

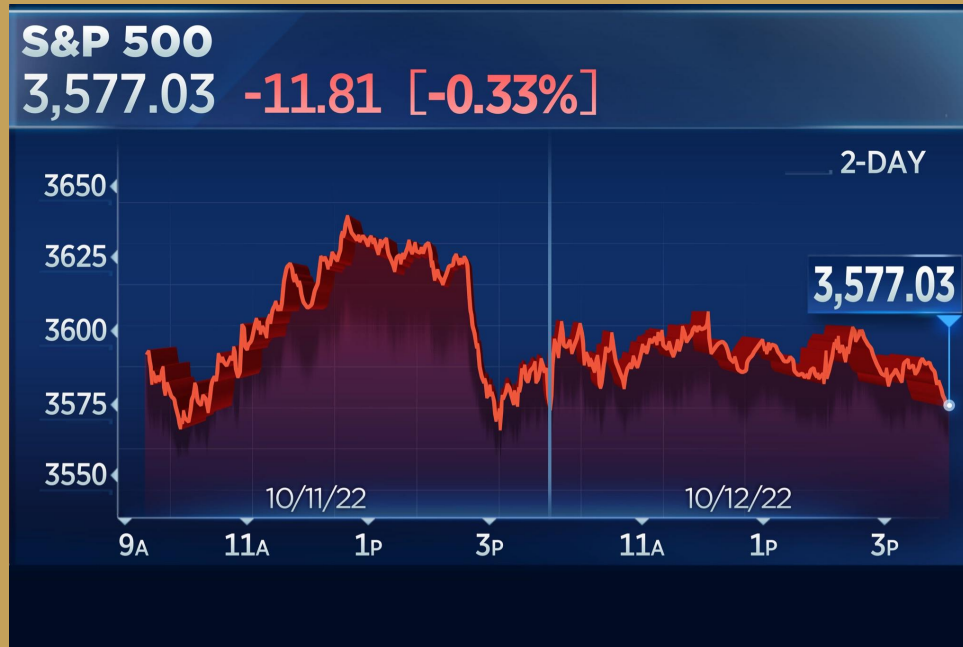
Cluster S&P500 según fundamentos financieros

Natalia De La Barra
Nicolás Estévez
Juan Ignacio García
Francisco Javier Hortal

Índice

1. Introducción a la problemática abordada
2. Análisis del dataset a trabajar y exploración inicial
3. Como se abordará la temática
4. Tratamiento/preprocesamiento de datos
5. Procedimiento experimental
6. Conclusiones
7. Propuesta de Trabajo pendiente

1. Introducción





80%

de los inversionistas pierden dinero en la bolsa

(The Annuity Expert, 2022)







