

investpy

September 1, 2019

1 investpy - a Python package for financial historical data extraction from Investing

por Álvaro Bartolomé del Canto @ alvarob96 en GitHub

1.1 Introducción

El paquete de Python [investpy](#) ha sido creado por [Álvaro Bartolomé del Canto](#), el cual es la plataforma de extracción de datos principal de su Trabajo de Fin de Grado (TFG) titulado "**Machine Learning for Stock Investment Recommendations**" el cual se basa en el uso de técnicas de Inteligencia Artificial para la predicción del comportamiento futuro de las distintas acciones de las empresas que forman el mercado continuo español, de modo que combinando el Machine Learning con las técnicas de análisis técnico propias del campo de la inversión, se propone una plataforma que identificará señales de compra/venta a partir del análisis y predicción de una serie temporal de valores **OHLC** (*Open High Low Close*).

Por tanto, a lo largo de este informe se describirá el uso principal que se le ha dado a **investpy** para el TCUE titulado "**Sistema de Recomendacion de Inversion en base a Predicciones de Mercado y Analisis de Opinion de Twitter**", cuya plataforma principal de extracción de datos será **investpy**.

investpy recupera la información de [Investing.com](#) y la formatea para poder ser usada por Python, en este caso en particular la información requerida en el proyecto es la que corresponde a las acciones (*equities*) del mercado continuo español.

1.2 Uso

```
In [1]: import investpy
```

```
In [2]: (investpy.__author__, investpy.__version__)
```

```
Out[2]: ('Alvaro Bartolome <alvarob96@usal.es>', '0.8.8')
```

Inicialmente, se puede recuperar el `pandas.DataFrame` que contiene la información sobre las acciones de las empresas indexadas en **investpy**, a partir de la cual se puede realizar el estudio.

Esta función es útil de cara a realizar un estudio completo del mercado, o bien, de cara a nuevos usuarios que desconozcan las empresas que forman del mercado continuo español de forma que este listado les de una idea de qué empresas son; junto con información adicional y/o de uso interno por **investpy**.

```
In [3]: equities = investpy.get_equities()
```

```
In [4]: equities
```

```
Out[4]:
```

	full_name	id	isin \
0	Actividades de Construcción y Servicios S.A.	442	ES0167050915
1	Abengoa S.A.	443	ES0105200416
2	Atresmedia Corp. de Medios de Com. S.A.	444	ES0109427734
3	Acerinox S.A.	445	ES0132105018
4	Banco Bilbao Vizcaya Argentaria S.A.	446	ES0113211835
5	Bankinter S.A.	447	ES0113679I37
6	Bolsas y Mercados Españoles	448	ES0115056139
7	Acciona S.A.	449	ES0125220311
8	CaixaBank S.A.	452	ES0140609019
9	Enagás S.A.	453	ES0130960018
10	Fomento de Construcciones y Contratas S.A.	455	ES0122060314
11	Naturgy Energy Group S.A.	456	ES0116870314
12	Siemens Gamesa Renewable Energy S.A.	457	ES0143416115
13	Grifols S.A.	458	ES0171996087
14	Ferrovial S.A.	459	ES0118900010
15	Red Eléctrica de España S.A.U	460	ES0173093024
16	Industria de Diseño y Textil S.A.	461	ES0148396007
17	Repsol S.A.	462	ES0173516115
18	Iberdrola S.A.	463	ES0144580Y14
19	Indra Sistemas S.A.	465	ES0118594417
20	Mapfre S.A.	466	ES0124244E34
21	Telefónica S.A.	469	ES0178430E18
22	Mediaset España Comunicación S.A.	471	ES0152503035
23	Sacyr SA	472	ES0182870214
24	Banco de Sabadell S.A.	473	ES0113860A34
25	Banco Santander S.A.	474	ES0113900J37
26	Inmobiliaria Colonial S.A.	9060	ES0139140174
27	Técnicas Reunidas S.A.	10526	ES0178165017
28	Obrascon Huarte Lain SA	13808	ES0142090317
29	International Consolidated Airlines Group SA	13809	ES0177542018
..
205	Greenalia	1055914	ES0105293007
206	Prevision Sanitaria Nacional Gestion SOCIMI SA	1057007	ES0170756037
207	Student Properties Spain SOCIMI SA	1057278	ES0105294005
208	Heref Habaneras SOCIMI SA	1061411	ES0105103008
209	Robot	1067156	ES0105225009
210	Tempore Properties Socimi	1073101	ES0105323002
211	Serrano 61 Desarrollo Socimi	1073221	ES0105317004

