

***Rotman***

# INTRO TO DATA VISUALIZATION

Part II Intro to Matplotlib – From Default to Publication-Ready

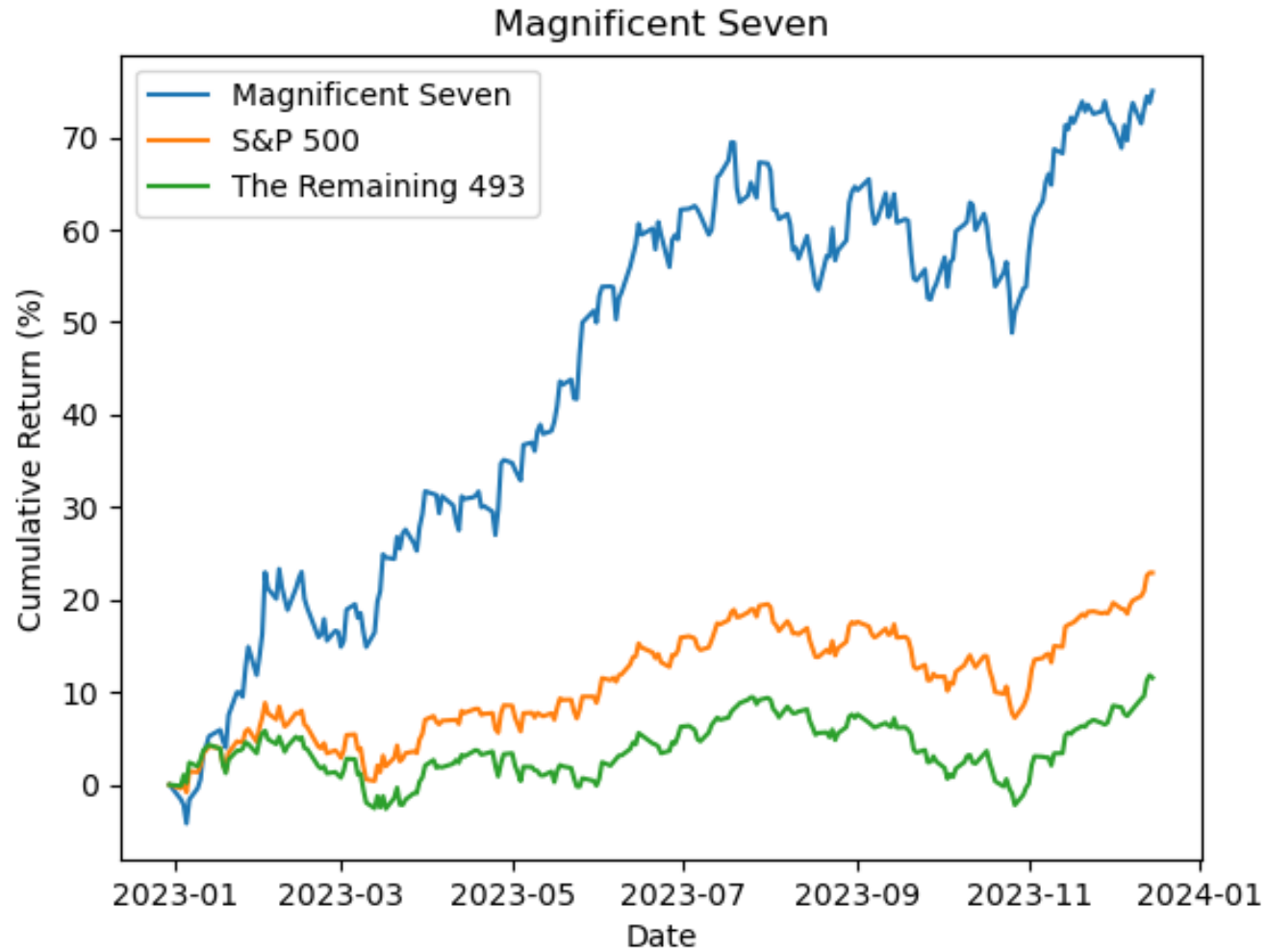
September 4, 2024 Prepared by Jay Cao / [TDMDAL](https://tdmdal.github.io)

