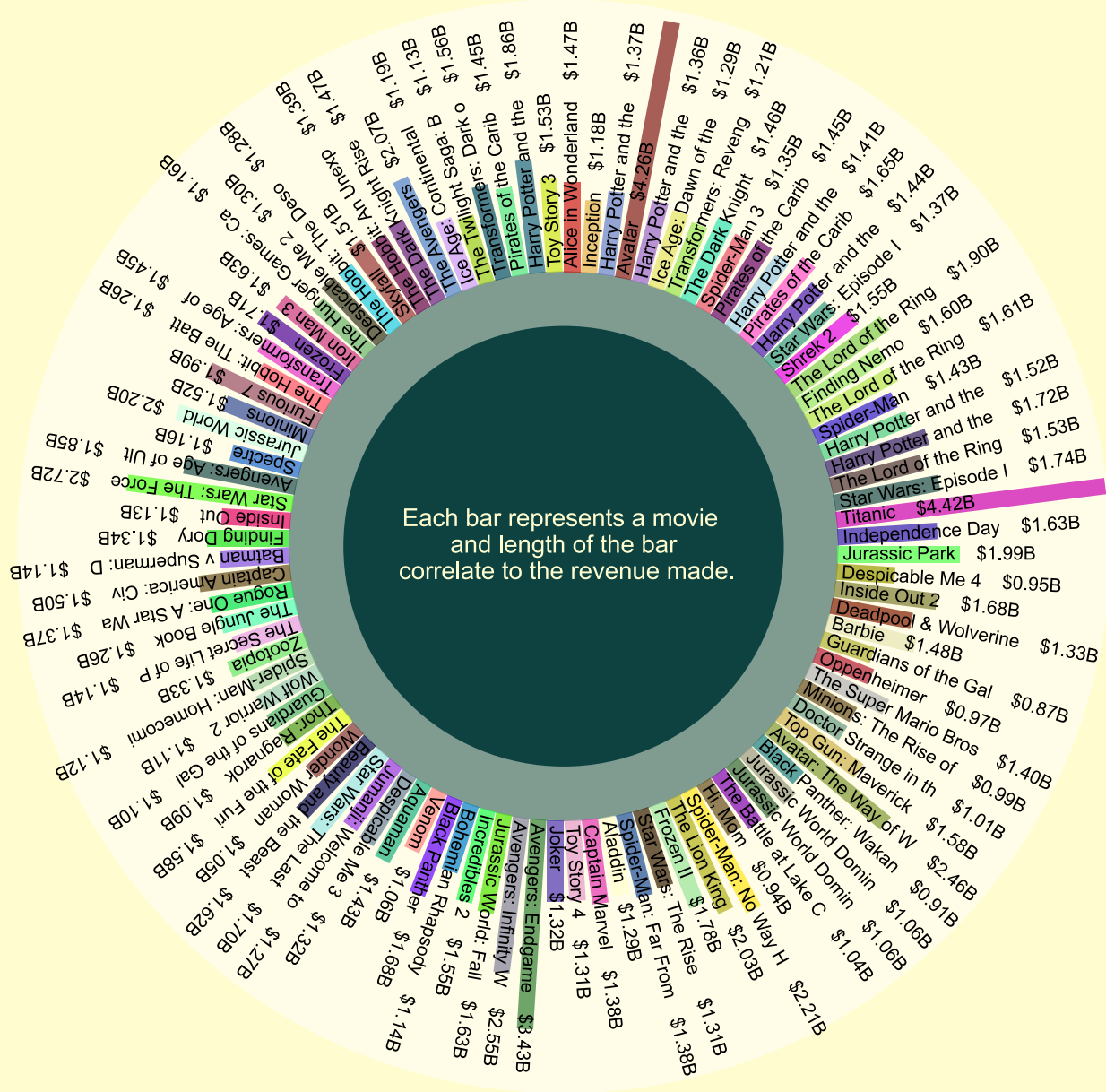


PYF200 Python Fundamentals



STUDENT MAME: MOHAMED ALI
STUDENT ID: 0
DATE: 04/12/2024

Table of Contents

Introduction	1
Recap: Assessment 2	2
Adjusting movie budget and revenue	3
DATA VISUALISATION	
Timeline of Movie Releases and Earnings	5
Top 10 Movies With the Highest Revenues	6
Visualising the Top 100 Highest Grossing Movies	7
Total Revenues by Genre and the Year	8
Genre Count By Year	9
Total Genre Counts	10
Changes in Total Revenues and Budgets Across the Years	11 – 12
References	13 – 14

Introduction

This project explores top 100 movies with the highest revenues. The revenue and budget of each movie were adjusted for inflation using the United States Consumer Price Index (CPI) and with a base year of 2024. The aim is to analyse factors such as budget, revenue made, and genres using CPI adjusted data to make meaningful comparisons. The data of the top 100 movies by revenue will be visually analysed to identify trends and patterns.

Recap: Assessment 2

In the last phase of this project, I have developed a hypothesis that genres such as Action, Adventure, and Drama are popular and therefore will perform best in terms of financial success. The data of the top 100 movies with highest revenues were scraped from the TMDB API and saved into a local Json file. The data was cleaned up, and basic analyses were performed.

Adjusting movie budget and revenue

A CSV file containing the CPI of the United States (Fred, 2024) was retrieved to extract the CPI value for each year in the data. A CPI calculation formula was used to calculate the inflation for a base year (How to Use the Consumer Price Index for Escalation : U.S. Bureau of Labor Statistics, 2023).

* Base Year = 2024.

* $\text{CPI} = (\text{Base Year Inflation} - \text{Target Year Inflation}) / \text{Target Year Inflation}$.

A CPI value was calculated for each year in the CSV file. Using pandas, the CSV file was merged with the movies DataFrame on the "year" column. This created a CPI column in the DataFrame. Subsequently, adjustedBudget and adjustedRevenues columns were created.

