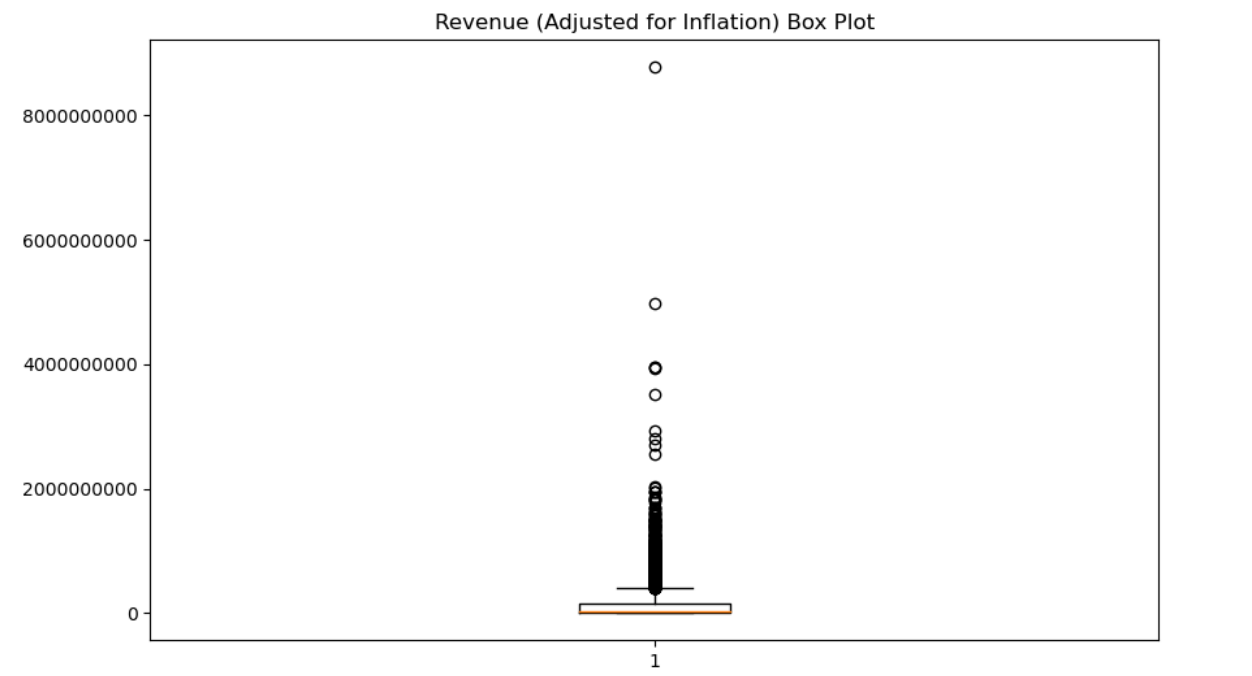
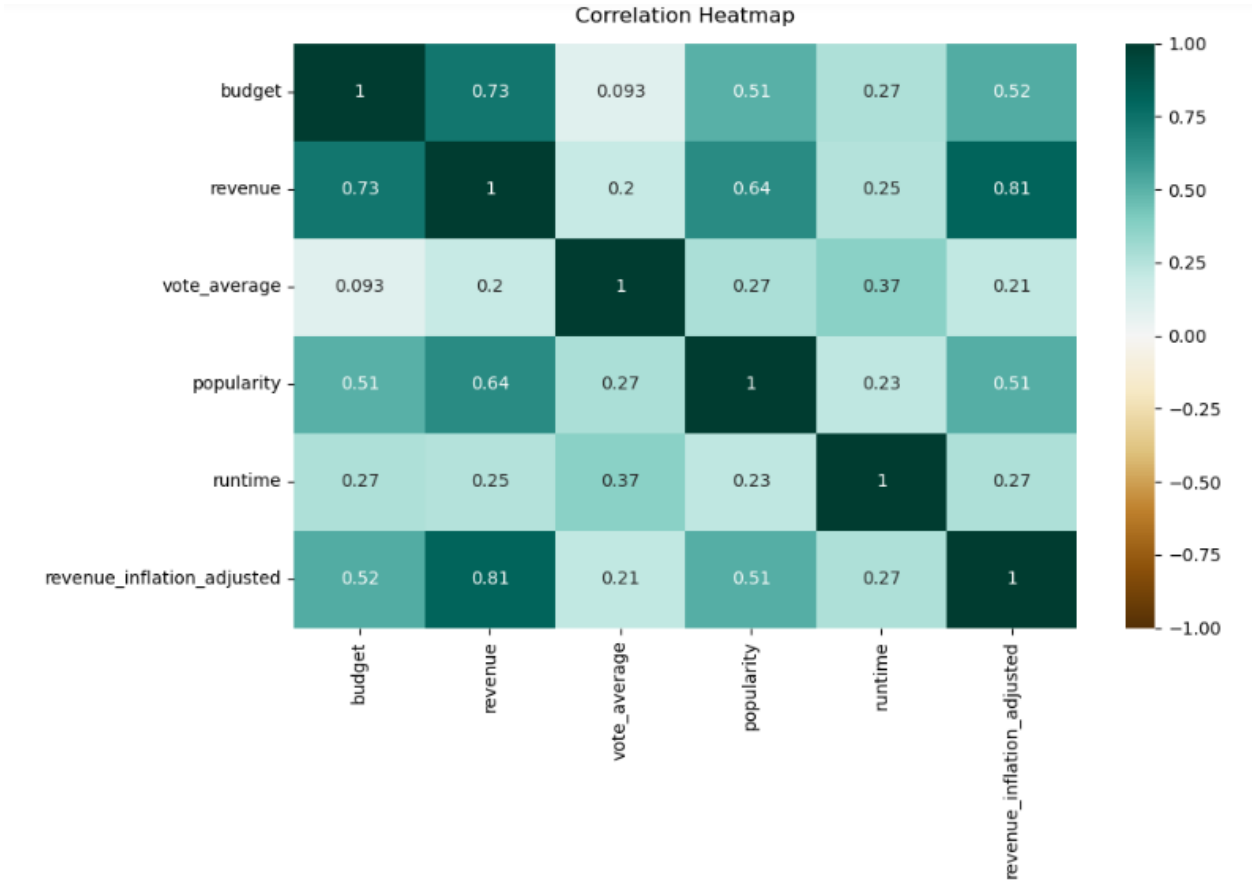
Introduction
In this project, I explore some variables that affect the revenue of movies.