2. Nuestro Dataset



Primer dataset

sp500_companies

16 columnas

6 columnas cuantitativas

10 columnas cualitativas

495 filas

Segundo dataset

5yr_fundamentals_S&P500

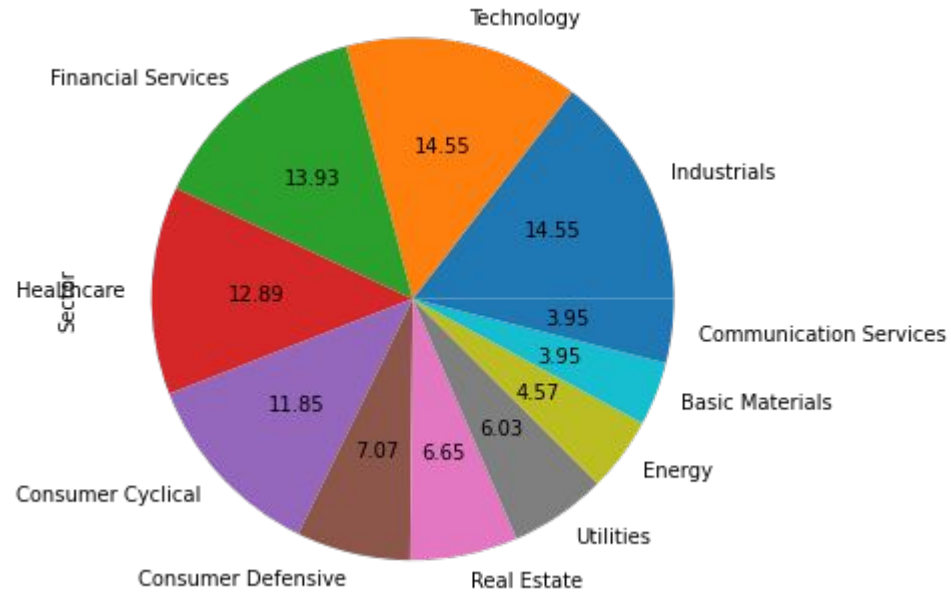
9 columnas

8 columnas cuantitativas

1 columnas cualitativas

2441 filas

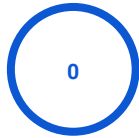
Distribución según Sector





3. Como se abordará la temática

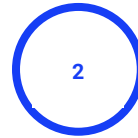




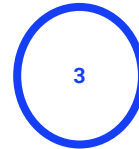
**Análisis de los datos de
ambos dataset**



**Preprocesamiento de los
datos**



**Reducción de
dimensionalidad y
graficar**



**Clustering mediante
Kmeans, Mini batch,
Dbscan y GMM**

Fundamentación

PCA:

- Condensar información
- Seleccionar información más relevante
- Mejor desempeño al clusterizar

K-Means y MiniBatch:

- Ampliamente utilizados
- Encontrar grupos ocultos
- Confirmar o descartar teorías

DBSCAN:

- Clasificar grupos de distinto tamaño
- No requiere número de cluster previamente
- Resistente al ruido

GMM:

- Detectar clusters con formas complejas
- Nuevas etiquetas de clasificación



4. Tratamiento/preprocesamiento datos



8 pilares

$$\text{5 yr P/E ratio: } \frac{\text{Market Cap}}{\text{last 5 years net income}}$$

$$\text{5 yr revenue growth: } \frac{\text{revenue now} - \text{revenue 5 years ago}}{\text{revenue 5 years ago}}$$

$$\text{Shares Outstanding (S.O): } \frac{\text{S.O now} - \text{S.O 5 years ago}}{\text{S.O 5 years ago}}$$

$$\text{Free Cash Flow (F.C.F) Growth: } \frac{\text{F.C.F now} - \text{F.C.F 5 years ago}}{\text{F.C.F 5 years ago}}$$

$$\text{5 yr ROIC: } \frac{\text{5 year cashflow}}{\text{total debt and equity}}$$

$$\text{Net income growth: } \frac{\text{net income now} - \text{net income 5 years ago}}{\text{net income 5 years ago}}$$

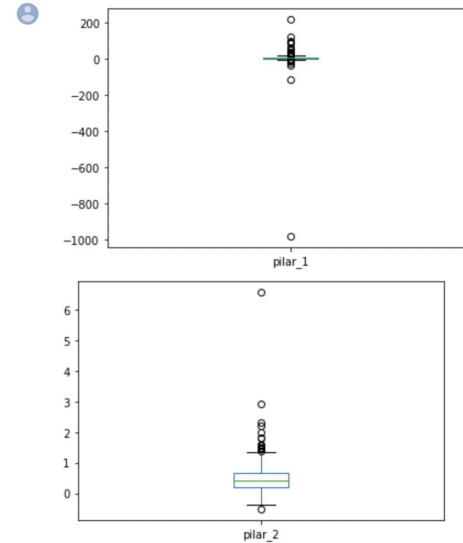
$$\text{Long Term Liabilities (L.T.L): } \frac{\text{L.T.L}}{\text{5 years free cash flow}}$$

$$\text{Price to F.C.F: } \frac{\text{Market cap}}{\text{5 years F.C.F}}$$

Eliminar outliers:

3% de los datos (15 acciones)

```
for i in columns:  
    plot_boxplot(df, i)
```

