

investpy

September 18, 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

```
[1]: import investpy
[2]: (investpy.__author__, investpy.__version__)
[2]: ('Alvaro Bartolome <alvarob96@usal.es>', '0.9.2')
```

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**.

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

```
[4]: equities.head()
```

```
[4]:
```

	country	name	full_name	\
0	argentina	Tenaris	Tenaris	
1	argentina	PETROBRAS ON	Petroleo Brasileiro - Petrobras	
2	argentina	GP Fin Galicia	Grupo Financiero Galicia B	
3	argentina	Ternium Argentina	Ternium Argentina Sociedad Anónima	
4	argentina	Pampa Energía	Pampa Energía S.A.	

	tag	isin	id	currency	symbol
0	tenaris?cid=13302	LU0156801721	13302	ARS	TS
1	petrobras-on?cid=13303	BRPETRACNOR9	13303	ARS	APBR
2	gp-fin-galicia	ARP495251018	13304	ARS	GGAL
3	siderar	ARSIDE010029	13305	ARS	TXAR
4	pampa-energia	ARP432631215	13306	ARS	PAMP

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.

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

```
[6]: choice(equities[equities['country'] == 'spain']['name'].tolist())
```

```
[6]: 'Adolfo Domínguez'
```

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.

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

```
[7]: df = investpy.get_recent_data(equity='mapfre',
                                   country='spain')
```

```
[8]: df.head()
```

```
[8]:
```

	Open	High	Low	Close	Volume	Currency
Date						
2019-08-19	2.373	2.394	2.365	2.383	2580000	EUR
2019-08-20	2.380	2.380	2.352	2.361	2750000	EUR
2019-08-21	2.363	2.385	2.362	2.378	2049999	EUR
2019-08-22	2.362	2.394	2.360	2.394	4310000	EUR
2019-08-23	2.400	2.405	2.379	2.389	4360000	EUR

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`.

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

```
[9]: df = investpy.get_historical_data(equity='mapfre',
                                     country='spain',
                                     from_date='01/01/2018',
                                     to_date='01/01/2019')
```

```
[10]: df.head()
```

```
[10]:
```

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

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

```
[11]: profile = investpy.get_equity_company_profile(equity='mapfre',
                                                    country='spain',
                                                    language='spanish')
```

```
[12]: from pprint import pprint
```

```
[13]: pprint(profile)
```

```
('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.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. MAPFRE cuenta con más de '
 '38.000 empleados y con aproximadamente 34 millones de clientes en todo el '
 'mundo.En 2015, sus ingresos superaron los 26.700 millones de euros y el '
 'beneficio neto se situó en 709 millones.')
```

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:

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

[15]: %matplotlib inline
sns.set(style='whitegrid')

[16]: plt.figure(figsize=(20, 10))

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

c:\users\usal\appdata\local\programs\python\python37\lib\site-packages\pandas\plotting_matplotlib\converter.py:103: FutureWarning: Using an implicitly registered datetime converter for a matplotlib plotting method. The converter was registered by pandas on import. Future versions of pandas will require you to explicitly register matplotlib converters.

To register the converters:

```
>>> from pandas.plotting import register_matplotlib_converters
>>> register_matplotlib_converters()
warnings.warn(msg, FutureWarning)
```



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']

