# 

# THE FINAL REPORT OF DATA VISUALIZATION

# Marvel vs DC Data Analysis

# PREPARED BY…

# MAAZ SHAHNAWAZ

# SUMAIYA AFZAL

# MUHAMMAD USAMA

**Table of Contents**

Introduction3

**Aim of project4**

**What is Data Visualization5**

Importing Library6

**More Information 9**

**Dropping unnecessary columns………………………………………………………………………………………………….……..10**

**Data type Conversion………………………………………………………………………………………………………………………..11**

**Mean of Every Unique Column………………………………………………………………………………………………….………12**

**Graph Showing Correlation…………………………………………………………………………………………………………….…13**

**Conclusion………………………………………………………………………………………………………………………………………..21**

**References……………………………………………………………………………………………………………………………………….22**

# Introduction

# 

# There are many ways to compare comic’s universes. Here I explain how I created a visualization to compare Marvel and DC universes using data.

My goal was to define a simple visualization to show differences between Marvel and DC in a clear and fast way. I plotted a graph for each Company and showed them side-by-side.

# Aim of Project

# MCU vs DC. Which one is better? Which has more high-rated movies? Analysis of Marvel and DC movies based on gross value.

Marvel Cinematic vs DC Universe, it’s a never-ending debate, right? Fans got crazy when you oppose any of these cinematic universes. But in the article, we are going to do a fight over Marvel vs DC based on some data. Data always tells the truth. So, let’s start this data war, with a cup of coffee.

# MCU vs DC

You can write the Python code in **Jupyter Notebook, Google Colab**, or any other preferred editor. I will recommend you Jupyter Notebook because I use it more often.

## ****What is Data Visualization?****

**Data Visualization techniques** are one of the key components of any analytics project. An end-to-end analytics use case involves ideation, requirement gathering, getting the raw data, analyzing the data, building a predictive model, deploying the model, and communicating the end result to the business.

Throughout this entire process, the analysis of data, and the communication of results to the business requires visualizing the raw data and understanding several inter-linked relations among the features. Python is the most preferred language which has several libraries and packages such as Pandas, NumPy, Matplotlib, Seaborn, and so on used to visualize the data.

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

**Import the libraries**

* **NumPy**

NumPy offers comprehensive mathematical functions, random number generators, linear algebra routines, Fourier transforms, and more.

* **Pandas**

Pandas is mainly used for data analysis. Pandas allows importing data from various file formats such as comma-separated values, JSON, SQL database tables or queries, and Microsoft Excel.

* **Matplotlib**

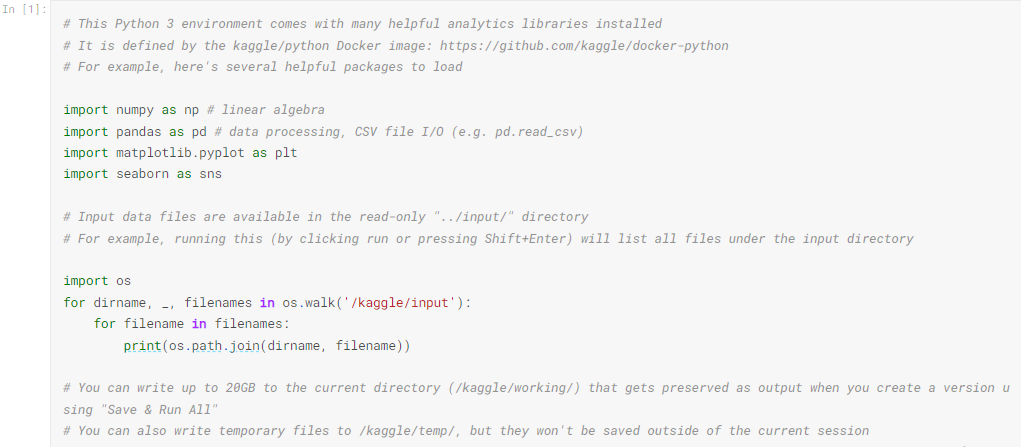
Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes easy things easy and hard things possible.