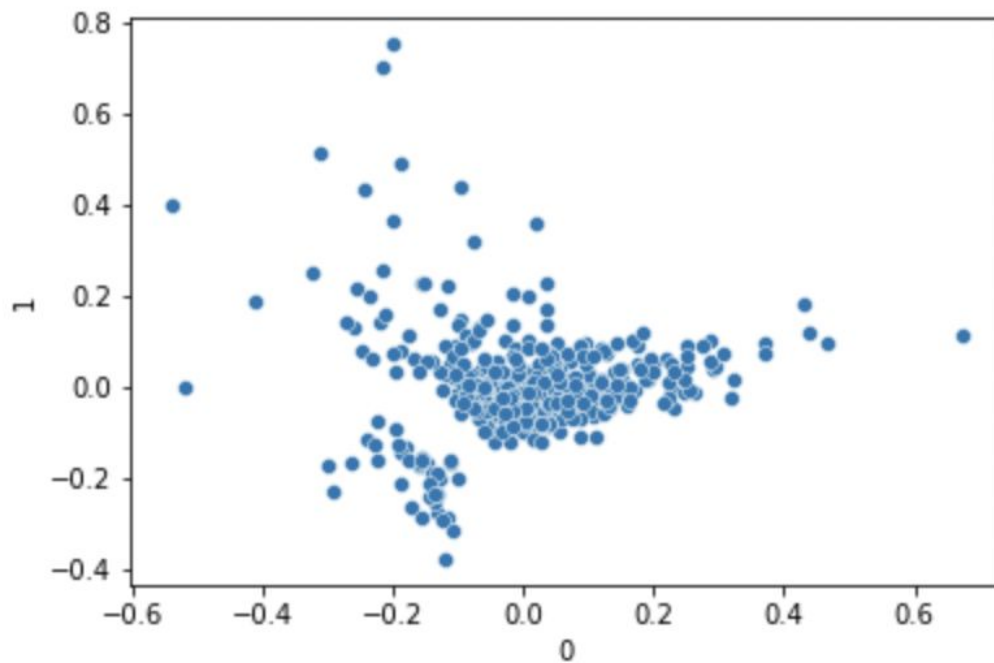




5. Procedimiento Experimental



PCA:



Explicación Varianza PCA

Ratio de la varianza explicada:

55%

en los primeros 2 componentes

Se decide reducir la cantidad de variables y utilizar las con mayor varianza y una combinación que incluya todos los factores financieros.

Análisis Variables

	pilar_1	pilar_2	pilar_3	pilar_4	pilar_5	pilar_6	pilar_7	pilar_8
count	466.000000	466.000000	466.000000	466.000000	466.000000	466.000000	466.000000	466.000000
mean	0.285878	0.291536	0.189362	0.494029	0.186275	0.329866	0.40678	0.405518
std	0.070877	0.121920	0.089354	0.055401	0.093332	0.076178	0.05479	0.077681
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000
50%	0.274452	0.269393	0.169457	0.489827	0.170002	0.327991	0.40408	0.401750
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.00000	1.000000

Se deciden utilizar los pilares 2, 3, 5 y 8

Explicación Varianza PCA con 4 variables

Ratio de la varianza explicada:

75%

en los primeros 2 componentes

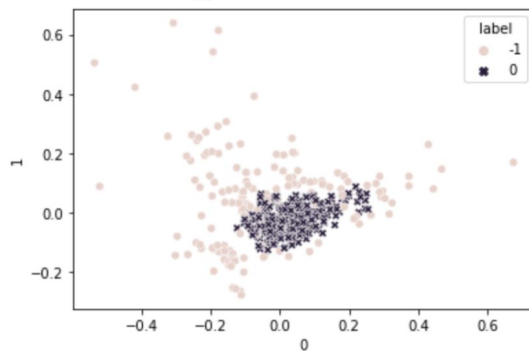
Pudiendo mantener las variables financieras del modelo.