* $\text{adjustedBudget} = \text{CPI} * \text{budget}$

* $\text{adjustedRevenues} = \text{CPI} * \text{revenue}$

NB: The revenue adjustment might be unnecessary, since almost all these movies are currently making revenues. These adjustments were made for demonstrational purposes

```
inflationDF = pd.read_csv("CPIAUCSL.csv")
inflationDF
```

```
yearly_inflations["CPI"] = (base_year_CPI - yearly_inflations["CPIAUCSL"]) / yearly_inflations["CPIAUCSL"]
yearly_inflations
```

	year	CPIAUCSL
0	1947	22.1550
1	1948	24.1000
2	1949	23.8300
3	1950	23.9750
4	1951	25.9250
...
73	2020	259.0760
74	2021	271.3290
75	2022	294.9865
76	2023	304.3155
77	2024	313.2160

78 rows × 2 columns

	year	CPIAUCSL	CPI
0	1947	22.1550	13.137486
1	1948	24.1000	11.996515
2	1949	23.8300	12.143768
3	1950	23.9750	12.064275
4	1951	25.9250	11.081620
...
73	2020	259.0760	0.208973
74	2021	271.3290	0.154377
75	2022	294.9865	0.061798
76	2023	304.3155	0.029248
77	2024	313.2160	0.000000

78 rows × 3 columns

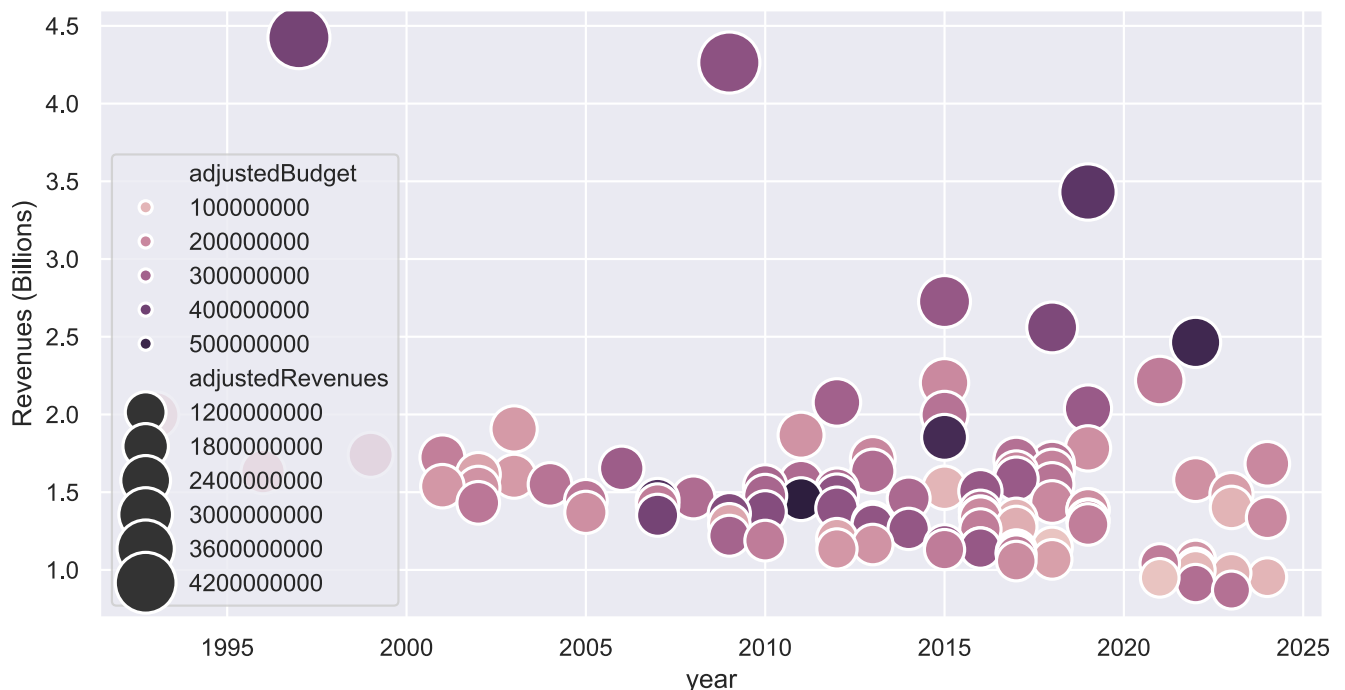
DATA VISUALISATION

Timeline of Movie Releases and Earnings

A scatter plot was created to visualise the relationship between release year and revenue for the top 100 movies with the highest revenue. Each data point in the plot represents a movie, the size of the movie correlates to its revenue while the colour (light to dark) represents the budget. This plot shows that most movies in the data were released between 2010 to 2019.

[22]:	title	genres	year	budget	revenue	CPIAUCSL	CPI	adjustedBudget	adjustedRevenues
0	Avatar	[Action, Adventure, Fantasy, Science Fiction]	2009	237000000	2923706026	214.76	0.46	345655072.22	4264108934.89
1	Avengers: Endgame	[Adventure, Science Fiction, Action]	2019	356000000	2799439100	255.55	0.23	436334699.02	3431158475.07
2	Avatar: The Way of Water	[Science Fiction, Adventure, Action]	2022	460000000	2320250281	294.99	0.06	488426961.91	2463636512.23
3	Titanic	[Drama, Romance]	1997	200000000	2264162353	160.30	0.95	390787273.86	4424029167.54
4	Star Wars: The Force Awakens	[Adventure, Action, Science Fiction]	2015	245000000	2068223624	237.58	0.32	323001631.05	2726692260.90
...
95	Inception	[Action, Science Fiction, Adventure]	2010	160000000	825532764	217.55	0.44	230362520.20	1188573800.13
96	Wonder Woman	[Action, Adventure, Fantasy]	2017	149000000	823970682	244.22	0.28	191096413.86	1056764043.33
97	Hi, Mom	[Drama, Comedy, Fantasy]	2021	59000000	822049668	271.33	0.15	68108252.34	948955359.77
98	Spider-Man	[Action, Science Fiction]	2002	139000000	821708551	179.80	0.74	242141401.56	1431436404.39
99	Independence Day	[Action, Adventure, Science Fiction]	1996	75000000	817400891	156.85	1.00	149768568.70	1632279486.61