Website: <https://tdmdal.github.io/mma-dv-2024/>

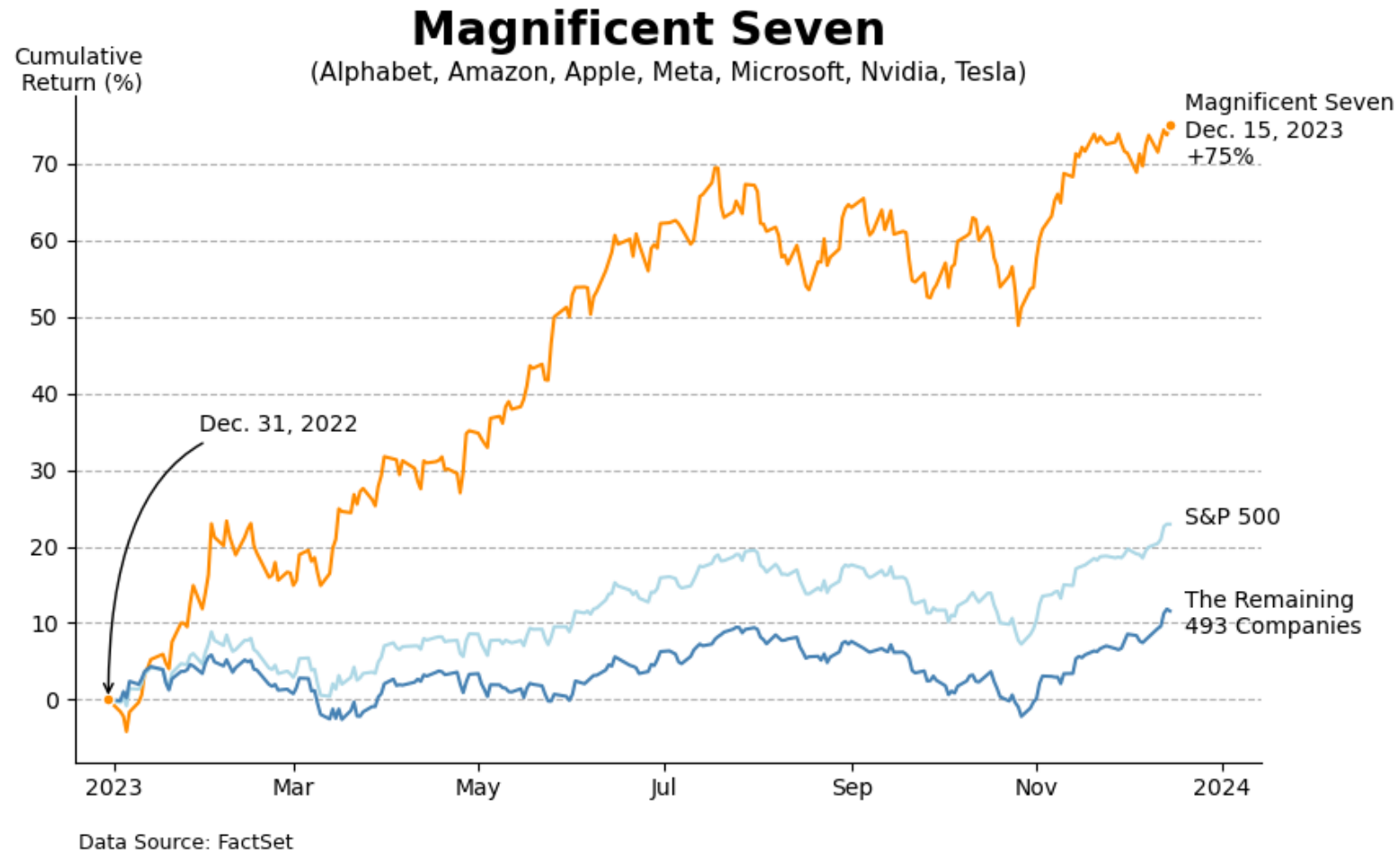


Rotman School of Management  
UNIVERSITY OF TORONTO

# Hands-on: From Matplotlib Default



