Ejercicio de Laboratorio 7. Apertura de DataSets

# Ejercicios

* Elabora un programa, en el lenguaje que quieras, que abra y cargue los contenidos del archivo adjunto.
* De cada columna numérica calcula el promedio, la varianza y la desviación estándar.
* Separa los datos en diferentes matrices, de acuerdo con la categoría de los datos y repite los cálculos del paso anterior.

# Código

## Función convertir\_a\_csv (archivo)



Esta función toma un archivo de entrada y lo convierte al formato CSV utilizando la biblioteca Pandas. Primero lee el archivo con ‘pd.read\_csv’, luego lo guarda como un archivo CSV con ‘to\_csv’. Después elimina la fila de los encabezados. Si existe algún error durante el proceso, se maneja adecuadamente.

#### Salida

**Archivo dataset.data | Archivo dataset.csv**

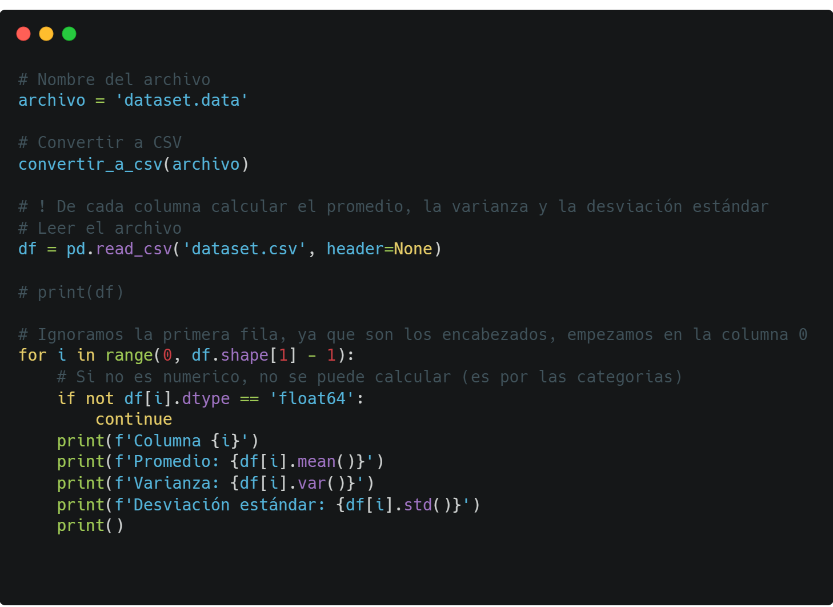
Imagen que contiene Forma

Descripción generada automáticamenteTabla

Descripción generada automáticamente con confianza media

## Función main ()

### Parte 1:



Ya con el archivo convertido a CSV, el código lee el archivo nuevamente con Pandas y realiza los cálculos estadísticos sobre cada columna numérica utilizando los métodos ‘mean ()’, ‘var ()’ y ‘std ()’. Esta función ignora las columnas que no son numéricas (por la columna de las categorías).

#### Salida

Texto

Descripción generada automáticamente

### Parte 2:



El código identifica las categorías en la última columna del dataframe y separa los datos en diferentes dataframes según la categoría. Luego, repite los cálculos estadísticos para cada categoría.

#### Salida

Texto

Descripción generada automáticamenteTexto

Descripción generada automáticamenteTexto

Descripción generada automáticamente

# Pruebas

Hicimos los cálculos en una hoja de Excel para corroborar que los cálculos son correctos.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Promedio** | **Varianza** | **Desv. Estándar** |
| **Col 0** | 5.843333333 | 0.685694 | 0.828066128 |
| **Col 1** | 3.057333333 | 0.189979 | 0.435866285 |
| **Col 2** | 3.758 | 3.116278 | 1.765298233 |
| **Col 3** | 1.199333333 | 0.581006 | 0.762237669 |
|  |  |  |  |
|  | **Iris-Setosa** | | |
|  | **Promedio** | **Varianza** | **Desv. Estándar** |
| **Col 0** | 5.006 | 0.124249 | 0.352489687 |
| **Col 1** | 3.428 | 0.14369 | 0.379064369 |
| **Col 2** | 1.462 | 0.030159 | 0.173663996 |
| **Col 3** | 0.246 | 0.011106 | 0.105385589 |
|  |  |  |  |
|  | **Iris-versicolor** | | |
|  | **Promedio** | **Varianza** | **Desv. Estándar** |
| **Col 0** | 5.936 | 0.266433 | 0.516171147 |
| **Col 1** | 2.77 | 0.098469 | 0.313798323 |
| **Col 2** | 4.26 | 0.220816 | 0.469910977 |
| **Col 3** | 1.326 | 0.039106 | 0.19775268 |
|  |  |  |  |
|  | **Iris-virginica** | | |
|  | **Promedio** | **Varianza** | **Desv. Estándar** |
| **Col 0** | 6.588 | 0.404343 | 0.635879593 |
| **Col 1** | 2.974 | 0.104004 | 0.322496638 |
| **Col 2** | 5.552 | 0.304588 | 0.551894696 |
| **Col 3** | 2.026 | 0.075433 | 0.274650056 |