representan. Este tipo de gráfico nos ayuda a identificar rápidamente las diferencias o similitudes entre las categorías, así como las tendencias o patrones que puedan existir. Además, nos permite realizar comparaciones entre distintos grupos o subconjuntos de datos de forma sencilla. En resumen, el gráfico de barras es una herramienta esencial para la presentación y visualización de datos, facilitando la interpretación y comprensión de la información de manera efectiva.

<Axes: xlabel='Precio', ylabel='Count'>



#Importar librerías para generación de gráficos a partir del análisis

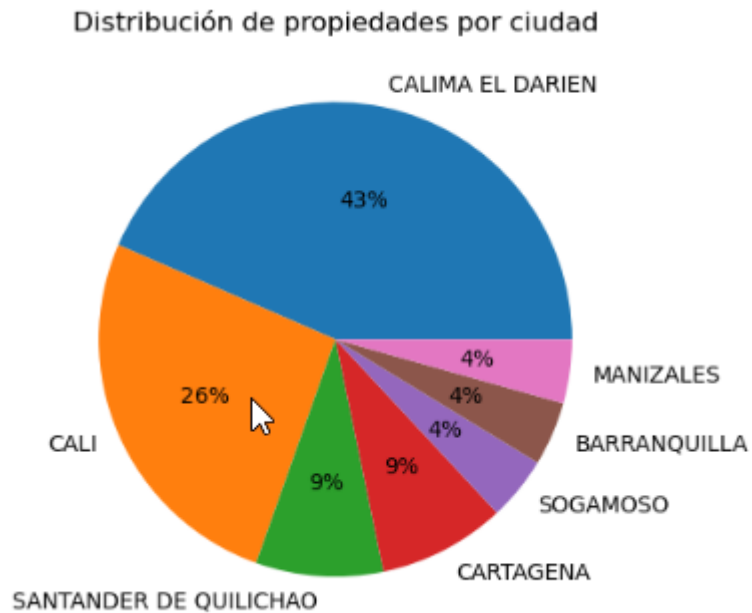
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

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

```
sns.histplot(df.Precio,color="b",bins=30,kde=True)
```

## Gráficos de torta

- El gráfico de torta, también conocido como gráfico de pastel, es una herramienta importante en el análisis de datos debido a su capacidad para representar la distribución porcentual de categorías o variables de manera visual y fácil de entender. Este tipo de gráfico se basa en la división de un círculo en segmentos proporcionales a los valores que representan, donde cada segmento representa una categoría y su tamaño está determinado por el porcentaje que representa dentro del conjunto total.



#Importar librerías para generación de gráficos a partir del analisis

```
import pandas as pd
```

```
import seaborn as sns
```

```
import matplotlib.pyplot as plt
```

#Distribución porcentual de Ciudad

```
propiedades_por_ciudad = df['Ciudad'].value_counts()
```

# Obtener las etiquetas y los valores para el gráfico de torta

```
etiquetas = propiedades_por_ciudad.index
```

```
valores = propiedades_por_ciudad.values
```

# Generar el gráfico de torta

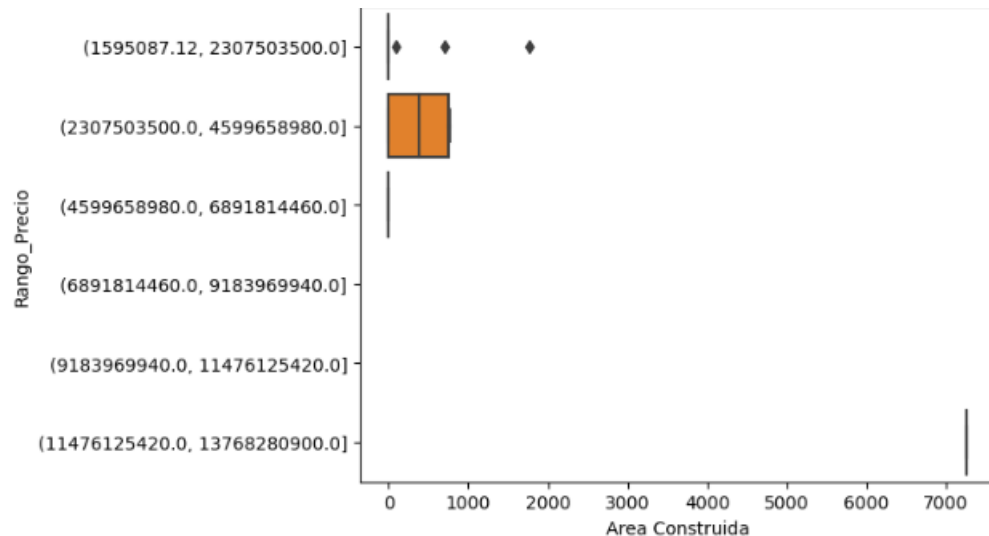
```
plt.pie(valores, labels=etiquetas, autopct='%0.0f%%')
```

```
plt.title('Distribución de propiedades por ciudad')
```

```
plt.show()
```

