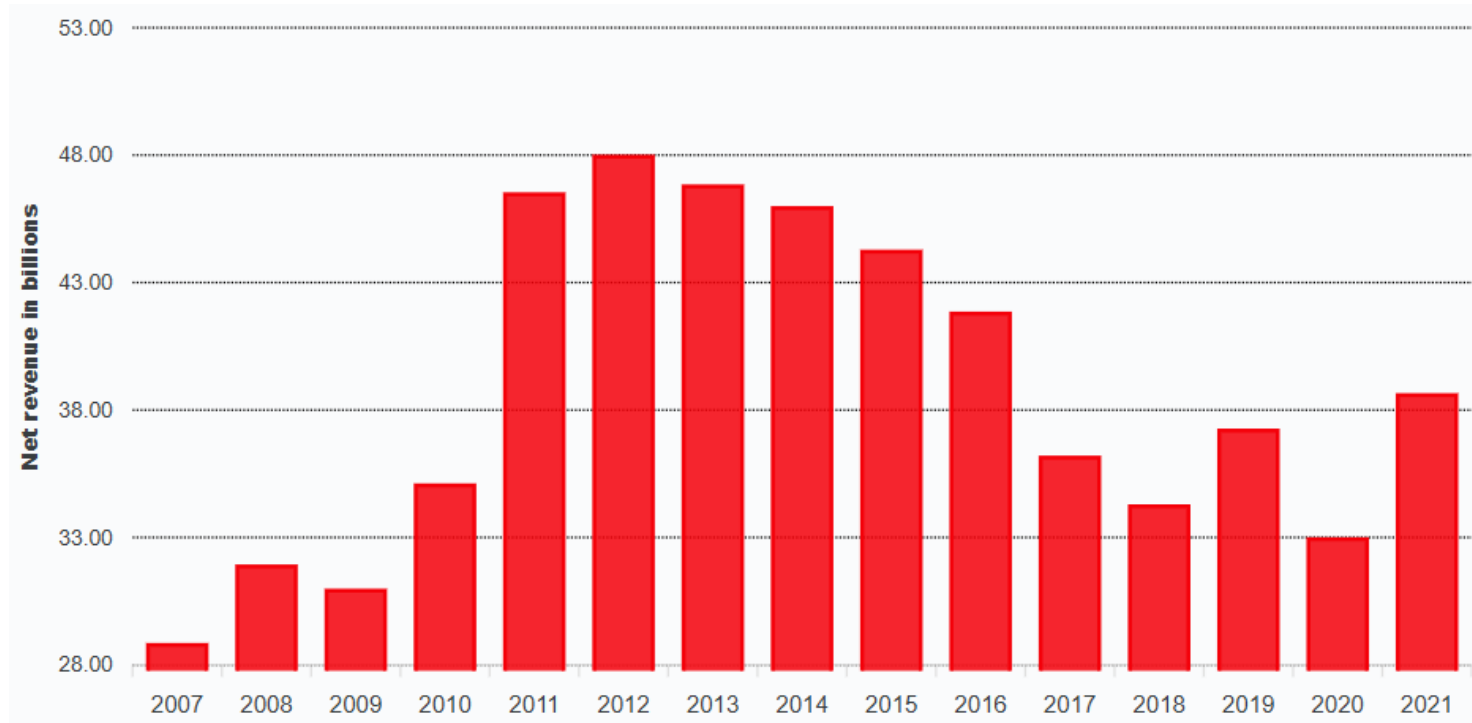# Data Analytics for Finance II

Luca Cereghetti
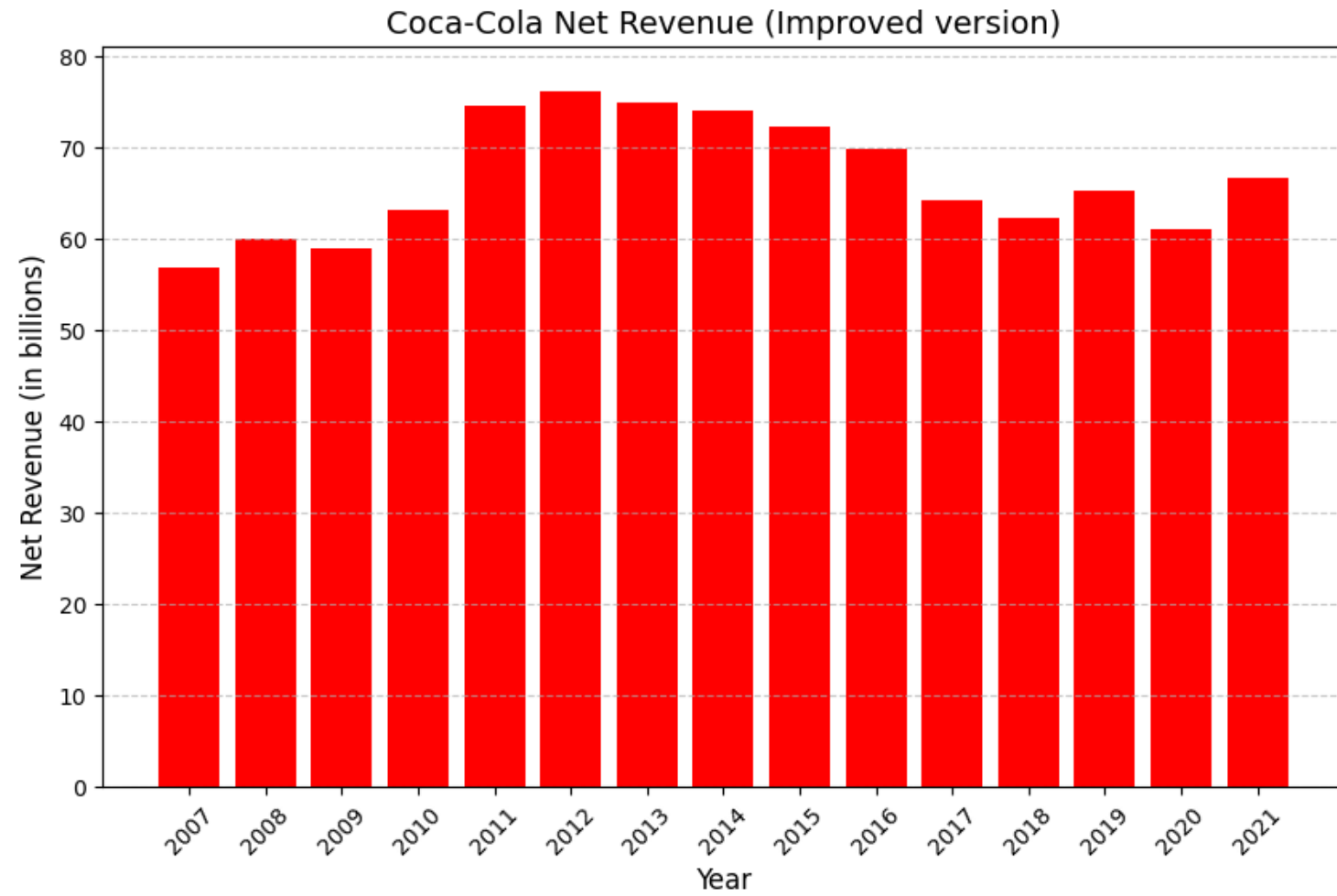
January 2025
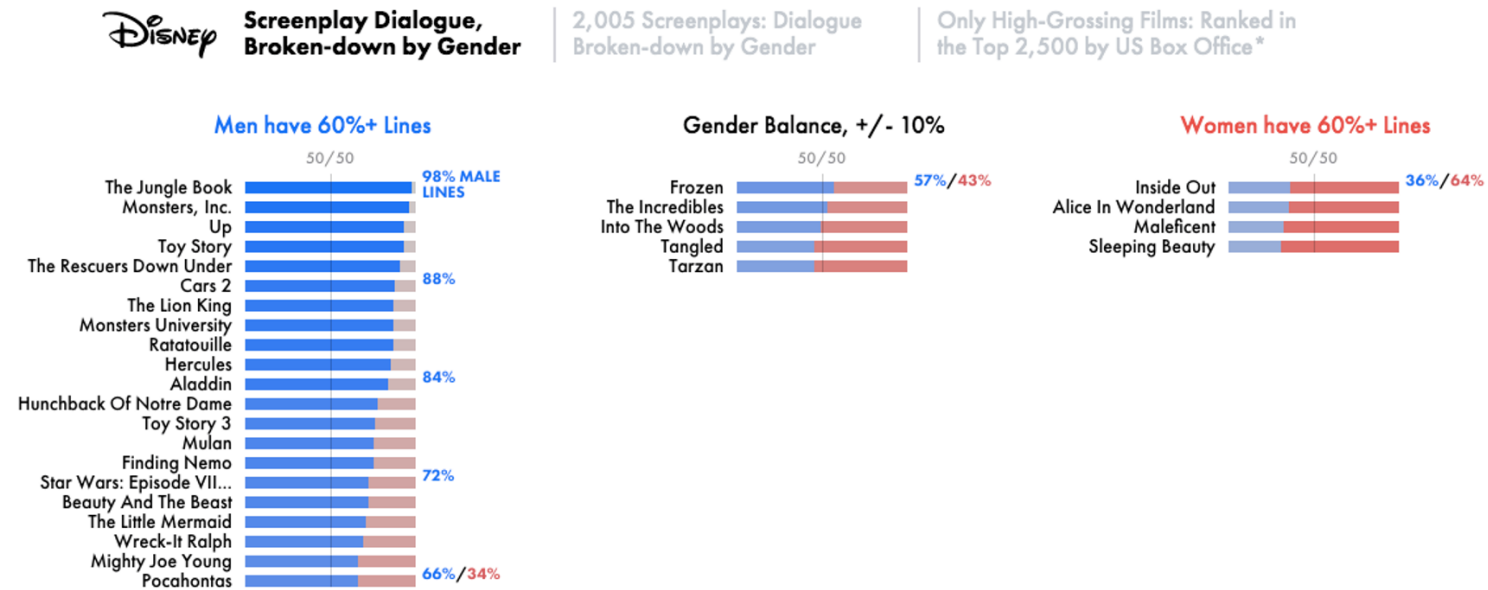
# 1 Bad and/or manipulative visualization



The bar chart highlights Coca-Cola's net revenue over the years, but it uses a deceptive visual technique. Instead of starting the Y-axis at zero, the chart begins at 28 billion, which greatly exaggerates the perceived growth in revenue. This creates an illusion of a dramatic increase in earnings that is not as significant when viewed in context. By truncating the vertical axis, the differences between the bars are visually amplified, misleading viewers into overestimating the extent of the changes. Such manipulative practices can distort data interpretation and influence opinions, particularly in financial presentations or marketing materials. It's important to use properly scaled axes to ensure accurate and unbiased representation of information, allowing viewers to make informed and fair assessments of the data presented.

## 2 Improved version of the "bad/manipulative" visualization



Coca-Cola Net Revenue (Improved version)

# 3  A particularly good visualization



**Disney Screenplay Dialogue, Broken-down by Gender**

2,005 Screenplays: Dialogue Broken-down by Gender

Only High-Grossing Films: Ranked in the Top 2,500 by US Box Office*

**Men have 60%+ Lines**

50/50

| Film | |
|---|---|
| The Jungle Book | 98% MALE LINES |
| Monsters, Inc. | |
| Up | |
| Toy Story | |
| The Rescuers Down Under | |
