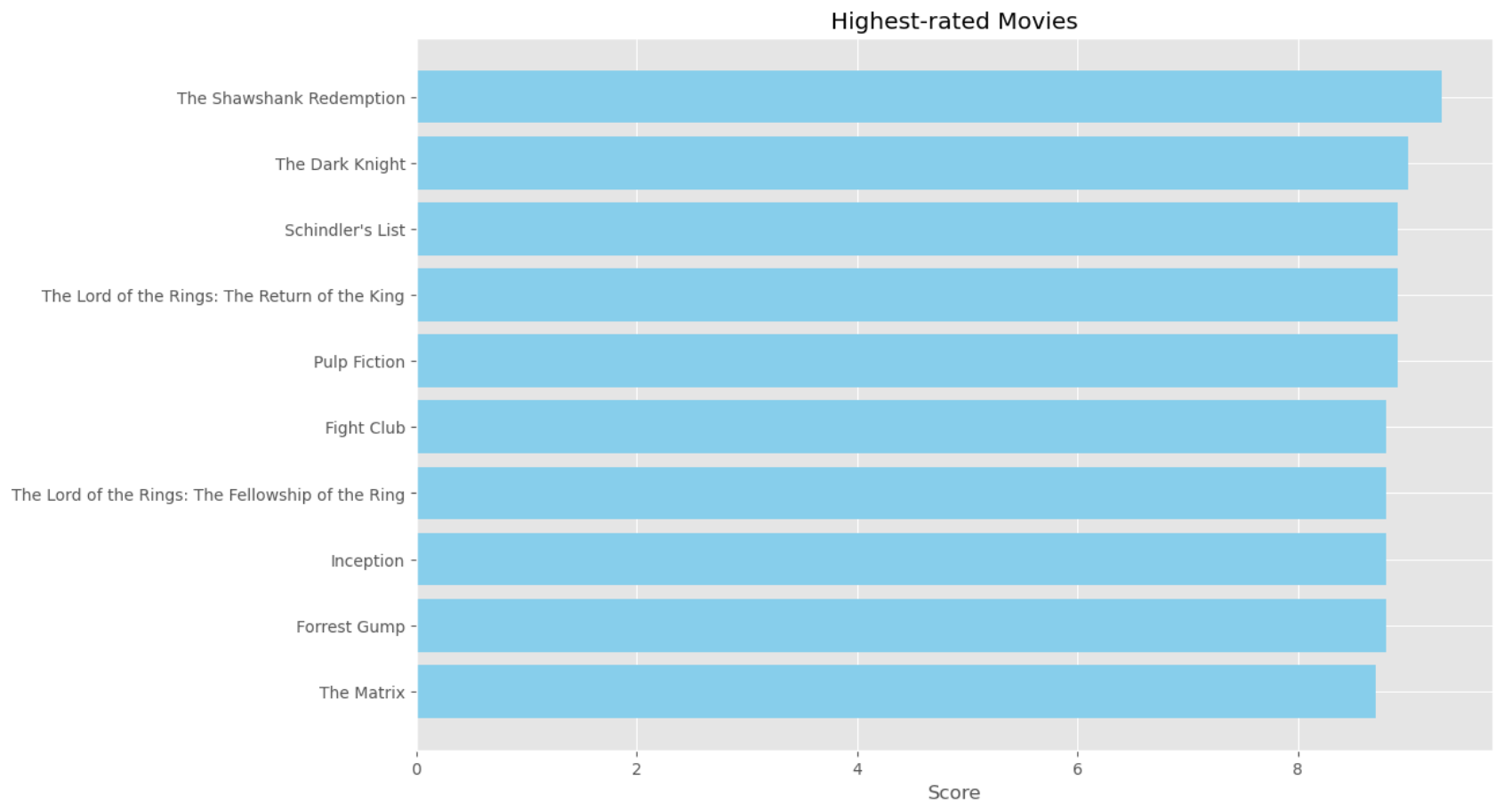
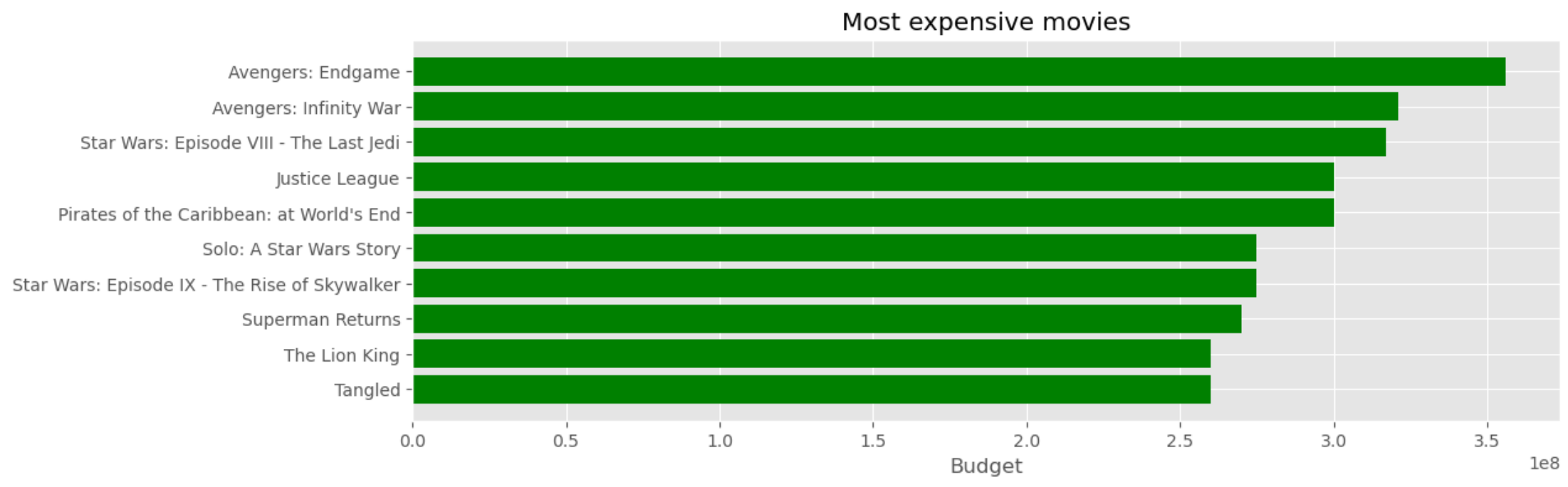
Movie data analysis for Microsoft

MARK K KURIA

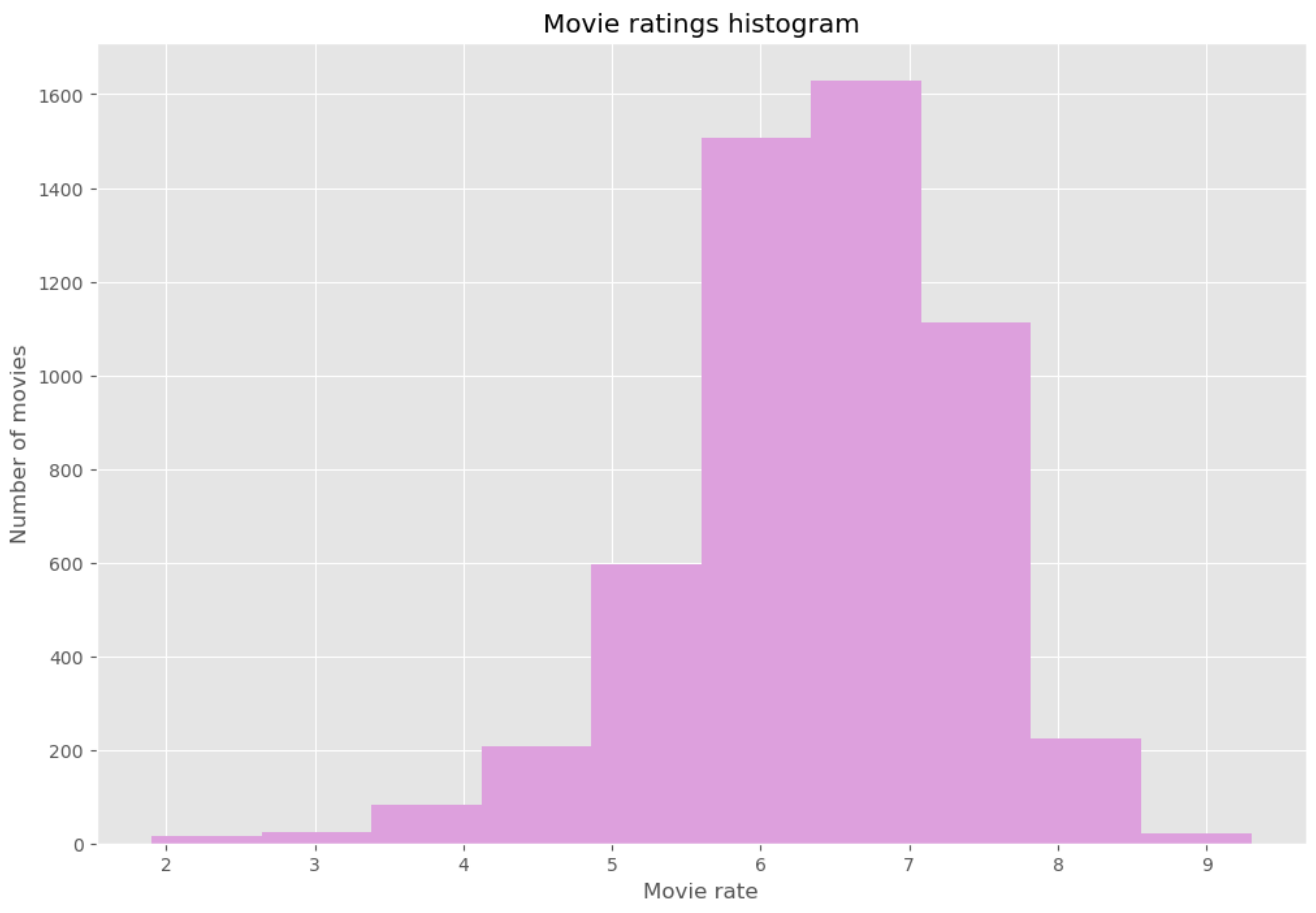
Here, the code generates a horizontal bar chart showing the names and scores of the top 10 highest-rated movies based on the 'score' column in the DataFrame df. The chart is colored in sky blue, and the y-axis is inverted to have the highest-rated movie at the top.
Here we can see that Shawshank redemption, a personal favorite of mine, is leading the lot and the rest of the 9 not too far apart in rating.



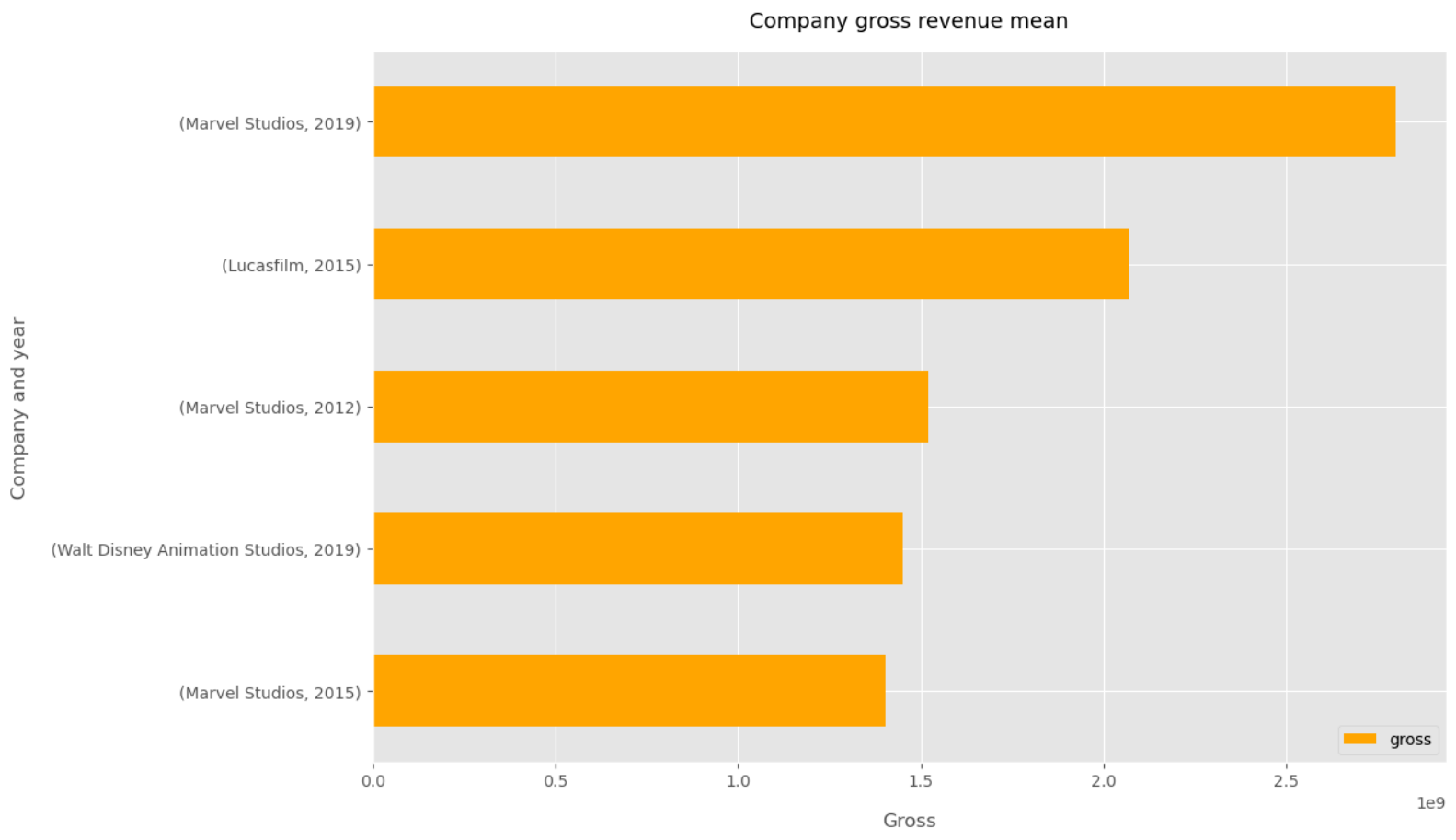
This diagram is focused on visualizing the budget information of movies.