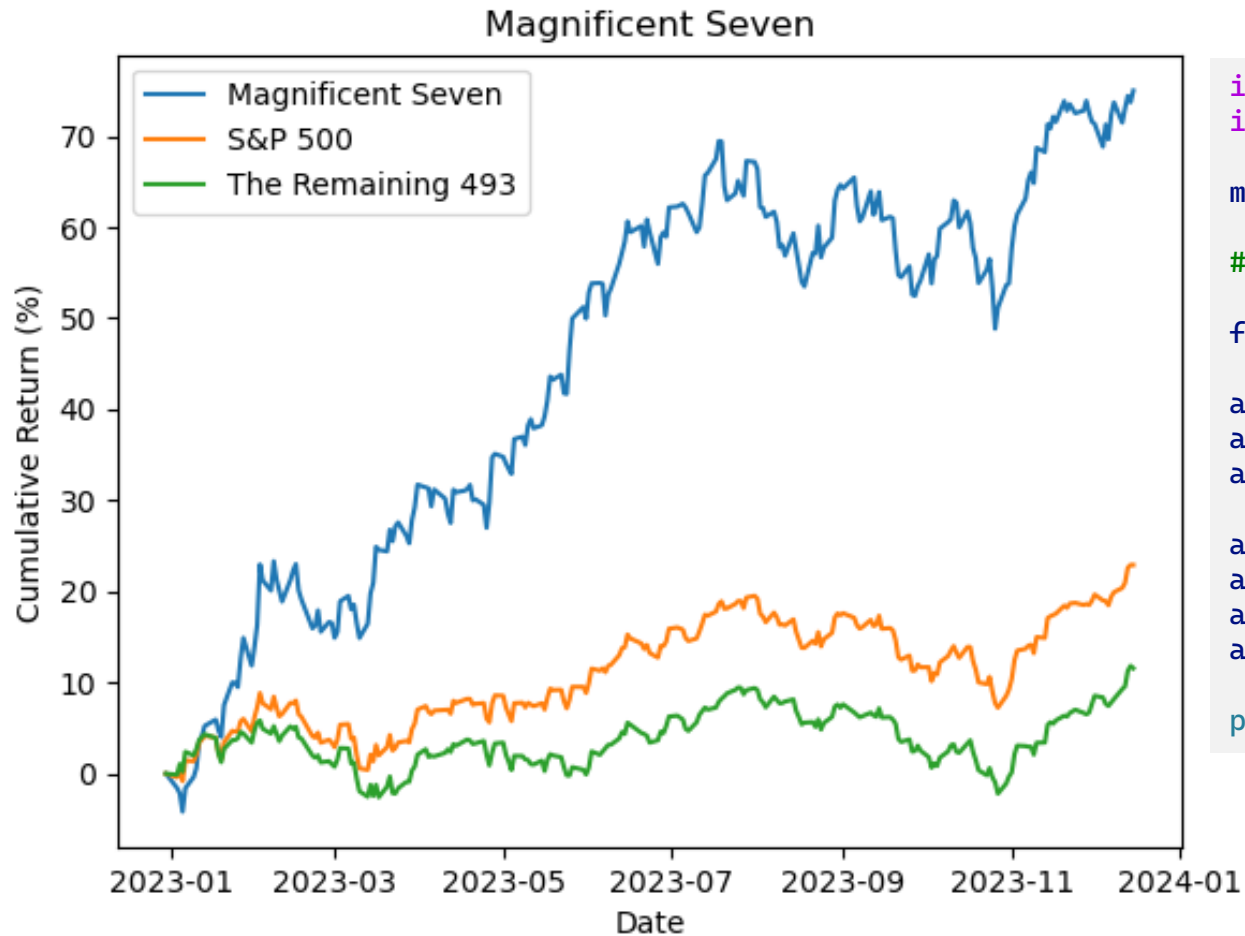
# Hands-on: To Publication-Quality



Inspiration Source 1: <https://www.nytimes.com/interactive/2024/01/22/business/magnificent-seven-stocks-tech.html>

