**Taller 1 – Python y datos:** Analítica Computacional de Datos

**Nombre:** Isabela Castillo Mercado

**Código:** 201813093

1. Se realizo la Instalación de Python 3.11

Texto

Descripción generada automáticamente

**(\*)** Aclaración: Mi usuario en este pc es fmerc.

**2.** Dentro de Visual Studio Code realizamos todo lo mencionado em el enunciado del taller y la instalación de los paquetes correspondientes.

Texto

Descripción generada automáticamente

Como se puede observar en la salida del terminal, tuve que especificar la versión python la que instalaba los paquetes, debido a que ya hacia uso de pyehon en Visual Studio Code y tenia la versión 3.92

Texto

Descripción generada automáticamente

Texto

Descripción generada automáticamente

Texto

Descripción generada automáticamente

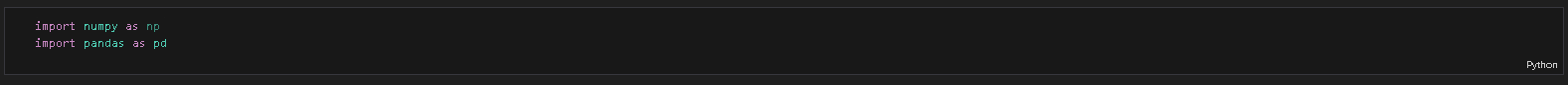
Texto

Descripción generada automáticamente

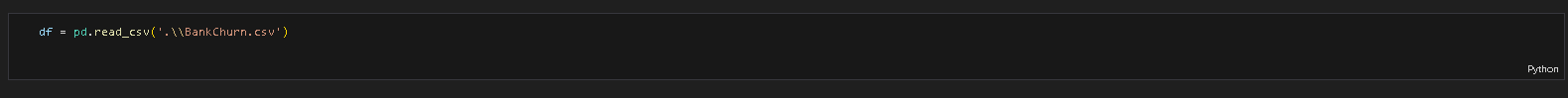
Texto

Descripción generada automáticamente

1. **Exploración de Datos en Python**



Se importan las librerías numpy como np y pandas como pd para su uso en el resto del cuaderno.



Se lee e importa el archivo CSV como un DataFrame utilizando la función read\_csv de la librería pandas (pd).

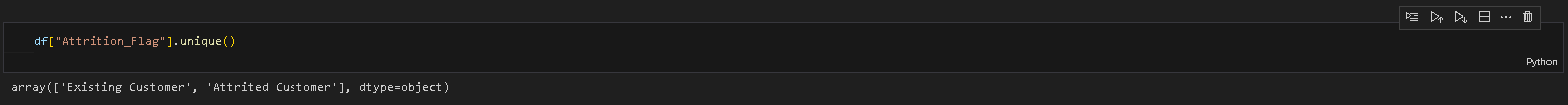


Muestra la dimensión de la data frame

Imagen que contiene camino, computadora, grande, hombre

Descripción generada automáticamente

La función head muestra las primeras 5 filas con todas las columnas del DataFrame, comenzando con el índice cero, ya que Python utiliza indexación basada en cero.

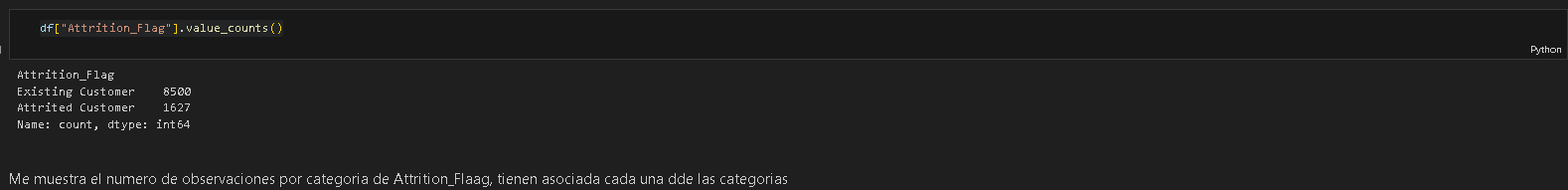


La función unique en Python muestra los valores únicos de una columna, eliminando duplicados. Dado que la columna Attrition\_Flag son 2 categorías, te devolverá solo los valores únicos.

