

Danilo Dominici
ddominici@sqlserverinfo.it

Come (non) diventare ricchi con SQL Server e Python

Chi sono



- Consulente indipendente, Trainer, Speaker, Autore
- Utilizzo SQL Server dalla versione 6.5
- Microsoft Certified Trainer dal 2000
- Microsoft Data Platform MVP dal 2014 al 2020
- Membro delle principali community italiane su SQL Server
- Fondatore di **SQL START!**
(spoiler: ven 11 giugno !!!)



Agenda



Un po' di teoria (finanziaria)



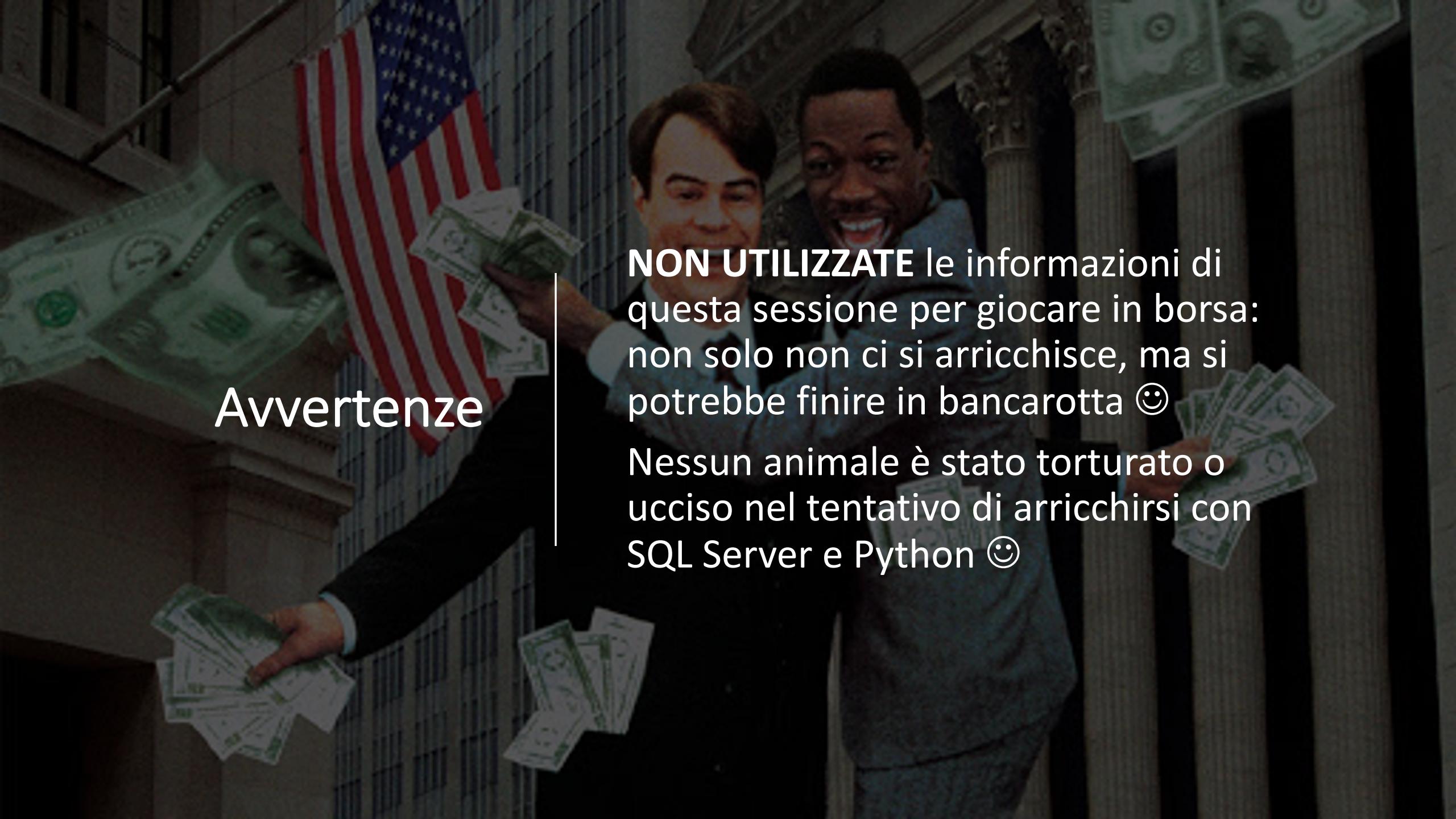
Python per l'analisi dei dati



Python e SQL Server (Machine Learning Services)



Let's play!