Inspiration Source 2: <https://www.wsj.com/finance/stocks/its-the-magnificent-sevens-market-the-other-stocks-are-just-living-in-it-5d212f95>

# Round 1 - Matplotlib Default



```
import pandas as pd
import matplotlib.pyplot as plt

mag7_df = pd.read_csv("mag7.csv", parse_dates=["date"])

# Round 1

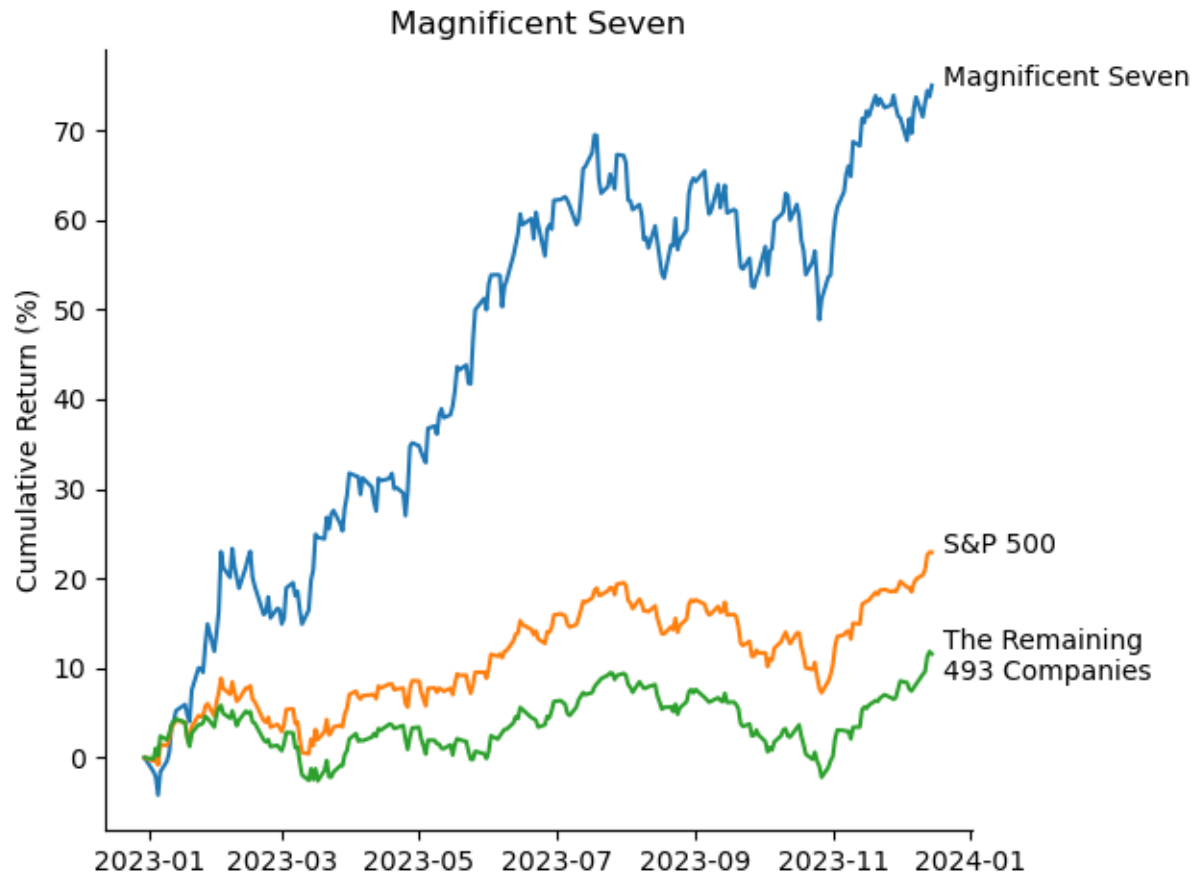
fig, ax = plt.subplots()

ax.plot(mag7_df["date"], mag7_df["mag7"])
ax.plot(mag7_df["date"], mag7_df["sp"])
ax.plot(mag7_df["date"], mag7_df["rest"])

ax.set_xlabel("Date")
ax.set_ylabel("Cumulative Return (%)")
ax.set_title("Magnificent Seven")
ax.legend(["Magnificent Seven", "S&P 500", "The Remaining 493"])

plt.show()
```