‘df.sort_values('budget', ascending=False) sorts the DataFrame df based on the 'budget' column in descending order (ascending=False). The result is stored in the variable budget_ranking’.
Here, a horizontal bar chart shows the names and budgets of the top 10 most expensive movies based on the 'budget' column in the DataFrame df. The chart is colored in green, and the y-axis is inverted to have the movie with the highest budget at the top. The figure size is also adjusted to provide a clearer visualization.
In this diagram we see the huge budget that Marvel allocated to its two highest selling movies and the fruits of the investment are seen in the Box Office records they shattered.



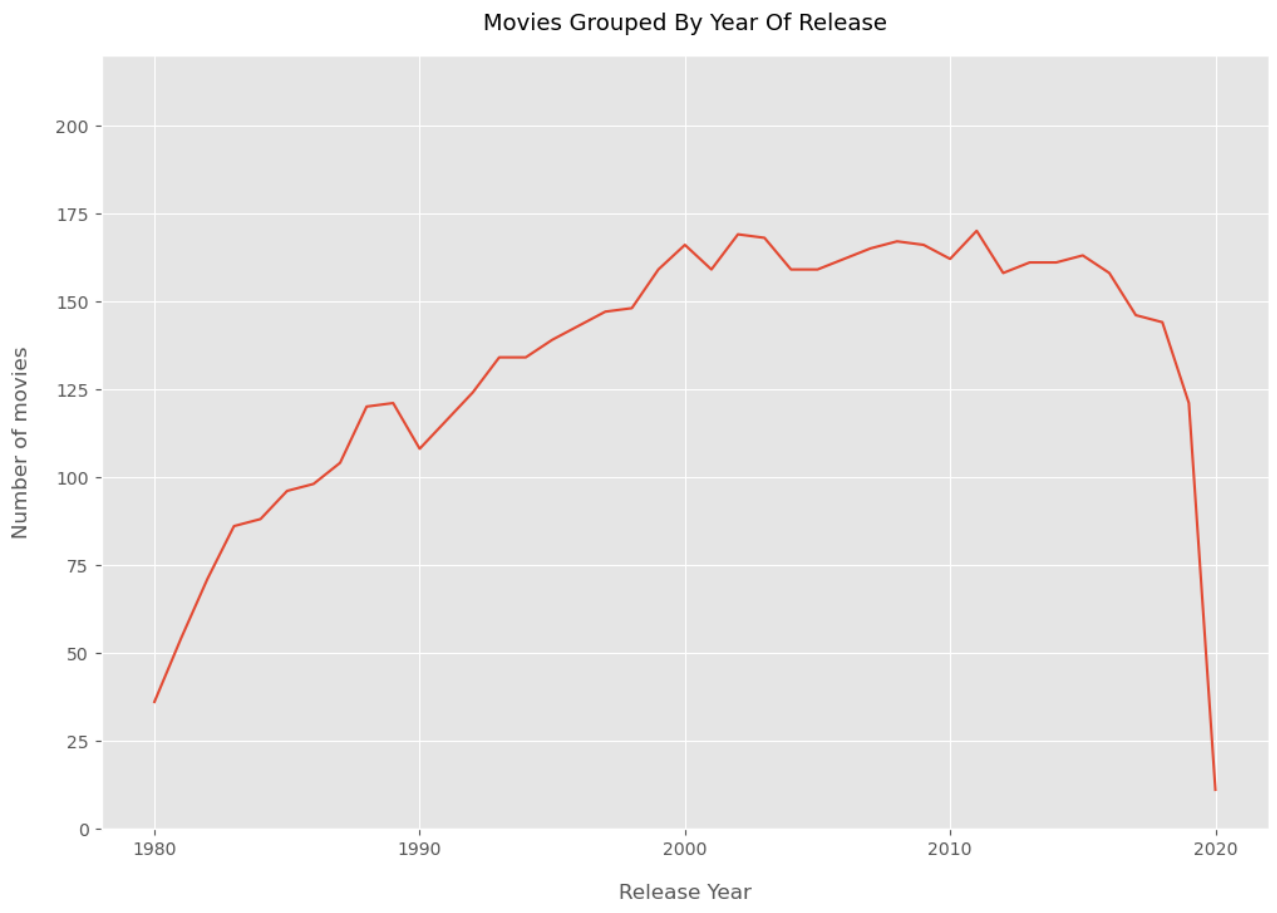