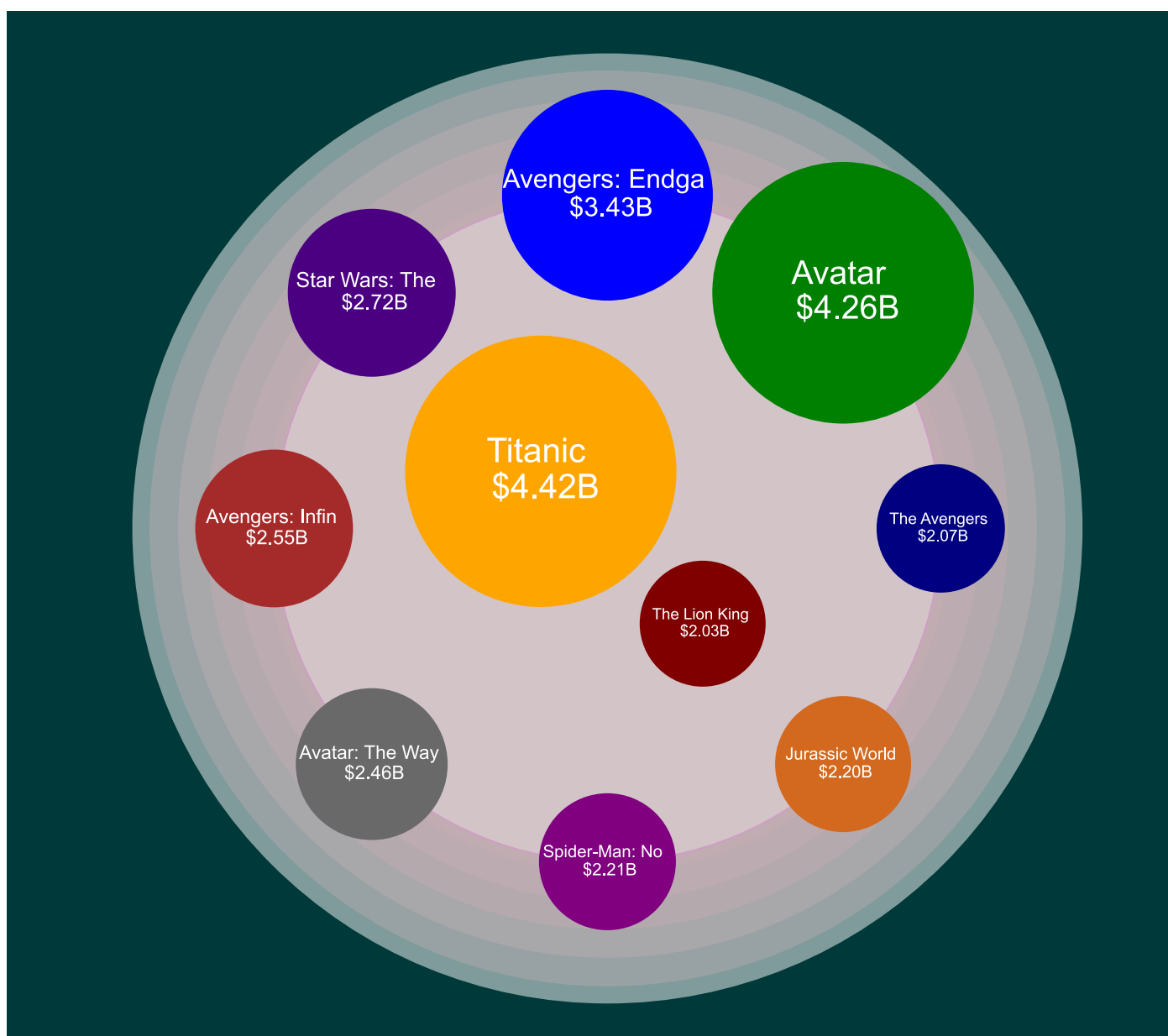
100 rows × 9 columns



Top 10 Movies With the Highest Revenues

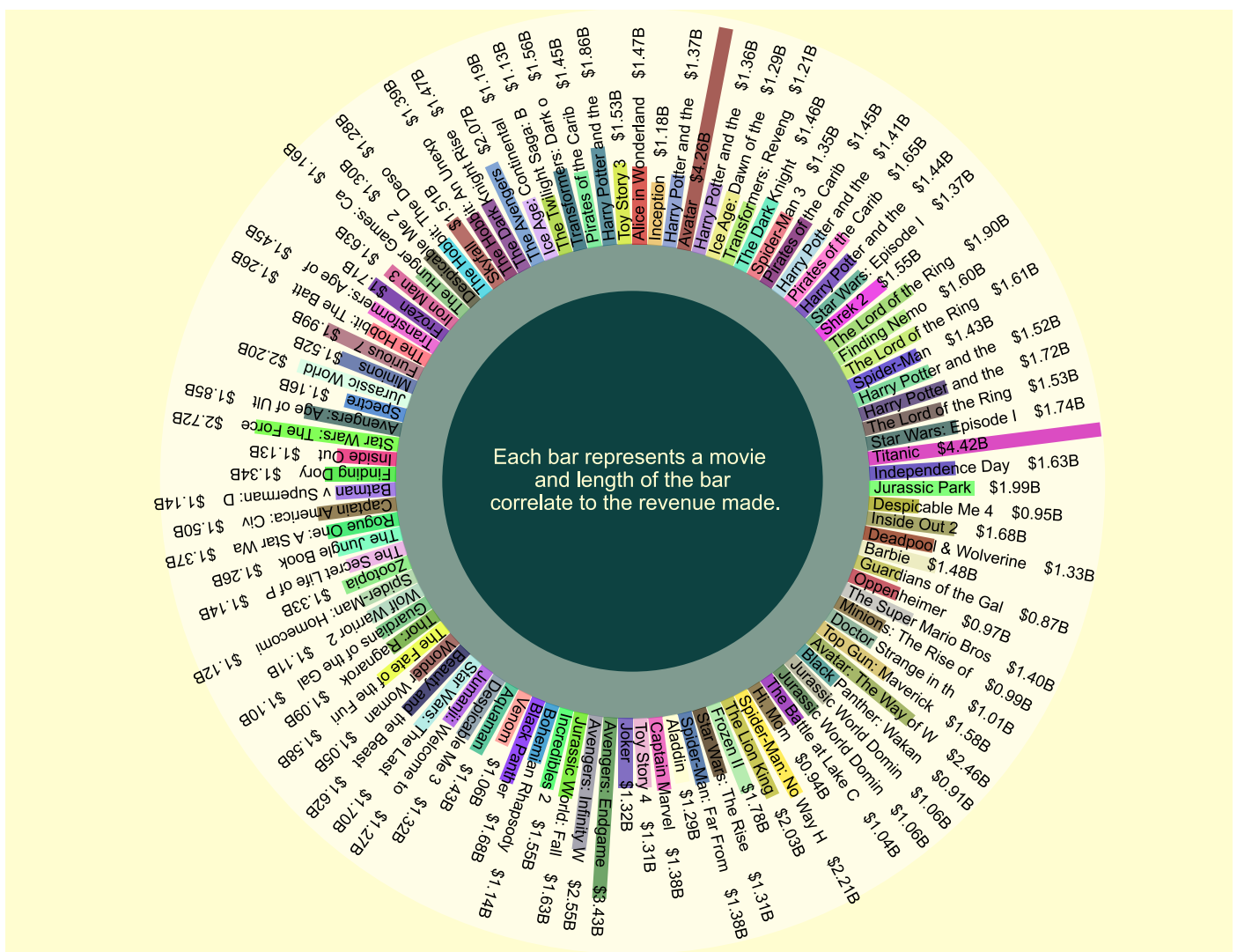
Using plotting tools, the revenues of the top 10 movies with the highest revenues were visualised. The following figure visualises each movie with a circle of a different colour. The size of the circle corresponds to revenues the movie made.