* **Seaborn**

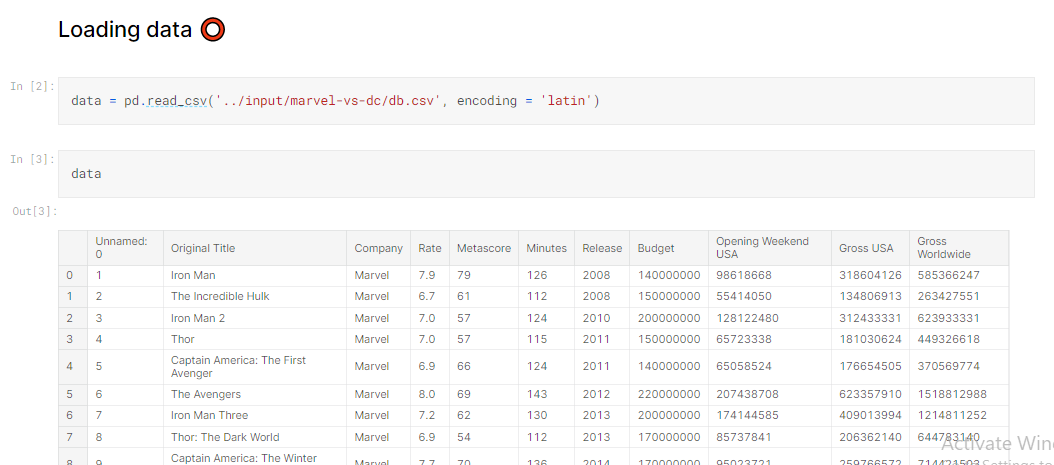
Seaborn is an open-source Python library built on top of Matplotlib. It is used for data visualization and exploratory data analysis. Seaborn works easily with data frames and the Pandas library.

* **Os**

The OS module in Python provides functions for interacting with the operating system. OS comes under Python's standard utility modules. This module provides a portable way of using operating system-dependent functionality.

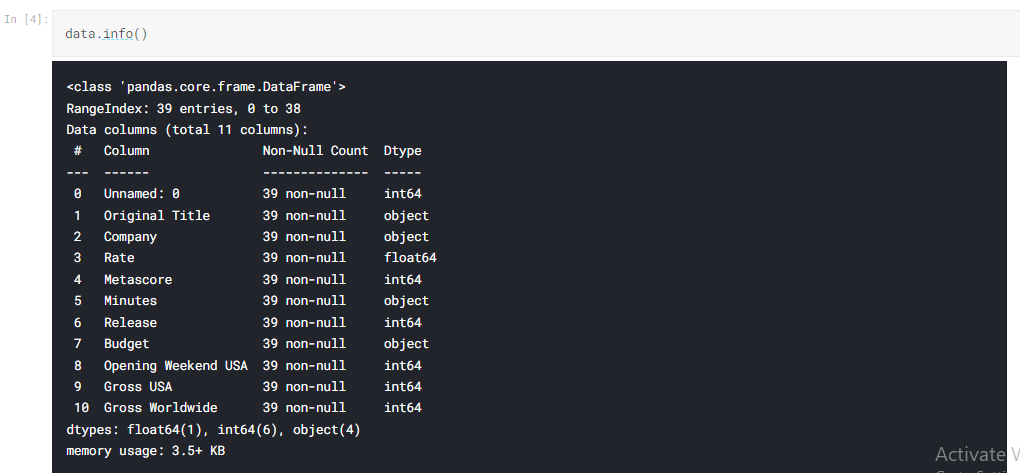


* **Let’s load the data and take a sneak peek at the data.**

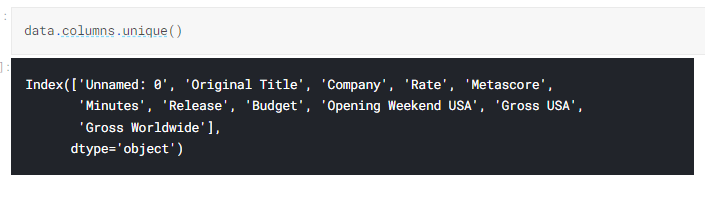


We have names of movies, year of release, genre, IMDB rating, IMDB gross, entity, and so on.

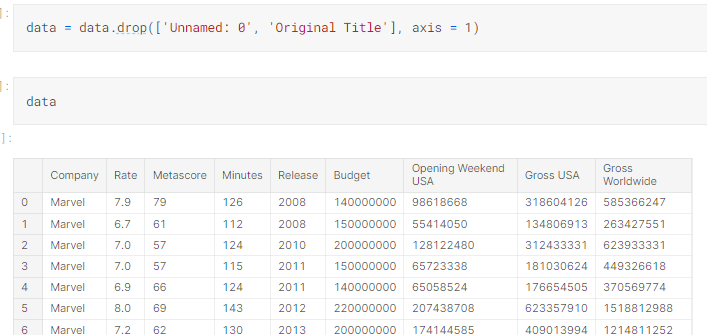
* **Gather some more information of data.**



Check out the null values in each column.  
After that, get more information about our dataset with the type of each column attributes.

