

Bruno Emilio Sánchez Rey

Contexto

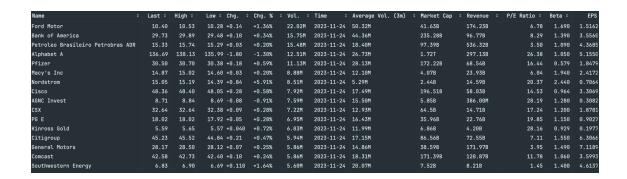
Como equipo de científicos de datos debemos encontrar un top 10 empresas que cumplan los siguientes requisitos:

- Que sea de creación posterior al año 2000
- Que haya pagado dividendos alguna vez en los últimos años
- Que, actualmente, tenga un PER (Price To Earnings Ratio) inferior a 30
- Que, actualmente, tenga beneficio por acción (EPS, Earnings Per Share) positivo.

Realización del proyecto

Para comenzar el proyecto lo primero que se ha realizado es buscar en diferentes webs de acciones los datos que mejor nos venían. Es por ello por lo que en este caso decidí utilizar dos: marketbeat.com e investing.com. La idea de usar dos fuentes es que en caso de tener muchos casos para nuestro top 10, podríamos ver cuales aparecen en ambas y así ver cuáles son más fiables para el top.

Primero scrapeamos investing.com con selenium para limpiarnos las coockies y luego con pandas obtener la tabla que deseamos. Tenemos que usar selenium pues debemos ir seleccionando los botones que en concreto queremos. Después de scrapear todo nos quedará el siguiente DF:



Como podemos observar nos queda un DF con las siguientes columnas: Name, precio de las acciones (last, high, low, change), cantidad de acciones, última vez que estuvo activo, el valor total de las acciones, las ganancias de la empresa y luego 3 datos económicos de los cuales dos usaremos más adelante para obtener nuestro top 10.

Ahora que ya tenemos nuestro primer dataset, scrapeamos la segunda web:

MarketBeat.com. En este caso debemos seleccionar algún filtro para obtener el dataset
que queremos por lo que con selenium realizaremos todos los pasos pertinentes.

Una vez tengamos ya la tabla deseada simplemente la guardamos obteniendo el siguiente DF:

^ Name :	Sector	Current Price	Market Cap	Dividend Payment Schedule	Consensus Price Target	Upside/Downside	Consensus Rating	≎ Current Year EPS Es ≎	P/E Ratio (Trailing) ‡
8 AAPLApple	Computer and Technology			Quarterly	\$198.25		Moderate Buy 2.77	\$6.56	31.84
1 MSFTMicrosoft	Computer and Technology			Quarterly	\$388.84		Moderate Buy 2.89		36.54
5 NVDANVIDIA	Computer and Technology				\$588.38		Moderate Buy 2.92		63.03
10 LLYEli Lilly and Company	Medical	\$601.10	\$570.638				Moderate Buy 2.81		108.89
11 TSMTaiwan Semiconductor Manufacturing	Computer and Technology			Quarterly					17.60
12 UNHUnitedHealth Group	Medical		\$506.038	Quarterly	\$576.98		Moderate Buy 2.94		23.75
							Moderate Buy 2.72		30.68
14 NVONovo Nordisk A/S	Medical						Moderate Buy 2.75		43.76
15 JPMJPMorgan Chase & Co.				Quarterly	\$168.80		Moderate Buy 2.56	\$16.62	9.17
17 WMTWalmart									25.88
18 XOMExxon Mobil	Oils/Energy	\$184.57	\$414.488	Quarterly	\$139.26	24.57%	Moderate Buy 2.65	\$9.32	10.38
19 AVGOBroadcom	Computer and Technology	\$978.87			\$858.05		Moderate Buy 2.74		30.10
22 MAMastercard			\$386.838	Quarterly	\$450.09		Moderate Buy 2.98		35.93
	Medical			Quarterly					11.32
24 PGProcter & Gamble							Moderate Buy 2.67		24.57
28 ORCLOracle	Computer and Technology						Moderate Buy 2.58		34.60
30 HDHome Depot	Retail/Wholesale			Quarterly			Moderate Buy 2.58		19.94
34 CVXChevron	Oils/Energy			Quarterly			Moderate Buy 2.82		10.76
37 COSTCostco Wholesale	Retail/Wholesale	\$591.36	\$261.828	Quarterly	\$586.43		Moderate Buy 2.77		41.76
38 MRKMerck & Co., Inc.	Medical	\$101.87					Moderate Buy 2.76		56.59
40 KOCoca-Cola	Consumer Staples						Moderate Buy 2.73	\$2.68	23.52
42 ABBVAbbVie	Medical			Quarterly			Moderate Buy 2.50		37.99
43 BACBank of America	Finance	\$29.73	\$235,288	Quarterly	\$35.54		Hold 2.47	\$3.43	8.33

Como podemos observar este dataset tiene nuevas columnas muy interesantes como los pagos por dividendos y otra información económica que nos ayudará a tener una mejor conclusión. También no damos cuenta de que las dos tablas tienen diferentes nombres de empresas (el segundo DF tiene distintos nombres que el primero pues su 'Name'. Es una clave y nombre) por lo que almacenamos todas las claves en una lista y las borramos.