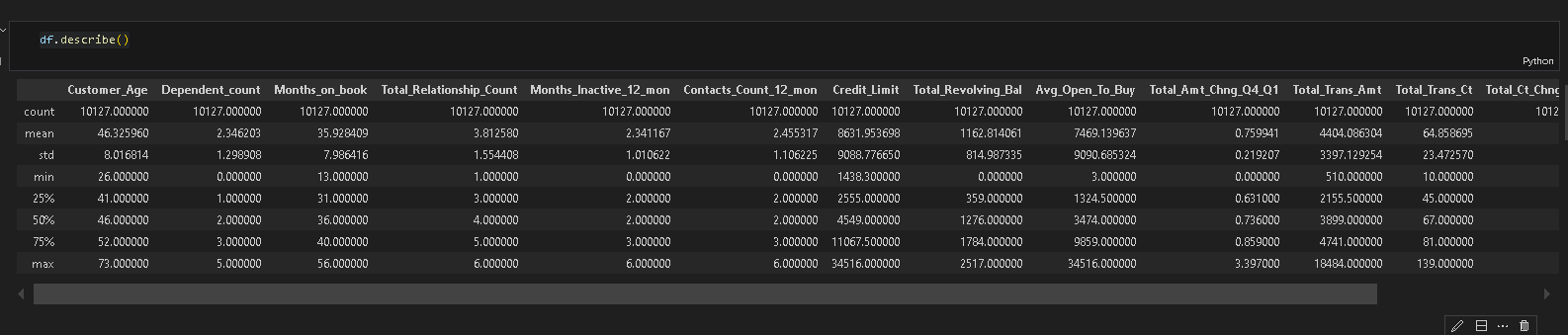
Imagen de la pantalla de un computador

Descripción generada automáticamente con confianza baja

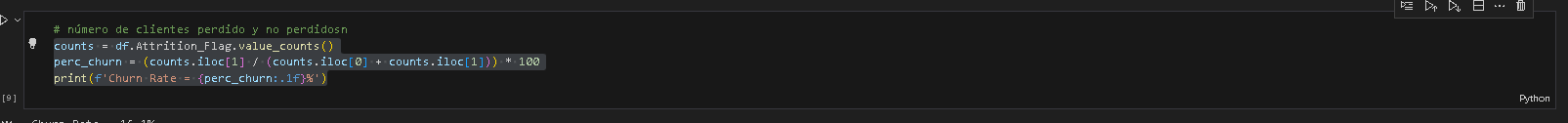
La función groupby agrupa por categoría y cuenta la cantidad de filas que tiene cada categoría en la columna Attrition\_flag.



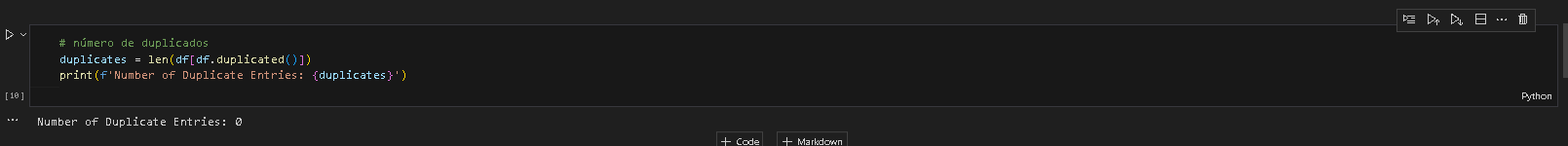
La función value\_counts muestra el número de observaciones por categoría en Attrition\_Flag, asociando cada una con su respectiva frecuencia.



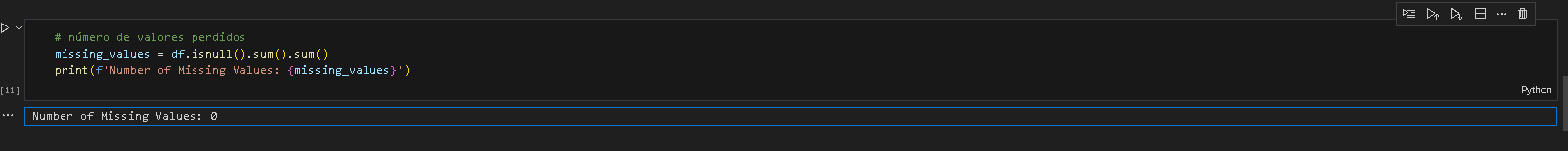
La función describe proporciona un análisis descriptivo de cada columna con atributos cuantitativos, mostrando el número de observaciones (count), el promedio (mean), la desviación estándar (std), el valor mínimo (min), los cuartiles del 25%, 50% y 75%, y el valor máximo (max) dentro de la columna.