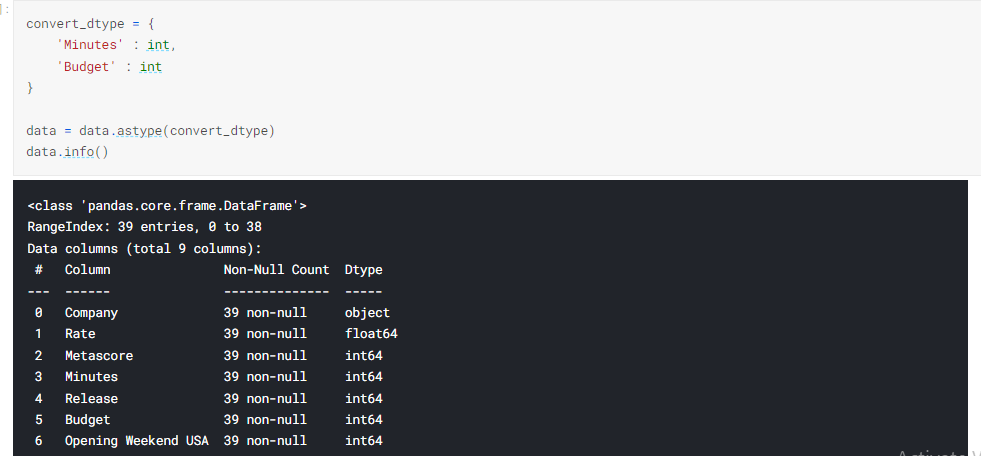


Checking the columns and using Unique to avoid repeated columns.



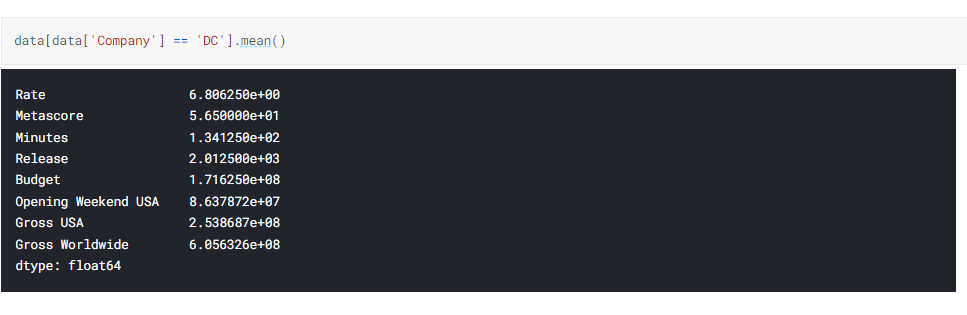
Dropping unnecessary columns.

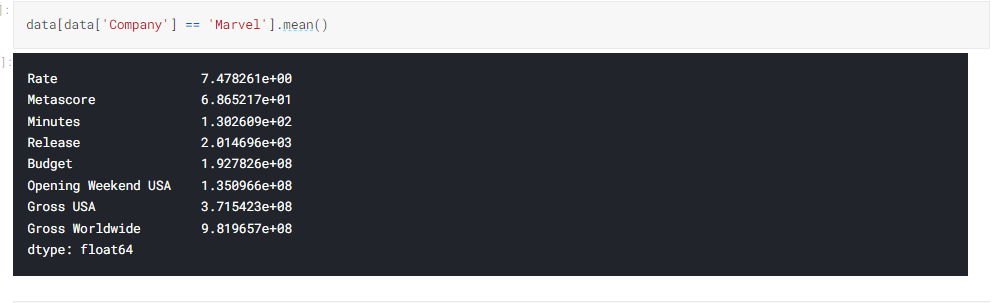
All columns except 'Company' type should be converted to either float or int type.



Converted

Now find out the mean or average of each rating, metascore, length, budget etc.