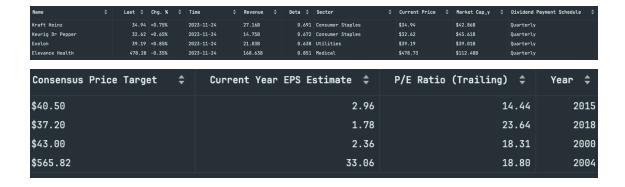
Ahora que tenemos la lista con las claves, utilizamos BeautifulSoup para scrapear las fechas de creación de cada una de las empresas:

```
siglas = []
nombres = []
for nombre in MB['Name']:
    primera_minuscula = next((i for i, c in enumerate(nombre) if c.islower()), None)
    if primera_minuscula is not None:
    # Almacena los caracteres hasta dos posiciones antes de la primera letra minúscula
    siglas.append(nombre):max(0, primera_minuscula - 1)])
    nombres.append(nombre)
```

Con este código obtendremos muchas de las fechas de creación de las empres as las cuales añadiremos junto a sus respectivas filas en una columna 'Year'.

Limpiamos bien los DF y finalmente los unimos por un inner join quedándonos un pequeño dataset de 88 empresas que cumplen todas las características menos los años.

Al fijarnos nos damos cuenta de que únicamente 4 empresas fueron creadas después del año 2000 por lo que tendríamos el top 4 del top 10 empresas.



Para las otras 6 volveremos un paso atrás donde obtendremos del DF de MarketBeat un total de 30 empresas. Este DF lo guardamos en SQL con el siguiente código:

```
#GUardamos todo en una base de datos SQL

engine = create_engine('postgresql://admin:password@localhost/top10_USA')

Executed at 2023.11.26 23:30:33 in 2s 815ms

MB.to_sql('marketbeat', con=engine, index= False ,if_exists='replace')

Frequent at 2023.11.26 23:21.20 is 212ms
```

Junto a Docker creamos y runeamos una imagen la cual nos ayudará a poder crear la base de datos. Ahora simplemente seleccionamos las mejores 10 empresas las cuales serán aquellas con un mayor EPS. Esta conclusión sale de que si lo que estamos buscando como inversor es obtener la mayor cantidad de dinero posible, nos centraremos en la estimación de ganancias por acción de este año por empresa viendo que el top 10 será:

Name	‡	Current Year EPS Estimate 💠
Kraft Heinz		2.96
Keurig Dr Pepper		1.78
Exelon		2.36
Elevance Health		33.06

ConocoPhillips	9.09
LyondellBasell Industries	8.71
Digital Realty Trust	6.61
Quanta Services	6.45
Philip Morris International	6.16
Skyworks Solutions	5.98

Donde ordenado será algo como:

- 1. Elevance Health
- 2. ConocoPhillips
- 3. LyondellBasell Industries
- 4. Digital Realty Trust
- 5. Quanta Services
- 6. Philip Morris International
- 7. Skyworks Solutions
- 8. Kraft Heinz
- 9. Exelon
- 10. Keurig Dr Pepper

Es decir de las 1000 empresas que teníamos al principio, únicamente 30 cumplían los requisitos necesarios.

Ahora hablando de indicadores fundamentales tenemos:

- 1. P/E Ratio (Trailing): Relación Precio/Ganancias. Mide la relación entre el precio actual de las acciones y las ganancias por acción del último año.
- 2. Current Price: El precio actual de la acción en el mercado.
- 3. Market Cap: Capitalización de mercado. Es el valor total de mercado de todas las acciones de una empresa.
- 4. Dividend Payment Schedule: Programa de pago de dividendos. Indica la frecuencia con la que la empresa paga dividendos (por ejemplo, trimestralmente, anualmente).
- 5. Current Year EPS Estimate: Estimación de las ganancias por acción para el año en curso.

Indicadores técnicos:

 Consensus Price Target: Precio objetivo consensuado. Es la estimación promedio del precio al que se espera que una acción alcance en el futuro, según analistas financieros.