	title	genres	year	budget	revenue	CPIAUCSL	CPI	adjustedBudget	adjustedRevenues
3	Titanic	[Drama, Romance]	1997	200000000	2264162353	160.30	0.95	390787273.86	4424029167.54
0	Avatar	[Action, Adventure, Fantasy, Science Fiction]	2009	237000000	2923706026	214.76	0.46	345655072.22	4264108934.89
1	Avengers: Endgame	[Adventure, Science Fiction, Action]	2019	356000000	2799439100	255.55	0.23	436334699.02	3431158475.07
4	Star Wars: The Force Awakens	[Adventure, Action, Science Fiction]	2015	245000000	2068223624	237.58	0.32	323001631.05	2726692260.90
5	Avengers: Infinity War	[Adventure, Action, Science Fiction]	2018	300000000	2052415039	251.12	0.25	374188821.10	2559969212.86
2	Avatar: The Way of Water	[Science Fiction, Adventure, Action]	2022	460000000	2320250281	294.99	0.06	488426961.91	2463636512.23
6	Spider-Man: No Way Home	[Action, Adventure, Science Fiction]	2021	200000000	1921847111	271.33	0.15	230875431.67	2218536406.79
8	Jurassic World	[Action, Adventure, Science Fiction, Thriller]	2015	150000000	1671537444	237.58	0.32	197756100.64	2203711513.34
10	The Avengers	[Science Fiction, Action, Adventure]	2012	220000000	1518815515	229.00	0.37	300910142.93	2077395425.91
9	The Lion King	[Adventure, Drama, Family, Animation]	2019	260000000	1663000000	255.55	0.23	318671409.40	2038271360.87



Visualising the Top 100 Highest Grossing Movies

Using plotting tools, the top 100 movies were visualised in terms of revenue. Each bar in the following figure represents a movie in the data, and the length of the bar corresponds to the revenue made by the movie. The bars were coloured with different RGB colours to create contrast. The figure shows Titanic and Avatar have the highest revenues respectively, while Guardians of the Galaxy and Black Panther: Wakanda Forever have the lowest revenues respectively. The data was sorted by year to get a random sequence of high and low revenues.

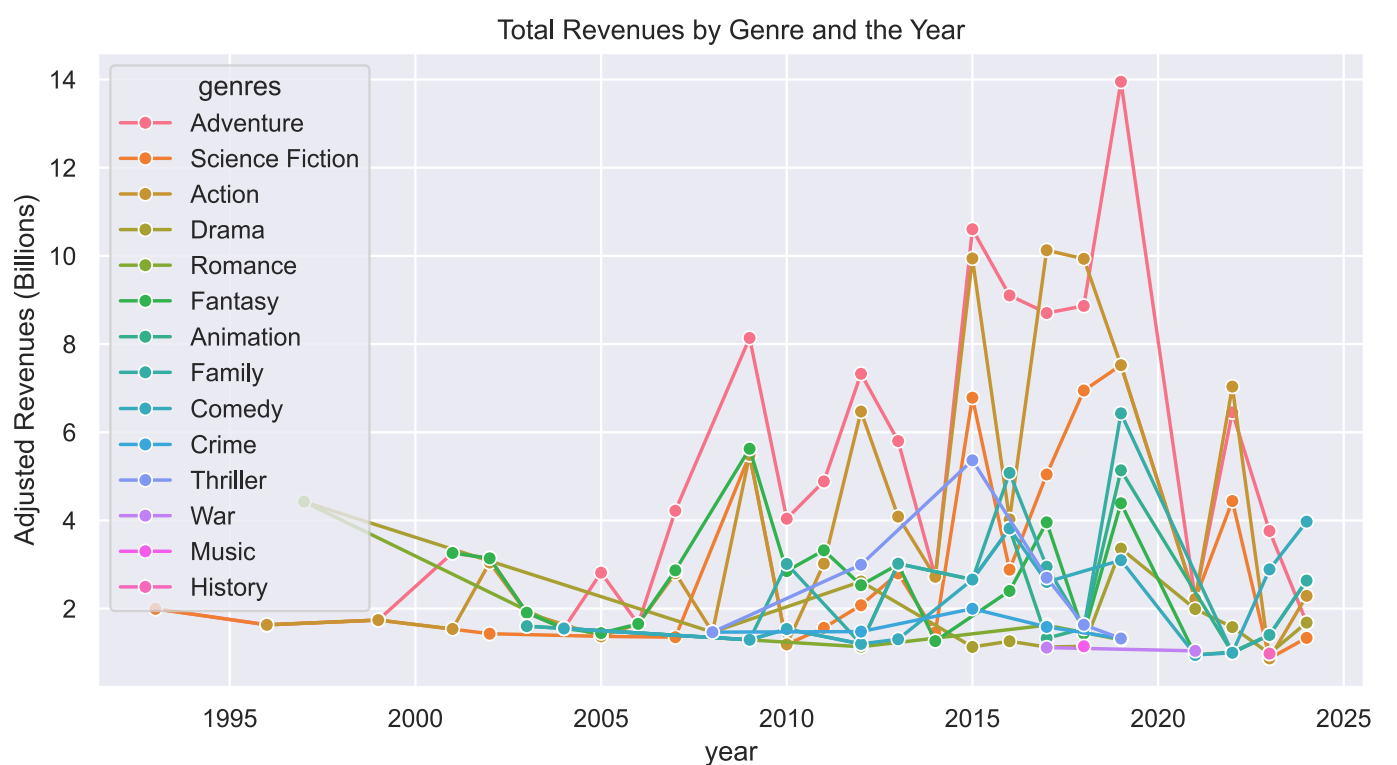
[illegible]