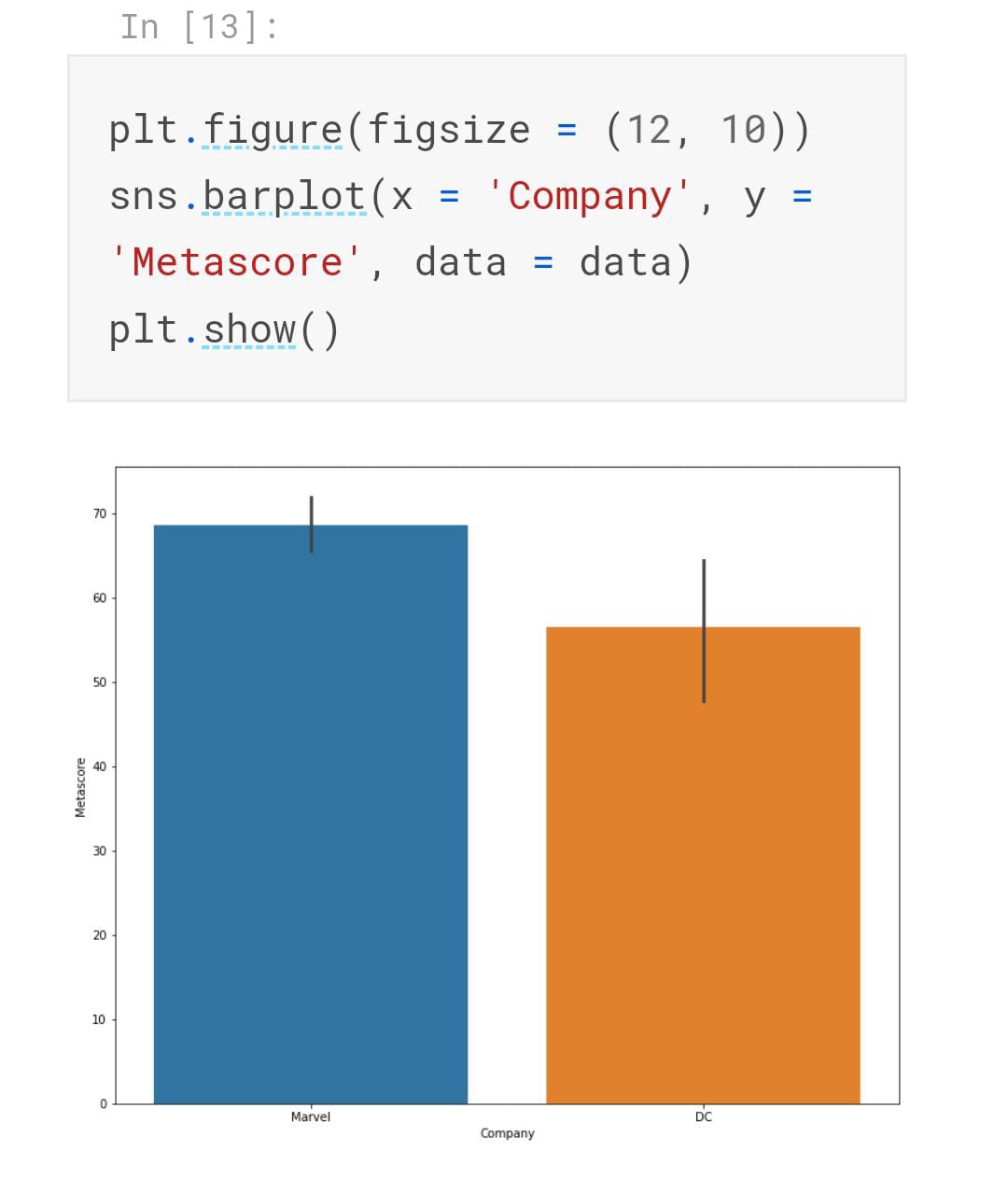


The average rating of DC movies is 6.886 and for Marvel movies, it’s 7.47. DC has one of the highest-rated movies of all time.

