

LENGUAJE PYTHON

Sesión No.3: Exploración y Análisis de datos

Jefferson A. Peña Torres

jefferson.amado.pena@correounivalle.edu.co

Escuela de Ingeniería de sistemas y computación (EISC)
Universidad del Valle, Cali
Enero 2020

Exploración y Análisis de datos: Archivos

Existen varios formatos de archivo de datos. Los formatos de archivos de datos a menudo son un problema porque realmente no existe un estándar.

- El formato más básico TXT, se estructura como una secuencia de líneas de texto electrónico.
- Por su simplicidad, los archivos CSV (valores separados por comas) son muy populares, pero esto no es realmente estandarizado. En la práctica, los usuarios tienen que adaptar sus lectores de archivos a cada caso específico.
- Otros formatos
 - HD5F (.hd5f)
 - JSON (.json)
 - VCARD (.vcf)
 - SQLite (.sqlite)
 - Excel (.xls, .xlsx)
 - OpenDocument (.ods)
 - TextoEnriquecido (.rtf)
 - Access (.dbf)

Exploración y Análisis de datos: Archivos

- **Crear, Adicionar, Leer, Leer linea a linea**

- Modos

- "r", "w", "x", "a", "t", "b", "+"

- Por defecto se pueden abrir archivos con la función

```
my_file = open("texto.txt","w+")
```

- **import** csv

- np.loadtxt(), np.genfromtxt(), np.fromfile()

- pd.read_table(), pd.read_csv(), pd.read_excel(), pd.read_json(), pd.read_spss(),

Exploración y Análisis de datos: Archivos

- **Crear, Adicionar, Leer, Leer linea a linea**

- Modos

- "r", "w", "x", "a", "t", "b", "+"

- Por defecto se pueden abrir archivos con la función

- `import csv`

- `np.loadtxt()`, `np.genfromtxt()`, `np.fromfile()`

- `pd.read_table()`, `pd.read_csv()`, `pd.read_excel()`, `pd.read_json()`, `pd.read_spss()`,

Ruta del archivo

Modo

```
my_file = open("texto.txt","w+")
```

Exploración y Análisis de datos: Archivos

Para crear un archivo:

- `my_file = open("texto.txt", "w+")`

Para adicionarle contenido a un archivo

- `my_file = open("texto.txt", "a+")`

Para leer todo el contenido del archivo

- `my_file = open("texto.txt", "r")`

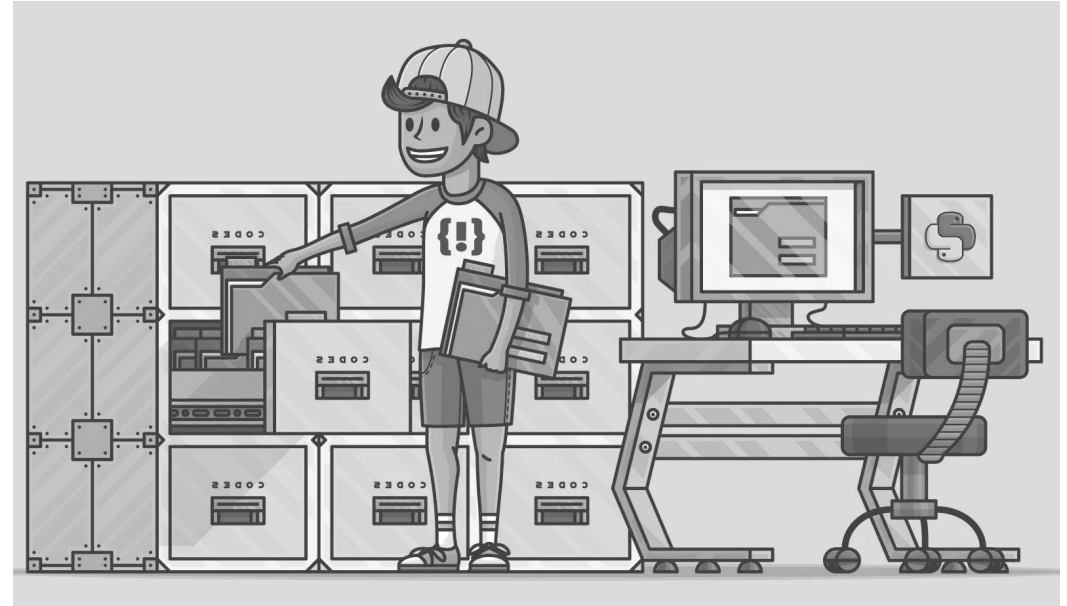
Para leer todo el contenido del archivo

- `my_file = open("texto.txt", "r")`
- `my_file.readlines()`



Exploración y Análisis de datos: Archivos

- Python posee una función llamada *open* que se encuentra por defecto en el lenguaje para abrir archivos.
- Utilizando *numpy*, se puede cargar desde un archivo una matriz numérica.
- Utilizando *pandas* puedo cargar un Dataframe

