* Question: In which months do movie releases make the most money?
* Background & Motivation:
  + At the fundamental level, films are investments for production companies. Every decision these companies make when preparing for a release is a strategic effort to maximize profits, including when they decide to release movies. In my data analysis, I sought to find out which months have the highest opening weekend revenue.
* Questions:
  + Which months usually have the highest opening weekend revenue?
  + In which months are movies released more frequently?
  + Are the trends the same for each year represented?
  + Are the trends the same for the four most represented distributors in the dataset?
* Data:
  + Box Office Mojo is a database managed by IMDB that contains information regarding the box office success of different movies. I specifically used their table containing the Top 1000 Opening Weekend Box Office Revenues. Scraping the data using a Python script resulted in the data below. (INSERT SCREENSHOT OF DATASET).
  + Data current as of November 15th, 2021
  + 1000 rows, each containing the information for one unique movie
  + Columns:
    - Rank: Ranking of Movie based on Opening Weekend Revenue (Numerical)
    - Release: Title of the Movie (Categorical)
    - Opening: Opening Weekend Revenue (Numerical)
    - Total Gross: Total Gross of the Film (Numerical)
    - % of Total: Opening Weekend Revenue as a percentage of the Total Gross (Numerical)
    - Theaters: # of Theaters the Film Released in (Numerical)
    - Average: Average Opening Weekend Revenue per Theater (Numerical)
    - Date: Date of Release (Categorical)
    - Distributor: Distributor of the Film (Categorical)
* Cleaning:
  + For the purposes of my research, I selected the “Release”, “Opening”, “Date” and “Distributor” columns of the table
  + Upon inspection of my data, I noticed that the films “Scary Movie 4”, “Halloween”, and “1408” were missing entries in the “Distributor” column. Through research, I found out that the Weinstein Company Distributed each of these films, so I replaced each null value with “The Weinstein Company”.
  + I modified the “Opening” column so that it contains integers instead of strings and the “Date” column so that each date became a DateTime object.
  + I also created “Month” and “Year” columns based on the month and year a film released.
* Visualization:
  + Part I: Analyzing the entire dataset
    - I first wanted to see the general distribution of my dataset, so I decided to create a scatter plot of the Opening Weekend Revenue vs Month of Release.
      * (1st Scatterplot)
    - There are two peaks in opening weekend revenue, the first around May and the second around November. To get a better idea of which month typically has the highest revenue, I decided to look at the median revenue for each month.
      * (Bar Graph of the Median Revenue vs Month of Release) \*\*\*\*Try to order by Revenue or use color\*\*\*\*
    - As depicted in the graph, the movies in my dataset that were released in May, June, July and November typically made the most money in their opening weekend. I assumed that studios probably use this historical trend when choosing release months, so I decided to see which release months are most common in the dataset.
      * (Screenshot of value counts for months)
    - As shown, June, July, November and May are the most common release months for films, aligning with my theory.
  + Part 2: Breaking the data up by year
    - To continue my analysis of this trend, I looked at data from the years 2010-2019 to see if the pattern was consistent for that decade.
      * (Year Bar Graphs)
    - In most of the years represented, there are noticeable peaks in revenue in the summer months and the late fall/early winter months. I then looked the distribution of the release months.
      * (Year Histograms)
    - In most years, more movies also released during the summer and late fall/early winter months.
  + Part 3: Looking at the Top 4 Studios Represented
    - When I looked at the Top 4 Studios represented in this dataset, Warner Bros., Universal Pictures, Walt Disney Studios Motion Pictures, and Twentieth Century Fox, the trends remained the same.
  + Conclusions:
    - Based on my analysis, movies released during the summer months and the late fall/early winter months tend to perform the best. Studios also seem to release more movies during these times. These trends show that studios are avoiding “Dump Months”. These months, including January and February, carry lower critical and commercial expectations. Studios often fill these months with movies that performed poorly for test audiences to fill contractual obligations. (<https://www.theatlantic.com/entertainment/archive/2012/01/january-dumping-ground-for-terrible-movies-like-contraband/251326/>) Meanwhile, the summer and early winter are good times to release films because kids and young adults are typically out of school in the United States. Studios also have a long history of saving their best films for the end of the calendar year, just in time for award season.
  + Future Research:
    - While my dataset follows accepted trends in the movie industry, it is very limited in scope. For instance, it only examines the 1000 movies that had record-breaking opening weekends. It would be more accurate to conduct on a wider range of films that weren’t necessarily this successful to see how a more standard films perform month to month. (<https://web.archive.org/web/20140104114025/http://www.nytimes.com/2013/01/20/magazine/how-to-survive-januarys-dearth-of-good-movies.html>)
    - While my dataset was international, it would be interesting to see how these trends vary from country to country. In my research, I read that January is starting to be a more successful month for movies in China, seeing far more domestic releases during the month. (<https://www.latimes.com/business/hollywood/la-fi-ct-china-box-office-new-year-20180222-story.html>)
    - These trends are also more representative of the time before the COVID-19 Pandemic, so in the coming years it will be interesting to see if these trends change.