Análisis de sensibilidad

```
Silhouette score for k(clusters) = 2 is 0.2865625875019962  
Silhouette score for k(clusters) = 3 is 0.34530552426107236  
Silhouette score for k(clusters) = 4 is 0.40532303243594053  
Silhouette score for k(clusters) = 5 is 0.4124128468082958  
Silhouette score for k(clusters) = 6 is 0.29490622259130905  
Silhouette score for k(clusters) = 7 is 0.286903921684372  
Silhouette score for k(clusters) = 8 is 0.3010495297861177
```

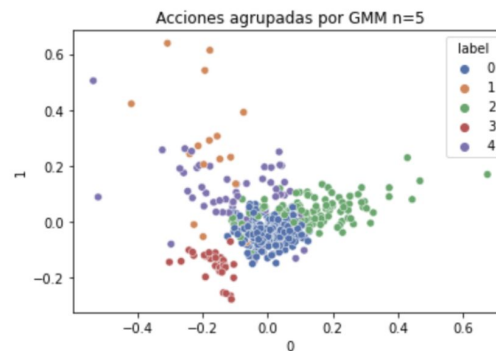
Se eligen 5 grupos

DBSCAN



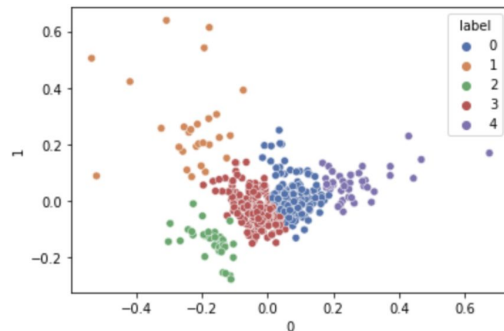
Silhouette score: 0.2948638251137764

GMM



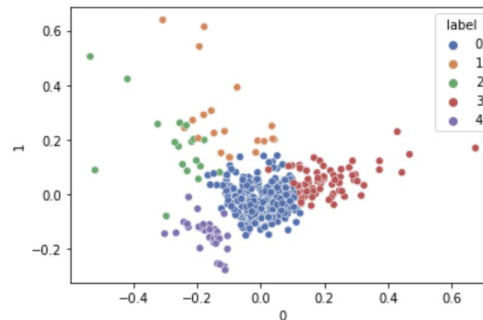
Silhouette score: 0.2826617842065334

Mini Batch K-means



Silhouette score: 0.28528898920562956

K-Means

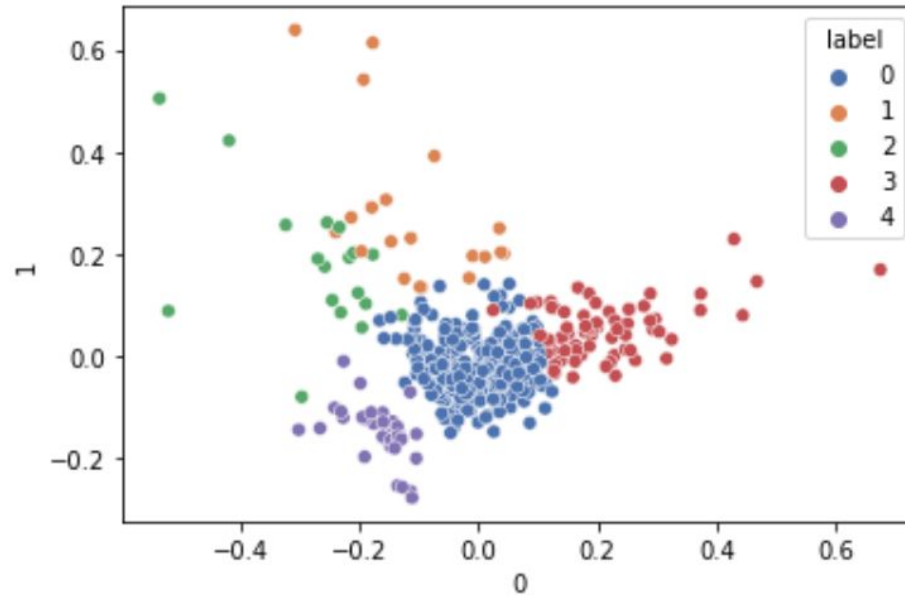


Silhouette score: 0.41241284680829576

Elección de Método clusterización

Método de clusterización	Score
K-means	0.41241284680829576
Mini batch	0.28528898920562956
GMM	0.2826617842065334
DBSCAN	0.2948638251137764

