



Pandas

Grupo 7

Agostinelli Gino 46795

Aguirre Gabriel 44899

ANALITICA DE DATOS

La utilización de internet y del resto de sistemas y tecnologías de comunicación (TIC), genera ingentes cantidades de datos de todo tipo.

Analítica de datos (DA) es el proceso de examinar conjuntos de datos para encontrar tendencias y sacar conclusiones sobre la información que contienen.





¿Qué es Pandas?

pandas es una biblioteca de software para manipulación y análisis de datos

- escrita como extensión de NumPy
- DataFrame para manipulación de datos
- Permite reordenar, dividir y combinar conjuntos de datos
- Escribe datos en distintos formatos
- Series temporales
- Línea de tiempo

	another_column	binary_solo	column_1	float_column
0	this	True	1	0.1000
1	column	False	2	0.5000
2	has	True	3	33.0000
3	strings	True	4	48.0000
4	inside!	False	5	42.5558

2008: Comienza el desarrollo de pandas

2012: Es publicada la primera edición de Python for Data Analysis

2018: Primer sprint presencial de desarrolladores del núcleo

2009: Pandas se convierte en un software de código abierto

2015: pandas se convierte en un proyecto patrocinado por NumFOCUS

Combinación con otras librerías



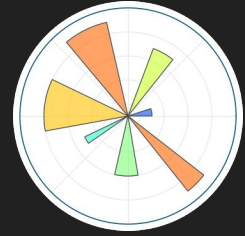
Pandas

Cargar datos



Numpy

Aplicar transformaciones



Matplotlib

Visualizar

comparación con pyspark



Pandas

- Procesamiento normal
- Dataframe mutable
- Fácil manipulación de datos a comparación de Spark
- Sin tolerancia a fallos
- Lo ideal es trabajar con menos de 2 mill registros

Spark

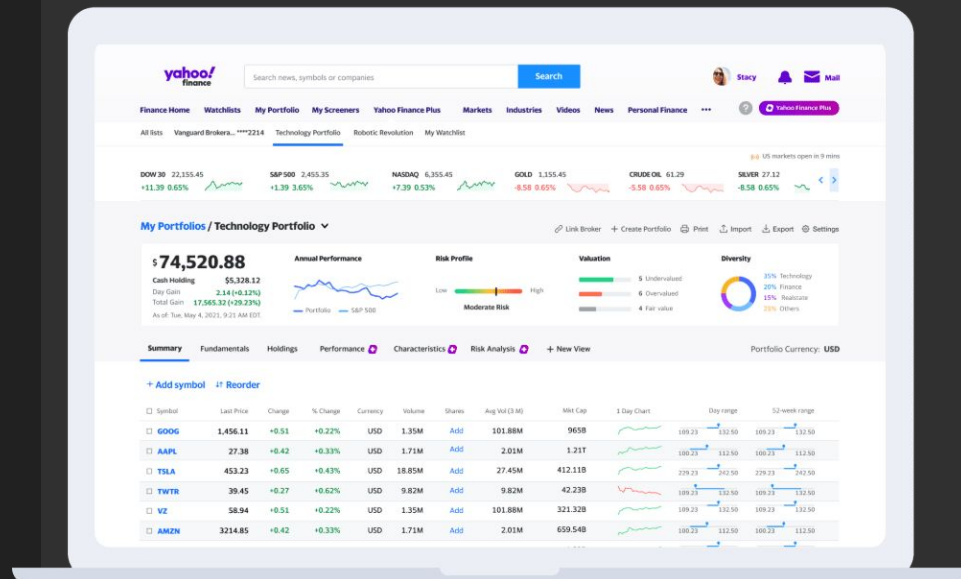
- Procesamiento distribuido
- Dataframe inmutable
- Manipulación de datos más complicada que Pandas
- Tolerancia a fallos
- Soporta grandes volúmenes de datos



Yahoo! Finance es un servicio de Yahoo! que proporciona información financiera y comentarios.

Ofrece información financiera, incluyendo cotizaciones de bolsa, índices bursátiles, comunicados de prensa corporativos y financieros, y foros de discusión.

También ofrece algunas herramientas para el manejo organizado de finanzas personales.



DEMO

- yfinance(API Yahoo finance)
- pandas
- matplotlib
- series de tiempo



+



Gracias!