212	Ap67 Socimi	1076567	ES0105325007
213	Tier1 Technology S.A.	1084842	ES0105344008
214	Torbel Investments 2015 Socimi SA	1087966	ES0105358008
215	Quid Pro Quo Alquiler Seguro Socimi	1088135	ES0105357000
216	Excem Capital Partners Sociedad de Inversión R...	1089239	ES0105367009
217	Castellana Properties Socimi	1089443	ES0105360004
218	Testa Residencial Socimi	1089467	ES0105210019
219	Berkeley Energia Ltd	1093905	AU000000BKYO
220	Alquiber Quality SA	1093991	ES0105366001
221	Go Madrid Benz Socimi SA	1096060	ES0105354007
222	Arima Real Estate SOCIMI SA	1107547	ES0105376000
223	Amrest Holdings SE	1111709	ES0105375002
224	Atom Hoteles Socimi	1112818	ES0105318002
225	Solarpack Corporacion Tecnologica SA	1115133	ES0105385001
226	Arrienda Rental Properties SOCIMI SA	1116351	ES0105383006
227	Vivenio Residencial Socimi SA	1116352	ES0105384004
228	Almagro Capital Socimi	1117970	ES0105389003
229	Inmobiliaria Park Rose Iberoamericana SOCIMI SA	1118084	ES0105391009
230	Euripo Properties Socimi	1122284	ES0105387007
231	Proeduca Altus SA	1123486	ES0105400008
232	Meridia Real Estate III Socimi SA	1123955	ES0105313003
233	Iffe Futura SA	1123956	ES0171613005
234	Compania Espanola de Viviendas en Alquiler SA	1130889	ES0132955008

	name	tag
0	ACS	acs-cons-y-serv
1	Abengoa	abengoa
2	Atresmedia	atresmedia
3	Acerinox	acerinox
4	BBVA	bbva
5	Bankinter	bankinter
6	BME	bolsas-y-mer-esp
7	Acciona	acciona-sa
8	CaixaBank	caixabank-sa
9	Enagás	enagas
10	FCC	fcc
11	Naturgy Energy	gas-natural-sdg
12	Siemens Gamesa	gamesa
13	Grifols	grifols
14	Ferrovial	grupo-ferrovial
15	Red Eléctrica	red-electrica
16	Inditex	inditex
17	Repsol	repsol-ypf
18	Iberdrola	iberdrola
19	Indra	indra-sistemas
20	Mapfre	mapfre
21	Telefónica	telefonica
22	Mediaset	mediaset-esp

23	Sacyr	sacyr-valle
24	Banco Sabadell	bco-de-sabadell
25	Santander	banco-santander
26	Colonial	inmob-colonial
27	Técnicas Reunidas	tecnicas-reunidas
28	OHL	ohl
29	IAG	intl.-cons.-air-grp?cid=13809
..
205	Greenalia	greenalia
206	Prevision Sanitaria	prevision-sanitaria-nacional
207	Student Properties	student-properties
208	Heref Habaneras	heref-habaneras
209	Robot	robot
210	Tempore Properties	tempore-properties-socimi
211	Serrano 61 Desarrollo	serrano-61-desarrollo-socimi
212	Ap67 Socimi	ap67-socimi
213	Tier1	tier-1-technology
214	Torbel Investment	torbel-investments-2015-socimi
215	Quid Pro Quo Alquiler	quid-pro-quo-alquiler-seguro-socimi
216	Excem Socimi	excem-capital-partners-sociedad
217	Castellana Properties	castellana-properties-socimi
218	Testa Residencial	testa-residencial-socimi
219	Berkeley Energy	berkeley-resources-limited?cid=1093905
220	Alquiber Quality	alquiber-quality
221	Go Madrid Benz	go-madrid-benz
222	Arima Real Estate	arima-real-estate
223	Amrest	amrest-holdings-se?cid=1111709
224	Atom Hoteles	atom-hoteles-socimi
225	Solarpack	solarpack-corporacion-tecnologica
226	Arrienda Rental	arrienda-rental
227	Vivenio Residencial	vivenio-residencial
228	Almagro Capital	almagro-capital-socimi
229	Inmobiliaria Park Rose	inmobiliaria-park-rose
230	Euripo Properties	euripo-properties-socimi
231	Proeduca Altus	proeduca-altus
232	Meridia RE III	meridia-re-iii
233	Iffe Futura	iffe-futura
234	Viviendas en Alquiler	viviendas-en-alquiler