This code generates a histogram of movie ratings from the 'score' column in the DataFrame df. The x-axis represents the movie ratings, and the y-axis represents the number of movies falling within each rating interval. The histogram is colored in plum, and labels and a title are added for better interpretation.
In this diagram, we can derive that a lot of movies ever rated generally tend to belong in the ‘ above average’ category, that is 6 and 7. The graph also shows that as much as theres very few terrible movies out there, theres equally a very low amount of outstanding movies. I would say that’s a pretty normal graph.



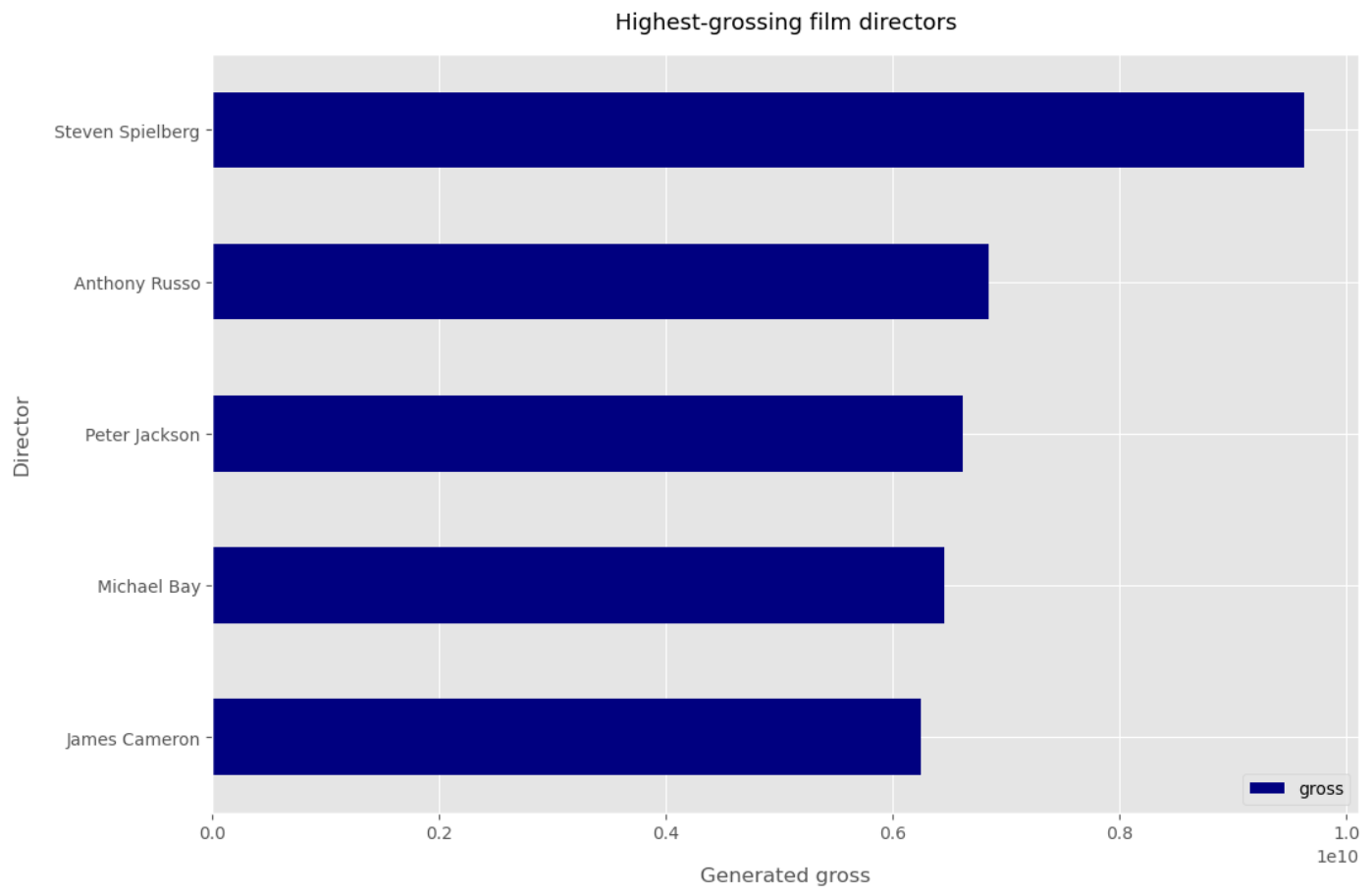
The code here generates a horizontal bar chart representing the top 5 company-year combinations with the highest