## Gráficos de Caja y Bigotes

- El gráfico de caja y bigotes (boxplot) se utiliza comúnmente para representar la distribución de variables numéricas, no de categorías como en el caso de los gráficos de torta. El gráfico de caja y bigotes es una herramienta valiosa en el análisis exploratorio de datos, ya que proporciona información sobre la mediana, los cuartiles, los valores atípicos y la dispersión de una variable.



#Importar librerías para generación de gráficos a partir del analisis

```
import pandas as pd
```

```
import seaborn as sns
```

```
import matplotlib.pyplot as plt
```

#Gráficos de Caja y Bigotes

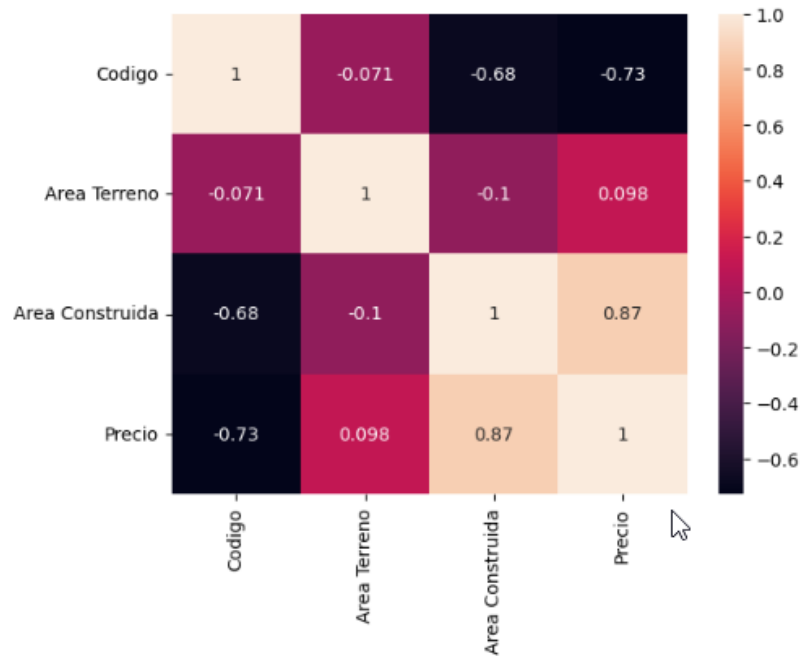
# establecer la relación entre el Area Construida" y el Rango\_Precio

```
fumador_valor = sns.boxplot(x=df["Area Construida"],  
y=df["Rango_Precio"])
```

## Gráficos de Correlación

- El gráfico de correlación es importante porque proporciona información valiosa sobre cómo se relacionan dos variables entre sí. Al visualizar la correlación, podemos determinar si existe una relación lineal positiva (a medida que una variable aumenta, la otra también lo hace), una relación lineal negativa (a medida que una variable aumenta, la otra disminuye) o una falta de relación lineal entre las variables. Esta información es esencial en el análisis de datos, ya que nos permite comprender la dependencia o independencia de las variables y su posible influencia mutua.

[46]: <Axes: >



#Importar librerías para generación de gráficos a partir del análisis

```
import pandas as pd
```

```
import seaborn as sns
```

```
import matplotlib.pyplot as plt
```

#Establecer la correlación entre las variables que permiten identificar una cercanía.

```
correlacion = df.corr(numeric_only=True)
```

```
correlacion
```

```
sns.heatmap(correlacion, xticklabels=correlacion.columns, yticklabels=correlacion.columns, annot=True)
```



# Conclusión

- En conclusión, los gráficos de frecuencia, barras, torta, caja y bigotes, y correlación son herramientas fundamentales en el análisis exploratorio de datos. Estos gráficos nos permiten visualizar y comprender la distribución, las relaciones y los patrones presentes en nuestros datos.

An abstract composition of various geometric shapes. In the top left, a green-outlined triangle points towards the top right. To its right is a solid blue circle. Below the triangle is a blue-outlined ring. In the center is a large orange semi-circle. To the right of the semi-circle are two vertical yellow dashes. In the bottom left is a large solid orange circle. Above it are three yellow dashes of varying lengths and orientations. In the bottom right is a green-outlined square.