Total Revenues by Genre and the Year

The genres column in the DataFrame (data) was exploded using pandas, then was grouped by “genres” and “year”. This allowed to analyse the “genres” factor in the data. Each line in the following line plot figure represents a genre, the figure demonstrates how each genre performed in terms of revenue over the years. The figure demonstrates how movies with Adventure genre performed well over the years.

	year	genres	title	budget	revenue	CPIAUCSL	CPI	adjustedBudget	adjustedRevenues
0	1993	Adventure	Jurassic Park	63000000	920100000	144.40	1.17	136652409.97	1995775911.36
1	1993	Science Fiction	Jurassic Park	63000000	920100000	144.40	1.17	136652409.97	1995775911.36
2	1996	Action	Independence Day	75000000	817400891	156.85	1.00	149768568.70	1632279486.61
3	1996	Adventure	Independence Day	75000000	817400891	156.85	1.00	149768568.70	1632279486.61
4	1996	Science Fiction	Independence Day	75000000	817400891	156.85	1.00	149768568.70	1632279486.61
...
158	2024	Animation	Inside Out 2Despicable Me 4	300000000	2635777087	626.43	0.00	300000000.00	2635777087.00
159	2024	Comedy	Inside Out 2Deadpool & WolverineDespicable Me 4	500000000	3972593199	939.65	0.00	500000000.00	3972593199.00
160	2024	Drama	Inside Out 2	200000000	1682636477	313.22	0.00	200000000.00	1682636477.00
161	2024	Family	Inside Out 2Despicable Me 4	300000000	2635777087	626.43	0.00	300000000.00	2635777087.00
162	2024	Science Fiction	Deadpool & Wolverine	200000000	1336816112	313.22	0.00	200000000.00	1336816112.00

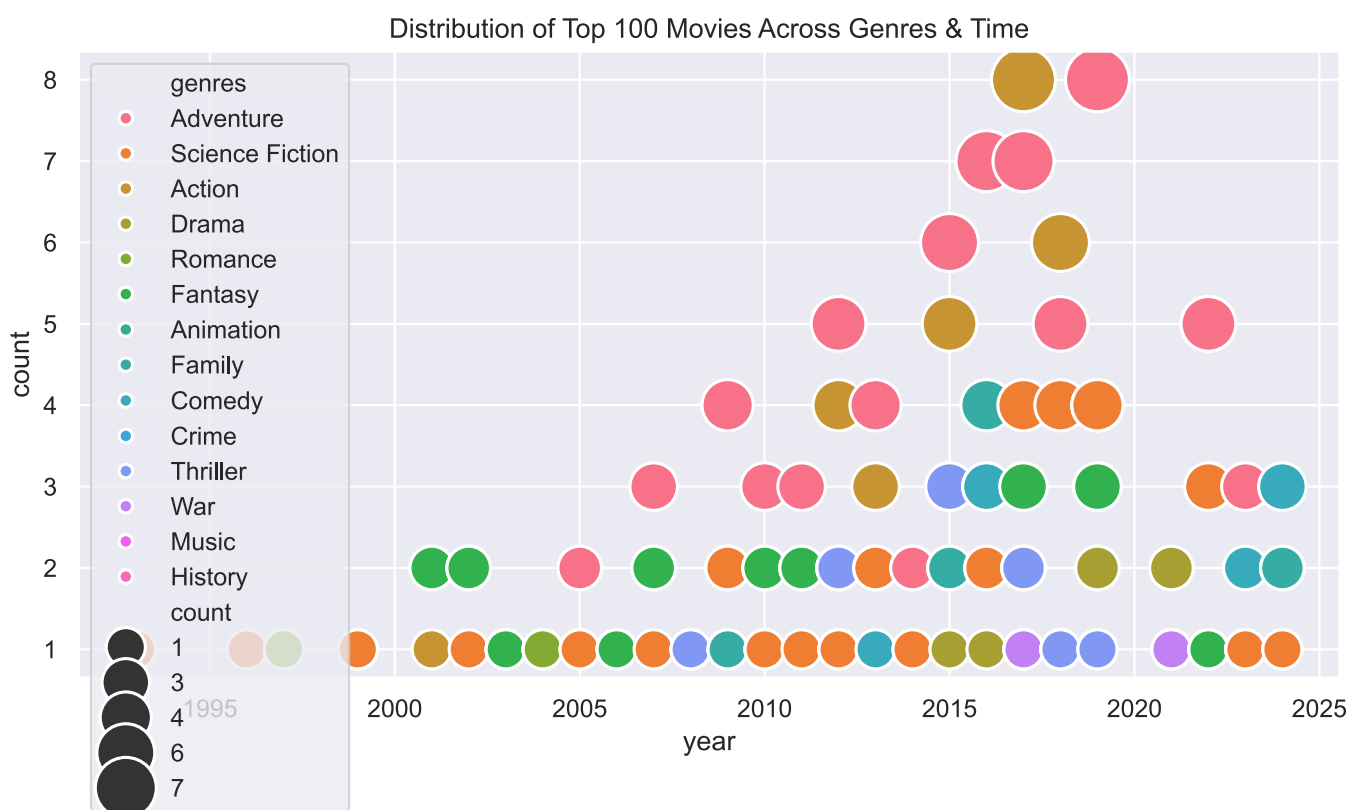
163 rows × 9 columns



Genre Count By Year

Additionally, the number of times each genre appeared in movies released each year was analysed. The following scatter plot shows the number of times each genre appeared in these movies each year. Colour represents genre and the size represents the number of times it appeared in a year (count). The figure demonstrates how adventure genre has been well represented in the movies over the years.

	year	genres	count
0	1993	Adventure	1
1	1993	Science Fiction	1
2	1996	Action	1
3	1996	Adventure	1
4	1996	Science Fiction	1
...
158	2024	Animation	2
159	2024	Comedy	3
160	2024	Drama	1
161	2024	Family	2
162	2024	Science Fiction	1