Avvertenze

NON UTILIZZATE le informazioni di questa sessione per giocare in borsa: non solo non ci si arricchisce, ma si potrebbe finire in bancarotta ☺

Nessun animale è stato torturato o ucciso nel tentativo di arricchirsi con SQL Server e Python ☺

Un po' di teoria

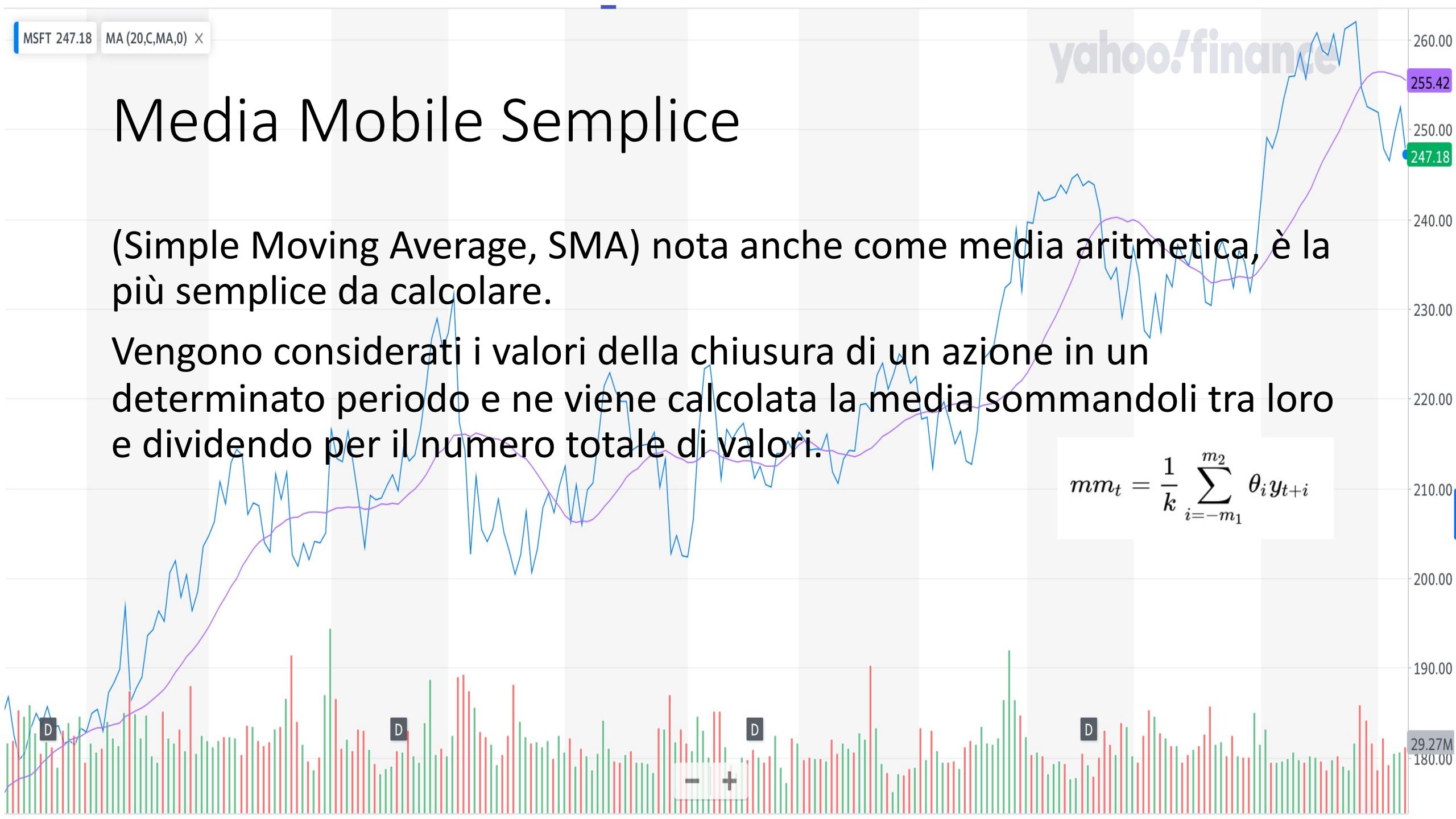
Districarsi con la terminologia dei mercati finanziari

Media Mobile Semplice

(Simple Moving Average, SMA) nota anche come media aritmetica, è la più semplice da calcolare.