K-Means



Silhouette score: 0.41241284680829576



6. Conclusiones



Perfiles de empresas

```
df_p = df_n_kmeans.groupby(['label']).mean()  
df_p
```

	pilar_1	pilar_2	pilar_3	pilar_4	pilar_5	pilar_6	pilar_7	pilar_8
label								
0	0.280454	0.273127	0.175532	0.490694	0.167518	0.342826	0.404705	0.411484
1	0.397046	0.255881	0.441459	0.527189	0.273629	0.399434	0.400556	0.580470
2	0.270376	0.190475	0.299430	0.498082	0.534326	0.328163	0.405190	0.387113
3	0.284280	0.508369	0.178881	0.505342	0.142229	0.322368	0.405675	0.404745
4	0.283403	0.109455	0.139363	0.481661	0.215786	0.194843	0.431007	0.269599

Grupo 0: Empresas promedio

Características: Crecimiento relativamente bajo, poca emisión de acciones, deuda relativamente baja.

Número de empresas: 321

Industrias principales: Servicios Financieros (16%), Industria (16%), Salud (14%)

Ejemplos: IBM, Caterpillar, 3M



Grupo 1: Crecimiento

Características: Tienen un Market Cap muy alto en relación a sus utilidades y free cash flow, un gran crecimiento en los ingresos y utilidades pero una deuda a largo plazo elevada en comparación a el free cash flow.

Número de empresas: 20

Industrias principales: Tecnología (40%), Salud (20%)

Ejemplos: Amazon, Tesla, Dexcom

The Amazon logo, featuring the word "amazon" in a bold, black, sans-serif font, with a curved orange arrow underneath it pointing from the 'a' to the 'z'.

TESLA

The Dexcom logo, featuring the word "Dexcom" in a bold, green, sans-serif font.

CONTINUOUS GLUCOSE MONITORING

Grupo 2: Diluidor

Características: Poca deuda en comparación al free cash flow. Mayor crecimiento de los ingresos que utilidades. Alta emisión de nuevas acciones.

Número de empresas: 18

Industrias principales: Industria (27%), Tecnología (16%)

Ejemplos: L3Harris Technologies, Diamondback Energy



Grupo 3: Multiplicadores de la inversión

Características: Dan un muy buen retorno por el capital invertido, pero los ingresos no crecen mucho. Emiten pocas acciones.

Número de empresas: 70

Industrias principales: Tecnología (31%), Consumo Cíclico (17%)

Ejemplos: Apple, Microsoft, Nike



Grupo 4: Cash Maker

Características: Muy baja deuda. Crecimiento bajo en utilidades e ingresos (más bajo en ingresos). Gran crecimiento en el free cash flow. Malos retornos en el capital invertido.

Número de empresas: 37

Industrias principales: Energía/*Utilities*(54%), Consumo Cíclico (12%)

Ejemplos: Entergy Corp, Boeing, Goldman Sachs



**Goldman
Sachs**

Recomendaciones

1

Invertir en 3 y 4

Los grupos Multiplicador de la inversión y Cash Maker tienen los mejores fundamentos financieros.

2

Descartar 2

Las empresas que diluyen a sus accionistas tienen malos fundamentos financieros y no se recomiendan.

3

Riesgo en 1

Las empresas de crecimiento se recomiendan si se está dispuesto a tener un gran riesgo a cambio de grandes retornos.

4

Estudiar más 0

Se recomienda analizar de mejor manera las empresas promedio, ya que no hay diferencias notorias en los fundamentos que nos permitan dar conclusiones.