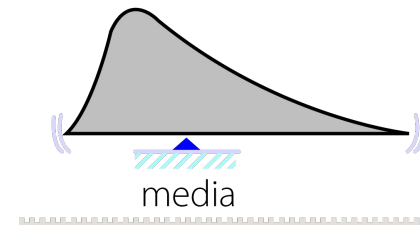
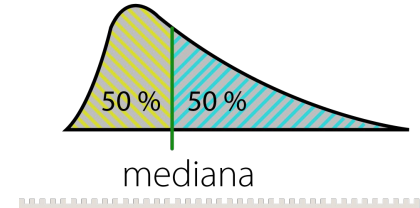
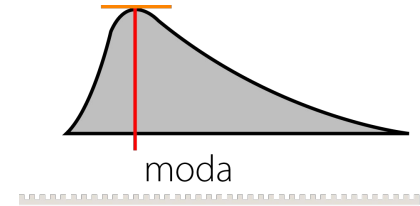
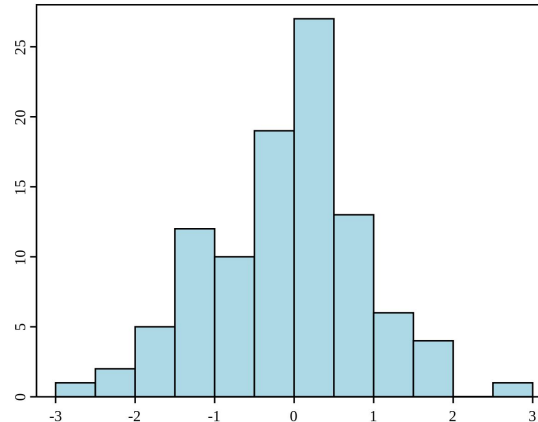


Exploración y Análisis de datos: Análisis Descriptivos

Parametros estadísticos básicos

- media, mediana, varianza, cuasivarianza, coef.-variación, curtosis
- Histograma
- Diagrama de dispersión
- Diagrama cuantil-cuantil
- Constrase
 - Shapiro-Wilks
 - Grubbs

$$S^2 = \sigma^2 = \frac{\sum_{i=1}^n x_i^2 \cdot f_i}{N} - \left(\bar{X}\right)^2$$



Exploración y Análisis de datos: Análisis Descriptivos

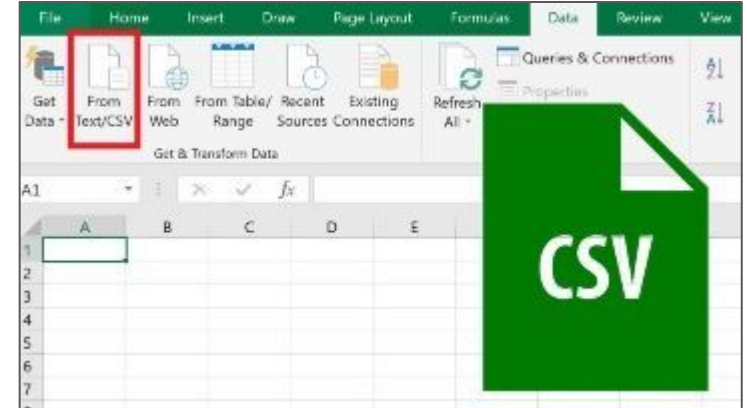
Ejercicio No. 1: Leer un .csv usando numpy

Ejercicio No. 2: Leer un .csv usando pandas

Ejercicio No. 3: Cree una tabla de frecuencias

Ejercicio No. 4: Encuentre las medidas de tendencia central, dispersión y posición para tres de las variables del conjunto de datos

Ejercicio No. 5: Apoye su análisis con gráficos, histograma, diagramas de caja, etc





GRACIAS

Jefferson A. Peña Torres

jefferson.amado.pena@correounivalle.edu.co

Escuela de Ingeniería de sistemas y computación (EISC)
Universidad del Valle, Cali
Enero 2020