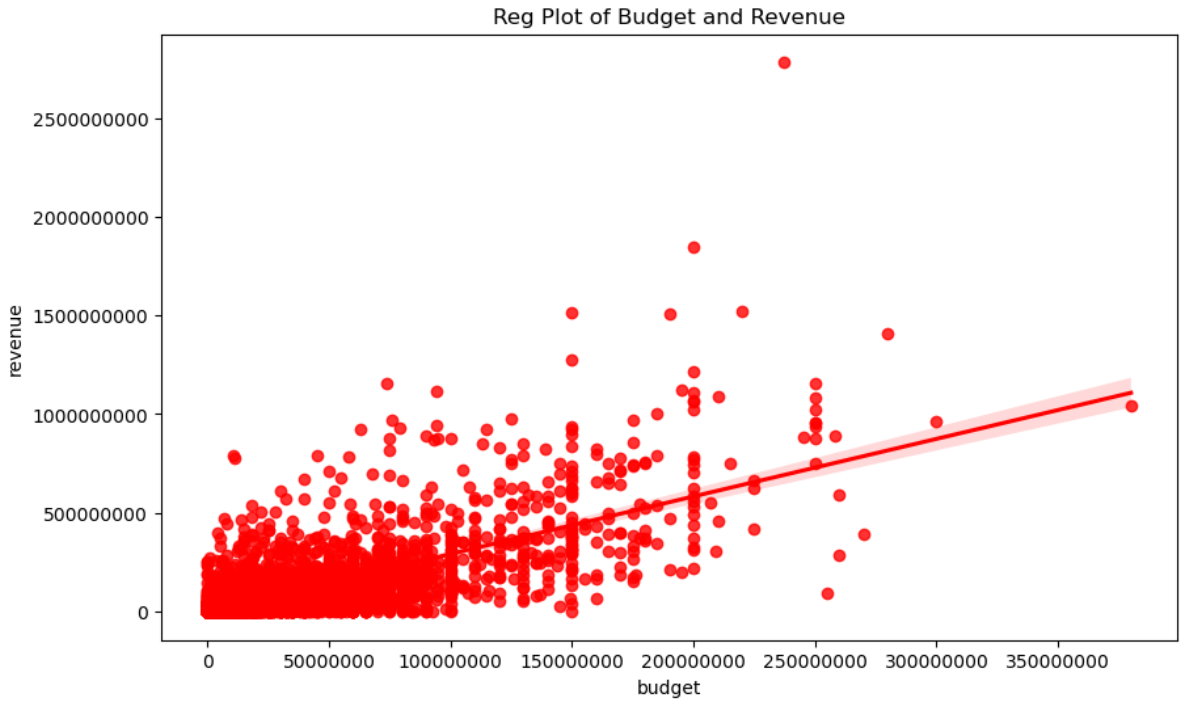
Methodology
- Data Collection and Sources. I obtained the data for this project from the Movies database created by Database Star. I queried the movie table in the database and exported the data to a CSV. I obtained the Consumer Price Index (CPI) data from the U.S. Bureau of Labor and Statistics. The CSV files can be found at GitHub.
- Data Wrangling. I utilized Python's Pandas library to ingest the CSV files. I did not have to do any data cleaning as there were no duplicate values, null values, or other issues in the numeric columns. I adjusted the revenue data to account for inflation - using the script in the Adjusting Revenue for Inflation notebook - and calculated the revenue in August 2023 dollars. If I had not adjusted for inflation, it could have affected the analysis.
- Data Analysis. I calculated the Pearson correlation coefficient for the numeric values in the Movies table. I calculated basic statistics for the 'budget' column, which positively correlated with 'revenue', and the 'revenue_inflation_adjusted', which positively correlated with 'popularity.' As the 'budget' column is not adjusted for inflation, I compared it to the non-adjusted 'revenue' column rather than the 'revenue_inflation_adjusted' column.
- Data Visualization. I used Python's Seaborn and Matplotlib libraries to visualize the data in a box plot, regression plot, heat map, a strip plot, and a distribution plot.
- Presentation.