# Round 2

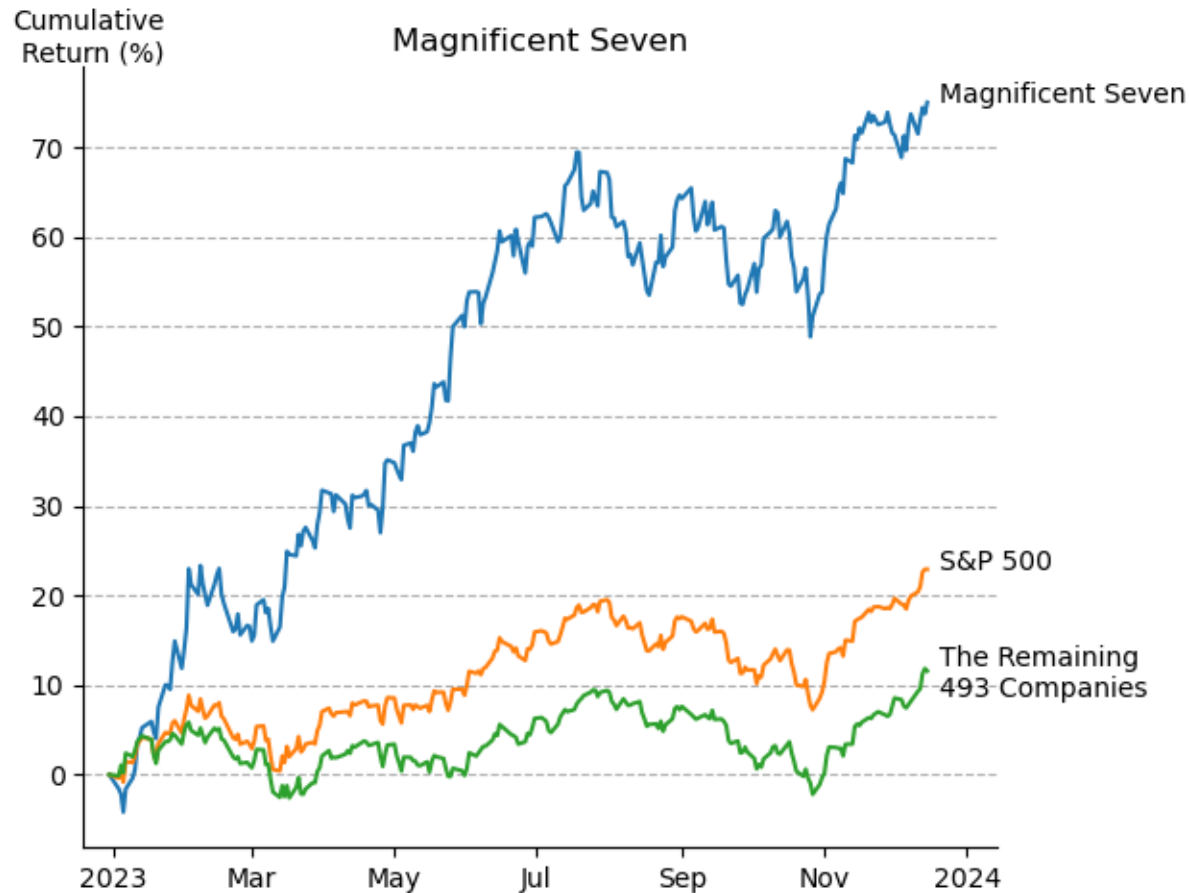


```
...  
  
# Round 2  
  
# remove x-axis label  
# remove legend  
# annotate at the end of each line  
# hide top and right spines  
  
# add annotations  
ax.annotate(  
    text="Magnificent Seven",  
    xy=(mag7_df["date"].iloc[-1], mag7_df["mag7"].iloc[-1]),  
    xytext=(  
        mag7_df["date"].iloc[-1] + pd.Timedelta("5 day"),  
        mag7_df["mag7"].iloc[-1],  
    ),  
)  
  
...  
  
# hide top and right spines  
ax.spines["top"].set_visible(False)  
  
...  
  
# use tight_layout() to automatically adjusts subplot  
plt.tight_layout()  
  
...
```