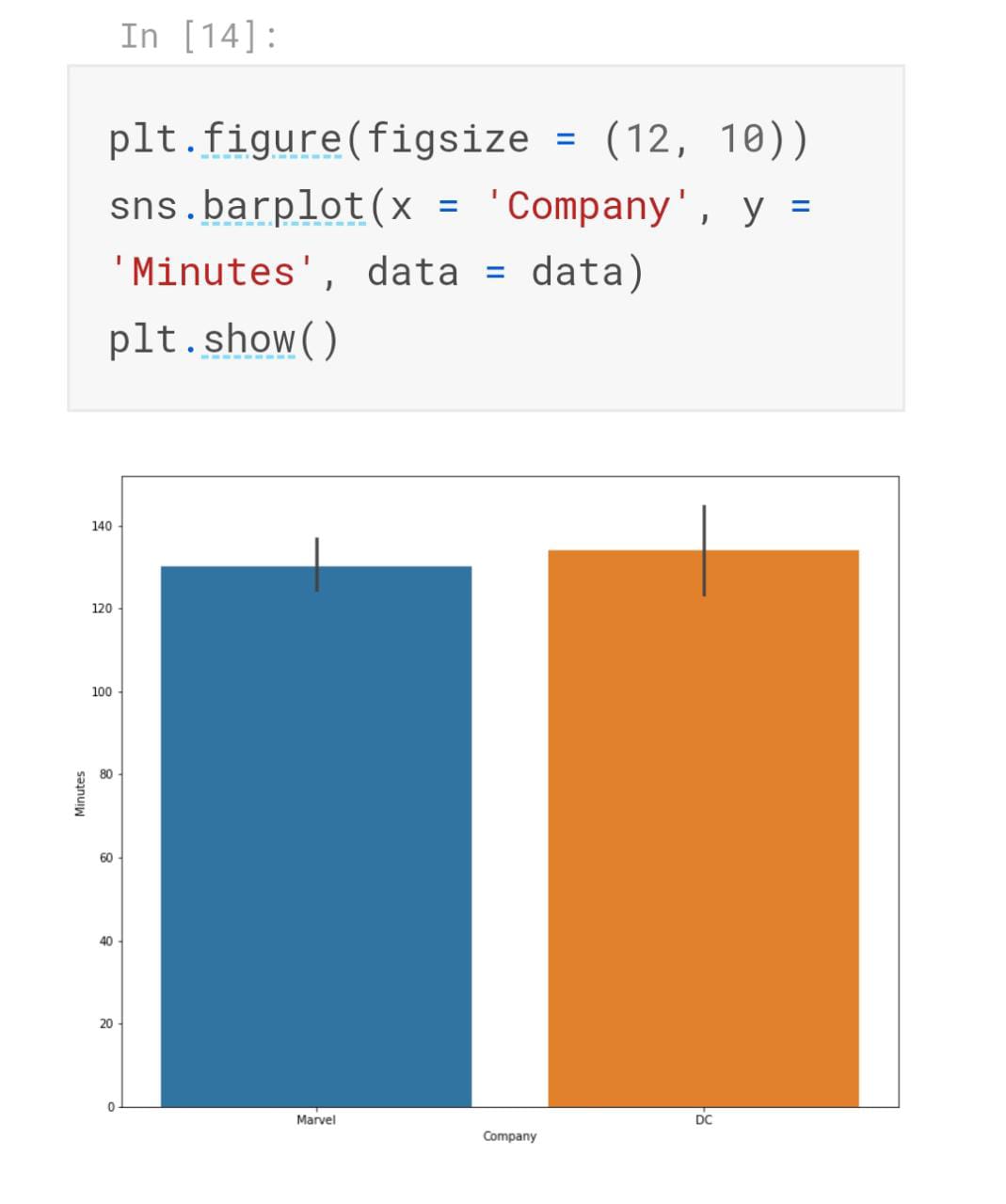
# Graphs showing correlation 📈

# 

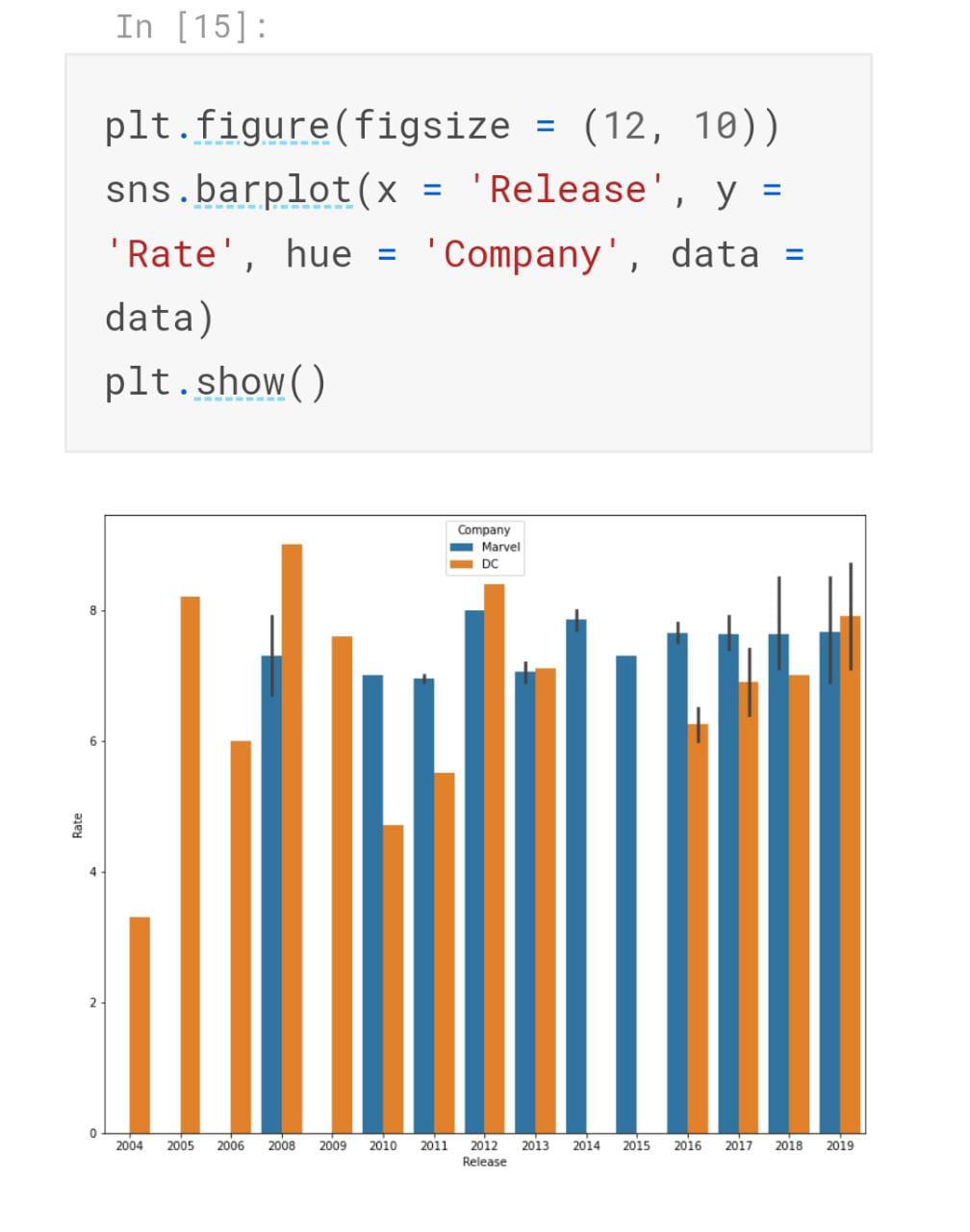
The above graph clearly shows the rating of Marvel movies is greater than DC.  In the ratings game, Marvel wins by a large margin: 66% of Marvel films are certified fresh compared to 54% of DC films. Between the box office numbers and ratings, Marvel is still coming out on top.