[235 rows x 5 columns]

Ahora a modo de ejemplo, se seleccionará una acción aleatoria de las previamente listadas a partir de su nombre, que es la entrada requerida por las funciones de extracción de datos de los valores de mercado de dicha acción tanto recientes como históricos.

```
In [5]: from random import choice
```

```
In [12]: choice(equities['name'].tolist())
```

```
Out[12]: 'Mapfre'
```

De este modo, haremos uso de la siguiente función para recuperar la información reciente de los valores **OHLC**, junto con el **Volumen**, en forma de `pandas.DataFrame` de la acción previamente seleccionada de forma aleatoria.

```
In [14]: params = {
        'equity': 'Mapfre',
        'as_json': False,
        'order': 'ascending',
        'debug': True
    }
```

Referencia: https://investpy.readthedocs.io/main_api.html#investpy.get_recent_data

```
In [17]: df = investpy.get_recent_data(equity=params['equity'],
        as_json=params['as_json'],
        order=params['order'],
        debug=params['debug'])
```

```
INFO:investpy:Searching introduced equity on Investing.com
INFO:investpy:Mapfre found on Investing.com
INFO:investpy:Request sent to Investing.com!
INFO:investpy:Request to Investing.com data succeeded with code 200!
INFO:investpy:Data parsing process starting...
INFO:investpy:Data parsing process finished...
```

```
In [22]: df.head()
```

```
Out[22]:
```

	Close	High	Low	Open	Volume
Date					
2018-01-02	2.708	2.713	2.665	2.688	4240000
2018-01-03	2.702	2.722	2.680	2.718	4139999
2018-01-04	2.751	2.768	2.710	2.710	4960000
2018-01-05	2.810	2.819	2.767	2.767	7660000
2018-01-08	2.793	2.829	2.787	2.824	4670000

Del mismo modo, **investpy** permite recuperar información histórica de una acción de un marco temporal a determinar por el usuario, entre el cual se recuperará la información de **Investing**. Tal y como se ha explicado previamente, la información resultante en este caso será en forma de `pandas.DataFrame`.

```
In [19]: params = {
        'equity': 'Mapfre',
        'from_date': '01/01/2018',
        'to_date': '01/01/2019',
        'as_json': False,
        'order': 'ascending',
        'debug': True
    }
```

Referencia: https://investpy.readthedocs.io/main_api.html#investpy.get_historical_data

```
In [20]: df = investpy.get_historical_data(equity=params['equity'],
                                           from_date=params['from_date'],
                                           to_date=params['to_date'],
                                           as_json=params['as_json'],
                                           order=params['order'],
                                           debug=params['debug'])
```

```
INFO:investpy:Searching introduced equity on Investing.com
INFO:investpy:Mapfre found on Investing.com
INFO:investpy:Request sent to Investing.com!
INFO:investpy:Request to Investing.com data succeeded with code 200!
INFO:investpy:Data parsing process finished...
```

```
In [21]: df.head()
```

```
Out [21]:
```

	Close	High	Low	Open	Volume
Date					
2018-01-02	2.708	2.713	2.665	2.688	4240000
2018-01-03	2.702	2.722	2.680	2.718	4139999
2018-01-04	2.751	2.768	2.710	2.710	4960000
2018-01-05	2.810	2.819	2.767	2.767	7660000
2018-01-08	2.793	2.829	2.787	2.824	4670000

A través de **investpy** también será posible recuperar información adicional de cada una de las acciones de las empresas que forman el mercado continuo español, como lo es el perfil de empresa, que contiene una descripción detallada creada por la propia empresa sobre la acción de la misma y el ámbito al que se dedica.

Esto aporta un punto de vista sobre la empresa importante, sobretodo de cara a gente que no la conozca, ya que es la propia empresa la que escribe la información del perfil de empresa.

Referencia: https://investpy.readthedocs.io/main_api.html#investpy.get_equity_company_profile

```
In [23]: profile = investpy.get_equity_company_profile(equity='Mapfre',
                                                       language='spanish')
```

```
In [25]: from pprint import pprint
```

```
In [26]: pprint(profile)
```

```
{'desc': 'MAPFRE es una aseguradora global con presencia en los cinco '
          'continentes. Compañía de referencia en el mercado asegurador '
          'español, es el primer grupo asegurador multinacional en América '
          'Latina y uno de los 10 mayores grupos europeos por volumen de '
          'primas.\n'
          'Además, se sitúa entre las 20 primeras compañías de seguros de '
          'automóviles en Estados Unidos. En la actividad de Asistencia, es la '
          'tercera aseguradora mundial.\xa0\xa0\n'}
```

```
'MAPFRE cuenta con más de 38.000 empleados y con aproximadamente 34 '
'millones de clientes en todo el mundo.\n'
'En 2015, sus ingresos superaron los 26.700 millones de euros y el '
'beneficio neto se situó en 709 millones.\n'
'\xa0\n'
'\xa0',
'url': 'http://www.bolsamadrid.es/esp/aspx/Empresas/FichaValor.aspx?ISIN=ES0124244E34'}
```