Vengono considerati i valori della chiusura di un'azione in un determinato periodo e ne viene calcolata la media sommandoli tra loro e dividendo per il numero totale di valori.

$$mm_t = \frac{1}{k} \sum_{i=-m_1}^{m_2} \theta_i y_{t+i}$$



Media Mobile Esponenziale

(Exponential Moving Average, EMA) Viene generata mediante un calcolo più complesso, che assegna un peso differente ai valori di un dato periodo, maggiore ai più recenti e minore a quelli più vecchi.



Esempio di strategia basata sulle medie mobili

- Estraggo i dati storici
- Calcolo la media mobile esponenziale a 20gg (veloce)
- Calcolo la media mobile esponenziale a 50gg (lenta)
- Calcolo la media mobile esponenziale a 200gg (di lungo periodo)
- Quando la media veloce incrocia la media lenta verso l'alto e sono al di sopra della media di lungo periodo **compro**
- Quando la media veloce incrocia la media lenta verso il basso e sono al di sotto della media di lungo periodo **vendo**
- Tutto chiaro? ☺

Convergenza/divergenza della media mobile

(Moving Average Convergence Divergence, MACD) è un indicatore utile ad identificare gli aspetti generali del trend generale di un asset.

In particolare, si usa per valutare il momentum (cioè la velocità di cambiamento dei prezzi) e quindi è utile per capire quando investire o disinvestire.

Si calcola effettuando la differenza dei valori tra due medie mobili (la MACD line) e una media mobile esponenziale delle stesse medie mobili (linea di segnale) e individua la differenza tra le due linee come un istogramma che oscilla sopra e sotto un epicentro con valore 0.