Ref 1. <https://matplotlib.org/stable/users/explain/text/annotations.html>

Ref 2. [https://matplotlib.org/stable/api/\\_as\\_gen/matplotlib.axes.Axes.annotate.html](https://matplotlib.org/stable/api/_as_gen/matplotlib.axes.Axes.annotate.html)

# Round 3



```
import matplotlib.dates as mdates

...

# Round 3
# improve the x-axis tick labels
# improve the y-axis labels (move it to the top)
# add y grid

...

# set x-axis view limits
ax.set_xlim(mag7_df["date"].iloc[1] - pd.Timedelta("15 days"),
            mag7_df["date"].iloc[-1] + pd.Timedelta("30 days"))

# set x-axis tick labels
locator = mdates.AutoDateLocator()
formatter = mdates.ConciseDateFormatter(locator)
ax.xaxis.set_major_locator(locator)
ax.xaxis.set_major_formatter(formatter)

# set the y-axis label
ax.set_ylabel("Cumulative\nReturn (%)", loc="top",
              rotation=0, labelpad=-50)

...

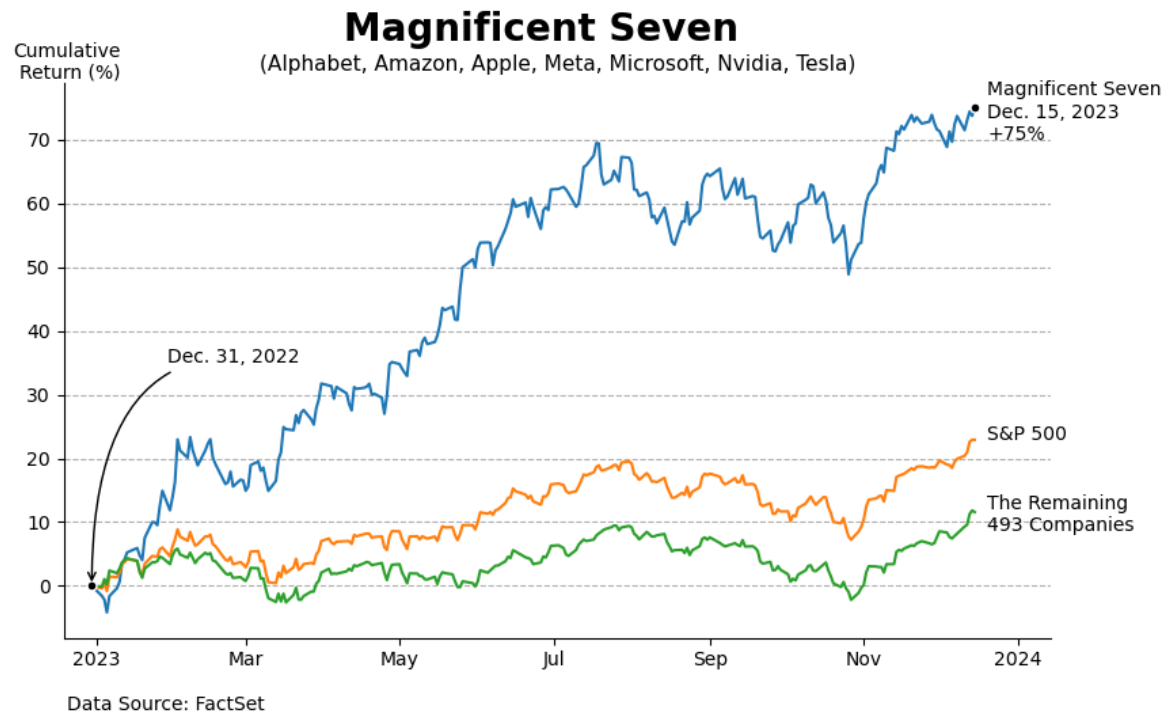
# add y grid
ax.grid(visible=True, axis='y', linestyle="--")

...
```

