

## Soutenance BIG DATA



PySpark



# Data Exploration et Visualisation

## Moving Average

## Descriptive Statistics

## Return Rate

Stats for : High

```
-----+-----+-----+-----+
|      avg(High)|min(High)|max(High)|stddev_samp(High)|
|-----+-----+-----+-----+
|1762.0071216958152| 758.76| 3552.25|667.2385315752687|
|-----+-----+-----+-----+
```

Stats for : Low

```
-----+-----+-----+-----+
|      avg(Low)|min(Low)|max(Low)|stddev_samp(Low)|
|-----+-----+-----+-----+
|1722.1011452099956| 747.7| 3486.69|644.7988093382762|
|-----+-----+-----+-----+
```

Stats for : Open

```
-----+-----+-----+-----+
|      avg(Open)|min(Open)|max(Open)|stddev_samp(Open)|
|-----+-----+-----+-----+
|1743.433881363487| 757.92| 3547.0| 657.115307092714|
|-----+-----+-----+-----+
```

Stats for : Close

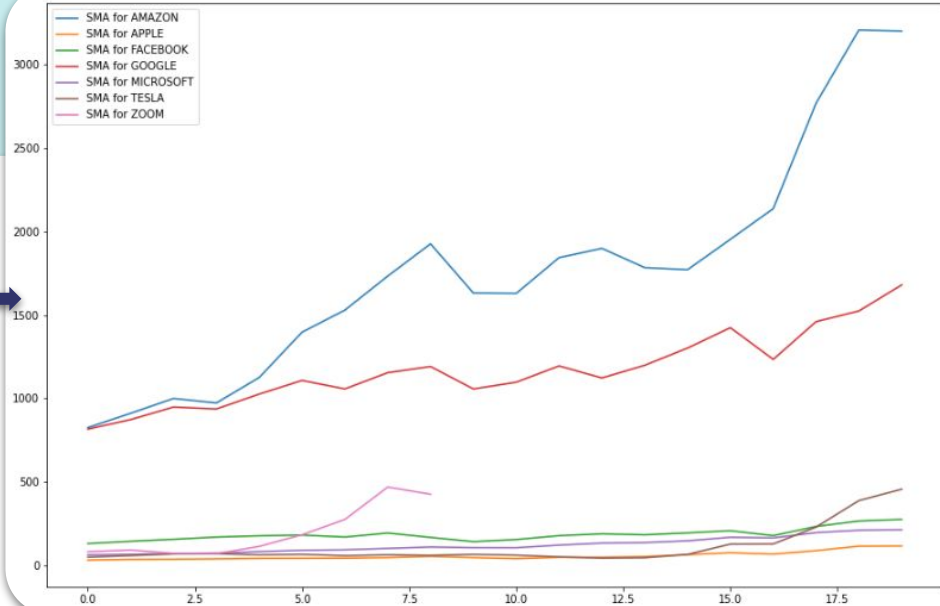
```
-----+-----+-----+-----+
|      avg(Close)|min(Close)|max(Close)|stddev_samp(Close)|
|-----+-----+-----+-----+
|1742.9566644206718| 753.67| 3531.45| 655.9576061129325|
|-----+-----+-----+-----+
```

Stats for : Volume

```
-----+-----+-----+-----+
|      avg(Volume)|min(Volume)|max(Volume)|stddev_samp(Volume)|
|-----+-----+-----+-----+
|4509728.05775076| 881300.0| 1.6565E7| 2179817.628631287|
|-----+-----+-----+-----+
```

Stats for : Adj Close

```
-----+-----+-----+-----+
|      avg(Adj Close)|min(Adj Close)|max(Adj Close)|stddev_samp(Adj Close)|
|-----+-----+-----+-----+
|1742.9566644206718| 753.67| 3531.45| 655.9576061129325|
|-----+-----+-----+-----+
```



