

ACTIVIDAD DE APRENDIZAJE 1

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 cookies 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:

Name	Last	High	Low	Chg.	Chg. %	Vol.	Time	Average Vol. (3m)	Market Cap	Revenue	P/E Ratio	Beta	EPS
Ford Motor	10.40	10.53	10.28	+0.14	+1.36%	22.02M	2023-11-24	50.32M	41.63B	174.23B	6.78	1.690	1.5162
Bank of America	29.73	29.89	29.48	+0.10	+0.34%	15.75M	2023-11-24	44.36M	235.28B	96.77B	8.29	1.390	3.5566
Petroleo Brasileiro Petrobras ADR	15.33	15.74	15.29	+0.03	+0.20%	15.48M	2023-11-24	18.40M	97.39B	536.32B	3.50	1.090	4.3685
Alphabet A	136.69	138.13	135.99	-1.80	-1.30%	12.51M	2023-11-24	26.73M	1.72T	297.13B	26.38	1.050	5.1558
Pfizer	30.50	30.70	30.38	+0.18	+0.59%	11.13M	2023-11-24	28.13M	172.22B	68.54B	16.44	0.579	1.8479
Macy's Inc	14.87	15.02	14.60	+0.03	+0.20%	8.88M	2023-11-24	12.10M	4.07B	23.93B	6.04	1.940	2.4172
Nordstrom	15.05	15.19	14.39	+0.84	+5.91%	8.51M	2023-11-24	5.29M	2.44B	14.59B	20.37	2.440	0.7064
Cisco	48.36	48.40	48.05	+0.28	+0.58%	7.92M	2023-11-24	17.49M	196.51B	58.03B	14.53	0.964	3.3069
AGNC Invest	8.71	8.84	8.69	-0.08	-0.91%	7.59M	2023-11-24	15.50M	5.85B	386.00M	28.19	1.280	0.3082
CSX	32.64	32.64	32.38	+0.09	+0.28%	7.22M	2023-11-24	12.93M	64.58	14.71B	17.24	1.200	1.8781
PB E	18.02	18.02	17.92	+0.05	+0.28%	6.95M	2023-11-24	16.43M	35.96B	22.76B	19.85	1.150	0.9027
Kinross Gold	5.59	5.65	5.57	+0.040	+0.72%	6.03M	2023-11-24	11.99M	6.86B	4.20B	28.16	0.929	0.1977
Citigroup	45.23	45.52	44.84	+0.21	+0.47%	5.94M	2023-11-24	17.15M	86.56B	72.55B	7.11	1.550	6.3066
General Motors	28.17	28.50	28.12	+0.07	+0.25%	5.86M	2023-11-24	14.86M	38.59B	171.97B	3.95	1.490	7.1189
Comcast	42.58	42.73	42.40	+0.10	+0.24%	5.86M	2023-11-24	18.31M	171.39B	120.87B	11.78	1.060	3.5993
Southwestern Energy	6.83	6.90	6.69	+0.110	+1.64%	5.60M	2023-11-24	20.07M	7.52B	8.21B	1.45	1.400	4.6137