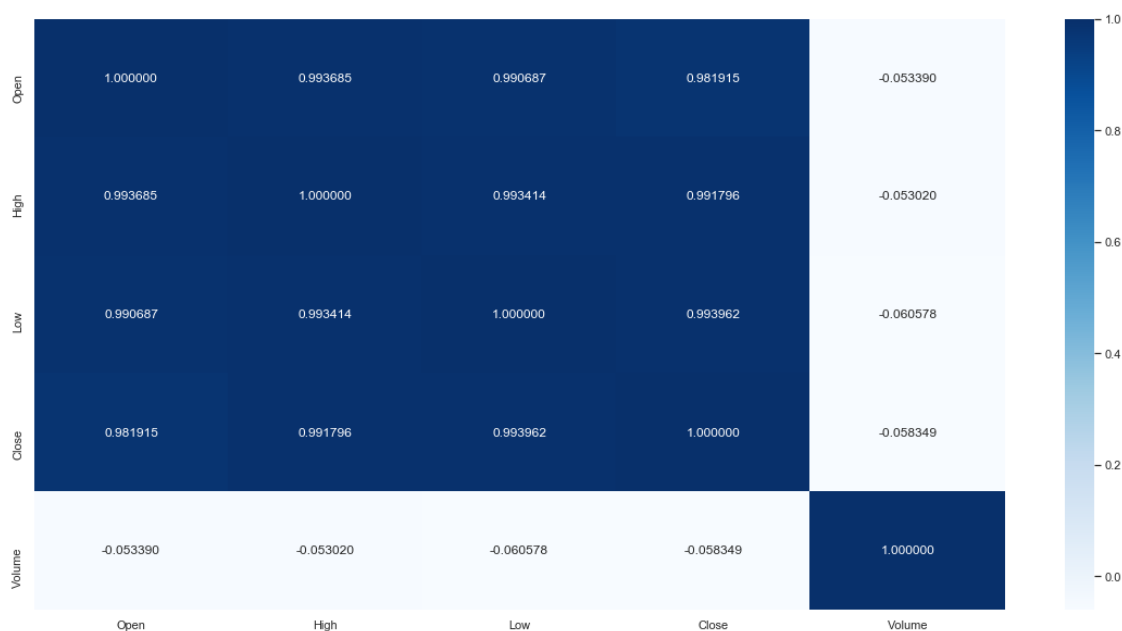
```
[17]: plt.figure(figsize=(20, 10))

corr = df.corr()

sns.heatmap(corr, xticklabels=corr.columns, yticklabels=corr.columns,
            cmap='Blues', annot=True, fmt='f')
```

INFO:numexpr.utils:NumExpr defaulting to 6 threads.

[17]: <matplotlib.axes._subplots.AxesSubplot at 0x21bda810b38>



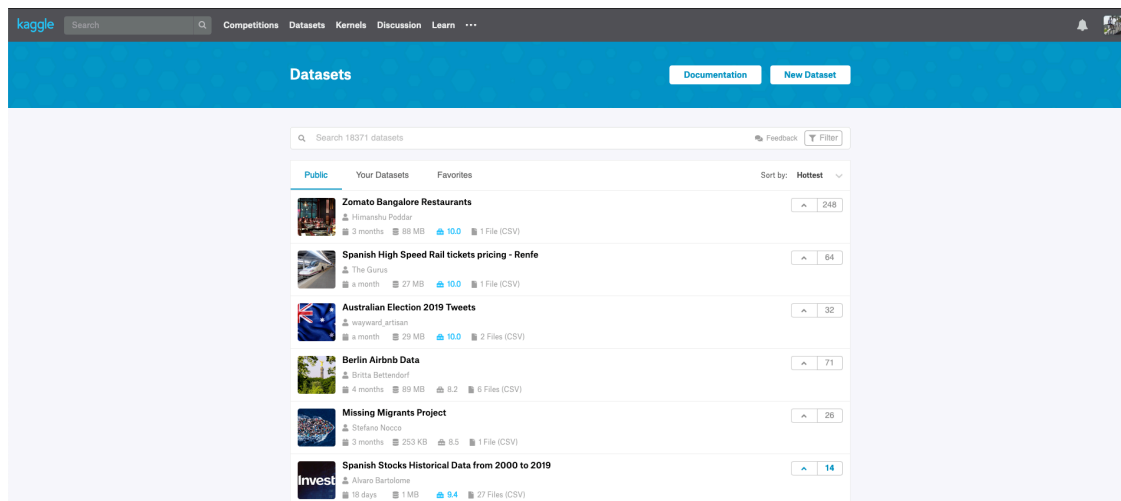
1.4 Información Adicional

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.

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

[18]:



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