**UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS**

**INVESTIGACION DE OPERACIONES 1**

****

**Gabriel Hernández - Código: 20181020059**

**Santiago Buitrago - Código: 20181020036**

**Kevin Santiago Garcia Cuellar - Código: 20181020074**

**Grupo 2**

**Bogotá D.C**

**2021**

**Librería Pandas**

**¿Qué es?**

Pandas es una librería para el análisis de datos que cuenta con las estructuras de datos que necesitamos para limpiar los datos en bruto y que sean aptos para el análisis (por ejemplo, tablas). Es importante señalar aquí que, dado que pandas lleva a cabo tareas importantes, como alinear datos para su comparación, fusionar conjuntos de datos, gestión de datos perdidos, etc., se ha convertido en una librería muy importante para procesar datos a alto nivel en Python (es decir, estadísticas ). Pandas fue diseñada originalmente para gestionar datos financieros, y como alternativo al uso de hojas de cálculo (es decir, Microsoft Excel)

La estructura de datos básica de pandas se denomina DataFrame, que es una colección ordenada de columnas con nombres y tipos, parecido a una tabla de base de datos, donde una sola fila representa un único caso (ejemplo) y las columnas representan atributos particulares. Cabe señalar aquí que elementos en distintas columnas pueden ser de diferentes tipos.

Por lo tanto, como resumen decir que pandas nos proporciona las estructuras de datos y funciones necesarias para el análisis de datos.

**Instalación de Pandas**

Ahora vamos a ver cómo podemos instalar pandas en nuestras máquinas y utilizarla para el análisis de datos. La forma más sencilla de instalar pandas y evitar problemas de dependencia es mediante el uso de Anaconda, que está dentro de los pandas. Como puedes leer en la página de descarga Anaconda:

*Anaconda is a completely free Python distribution (including for commercial use and redistribution). It includes more than 400 of the most popular Python packages for science, math, engineering, and data analysis*

La distribución Anaconda es multiplataforma, lo que significa que se puede instalar en OS X, Windows y Linux. Se puede elegir el instalador que más se ajuste a tu sistema operativo. Después de descargar el instalador, simplemente sigue los sencillos pasos del asistente de instalación y ya está. Todo lo que tenemos que hacer ahora con el fin de utilizar pandas es importar el paquete de la siguiente manera:

import pandas as pd

**DataFrame**

El DataFrame es una estructura de datos tabular que se compone de columnas y filas ordenadas. Para que todo sea más sencillo vamos a ver un ejemplo de creación de un DataFrame (tabla) de un diccionario de listas. El siguiente ejemplo muestra un diccionario que consta de dos keys, Name y Age, y su correspondiente lista de valores.

Al ejecutar el script anterior se obtiene la siguiente salida:  


Observa que el constructor del DataFrame ordena las columnas en orden alfabético. Si deseas cambiar el orden de las columnas, debes escribir lo siguiente debajo de data\_frame:



Para ver el resultado escribe: print data\_frame\_2

Decirte que no necesitas utilizar los labels por defecto 0, 1, 2, ..., y puedes usar a, b, c, ... en su lugar. En ese caso, puedes utilizar index en la script anterior de la siguiente manera:

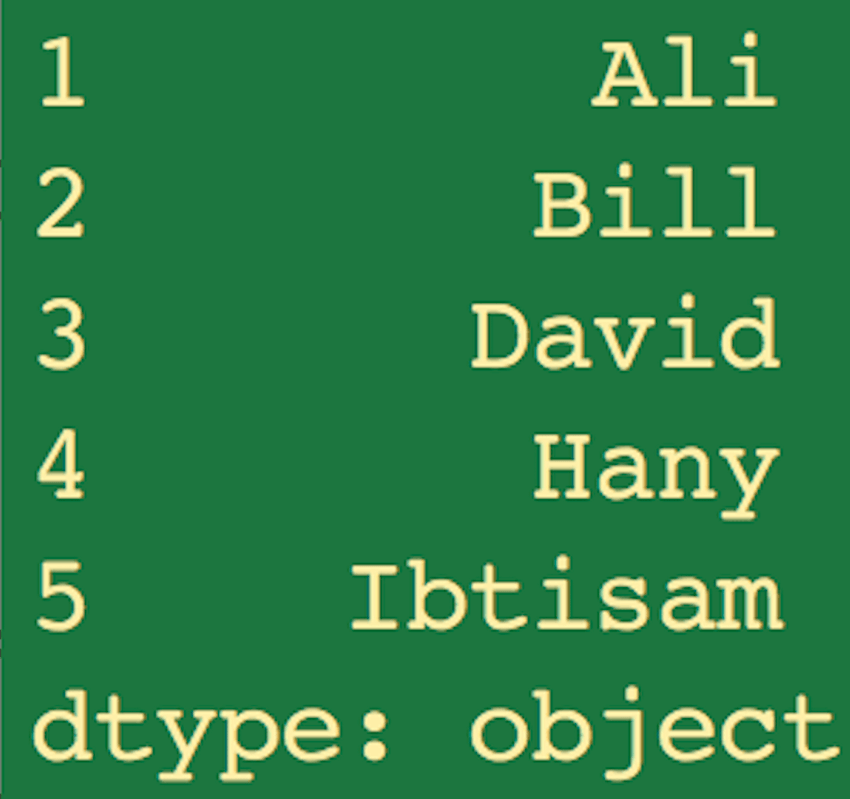


**Series**

Series es la segunda estructura de datos pandas de la que os voy a hablar. Series es un objeto unidimensional (1D) similar a la columna de una tabla. Si queremos crear un Series para un listado de nombres, podemos hacer lo siguiente**:**

****

La salida que genera es:



**Funciones:**

**Head y tail**

Las funciones head() y tail() nos permiten ver una muestra de nuestros datos, especialmente cuando tenemos muchísimos. El número por defecto de elementos a mostrar es 5, pero puedes mostrar tantos como quieras. No hay problema.

Digamos que tenemos un Series compuesto por 20,000 elementos aleatorios (números):

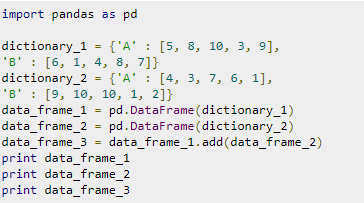


Si queremos ver los primeros cinco elementos, y los cinco últimos, podemos hacer lo siguiente:



**Add**

Vamos a ver un ejemplo de la función add. En esta ocasión veremos cómo añadir dos DataFrames**:**

****