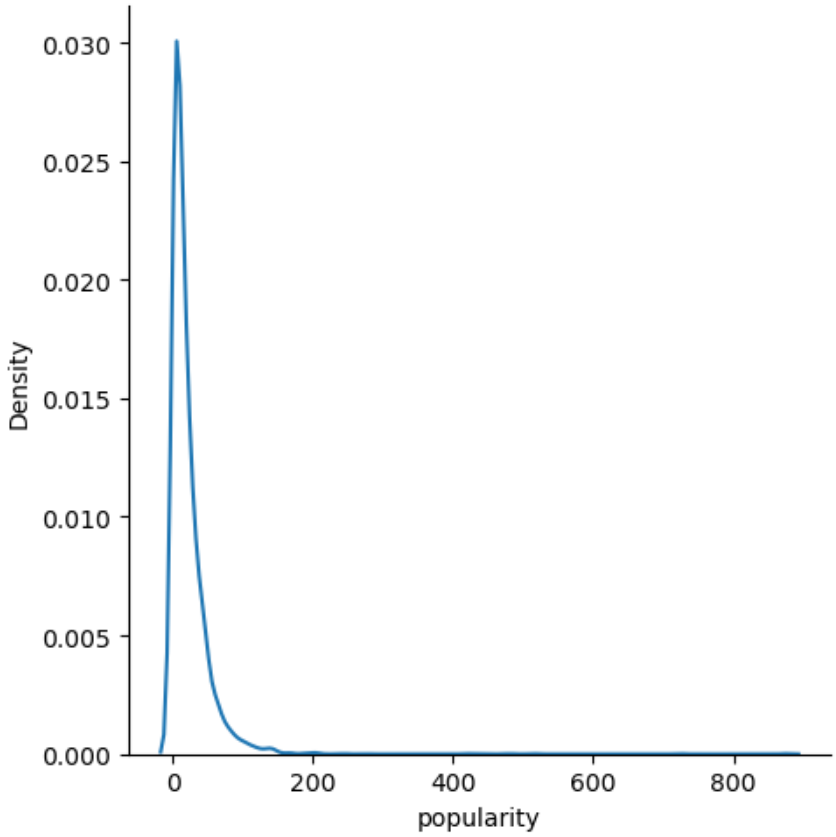
Results
'Revenue' positively correlated with 'budget' with a Pearson correlation coefficient value of 0.73 and a p-value of 0.00.