1.3 Visualización

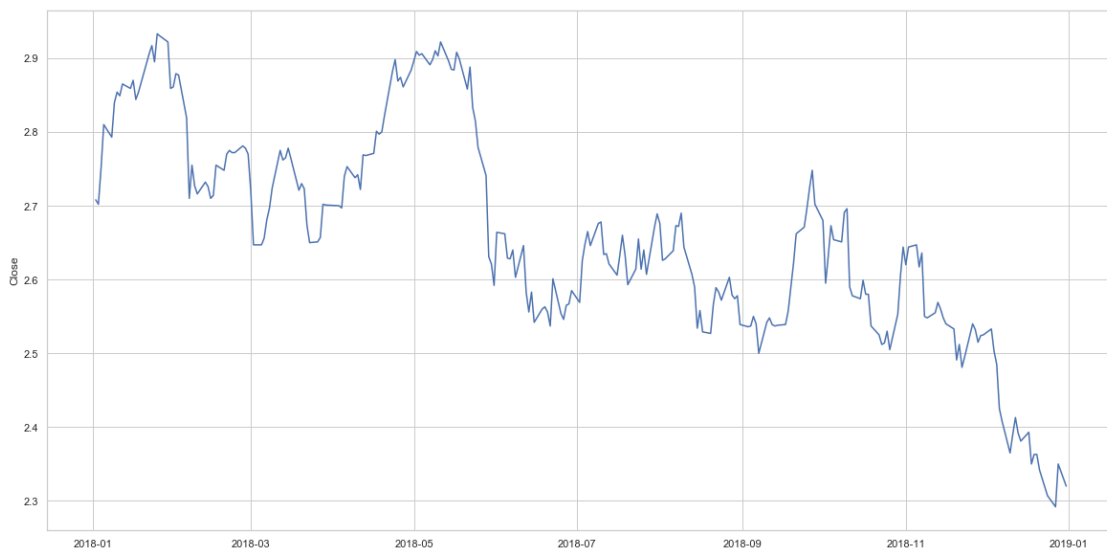
Por último, en cuanto a la visualización de los datos, de forma complementaria se hará uso de las bibliotecas de Python `matplotlib` y `seaborn`. Por tanto, se presenta la representación de las series temporales previamente extraídas de Investing a través de `investpy` de la forma:

```
In [29]: import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [44]: %matplotlib inline
sns.set(style='whitegrid')
```

```
In [48]: plt.figure(figsize=(20, 10))
```

```
ax = sns.lineplot(x=df.index.values, y='Close', data=df)
```