| Cars 2 | 88% |
| The Lion King | |
| Monsters University | |
| Ratatouille | |
| Hercules | |
| Aladdin | 84% |
| Hunchback Of Notre Dame | |
| Toy Story 3 | |
| Mulan | |
| Finding Nemo | |
| Star Wars: Episode VII... | 72% |
| Beauty And The Beast | |
| The Little Mermaid | |
| Wreck-It Ralph | |
| Mighty Joe Young | |
| Pocahontas | 66%/34% |

**Gender Balance, +/- 10%**

50/50

| Film | |
|---|---|
| Frozen | 57%/43% |
| The Incredibles | |
| Into The Woods | |
| Tangled | |
| Tarzan | |

**Women have 60%+ Lines**

50/50

| Film | |
|---|---|
| Inside Out | 36%/64% |
| Alice In Wonderland | |
| Maleficent | |
| Sleeping Beauty | |

This visualization effectively communicates the breakdown of dialogue by gender in various Disney films. It uses three clear columns to organize the data: films where men have over 60% of the dialogue, films with near gender balance, and films where women have over 60% of the dialogue. This logical grouping makes it easy for the viewer to interpret the information at a glance. The choice of color is particularly effective, with blue representing male dialogue and light red for female dialogue, ensuring quick visual differentiation. Additionally, the inclusion of specific percentages, such as the 50/50 balance line and the proportion of dialogue for each gender, provides clarity without overwhelming the viewer with excessive data. Overall, this visualization strikes a good balance between simplicity and informativeness, making it an excellent tool for understanding gender representation in Disney films.