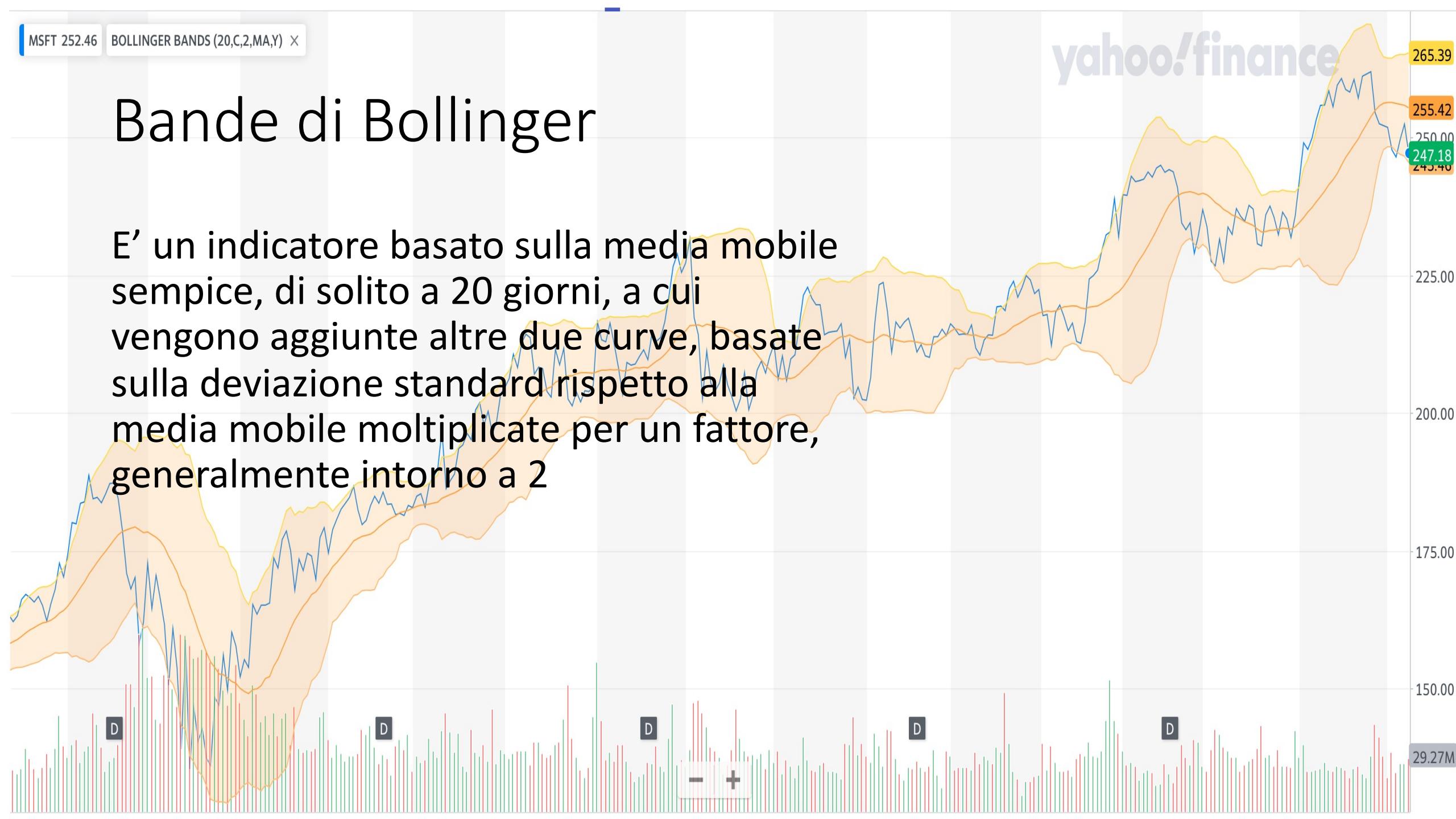


SMA, EMA e MACD



Bandi di Bollinger

E' un indicatore basato sulla media mobile semplice, di solito a 20 giorni, a cui vengono aggiunte altre due curve, basate sulla deviazione standard rispetto alla media mobile moltiplicate per un fattore, generalmente intorno a 2



Bande di Bollinger



Come usare le bande di Bollinger?

- Quando il prezzo esce dalla banda superiore e poi vi rientra, è un segnale di **vendita**
- Quando il prezzo esce dalla banda inferiore e vi rientra, è un segnale di **acquisto**