mean gross revenue. The chart is colored in orange, and labels and a title are added for better interpretation. The y-axis is inverted to have the highest mean gross at the top.
The graph generally analyzes that 2019 , 2015 and 2012 provided movies companies with some of their highest revenues. Marvel has three years in the top 5 in 2012, 2015 and 2019 where they generated the years highest revenue mean.



Here, this code creates a line plot to visualize the distribution of movies released each year. The x-axis represents the release years, the y-axis represents the number of movies released in each year, and labels and a title are added for better interpretation. The y-axis limits and x-axis ticks are adjusted for better visualization.
Since the year 1980 there was a incline in movies created up until 2010 where there became a nnoticeable decline in prodeuction. Specifically around 2020, where the world experience a pandemic which led to closing down of many companies and more impactfully cinemas.
Audiences could not watch as many mocies and eventually movies companies couldn’t make as may either.



Here, the code generates a horizontal bar chart that showcases the total gross revenue generated by the top 5 film directors. The chart is colored in navy, and labels and a title are added for better interpretation. The y-axis is inverted to have the director with the highest total gross at the top.
This bar chart shows us that out these 5 legendary movie directors, Steven Spielberg has created the movies which have generated the most gross revenue worldwide.



Finally, here our code generates a horizontal bar chart that visualizes the count of movies for the top 5 genres in the DataFrame df. Labels and a title are added for better interpretation. The chart represents the number of movies on the x-axis and the movie genres on the y-axis.

Well, this is a downer to most action movies lovers I'm sure. The data we've used here explains that comedy genre movies have been produced the most number of times which is generally also an indication that they have been the most successful. Adventure eg. Nat Geo and crime movies eg. Crime Investigation have the least number of created movies.