4

# 4 Climate change visualisation - LinkedIn post



Annual Emissions Over Time for Switzerland
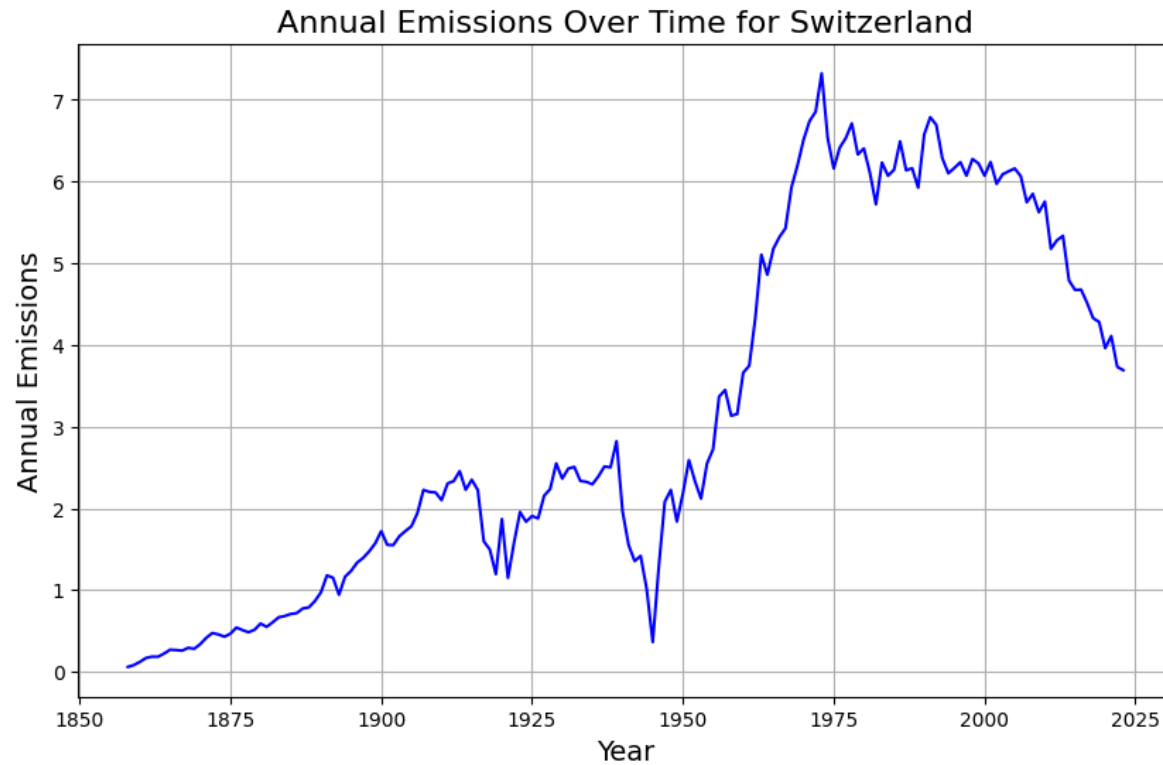
## Switzerland's CO$_2$ Emissions: A Journey Through Time

This graph highlights Switzerland's annual CO$_2$ emissions from 1850 to 2025. Emissions rose steadily, peaking in the 1970s, but have since declined thanks to cleaner energy, improved efficiency, and climate-conscious policies. While progress is clear, the journey to net-zero isn't over. Let's continue innovating and working together for a sustainable future. What's your take on Switzerland's progress and the next steps? Share your thoughts below!

#Sustainability #ClimateChange #DataVisualization #NetZero #CleanEnergy #Switzerland
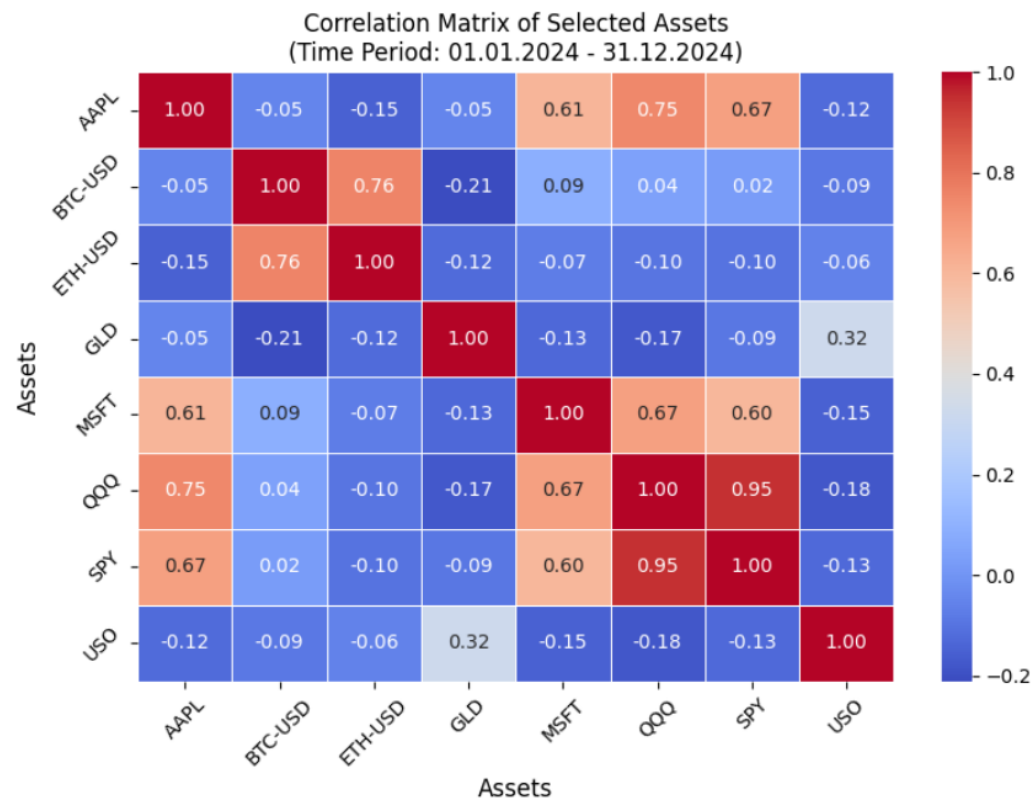
# 5 Black-and-white visualization