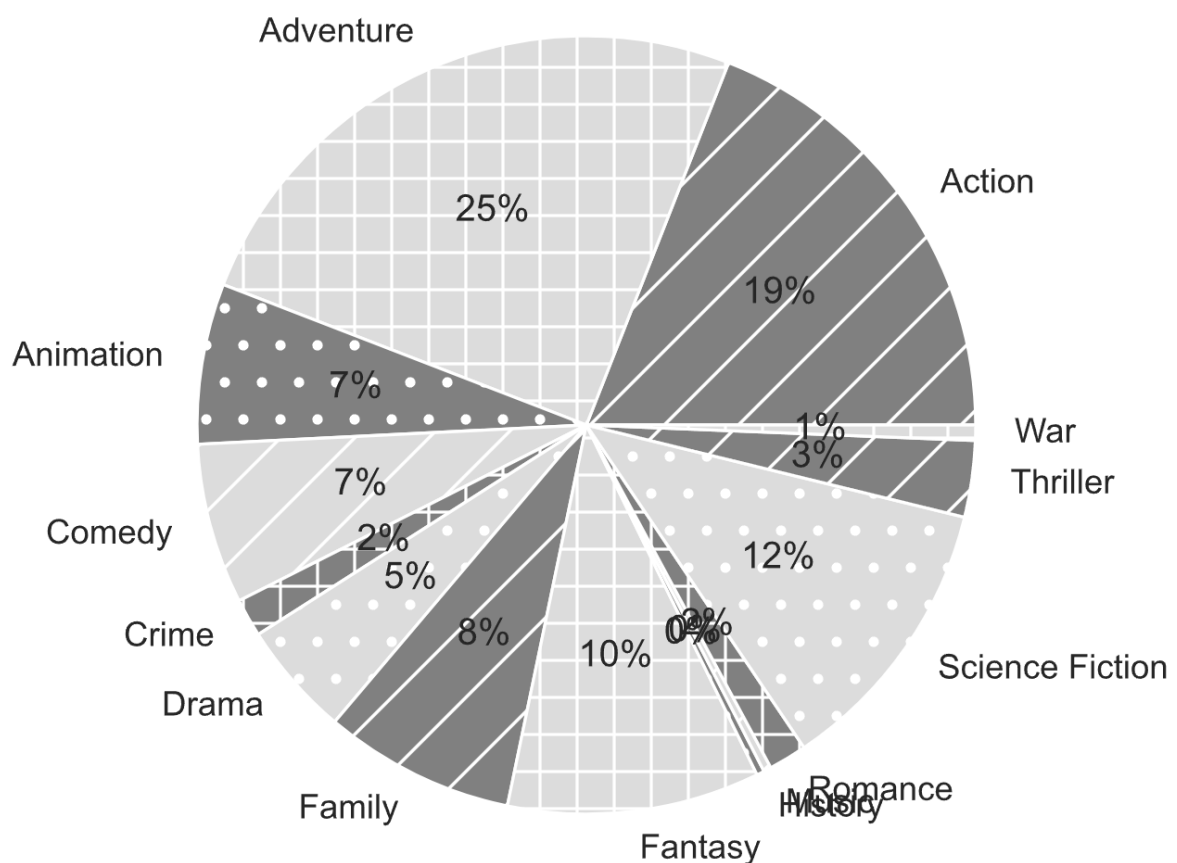


Total Genre Counts

Subsequently, the data was grouped by “genre” column again, to analyse the popularity of each genre. This produced the total count of each genre across the data set. The following pie chart demonstrates this in percentages.

	genres	year	count
0	Action	48271	60
1	Adventure	50260	79
2	Animation	28205	21
3	Comedy	26205	21
4	Crime	10071	5
5	Drama	24192	15
6	Family	28205	25
7	Fantasy	40233	33
8	History	2023	1
9	Music	2018	1
10	Romance	10049	5
11	Science Fiction	42246	37
12	Thriller	12089	10
13	War	4038	2