7.Propuesta de trabajo pendiente



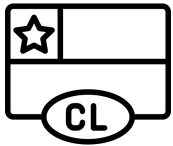
- Trabajar con datos de más empresas, es decir, ampliar la cantidad de datos con los que se trabaja.



- Clusterizar nuevamente las empresas del grupo promedio para poder analizarlas de mejor manera



- Extrapolar nuestro trabajo a las empresas que se transan en la bolsa de Santiago





Francisco Hortal Correa



Francisco Hortal Correa • Tú

Estudiante de Ingeniería Civil Industrial TI - Fundador de FJH Inversiones
ahora •

Con el objetivo de aumentar el conocimiento financiero a las personas comunes y corrientes decidimos utilizar las herramientas del curso Minería de Datos y generar grupos de empresas según fundamentos financieros. Todo esto gracias a métodos como Kmeans y PCA.

Autores: Natalia De La Barra, Nicolás Estévez, Juan Ignacio García, Francisco Hortal.

*Disclaimer: Esto no es consejo financiero.

Acciones y sus respectivos grupos

Grupo 1: AMZN, CL, CRM, CTCT, DIXM, ENPH, EPAM, ETSY, EVRG, MPWR, NOW, NVDA, OKE, PAVC, PKD, SEDG, SIVB, TSLA, VICI, VRTX

Grupo 2: BKR, CNC, DLJ, DXC, EQT, FANG, FIS, FISV, GPN, IFF, IR, LHX, LIN, NEM, PCG, RTX, TFC, WAB

Grupo 3: AAPL, ABC, ACN, ADBE, ADR, ALLE, AMAT, AMGN, AMP, ANET, AOS, AZO, BBY, BIBB, BKNG, CL, CLX, CSCO, CTSH, DFS, DPZ, FAST, FDS, FFIN, FINN, GILD, GWW, HD, HPQ, IDXX, INTU, ITW, JHLY, KLAC, KMB, LMT, LOW, LUCK, MA, MAS, MCK, MKTX, MDO, MCH, MSFT, MSI, MTCH, MTD, NKE, NTAQ, NVR, ORLY, PAYX, PFG, PGR, PM, PRU, QCOM, RHI, RUF, ROL, ROST, SGI, STX, SYF, TER, TXN, WAT, YUM, ZBRA

Grupo 4: AAL, ADM, AEE, AEP, AIG, ALB, ATO, AWK, BA, CCL, CMS, D, DTE, ED, EIX, ES, ETR, EXC, FE, GM, GS, HES, HWM, KMX, NCLH, NFLX, NI, PEG, PNW, PPL, RCL, SO, SRE, TMUS, UAL, WYNN, XEL



Recomendar



Comentar



Compartir



Enviar

Cluster S&P500 según fundamentos financieros

Natalia De La Barra
Nicolás Estévez
Juan Ignacio García
Francisco Javier Hortal

Bibliografía:

<https://www.annuityexpertadvice.com/why-is-the-stock-market-down-crashing/>

<https://www.kaggle.com/datasets/andrewmvd/sp-500-stocks>

<https://site.financialmodelingprep.com/developer/docs/>

<https://www.youtube.com/watch?v=1u6qvel9XnM>

Acciones y sus respectivos grupos

Grupo 0: A, AAP, ABBV, ABMD, ABT, ADI, AFL, AIZ, AJG, AKAM, ALGN, ALK, ALL, AMCR, AME, AMT, ANSS, AON, APA, APD, APH, APTV, ARE, ATVI, AVB, AVGO, AVY, AXP, BAC, BALL, BAX, BBWI, BDX, BEN, BIO, BK, BLK, BMY, BR, BRK-B, BRO, BSX, BWA, BXP, C, CAG, CAT, CB, CBOE, CBRE, CCI, CDNS, CDW, CE, CF, CFG, CHD, CHRW, CHTR, CINF, CMA, CMCSA, CME, CMG, CMI, COF, COO, COP, COST, CPB, CPRT, CPT, CRL, CSGP, CSX, CTAS, CTRA, CVS, CVX, DAL, DD, DE, DG, DGX, DHI, DHR, DIS, DISH, DLTR, DOV, DRI, DVA, DVN, EA, EBAY, ECL, EFX, EL, EMN, EMR, EOG, EQR, ESS, ETN, EW, EXPD, EXPE, EXR, F, FBHS, FCX, FDX, FITB, FLT, FMC, FRC, FTV, GD, GE, GIS, GL, GLW, GNRC, GOOG, GOOGL, GPC, GRMN, HAL, HAS, HBAN, HCA, HIG, HII, HLT, HOLX, HON, HPE, HRL, HSIC, HST, HSY, HUM, IBM, ICE, IEX, ILMN, INCY, INTC, INVH, IP, IPG, IQV, IRM, ISRG, IVZ, J, JBHT, JCI, JNJ, JNPR, JPM, K, KEY, KEYS, KHC, KIM, KMI, KO, L, LDOS, LEN, LH, LKQ, LLY, LNC, LNT, LUV, LVS, LW, LYB, MAA, MAR, MCD, MCHP, MCO, MDLZ, MDT, MET, MGM, MHK, MKC, MLM, MMC, MMM, MNST, MOS, MPC, MRK, MRO, MS, MSCI, MTB, MU, NDAQ, NDSN, NLOK, NOC, NRG, NSC, NTRS, NUE, NWL, NWS, NWSA, NXPI, O, ODFL, OMC, ON, ORCL, OXY, PCAR, PEAK, PEP, PFE, PG, PH, PHM, PKG, PKI, PLD, PNC, PNR, POOL, PPG, PSA, PSX, PWR, PYPL, QRVO, RE, REG, REGN, RF, RL, RMD, ROK, ROP, RSG, SBAC, SBNY, SCHW, SEE, SHW, SJM, SLB, SNA, SNPS, SPG, STE, STT, STZ, SWK, SWKS, SYK, SYY, T, TAP, TDG, TDY, TECH, TEL, TFX, TGT, TJX, TMO, TPR, TRMB, TROW, TRV, TSCO, TSN, TT, TTWO, TXT, TYL, UDR, UHS, ULTA, UNH, UNP, UPS, URI, USB, VFC, VLO, VMC, VNO, VRSK, VTR, VTRS, VZ, WBA, WDC, WEC, WELL, WFC, WHR, WM, WMB, WMT, WRB, WRK, WST, WY, XOM, XRAY, XYL, ZBH, ZION, ZTS

Acciones y sus respectivos grupos

Grupo 1: AMZN, CI, CRM, CTLT, DXCM, ENPH, EPAM, ETSY, EVRG, MPWR, NOW, NVDA, OKE, PAYC, PXD, SEDG, SIVB, TSLA, VICI, VRTX

Grupo 2: BKR, CNC, DLR, DXC, EQT, FANG, FIS, FISV, GPN, IFF, IR, LHX, LIN, NEM, PCG, RTX, TFC, WAB

Grupo 3: AAPL, ABC, ACN, ADBE, ADP, ALLE, AMAT, AMGN, AMP, ANET, AOS, AZO, BBY, BIIB, BKNG, CL, CLX, CSCO, CTSH, DFS, DPZ, FAST, FDS, FFIV, FTNT, GILD, GWW, HD, HPQ, IDXX, INTU, ITW, JKHY, KLAC, KMB, LMT, LOW, LRCX, MA, MAS, MCK, MKTX, MO, MOH, MSFT, MSI, MTCH, MTD, NKE, NTAP, NVR, ORLY, PAYX, PFG, PGR, PM, PRU, QCOM, RHI, RJF, ROL, ROST, SPGI, STX, SYF, TER, TXN, WAT, YUM, ZBRA

Grupo 4: AAL, ADM, AEE, AEP, AIG, ALB, ATO, AWK, BA, CCL, CMS, D, DTE, ED, EIX, ES, ETR, EXC, FE, GM, GS, HES, HWM, KMX, NCLH, NFLX, NI, PEG, PNW, PPL, RCL, SO, SRE, TMUS, UAL, WYNN, XEL