Density Distribution of Tesla Daily Returns (01.01.2023 - 01.01.2025)
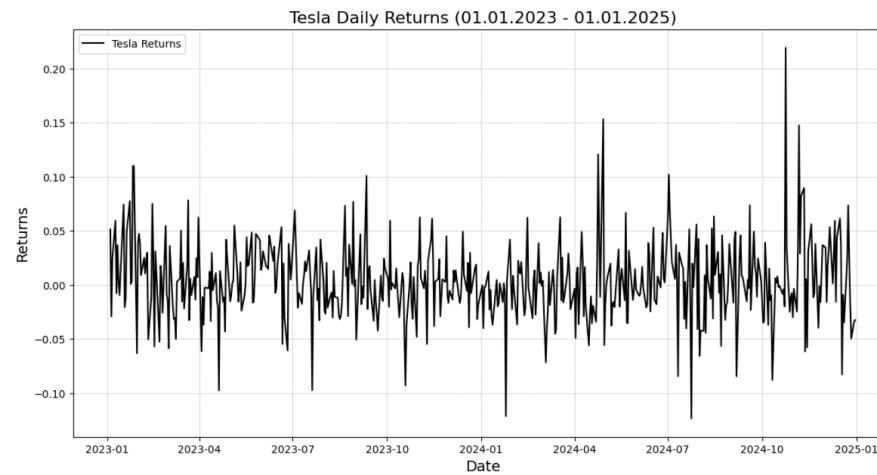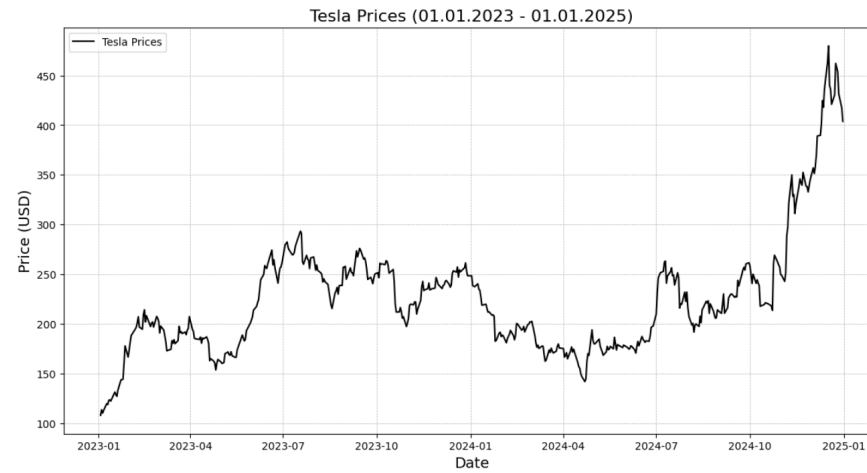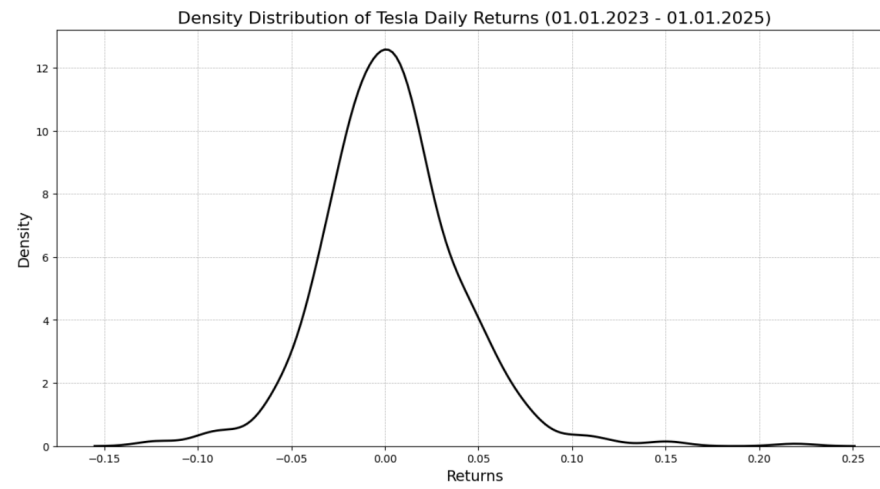
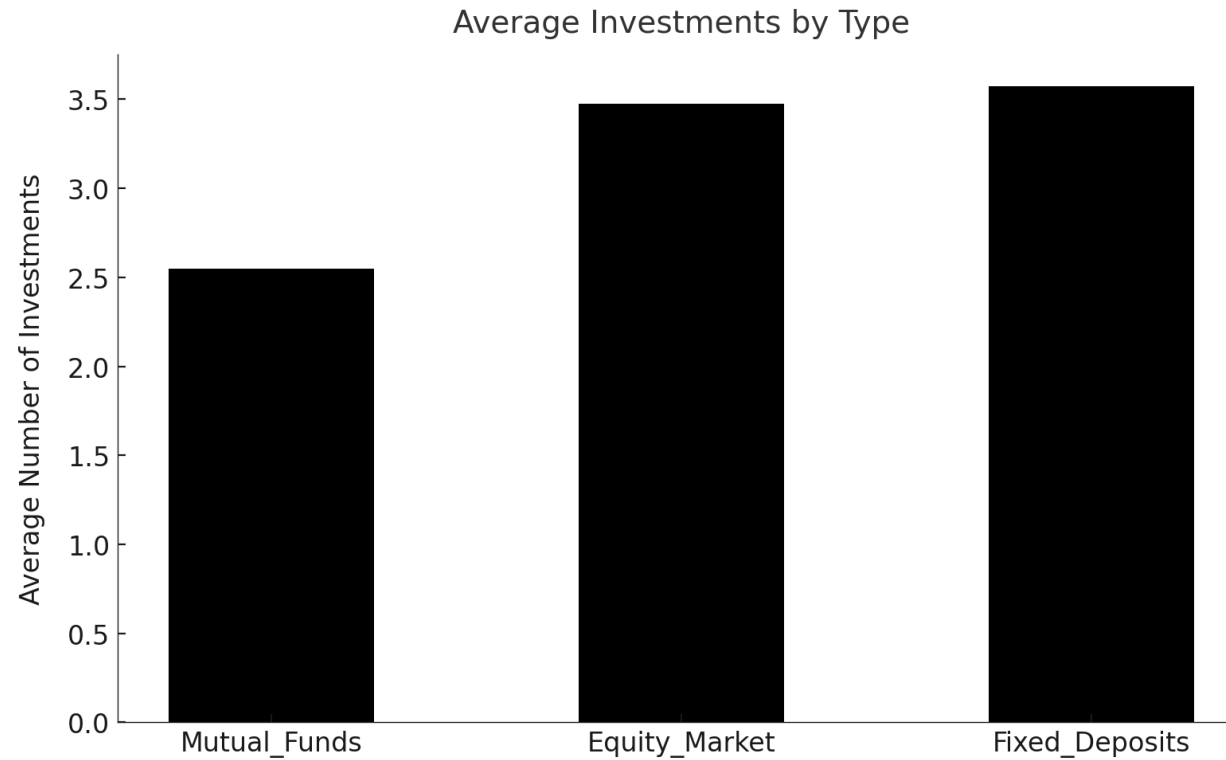# 6 Visualisation using colours