Demo

<https://finance.yahoo.com/quote/MSFT?p=MSFT&.tsrc=fin-srch>

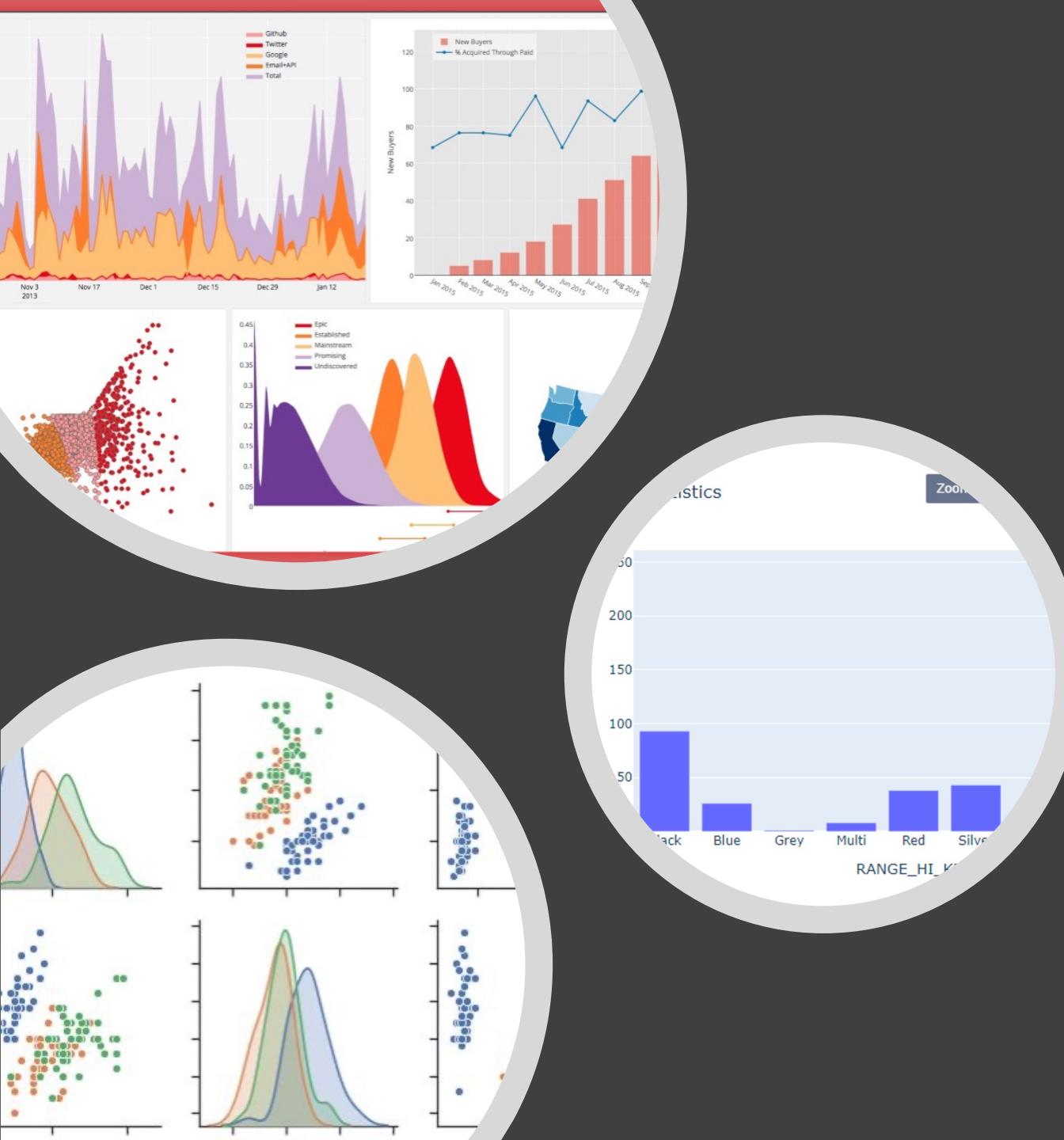
Python per l'analisi dei dati

Perché Python?

- 
- Facile da imparare
 - Leggibile
 - Scalabile
 - Tantissime librerie pronte all'uso
 - Facile integrarlo con altre applicazioni
 - Comunità di utilizzatori ed ecosistema molto attivi

Librerie per l'analisi e la visualizzazione dei dati

- Analisi
 - NumPy
 - Pandas
 - SciPy
 - Natural Language Toolkit (NLTK)
 - ...
- Visualizzazione
 - Matplotlib
 - Plotly
 - Seaborn
 - ...



Pandas dataframe

- E' una struttura che contiene dati bidimensionali e relative *etichette*
- Simili alle tabelle di un database relazionale o ai fogli di calcolo di Excel (righe e colonne)
- Può contenere vari tipi di dati bidimensionali (liste, coppie di valori, array, dizionari, serie - usando ad esempio la libreria NumPy)

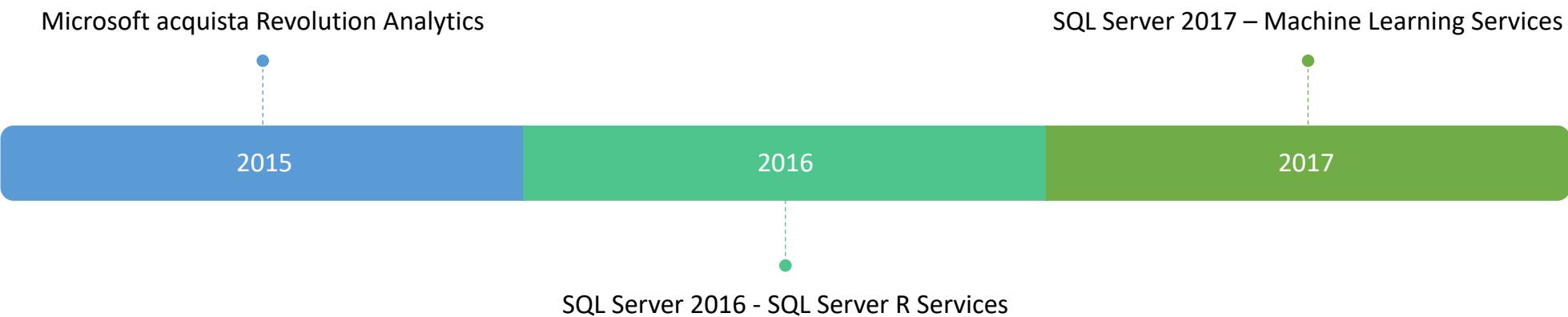
Demo

Pandas dataframe

Python e SQL Server

Machine Learning Services

Machine Learning Services



La distribuzione Python “Anaconda”