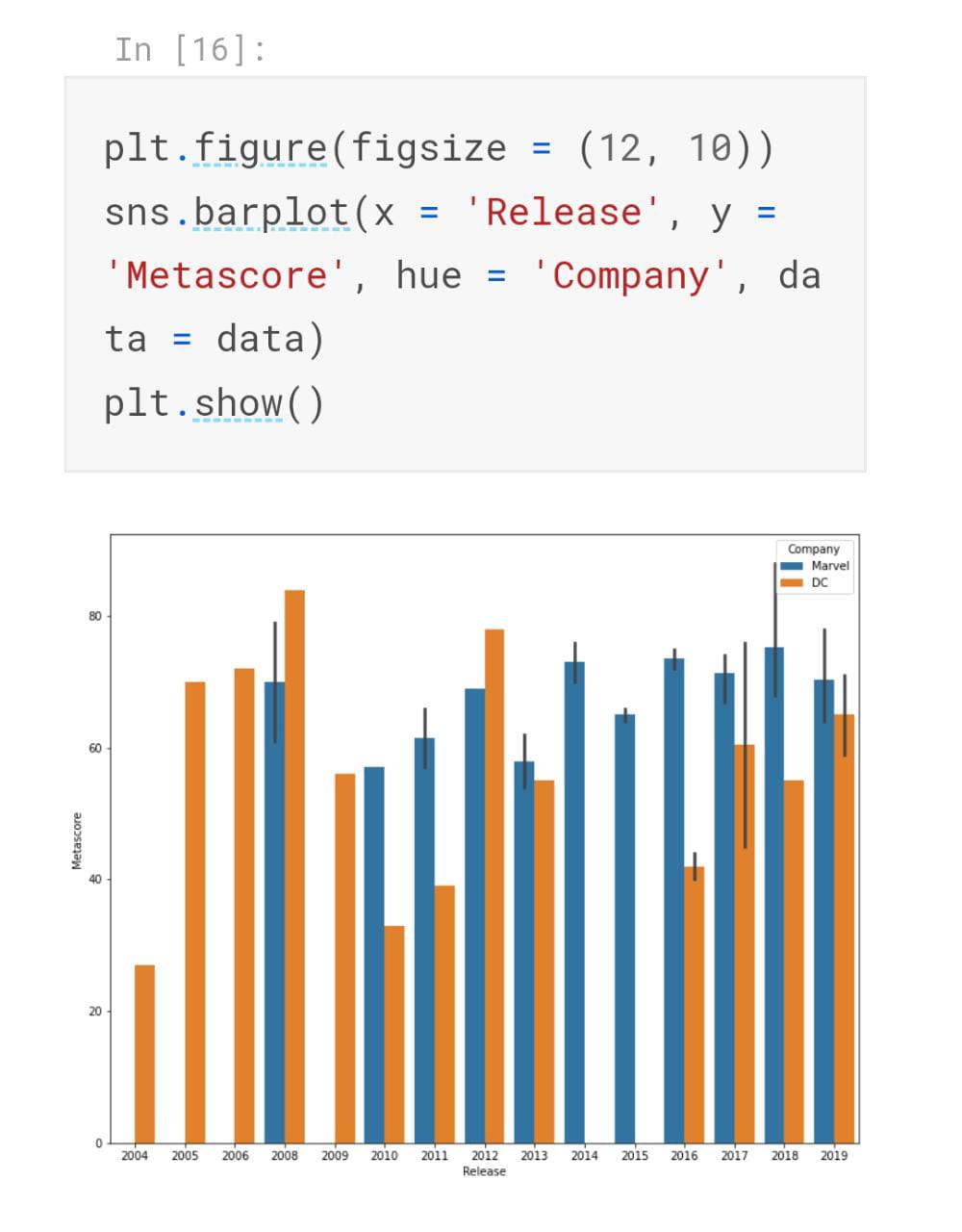
This graph shows the Metascore of Marvel is also higher than DC. According to critics, Marvel is better than DC, but only by a hair's breadth, with an average Metacritique score of 58.73 and over DC's 56.71.