The return rate for each year for AMAZON

Average return rate in 2017

Return rate in 2017 is 54.29992572735133%

Average return rate in 2018

Return rate in 2018 is 28.154434360334896%

Average return rate in 2019

Return rate in 2019 is 26.115207985258248%

Average return rate in 2020

Return rate in 2020 is 71.38338541666667%



## More Insights

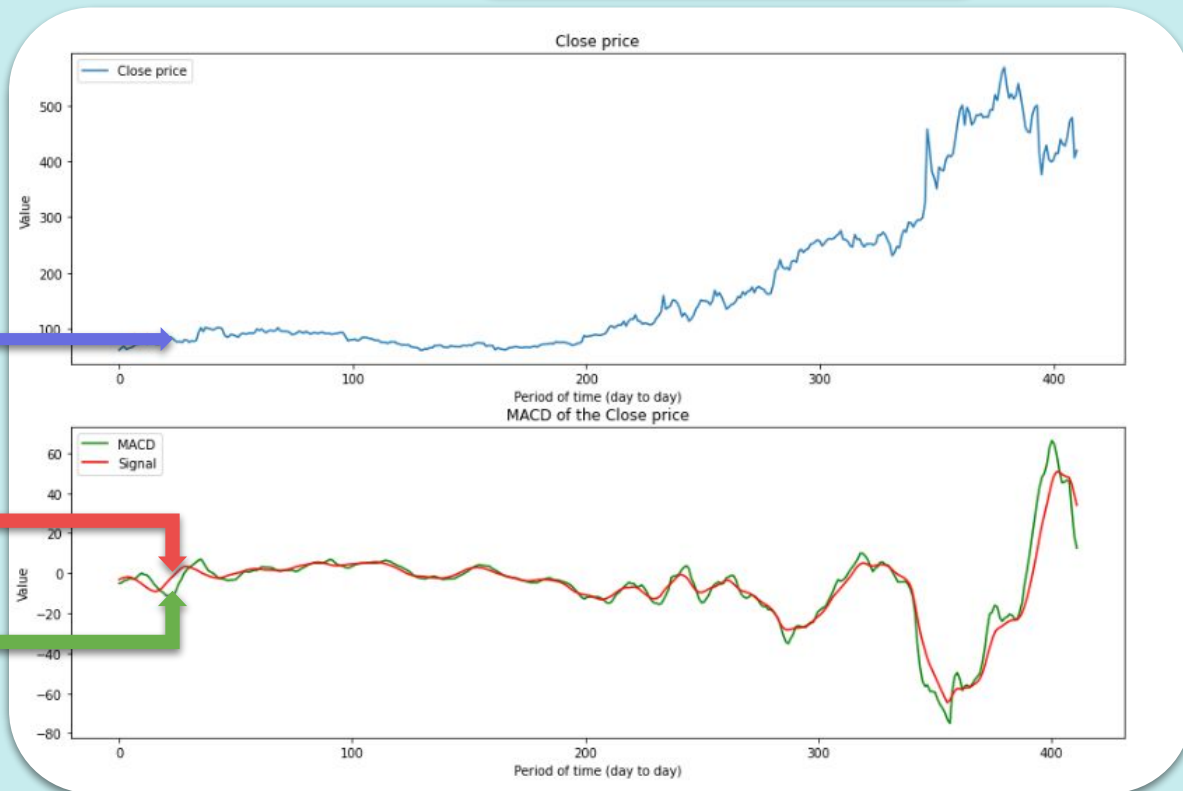
### Moving Average Convergence Divergence (MACD)