La función calcula la tasa de abandono como el porcentaje de Attrited Customer sobre el total de clientes (la suma de Existing Customer y Attrited Customer), multiplicado por 100. Utiliza iloc para seleccionar valores específicos en el DataFrame count (almacena el conteo por categoría de la columna Attrition\_Flag utilizando la función value\_counts), donde Existing Customer está en la fila 1 y Attrited Customer en la fila 0. Finalmente, imprime el valor calculado en perc\_churn junto con el texto "Churn Rate", formateado como porcentaje con un decimal.



Calcula el número de duplicados (Calcula el largo de un arreglo) que se encuentran en toda la data frame de igual manera imprimimos el resultado con la función `print`

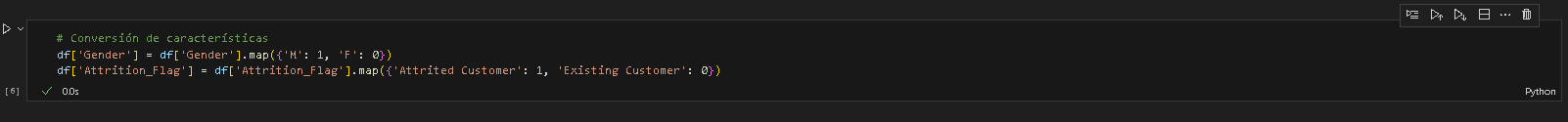


La función is\_null() identifica los valores nulos en cada celda del DataFrame, asignando un valor booleano: True para nulos (representado internamente como 1) y False para no nulos (representado como 0). Primero, se suman los valores nulos por columna, y luego se suman estas sumas para obtener el total de valores faltantes en todo el DataFrame. Finalmente, este total se imprime junto con un texto descriptivo usando print.

Patrón de fondo

Descripción generada automáticamente con confianza baja

Esta función cuenta y muestra la cantidad de columnas, filas y la frecuencia de cada tipo de dato en el DataFrame df.



Esta función convierte los valores categóricos de las columnas Gender y Attrition\_Flag en valores numéricos (1 y 0) en el DataFrame df.

Imagen que contiene Patrón de fondo

Descripción generada automáticamente

Esta función identifica las columnas categóricas, enteras y de punto flotante en df, aplica codificación one-hot (get\_dummies) a las categóricas, y luego imprime el nuevo número de columnas del DataFrame.

Histograma limite de Credito por genero

Gráfico, Histograma

Descripción generada automáticamente

Realice un histograma, segmentado por genero para tener una noción de si la concentración del límite de créditos se ve impactado por el género. Si bien podemos observar que en los valores más menores de crédito se concentran mas hombres que mujeres, de lo contrario en valores más altos de crédito se concentran más mujeres que hombres.

El análisis, no tiene un sustento estadístico, por lo que simplemente es una noción cualitativa de la grafica

El análisis de la BBDD BikePrices.csv se encuentra como archivo de soporte BikePrices.ipynb

1. **Números Aleatorios y Bondad de Ajuste en Python**

Explicación de cada celda de codigo del Notebook de data-gen.ipynb

Patrón de fondo

Descripción generada automáticamente con confianza baja

La función de numpy genera un arreglo de 1000 números (n), ajustado a una normal con media 3 y desviación estándar de 0.5. y después se imprime el arreglo Imagen que contiene Patrón de fondo

Descripción generada automáticamente

Realiza un analisis descriptivo de el arreglo vals que se genero

Imagen que contiene Interfaz de usuario gráfica

Descripción generada automáticamente

Imagen que contiene Texto

Descripción generada automáticamenteLa función realiza un histograma, con particiones de 30, y el comando plt llama las función de matplotlib para asignar titulo, eje x y eje y.El Histograma replica la forma de una Normal.