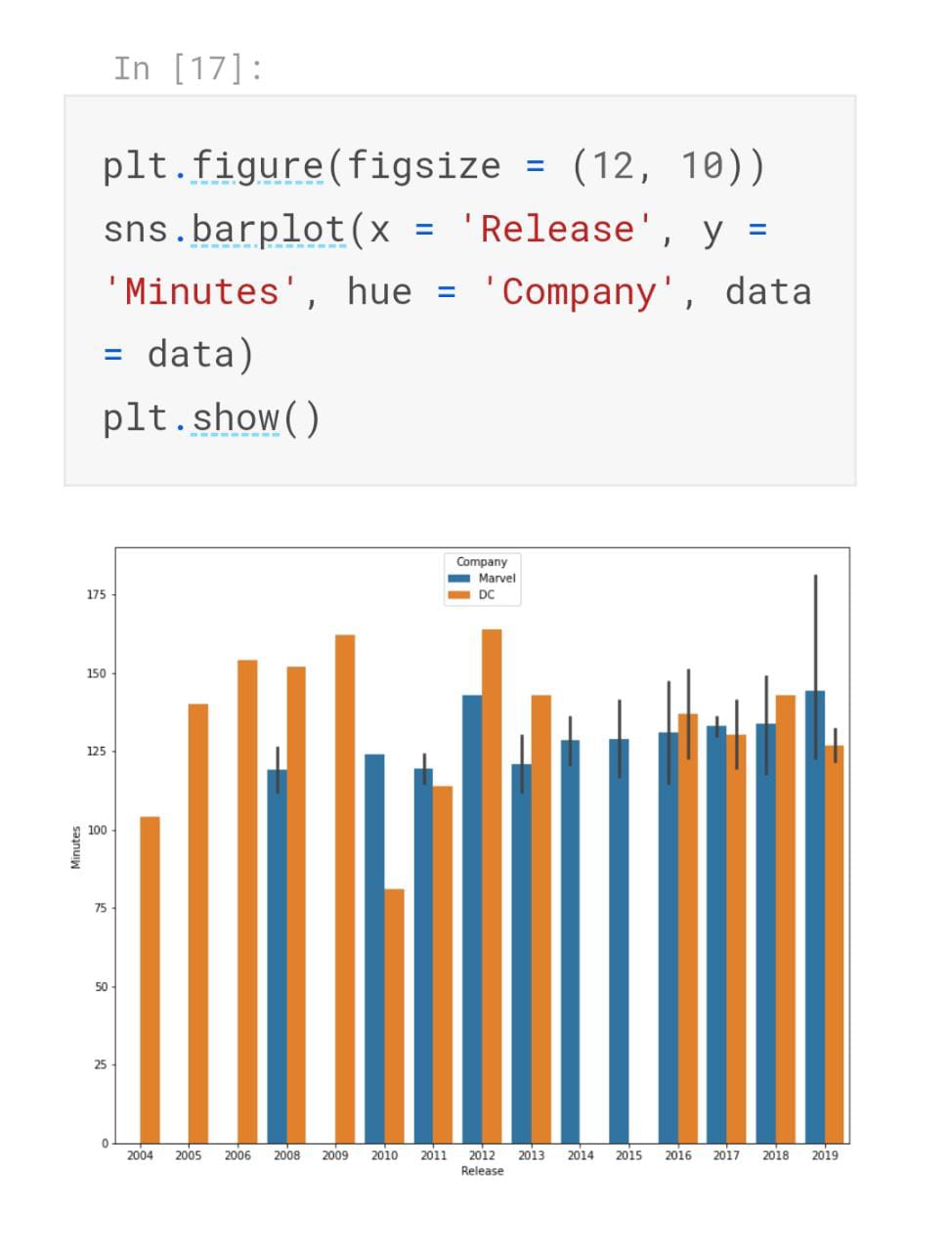
This graph shows the length of the movies. Here the length of DC movies is greater than Marvel.