Genre Popularity in Top 100 Highest Revenue Genreting Movies

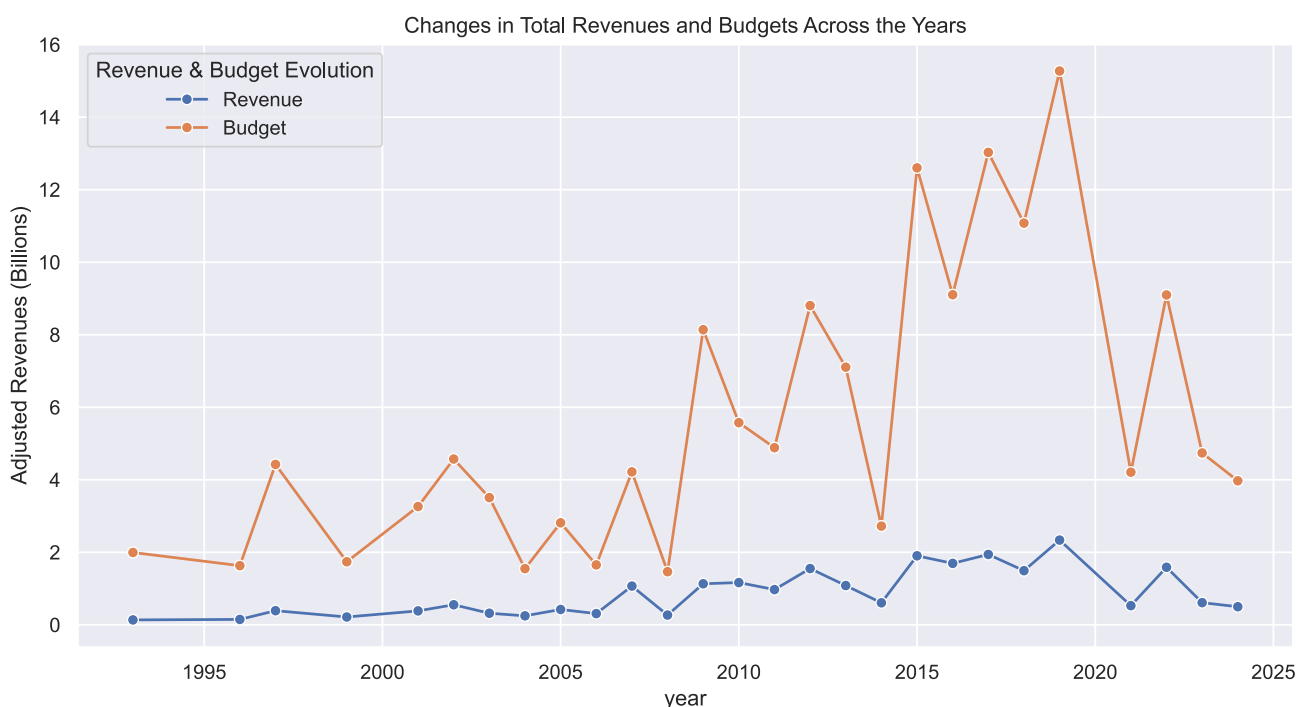


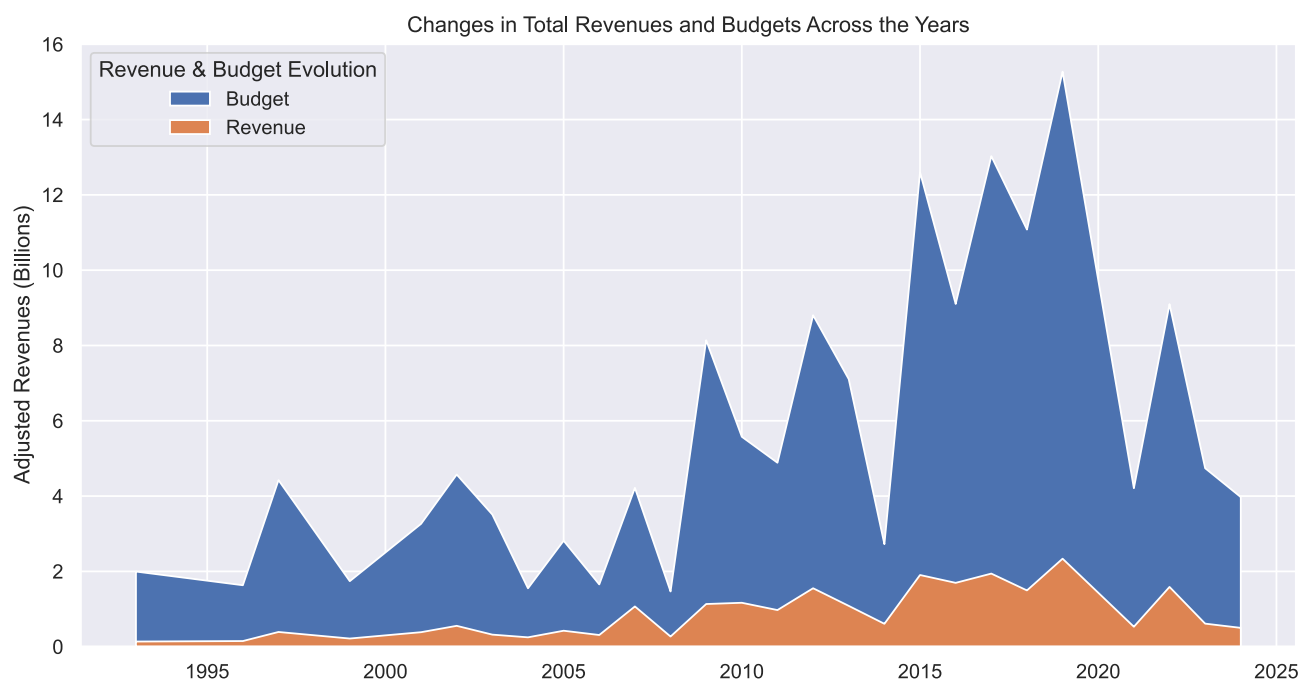
Changes in Total Revenues and Budgets Across the Years

The data was grouped into “year” column to analyse the changes in total revenues and budget across the years. The following figures demonstrate that there’s a strong correlation between the movie budgets and revenues over the years.

Correlation between budgets and revenues over the years: 0.9601709040673926

	year	title	genres	budget	revenue	CPIAUCSL	CPI	adjustedBudget	adjustedRevenues
0	1993	Jurassic Park	[Adventure, Science Fiction]	63000000	920100000	144.40	1.17	136652409.97	1995775911.36
1	1996	Independence Day	[Action, Adventure, Science Fiction]	75000000	817400891	156.85	1.00	149768568.70	1632279486.61
2	1997	Titanic	[Drama, Romance]	200000000	2264162353	160.30	0.95	390787273.86	4424029167.54
3	1999	Star Wars: Episode I – The Phantom Menace	[Adventure, Action, Science Fiction]	115000000	924317558	166.35	0.88	216530447.85	1740372998.18
4	2001	Harry Potter and the Philosopher's StoneThe Lo...	[Adventure, Fantasy, Adventure, Fantasy, Action]	218000000	1847843914	354.80	1.53	384899030.44	3262538215.15
5	2002	The Lord of the Rings: The Two TowersHarry Pot...	[Adventure, Fantasy, Action, Adventure, Fantas...	318000000	2624684433	539.40	2.23	553963781.98	4572264512.61
6	2003	The Lord of the Rings: The Return of the KingF...	[Adventure, Fantasy, Action, Animation, Family]	188000000	2059224515	367.60	1.41	320373275.30	3509151608.76
7	2004	Shrek 2	[Animation, Family, Comedy, Fantasy, Adventure...	150000000	935454538	189.00	0.66	248584126.98	1550260997.75
8	2005	Harry Potter and the Goblet of FireStar Wars: ...	[Adventure, Fantasy, Adventure, Action, Scienc...	263000000	1745921036	388.60	1.22	423961955.74	2814464247.10
9	2006	Pirates of the Caribbean: Dead Man's Chest	[Adventure, Fantasy, Action]	200000000	1065700000	201.85	0.55	310345305.92	1653674962.60
10	2007	Pirates of the Caribbean: At World's EndHarry ...	[Adventure, Fantasy, Action, Adventure, Fantas...	708000000	2794196111	622.26	1.53	1069128009.31	4219425601.40
11	2008	The Dark Knight	[Drama, Action, Crime, Thriller]	185000000	1004558444	214.57	0.46	270045252.24	1466358045.42