For this analysis, I selected assets from different categories to build a correlation matrix:

- **Stocks**: AAPL (Apple Inc.), MSFT (Microsoft)

- **Cryptocurrencies**: BTC-USD (Bitcoin), ETH-USD (Ethereum)

- **Commodities**: GLD (Gold ETF), USO (Crude Oil ETF)

- **Indices**: SPY (S&P 500 ETF), QQQ (Nasdaq 100 ETF)



Correlation Matrix of Selected Assets
(Time Period: 01.01.2024 - 31.12.2024)

# 7 Maximizing Clarity Through Tufte's Data-Ink Ratio Principle

## Average Investments by Type



The bar chart represents the average number of investments in Mutual Funds, Equity Market, and Fixed Deposits, based on data from 40 individuals. Each individual provided information on their investment preferences, indicating how many times they had invested in each type. The averages highlight the most and least popular investment options. The results show that the Equity Market had the highest average number of investments, followed by Fixed Deposits. Mutual Funds had the lowest average, indicating relatively less frequent investment. The visualization adheres to Tufte's principle of maximizing the data-ink ratio, focusing solely on essential data trends.

# 8 Network Graph



Network Graph of Correlations Between Magnificent 7
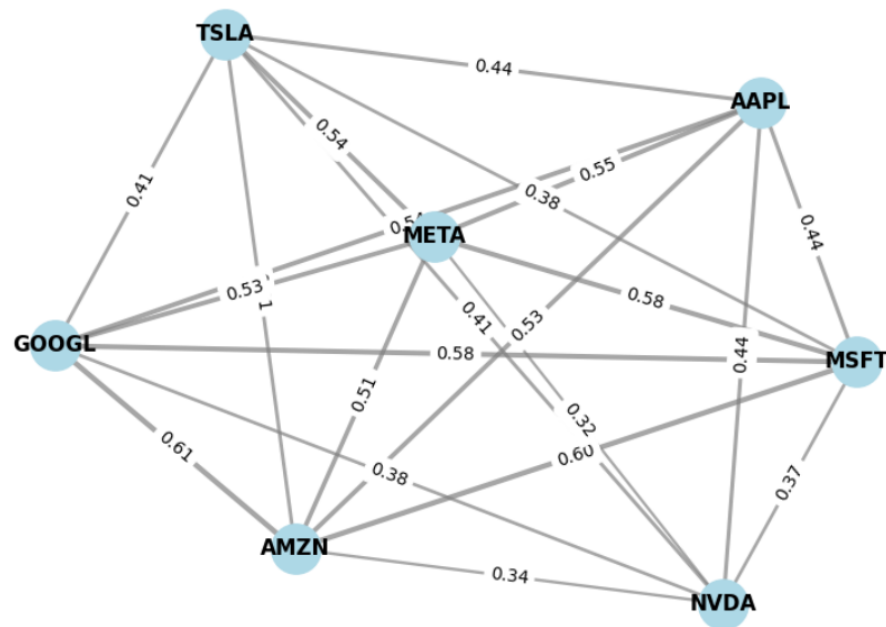
The network graph illustrates the correlations between the **Magnificent 7** companies: Apple (**AAPL**), Microsoft (**MSFT**), Amazon (**AMZN**), Alphabet (**GOOGL**), Meta Platforms (**META**), Tesla (**TSLA**), and NVIDIA (**NVDA**). These companies, spanning the technology, consumer discretionary, and communication services sectors, are known for driving significant growth in the stock market.

The thickness of the edges represents the strength of the correlation between the companies' stock returns, with values ranging from moderate (e.g., 0.32) to strong (e.g., 0.61). For example, the correlation between Amazon and Alphabet is notably high, reflecting their similar market trends. This graph highlights the interconnectedness of these dominant players in the U.S. market, underscoring their collective influence on major indices like the S&P 500 and Nasdaq 100.
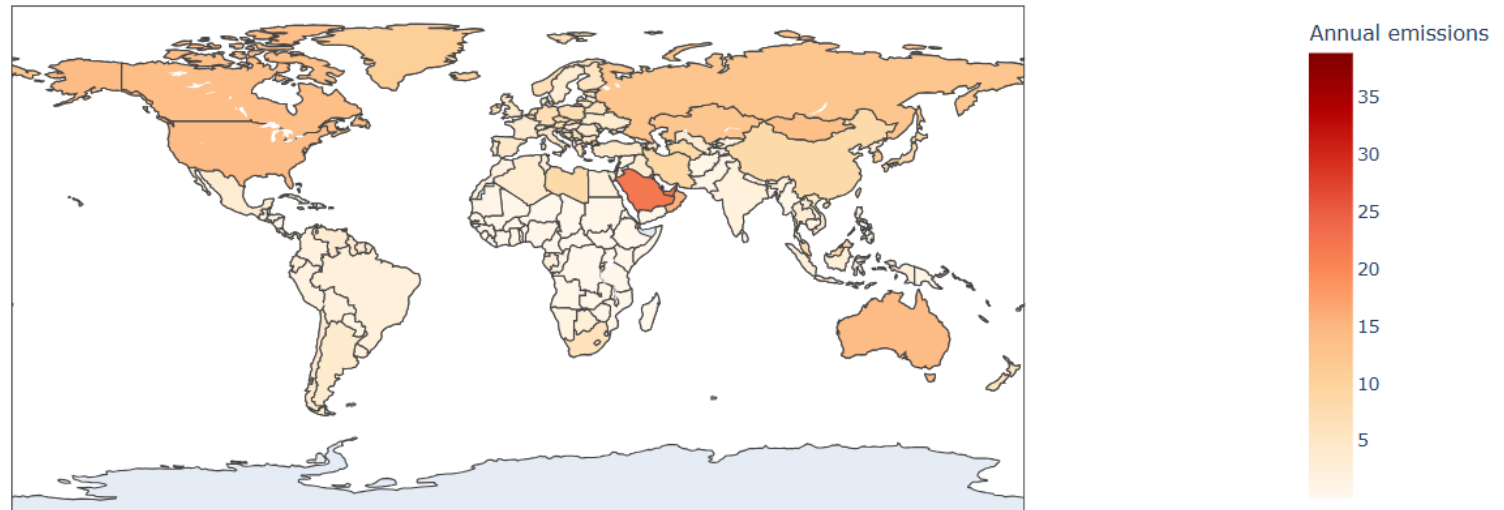
# 9   Visualization by hand