Ref 1. [https://matplotlib.org/stable/api/dates\\_api.html](https://matplotlib.org/stable/api/dates_api.html)

Ref 2. [https://matplotlib.org/stable/api/\\_as\\_gen/matplotlib.pyplot.grid.html](https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.grid.html)

# Round 4

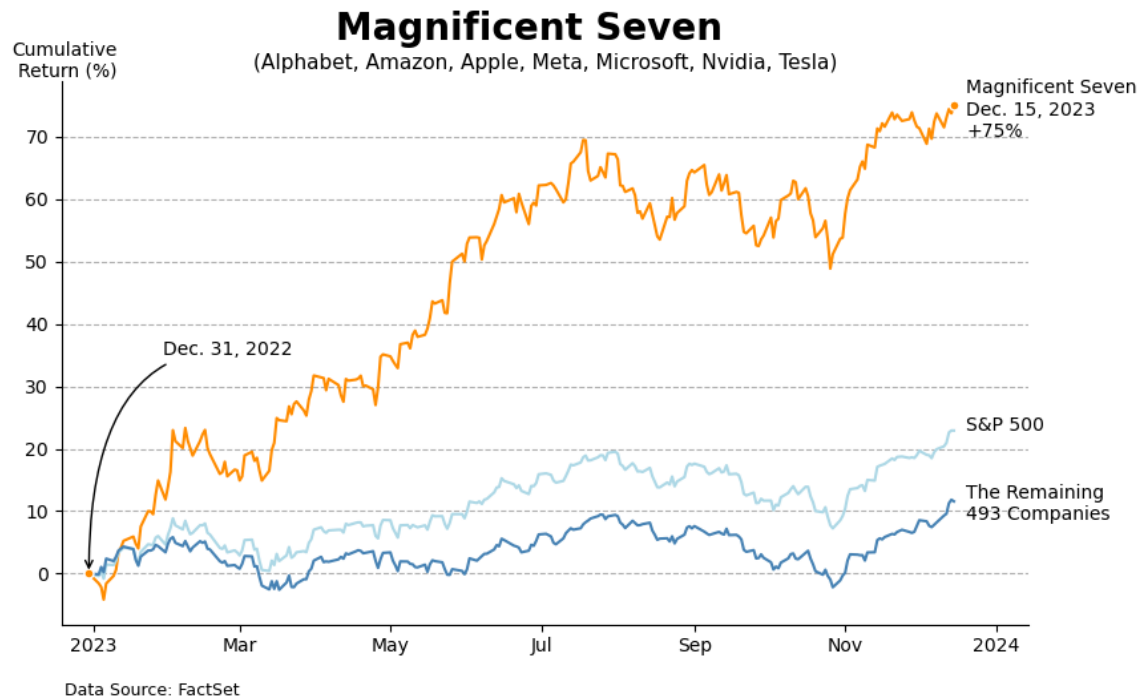


```
...  
  
# Round 4  
# add the first data point as a dot and annotate it  
# add the last data point of the mag7 as a dot and annotate it  
# improve title  
# add caption  
# adjust axes aspect ratio  
# add the figure size  
# remove tight_layout()  
  
# set ax aspect ratio and figure size  
ax.set_box_aspect(aspect=9/16)  
fig_width = 9*1.05  
fig_height = fig_width/16*9  
fig.set_figwidth(fig_width)  
fig.set_figheight(fig_height)  
  
# set title and subtitle  
fig.suptitle("Magnificent Seven", fontsize=20, fontweight='bold')  
ax.set_title("(Alphabet, Amazon, Apple, Meta, Microsoft, Nvidia, Tesla)",  
             fontdict={'fontsize': 11})  
  
# plot the first data point as a dot and annotate it  
ax.plot(mag7_df["date"].iloc[0], mag7_df["mag7"].iloc[0],  
        color="black", marker="o", markersize=5, markeredgcolor="white")  
ax.annotate(  
    text="Dec. 31, 2022",  
    xy=(mag7_df["date"].iloc[0], mag7_df["mag7"].iloc[0]),  
    xytext=(  
        mag7_df["date"].iloc[0] + pd.Timedelta("30 day"),  
        mag7_df["mag7"].iloc[0] + 35,  
    ),  
    arrowprops=dict(arrowstyle="->", connectionstyle="angle3, angleA=0, angleB=90"),  
)  
...  
  
# add caption  
# https://matplotlib.org/stable/api/_as_gen/matplotlib.figure.Figure.text.html  
fig.text(0.13, 0.01, "Data Source: FactSet")  
# plt.tight_layout()
```