Example : MACD of Zoom's  
Close price

Closing Price

Signal = EMA 9 on MACD

MACD = EMA 12 - EMA 26





## More Insights

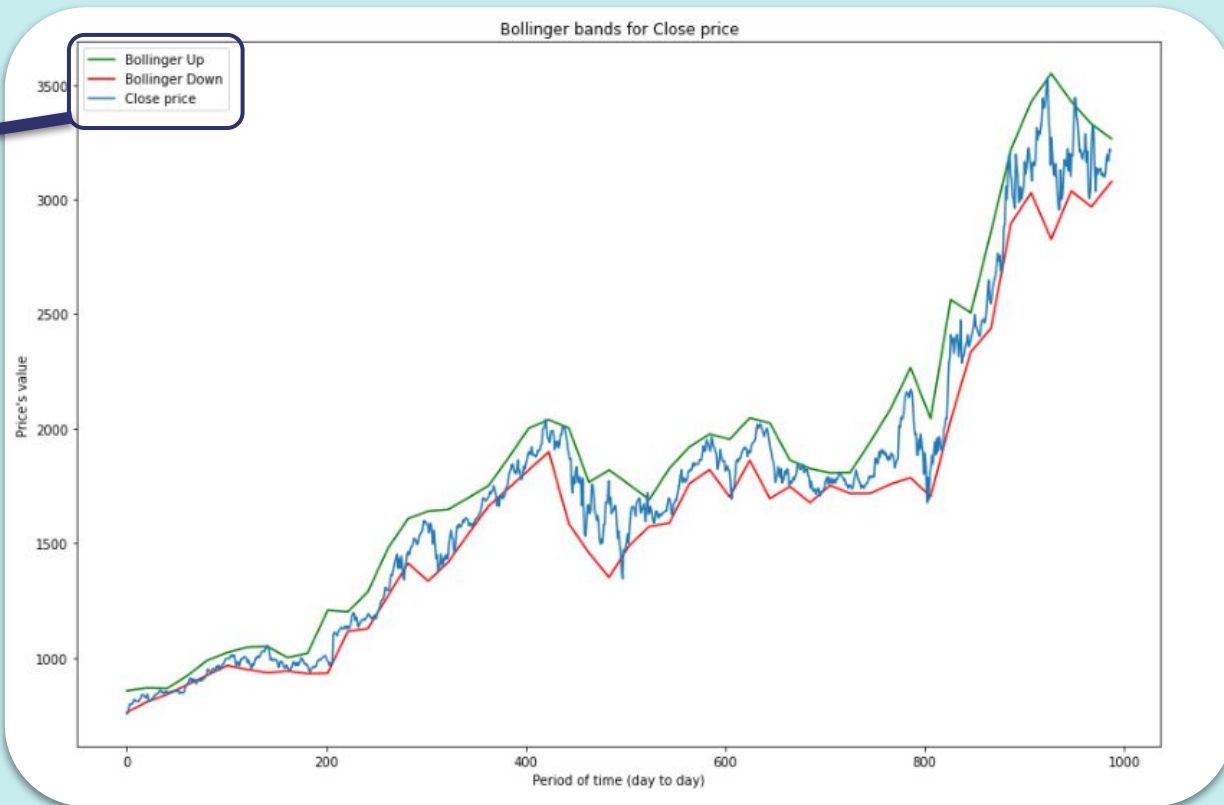
### Bollinger Bands

-  **Bollinger Up**
-  **Bollinger Down**
-  **Close Price**

**Bollinger Up :  $SMA + std^2$**

**Bollinger Down :  $SMA - std^2$**

**Example : Bollinger Bands on Amazon's  
Close price on a 20 period**

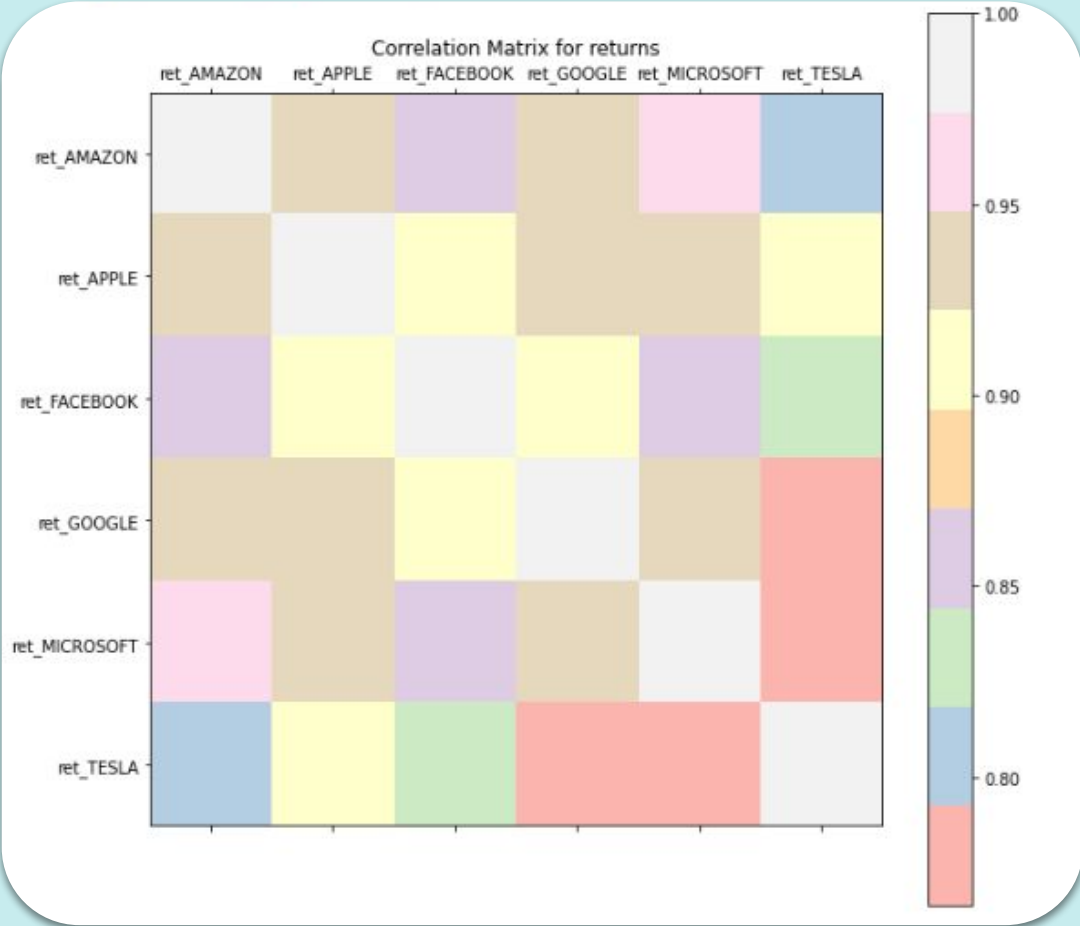




More Insights

## Portfolio Diversification

**Example : Correlation matrix on returns  
on Amazon, Apple, Facebook, Google,  
Microsoft and Tesla**



## Machine Learning : Regression Model

DataFrame for  
model

| Date                         | Close             | EMA               |
|------------------------------|-------------------|-------------------|
| 2017-01-17T00:00:00.000+0000 | 804.6099853515625 | 801.1410034179687 |
| 2017-01-18T00:00:00.000+0000 | 806.0700073242188 | 802.0371859463778 |
| 2017-01-19T00:00:00.000+0000 | 802.1749877929688 | 802.062240827576  |
| 2017-01-20T00:00:00.000+0000 | 805.02001953125   | 802.6000187736986 |
| 2017-01-23T00:00:00.000+0000 | 819.3099975585938 | 805.6381967345885 |
| 2017-01-24T00:00:00.000+0000 | 823.8699951171875 | 808.9530691677883 |
| 2017-01-25T00:00:00.000+0000 | 835.6699829101562 | 813.8106898482188 |
| 2017-01-26T00:00:00.000+0000 | 832.1500244140625 | 817.1451143147358 |
| 2017-01-27T00:00:00.000+0000 | 823.3099975585938 | 818.2660021772554 |

Split

Train 80%

Test 20%

Regression  
Model

RMSE : 39.44  
R2 : 0.96