# 10 Visualization with ChatGPT



Top 10 Films with Highest Returns
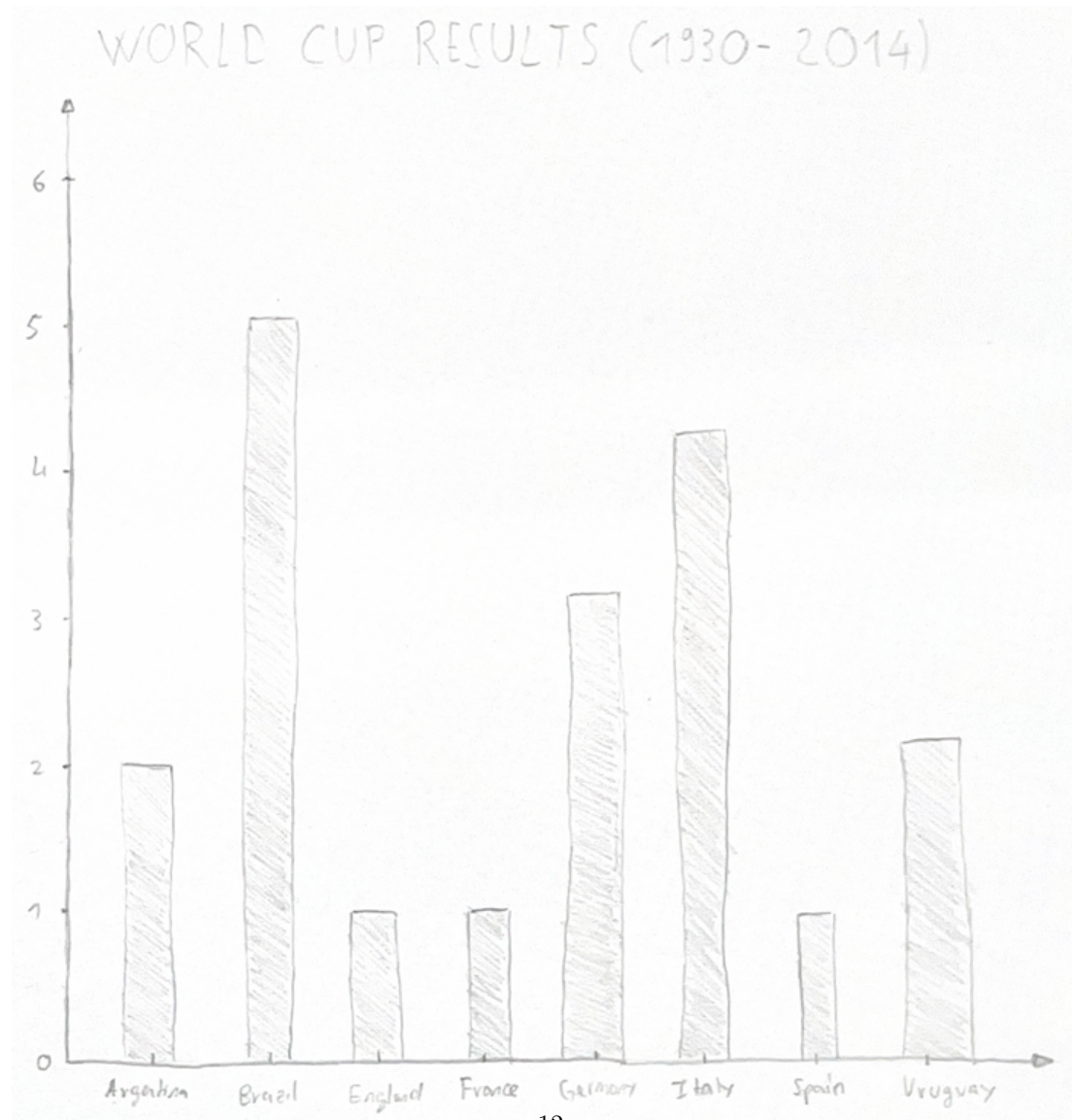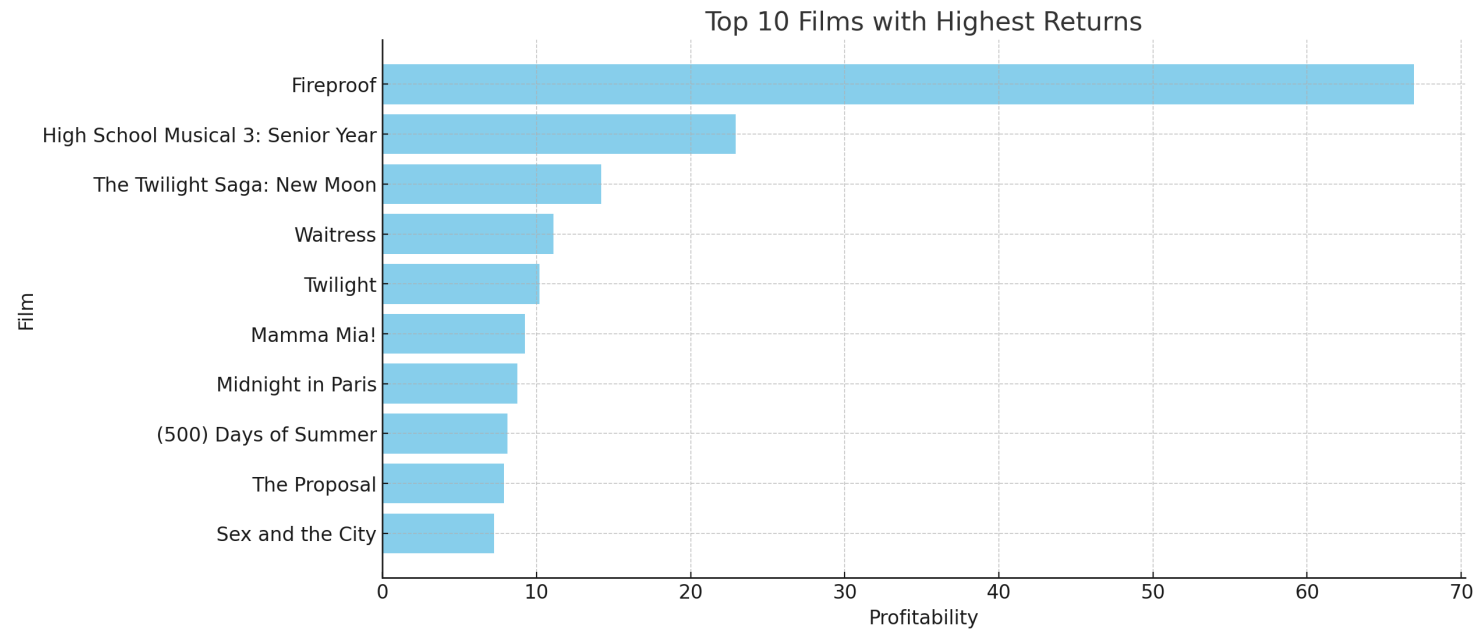
# 11   Data Map