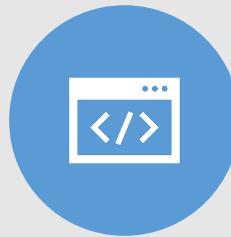
DISTRIBUZIONE DI
PYTHON PER LA
DATA SCIENCE



GESTORE PACKAGES
E AMBIENTE DI
LAVORO



100+ PACKAGES
INCLUSI



PYTHON VERSIONE
3.5.2



JUPYTER
NOTEBOOK

Include i
packages per
la Data
Science
più diffusi

NumPy – N-dimensional arrays,
random numbers

pandas – data manipulation,
DataFrame object

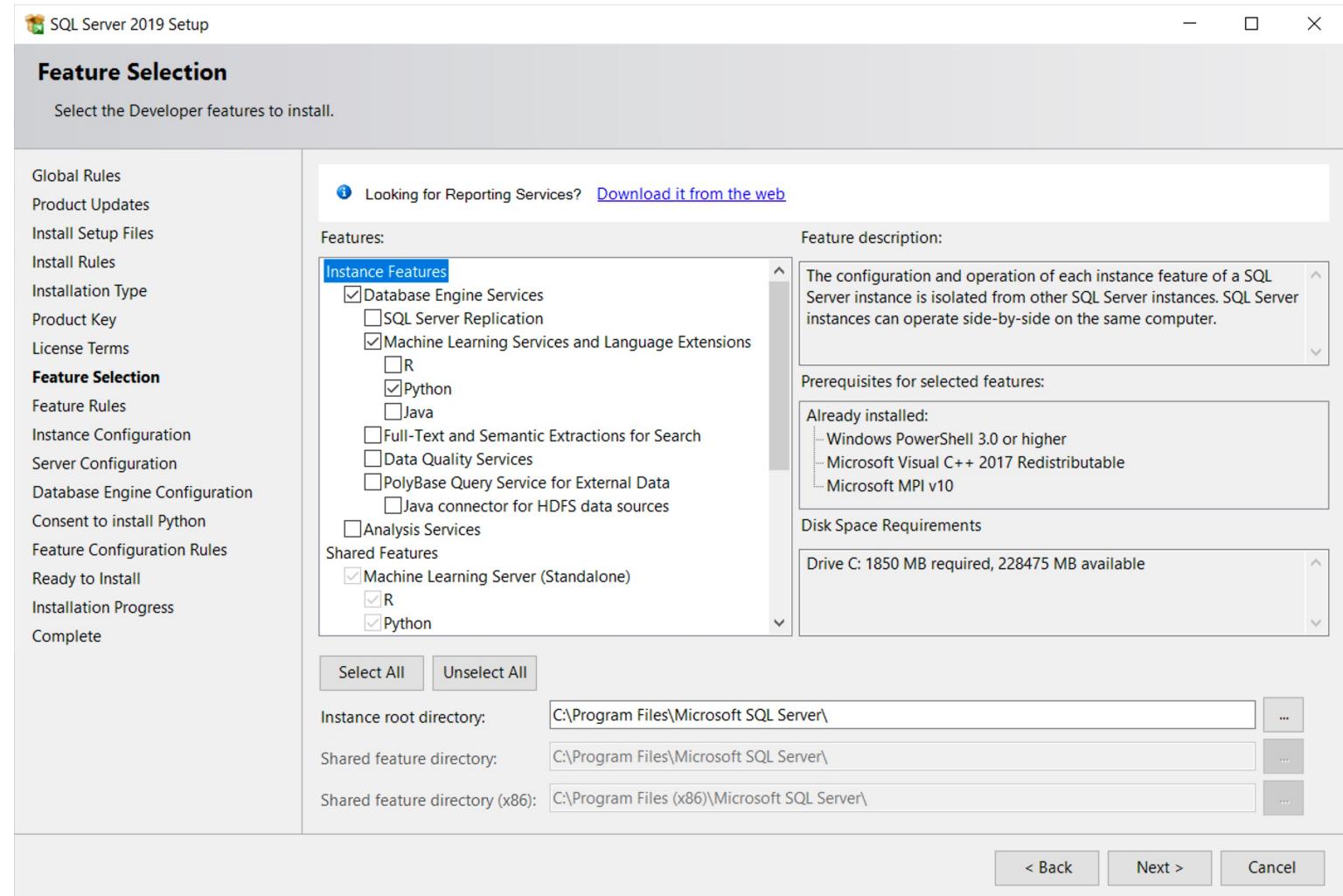
SciPy – scientific computing and
statistical methods