Esta función genera y visualiza la función de densidad de probabilidad de una distribución normal estándar (media 0, desviación estándar 1) usando matplotlib.

Imagen que contiene Interfaz de usuario gráfica

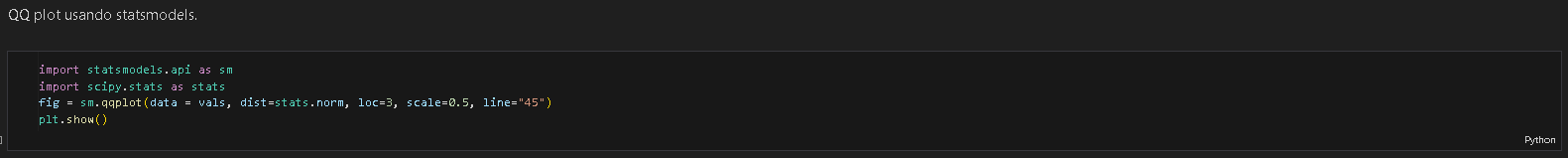
Descripción generada automáticamente

Esta función genera y visualiza la función de densidad de probabilidad de una distribución normal con media 3 y desviación estándar 1 usando matplotlib.

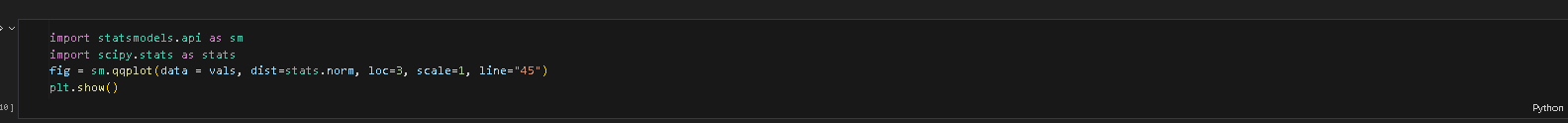
Imagen que contiene Patrón de fondo

Descripción generada automáticamente

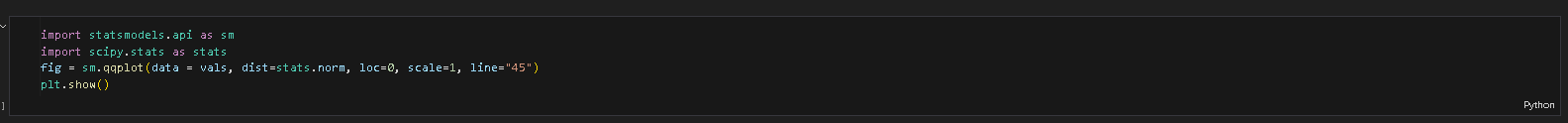
Esta función genera y visualiza la función de densidad de probabilidad de una distribución normal con media 3 y desviación estándar 0.5 usando matplotlib.



Este código genera un gráfico Q-Q (cuantiles-cuantiles) para comparar la distribución de los datos en vals con una distribución normal con media 3 y desviación estándar 0.5, utilizando statsmodels.



Este código genera un gráfico Q-Q (cuantiles-cuantiles) para comparar la distribución de los datos en vals con una distribución normal con media 3 y desviación estándar 1, utilizando statsmodels.



Este código genera un gráfico Q-Q (cuantiles-cuantiles) para comparar la distribución de los datos en vals con una distribución normal con media 0 y desviación estándar 1, utilizando statsmodels.

Imagen que contiene Interfaz de usuario gráfica

Descripción generada automáticamente

Este código crea un histograma de los datos en vals con 30 intervalos y superpone la función de densidad de una distribución normal con media 3 y desviación estándar 0.5, y luego muestra el gráfico con matplotlib.

**Cambio a distribución beta**

**Gráfico, Histograma

Descripción generada automáticamenteGráfico, Histograma

Descripción generada automáticamente**

Gráfico, Gráfico de líneas

Descripción generada automáticamenteGráfico, Histograma

Descripción generada automáticamente

Vemos que en el cambio de distribución la grafica un poco con tipo de rezago hacia la derecha, haciendo una contraste de estas graficas podemos observar que una grafica se ajusta a la normal y se observa como la distribución beta no se justa a la normal