| Date                         | Close             | EMA               | log_EMA           | features   | log_pred          | prediction        |
|------------------------------|-------------------|-------------------|-------------------|--|-------------------|-------------------|
| 2017-01-19T00:00:00.000+0000 | 802.1749877929688 | 802.062240827576  | 6.687186211872174 | Map(vectorType -> dense, length -> 1, values -> List(802.1749877929688)) | 6.746635919438638 | 851.1904673379695 |
| 2017-01-25T00:00:00.000+0000 | 835.6699829101562 | 813.8106898482188 | 6.701727771200435 | Map(vectorType -> dense, length -> 1, values -> List(835.6699829101562)) | 6.773805696480784 | 874.6341603583219 |
| 2017-01-27T00:00:00.000+0000 | 823.3099975585938 | 818.2660021772554 | 6.707187469776992 | Map(vectorType -> dense, length -> 1, values -> List(823.3099975585938)) | 6.763779779322975 | 865.9089628640434 |
| 2017-02-03T00:00:00.000+0000 | 801.489990234375  | 806.0948136663695 | 6.692201370409993 | Map(vectorType -> dense, length -> 1, values -> List(801.489990234375))  | 6.746080277293407 | 850.7176414138683 |
| 2017-02-13T00:00:00.000+0000 | 819.239990234375  | 810.0153913176689 | 6.697053249112708 | Map(vectorType -> dense, length -> 1, values -> List(819.239990234375))  | 6.760478355000563 | 863.0549437094351 |
| 2017-02-17T00:00:00.000+0000 | 828.0700073242188 | 817.5324230881939 | 6.706290563254883 | Map(vectorType -> dense, length -> 1, values -> List(828.0700073242188)) | 6.767640905547893 | 869.2588095977783 |



**MERCI DE VOTRE  
ATTENTION**