THANK YOU

References

- B, E. (2013, February 16). python - How to remove frame from a figure - Stack Overflow. <https://stackoverflow.com/questions/14908576/how-to-remove-frame-from-a-figure>
- Burton, J. "Fluffy." (2018, December 6). python - How to draw a rectangle on image - Stack Overflow. <https://stackoverflow.com/questions/37435369/how-to-draw-a-rectangle-on-image>
- Corralien. (2022, January 31). python - Pandas: Get cell value by row index and column name - Stack Overflow. <https://stackoverflow.com/questions/70931002/pandas-get-cell-value-by-row-index-and-column-name>
- Dnswlt. (2019, April 8). python - How do I create a multiline plot using seaborn? - Stack Overflow. <https://stackoverflow.com/questions/52308749/how-do-i-create-a-multiline-plot-using-seaborn>
- Fred. (2024, November 13). Consumer Price Index for All Urban Consumers: All Items in U.S. City Average (CPIAUCSL) | FRED | St. Louis Fed. <https://fred.stlouisfed.org/series/CPIAUCSL>
- Fuentes, A. (2014, February 5). python - percent label position in pie chart - Stack Overflow. <https://stackoverflow.com/questions/21572870/percent-label-position-in-pie-chart%20##%20autopct='%1.0f%'>
- Hayden, A. (2013, June 16). python - Pandas DataFrame Groupby two columns and get counts - Stack Overflow. <https://stackoverflow.com/questions/17679089/pandas-dataframe-groupby-two-columns-and-get-counts>
- How to calculate summary statistics — pandas 2.2.3 documentation. (n.d.). Retrieved December 4, 2024, from https://pandas.pydata.org/docs/getting_started/intro_tutorials/06_calculate_statistics.html#min-tut-06-stats
- How to rotate text in Matplotlib – Python - GeeksforGeeks. (2022, December 12). <https://www.geeksforgeeks.org/how-to-rotate-text-in-matplotlib-python/>
- How to Set Plot Background Color in Matplotlib? - GeeksforGeeks. (2022, December 13). <https://www.geeksforgeeks.org/how-to-set-plot-background-color-in-matplotlib/>
- How to Use the Consumer Price Index for Escalation : U.S. Bureau of Labor Statistics. (2023, June 8). U.S. Bureau of Labor Statistics. <https://www.bls.gov/cpi/factsheets/escalation.htm>
- ImportanceOfBeingErnes. (2019, February 22). Matplotlib: plt.text with user-defined circle radii - Stack Overflow. <https://stackoverflow.com/questions/54831344/matplotlib-plt-text-with-user-defined-circle-radii>
- ImportanceOfBeingErnest. (2020, February 26). python - Rotating Rectangles around point with matplotlib - Stack Overflow. <https://stackoverflow.com/questions/60413174/rotating-rectangles-around-point-with-matplotlib>
- JJAN. (2019, May 24). python - How to suppress scientific notation when printing float values? - Stack Overflow. <https://stackoverflow.com/questions/658763/how-to-suppress-scientific-notation-when-printing-float-values>
- Jtlz2. (2019, February 9). python - Use .corr to get the correlation between two columns - Stack Overflow. <https://stackoverflow.com/questions/42579908/use-corr-to-get-the-correlation-between-two-columns>
- McKinney, T. (2015, May 22). python - pandas extract year from datetime: df['year'] = df['date'].year is not working - Stack Overflow. <https://stackoverflow.com/questions/30405413/pandas-extract-year-from-datetime-dfyear-dfdate-year-is-not-working>
- McKinney, T. (2020, August 13). python - Putting text in top left corner of matplotlib plot - Stack Overflow. <https://stackoverflow.com/questions/8482588/putting-text-in-top-left-corner-of-matplotlib-plot>
- Nelamali, N. (2024, November 4). How to Rename Column by Index in Pandas - Spark By [Examples]. <https://sparkbyexamples.com/pandas/pandas-rename-column-by-index/>
- Nouri, B. (2014, March 15). python - How to pass a list of pandas columns to a pyplot chart? - Stack Overflow. <https://stackoverflow.com/questions/22425313/how-to-pass-a-list-of-pandas-columns-to-a-pyplot-chart>

Pandas Read CSV. (n.d.). W3School. Retrieved December 4, 2024, from https://www.w3schools.com/python/pandas/pandas_csv.asp

pandas.DataFrame.explode — pandas 2.2.3 documentation. (n.d.). Retrieved December 4, 2024, from <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.explode.html>

pandas.DataFrame.reset_index — pandas 2.2.3 documentation. (n.d.). Retrieved December 4, 2024, from https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.reset_index.html

pandas.DataFrame.sort_values — pandas 2.2.3 documentation. (n.d.). Retrieved December 4, 2024, from https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.sort_values.html

Płońska, A., & Płoński, P. (2022, November 21). 9 ways to set colors in Matplotlib | MLJAR. <https://mljar.com/blog/matplotlib-colors/>

python - How to pass a list of pandas columns to a pyplot chart? - Stack Overflow. (n.d.). Retrieved December 4, 2024, from <https://stackoverflow.com/questions/22425313/how-to-pass-a-list-of-pandas-columns-to-a-pyplot-chart>

S, S. (2019, March 19). python - How can I get the output of a matplotlib plot as an SVG? - Stack Overflow. <https://stackoverflow.com/questions/24525111/how-can-i-get-the-output-of-a-matplotlib-plot-as-an-svg>

seaborn.scatterplot — seaborn 0.13.2 documentation. (n.d.). Retrieved December 4, 2024, from <https://seaborn.pydata.org/generated/seaborn.scatterplot.html>

Tarek. (2018, June 12). pandas - How do I compare columns in different data frames? - Data Science Stack Exchange. <https://datascience.stackexchange.com/questions/33053/how-do-i-compare-columns-in-different-data-frames>

Text Rotation Mode — Matplotlib 3.9.3 documentation. (n.d.). Retrieved December 4, 2024, from https://matplotlib.org/stable/gallery/text_labels_and_annotations/demo_text_rotation_mode.html

Thompson, G. (2019, March 11). python - How to edit a seaborn legend title and labels for figure-level functions - Stack Overflow. <https://stackoverflow.com/questions/45201514/how-to-edit-a-seaborn-legend-title-and-labels-for-figure-level-functions>

Y, D. (2022, October 15). Write JSON data to a file in Python | Sentry. <https://sentry.io/answers/write-json-data-to-a-file-in-python/>