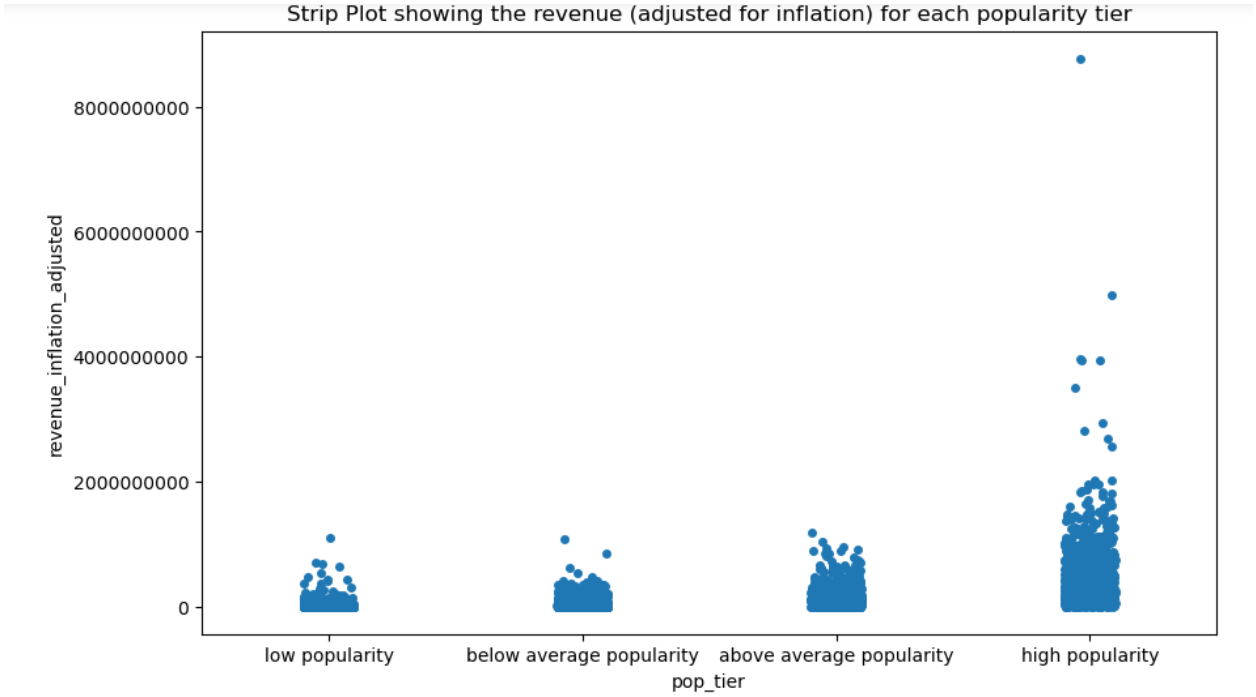
'Revenue_inflation_adjusted' positively correlated with 'popularity' with a Pearson correlation coefficient value of 0.51 and a p-value of approximately 0.00.


Revenue (Adjusted for Inflation) Box Plot
8000000000
6000000000
4000000000
2000000000
0
1


Correlation Heatmap
budget
revenue
vote_average
popularity
runtime
revenue_inflation_adjusted
1
0.73
0.093
0.51
0.27
0.52
0.73
1
0.2
0.64
0.25
0.81
0.093
0.2
1
0.27
0.37
0.21
0.51
0.64
0.27
1
0.23
0.51
0.27
0.25
0.37
0.23
1
0.27
0.52
0.81
0.21
0.51
0.27
1
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00


Reg Plot of Budget and Revenue
revenue
budget
2500000000
2000000000
1500000000
1000000000
500000000
0
0
50000000
100000000
150000000
200000000
250000000
300000000
350000000


Density
popularity
0.030
0.025
0.020
0.015
0.010
0.005
0.000
0
200
400
600
800



Findings & Implications

Findings

- Movies that generate high revenue usually (but not always) have a large budget as well. However, a large budget is certainly not a guarantee of a film producing high revenue.
- Films that are highly popular tend to produce the most revenue. There are not necessarily large differences in revenue in films that are not highly popular.

Implications

- In order to return a movie with high revenue, movie studios usually will need to invest a large amount of money in a high budget film. However, this is not a guarantee of success, i.e., box office bombs.
- Unsurprisingly, films that are more popular tend to return higher revenue. However, striving to make a film more appealing to more people does not necessarily result in higher revenue and the ultimate success of a film rests on a number of factors, especially the talent involved in its creation.