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	ASAPStaps	Computer and Technology	\$189.97	\$2.95T	Quarterly	\$188.25	4.51%	Moderate Buy 2.77	\$1.56	31.84
9	MSFTMicrosoft	Computer and Technology	\$277.43	\$2.81T	Quarterly	\$288.84	3.02%	Moderate Buy 2.89	\$11.13	36.54
5	WDAVWDZIA	Computer and Technology	\$477.76	\$1.18T	Quarterly	\$588.38	23.15%	Moderate Buy 2.92	\$9.71	43.83
10	LLVLLi Lilly and Company	Medical	\$601.10	\$579.63B	Quarterly	\$557.08	-7.34%	Moderate Buy 2.81	\$6.42	108.89
11	TSMTaiwan Semiconductor Manufacturing	Computer and Technology	\$97.83	\$507.39B	Quarterly	\$116.67	19.25%	Buy 3.00	\$6.97	17.48
12	UNHUnitedHealth Group	Medical	\$247.18	\$266.10B	Quarterly	\$276.99	5.45%	Moderate Buy 2.94	\$24.94	23.75
13	Veeva	Business Services	\$254.38	\$472.47B	Quarterly	\$272.67	8.40%	Moderate Buy 2.72	\$9.90	39.48
14	WYNovo Nordisk A/S	Medical	\$105.45	\$473.21B	Annually	\$46.85	-42.31%	Moderate Buy 2.75	\$2.42	43.76
15	JPMJPMorgan Chase & Co.	Finance	\$153.54	\$643.89B	Quarterly	\$168.80	9.94%	Moderate Buy 2.56	\$16.42	9.17
17	WMTWalmart	Retail/Wholesale	\$156.06	\$429.85B	Quarterly	\$179.08	14.70%	Moderate Buy 2.87	\$6.45	25.88
18	XOMExxon Mobil	Oil/Energy	\$184.57	\$434.48B	Quarterly	\$139.26	-24.57%	Moderate Buy 2.65	\$9.32	19.38
19	AMZNAmazon	Computer and Technology	\$978.87	\$404.85B	Quarterly	\$858.16	-12.34%	Moderate Buy 2.76	\$37.83	38.18
22	MRMustard	Business Services	\$412.58	\$386.83B	Quarterly	\$458.09	9.11%	Moderate Buy 2.98	\$12.16	35.93
23	JNJJohnson & Johnson	Medical	\$152.58	\$367.11B	Quarterly	\$149.19	-10.94%	Hold 2.43	\$10.10	11.32
24	PDPProcter & Gamble	Consumer Staples	\$151.38	\$356.79B	Quarterly	\$166.41	9.91%	Moderate Buy 2.67	\$6.42	24.57
28	ORCLOracle	Computer and Technology	\$116.25	\$318.45B	Quarterly	\$124.23	6.87%	Moderate Buy 2.58	\$6.48	24.48
38	HDHome Depot	Retail/Wholesale	\$158.78	\$309.23B	Quarterly	\$139.57	-9.29%	Moderate Buy 2.58	\$15.85	19.74
34	CEXChevron	Oil/Energy	\$144.90	\$275.53B	Quarterly	\$187.89	29.47%	Moderate Buy 2.82	\$11.59	18.76
37	COSTCostco Wholesale	Retail/Wholesale	\$291.36	\$262.82B	Quarterly	\$286.43	-1.63%	Moderate Buy 2.77	\$15.72	43.76
39	MONMerck & Co., Inc.	Medical	\$101.87	\$258.14B	Quarterly	\$125.14	22.85%	Moderate Buy 2.76	\$1.37	56.59
48	KSCoca-Cola	Consumer Staples	\$58.57	\$253.22B	Quarterly	\$66.20	13.02%	Moderate Buy 2.73	\$2.48	23.52
42	ABBVAbbVie	Medical	\$138.47	\$244.83B	Quarterly	\$167.69	20.93%	Moderate Buy 2.58	\$11.26	37.99
43	BACBank of America	Finance	\$29.73	\$235.28B	Quarterly	\$35.54	19.54%	Hold 2.47	\$1.43	8.31

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)
```

```

creacion = []
for sigla in siglas:
    try:
        # Crea la URL dinámica para la empresa actual
        url_empresa = f'https://www.marketbeat.com/stocks/NASDAQ/{sigla}'

        # Realiza la solicitud HTTP a la página de la empresa
        response = requests.get(url_empresa)
        response.raise_for_status() # Verifica si la solicitud fue exitosa

        # Crea un objeto BeautifulSoup para parsear el contenido HTML de la página
        soup = BeautifulSoup(response.text, 'html.parser')

        # Encuentra la fecha de creación utilizando el XPath proporcionado
        fecha_creacion_elemento = soup.select_one('div#tabContentCompanyProfile div:nth-of-type(4) div:nth-of-type(2) dl:nth-of-type(2) div:nth-of-type(12) dd strong')

        fecha_creacion = fecha_creacion_elemento.text if fecha_creacion_elemento else "No disponible"

        # Imprime la información o realiza cualquier otra acción que desees
        print(f"Fecha de creación de {sigla}: {fecha_creacion}")
        creacion.append(fecha_creacion)

    except Exception as e:
        print(f"No se pudo obtener la fecha de creación para {sigla} - Error: {e}")
        creacion.append("No disponible")

```

Con este código obtendremos muchas de las fechas de creación de las empresas 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.

Name	Last	Chg. %	Time	Revenue	Beta	Sector	Current Price	Market Cap_y	Dividend Payment Schedule
Kraft Heinz	34.94	+0.75%	2023-11-24	27.168	0.691	Consumer Staples	\$34.94	\$42.86B	Quarterly
Kaurig Dr Pepper	32.62	+0.65%	2023-11-24	14.758	0.472	Consumer Staples	\$32.62	\$45.61B	Quarterly
Exelon	39.19	+0.85%	2023-11-24	21.038	0.638	Utilities	\$39.19	\$39.01B	Quarterly
Elevance Health	478.28	-0.35%	2023-11-24	168.658	0.851	Medical	\$478.73	\$112.48B	Quarterly

Consensus Price Target	Current Year EPS Estimate	P/E Ratio (Trailing)	Year
\$40.50	2.96	14.44	2015
\$37.20	1.78	23.64	2018
\$43.00	2.36	18.31	2000
\$565.82	33.06	18.80	2004

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')
Executed at 2023.11.26 23:31:22 in 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:

1. 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.