Global Annual CO$_2$ Emissions by Country (2023)
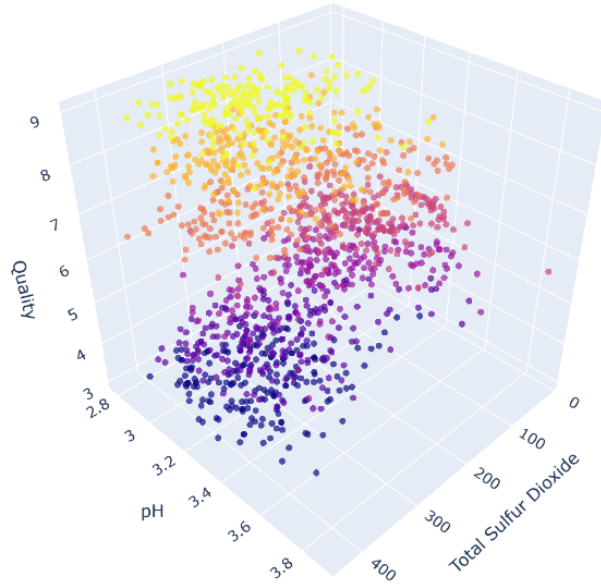
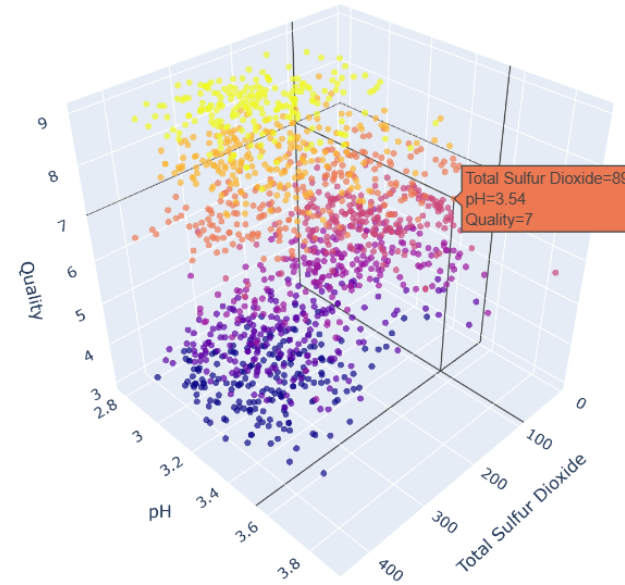# 12  Interactive Visualization



Figure 1: First Visualization



Figure 2: Second Visualization

Interactive Visualization

We developed an interactive visualization to analyze the relationship between wine quality and its chemical properties, focusing on pH and total sulfur dioxide. Using a 3D scatter plot, the axes represent total sulfur dioxide, pH, and wine quality, with a color gradient from purple to yellow indicating lower to higher quality. The graph allows users to interact dynamically by rotating to view patterns from different perspectives, zooming to focus on specific clusters or outliers, and hovering over points to see detailed values for pH, total sulfur dioxide, and quality.

# 13   Documentation of creation process
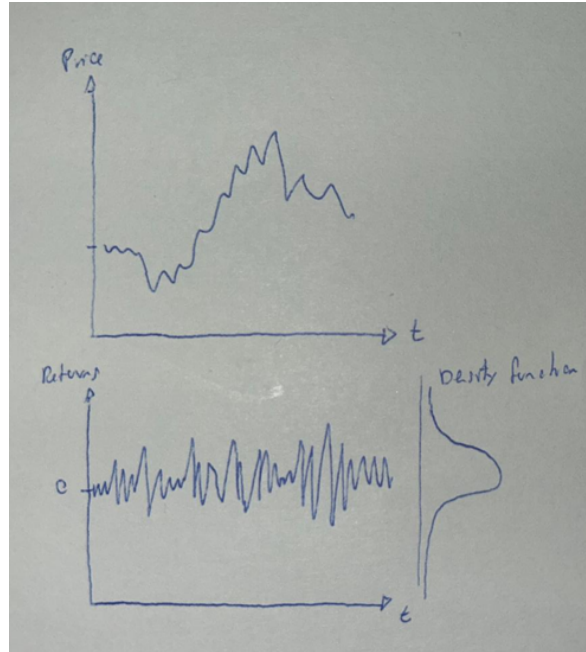


To create the black-and-white visualizations (Visualization number 5), I began by sketching the graphs by hand to conceptualize the project and identify the key aspects of the data to highlight. These initial sketches included plots of the asset price over time, its daily returns, and the density distribution of returns. After finalizing the idea, I selected Tesla as the asset for analysis and downloaded its historical price data from Yahoo Finance.