Ref 1. <https://matplotlib.org/stable/users/explain/text/annotations.html#customizing-annotation-arrows>

Ref 2. [https://matplotlib.org/stable/gallery/subplots\\_axes\\_and\\_figures/figure\\_title.html](https://matplotlib.org/stable/gallery/subplots_axes_and_figures/figure_title.html)

# Round 5



```
...  
  
# Round 5  
# adjust colors and fonts  
  
...  
  
# plot the data  
ax.plot(mag7_df["date"], mag7_df["mag7"], color="darkorange")  
ax.plot(mag7_df["date"], mag7_df["sp"], color="lightblue")  
ax.plot(mag7_df["date"], mag7_df["rest"], color="steelblue")  
  
# set title and subtitle  
fig.suptitle("Magnificent Seven", fontsize=20, fontweight='bold')  
ax.set_title("(Alphabet, Amazon, Apple, Meta, Microsoft, Nvidia, Tesla)",  
             fontdict={'fontsize': 11})  
  
...  
  
# plot the first data point as a dot and annotate it  
ax.plot(mag7_df["date"].iloc[0], mag7_df["mag7"].iloc[0],  
        color="darkorange", marker="o", markersize=5, markeredgcolor="white")  
  
...  
  
# plot the last data point of magnificent seven as a dot  
ax.plot(mag7_df["date"].iloc[-1], mag7_df["mag7"].iloc[-1],  
        color="darkorange", marker="o", markersize=5, markeredgcolor="white")  
  
...  
  
# add caption  
fig.text(0.13, 0.01, "Data Source: FactSet",  
         fontdict={'fontsize': 9, 'fontweight': 'light'})  
  
...
```

Ref 1. [https://matplotlib.org/stable/gallery/color/named\\_colors.html](https://matplotlib.org/stable/gallery/color/named_colors.html)

Ref 2. [https://matplotlib.org/stable/users/explain/text/text\\_intro.html](https://matplotlib.org/stable/users/explain/text/text_intro.html)