scikit-learn – machine learning

matplotlib – plotting and graphics

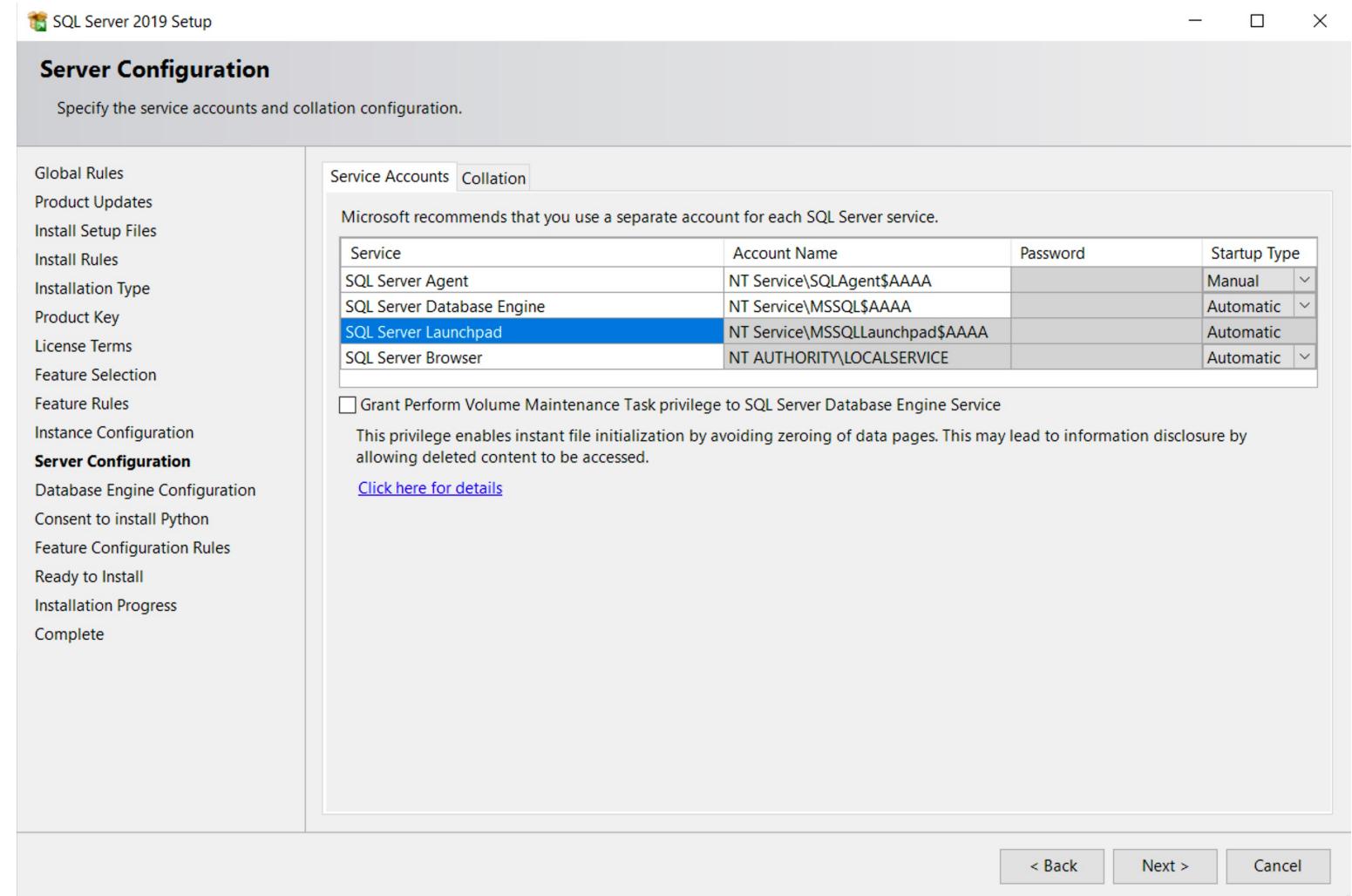
Installare i Machine Learning Services

step 1



Installare i Machine Learning Services

step 2



Installare i Machine Learning Services

step 3

Post-installazione

Abilito gli *external scripts*

```
EXEC sp_configure 'external  
scripts enabled', 1;  
RECONFIGURE;
```