Using Python, I employed the 'matplotlib.pyplot' library to generate the visualizations. The first graph represents Tesla's price over time, showing its overall trend. The second plot illustrates daily returns to emphasize volatility, while the third shows the density distribution of returns, capturing their statistical properties. The black-and-white theme was chosen to ensure simplicity and clarity, making the graphs suitable for reports and publications. This structured process, from sketches to the finalized plots, facilitated a clear and effective representation of the data.
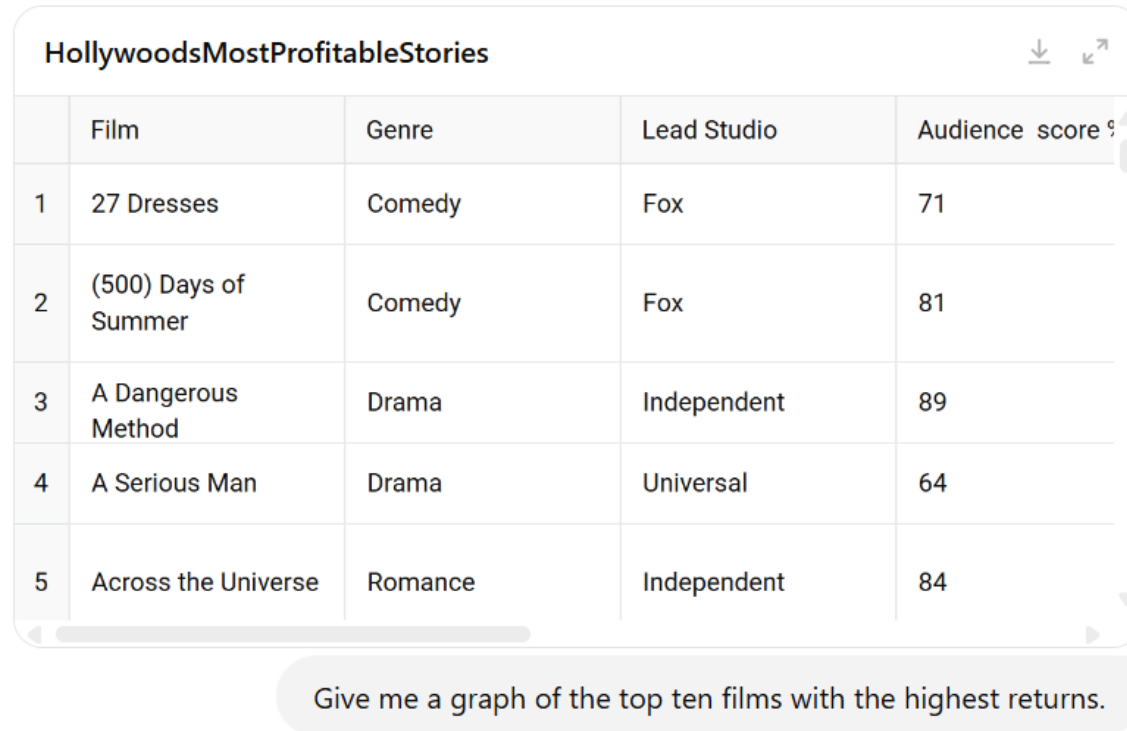
# 14 References

1. Task 1 "bad and/or manipulative visualization": https://wpdatatables.com/misleading-data-visualization-examples/

2. Task 2 "Improved version of the 'bad/manipulative' visualization": https://wpdatatables.com/misleading-data-visualization-examples/

3. Task 3 "A particularly good visualization": https://www.tableau.com/visualization/data-visualization-examples

4. Task 4 "Climate change visualisation": https://ourworldindata.org/co2-emissions

5. Task 5 "Black-and-white visualization": financial data using Python and the `yfinance` library.

6. Task 6 "Visualisation using colours": financial data using Python and the `yfinance` library.

7. Task 7 "Maximizing Clarity Through Tufte's Data-Ink Ratio Principle": https://www.kaggle.com/datasets/nitindatta/finance-data?resource=download

8. Task 8 "Network Graph": financial data using Python and the `yfinance` library.

9. Task 9 "Visualization by hand": https://public.tableau.com/app/learn/sample-data

10. Task 10 "visualization with ChatGPT": https://public.tableau.com/app/learn/sample-data

11. Task 11 "Climate change visualisation": https://ourworldindata.org/co2-emissions

12. Task 12 "Interactive Visualization": https://www.kaggle.com/datasets/taweilo/wine-quality-dataset-balanced-classification

# 15   Appendix

## 15.1   Visualization with ChatGPT screen chat

| | Film | Genre | Lead Studio | Audience score |
|---|---|---|---|---|
| | | | | |
| 1 | 27 Dresses | Comedy | Fox | 71 |
| 2 | (500) Days of Summer | Comedy | Fox | 81 |
| 3 | A Dangerous Method | Drama | Independent | 89 |
| 4 | A Serious Man | Drama | Universal | 64 |
| 5 | Across the Universe | Romance | Independent | 84 |

**HollywoodsMostProfitableStories**

Give me a graph of the top ten films with the highest returns.

## Top 10 Films With Highest Returns
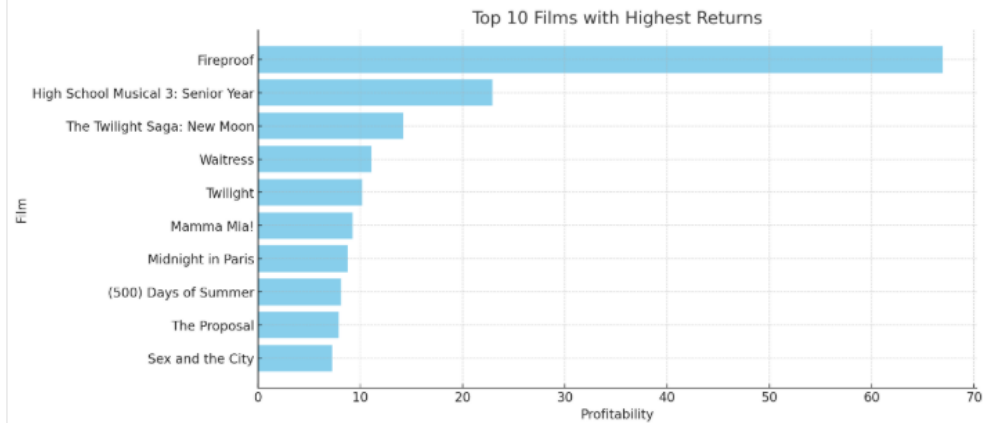


Here is a graph showing the top 10 films with the highest returns based on their profitability. Let me know if you'd like further analysis or a different visualization! [>-]

Message ChatGPT