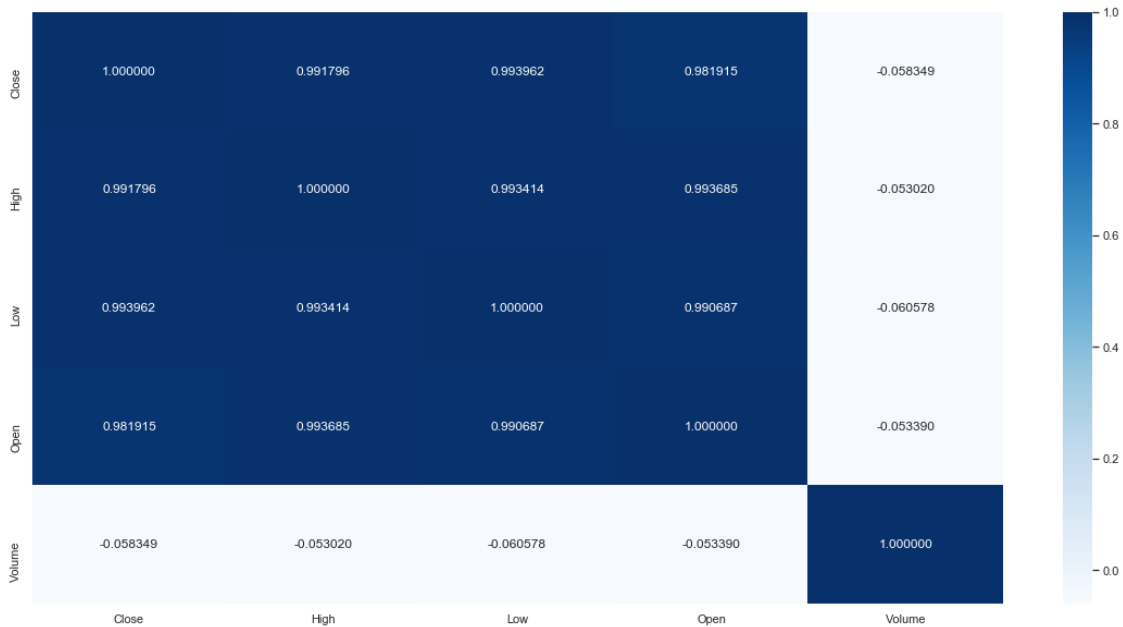
También se presentará la matriz de correlación entre las distintas variables del conjunto de datos, es decir, entre las columnas del pandas.DataFrame que son: `['Open', 'High', 'Low', 'Close', 'Volume']`

```
In [47]: plt.figure(figsize=(20, 10))
```

```
corr = df.corr()
```

```
sns.heatmap(corr, xticklabels=corr.columns, yticklabels=corr.columns, cmap='Blues', an
```

```
Out[47]: <matplotlib.axes._subplots.AxesSubplot at 0x1a1e525f98>
```



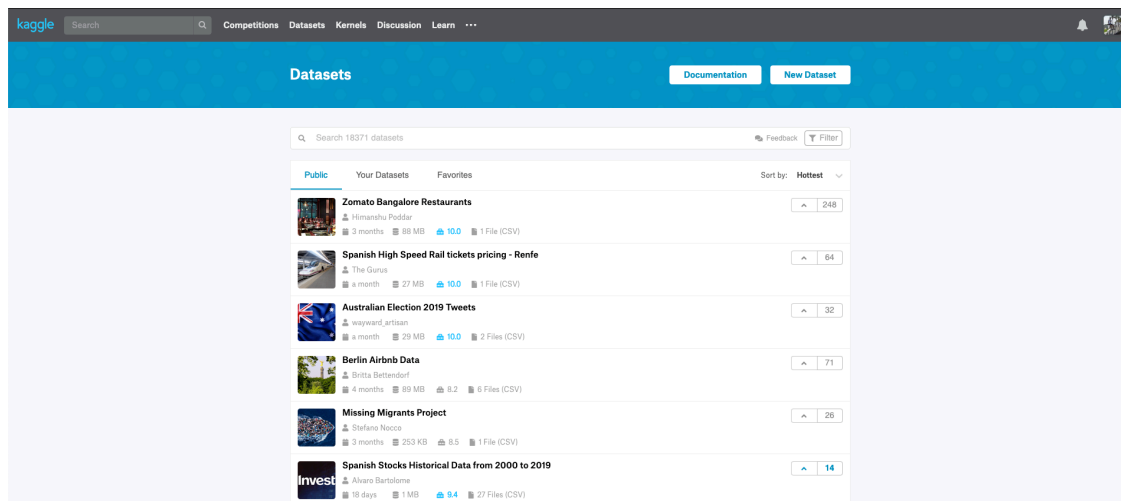
1.4 Información Adicional

Además, en [Kaggle](#) se ha subido un dataset que contiene los valores históricos de las principales acciones del mercado continuo español, a partir del cual varios usuarios han creado sus propios **Kernels** sobre el manejo y/o uso de dicha información para distintos propósitos/objetivos.

En base a dicho [dataset](#) subido a **Kaggle**, destacar que durante los meses de Mayo/Junio de 2019 se posicionó en el **Top 10 Hottests Datasets**, alcanzando la posición 6 del mundo en dicho ranking.

```
In [50]: from IPython.display import Image
         Image(filename='resources/top6-dataset.png')
```

```
Out[50]:
```

Finalmente, el dataset alcanzó 42 votos.

1.5 Referencias

Las referencias utilizadas a lo largo de este informe se pueden encontrar en la documentación creada y generada utilizando [sphinx](#), la cual está alojada en [Readthedocs](#).

1.6 Licencia

[MIT License](#) MIT License

Copyright (c) 2018-2019 Álvaro Bartolomé

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

1.7 Autor

Álvaro Bartolomé del Canto, 2018-2019