Configurare la memoria in maniera adeguata

Installare ulteriori packages

```
Administrator: Command Prompt - conda install pyodbc
C:\Program Files\Microsoft SQL Server\MSSQL15.MSSQLSERVER\PYTHON_SERVICES\Scripts>conda install pyodbc
Collecting package metadata (current_repodata.json): done
Solving environment: /
The environment is inconsistent, please check the package plan carefully
The following packages are causing the inconsistency:

- defaults/win-64::blaze==0.11.3=py37_0
- defaults/win-64::numba==0.42.0=py37hf9181ef_0
done

## Package Plan ##

environment location: C:\Program Files\Microsoft SQL Server\MSSQ

added / updated specs:
- pyodbc

The following packages will be downloaded:

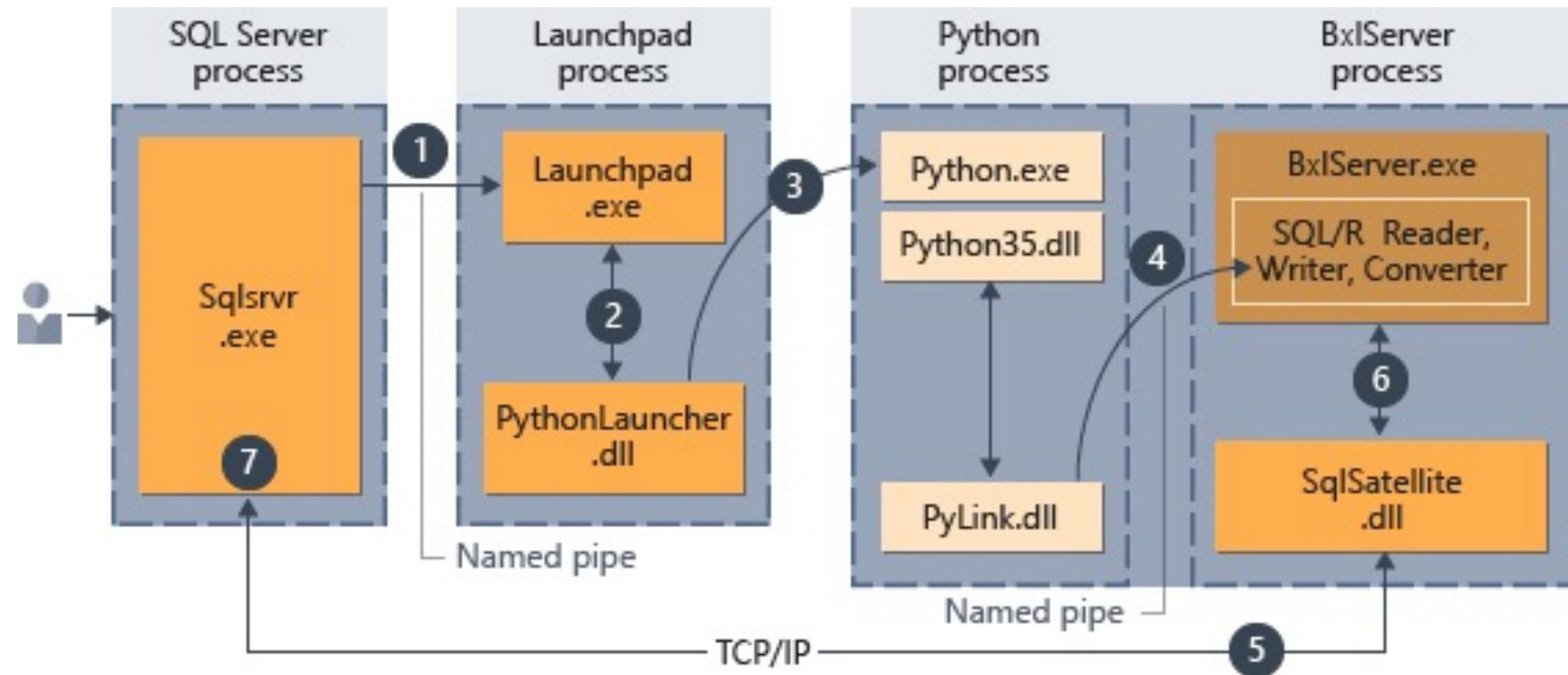
          package           build
ca-certificates-2021.1.19 | haa95532_0   122 KB
certifi-2020.12.5        | py37haa95532_0 141 KB
conda-4.9.2               | py37haa95532_0  2.9 MB
tbb-2018.0.5              | he980bc4_0    150 KB
                                           Total: 3.3 MB

The following NEW packages will be INSTALLED:
tbb                         pkgs/main/win-64::tbb-2018.0.5-he980bc4_0

The following packages will be UPDATED:
ca-certificates             2019.11.27-0 --> 2021.1.19-haa95532_0
certifi                      2019.3.9-py37_0 --> 2020.12.5-py37haa95532_0
conda                        4.7.12-py37_0 --> 4.9.2-py37haa95532_0

Proceed ([y]/n)?
```

Come funziona SQL Server + Python



Demo

Python + Pandas Dataframe + SQL Server

Demo

Reporting Services

Demo

Python e le statistiche di SQL Server