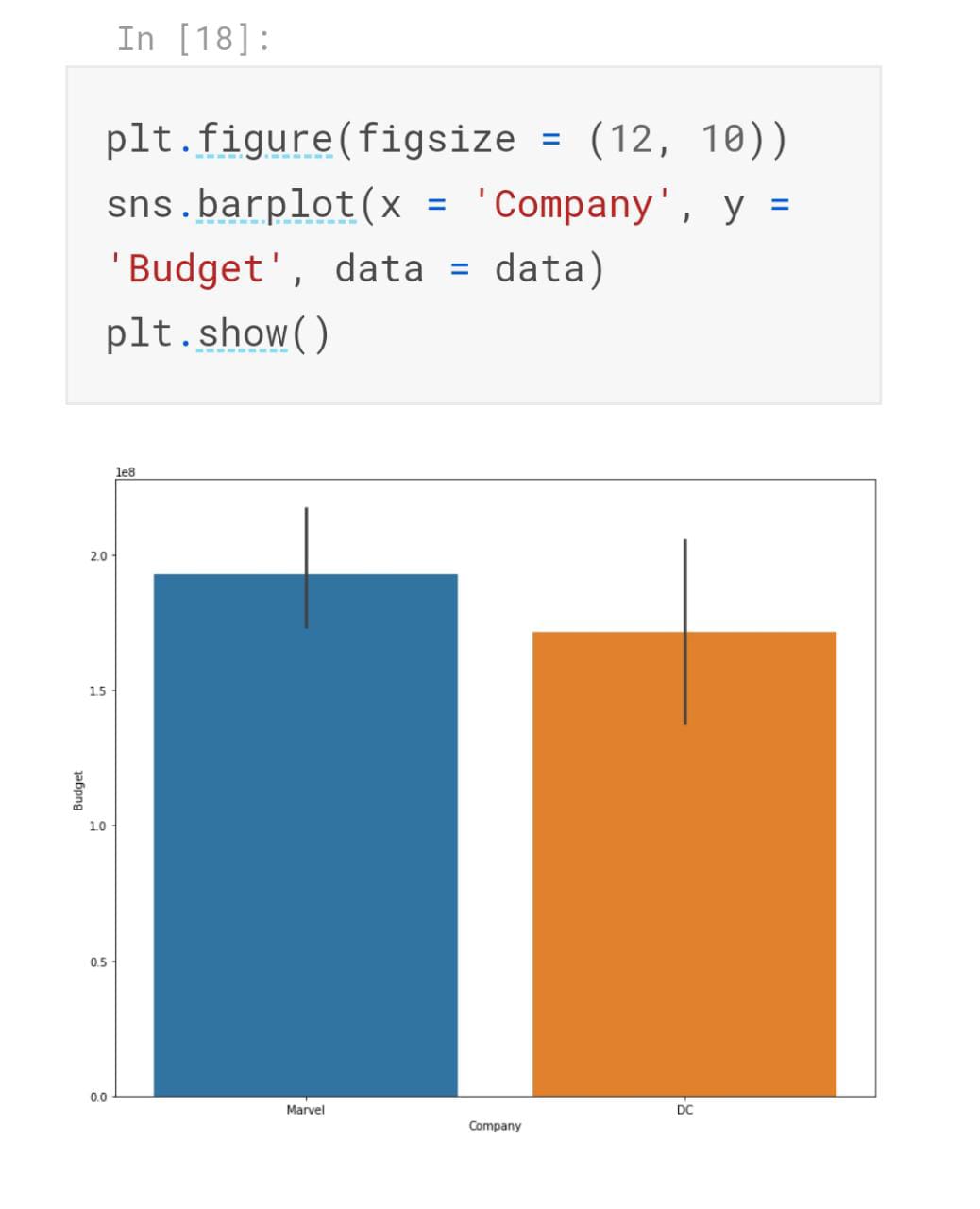


This graph shows rating of each movies in the year 2004 to 2019.The Dark Knight is the Top-rated DC movie. It has an IMDB rating of 9. If you didn’t watch it yet then do watch. You will witness the legendary act of Sir Heath Ledger. This movie shows that what DC Universe is capable of.

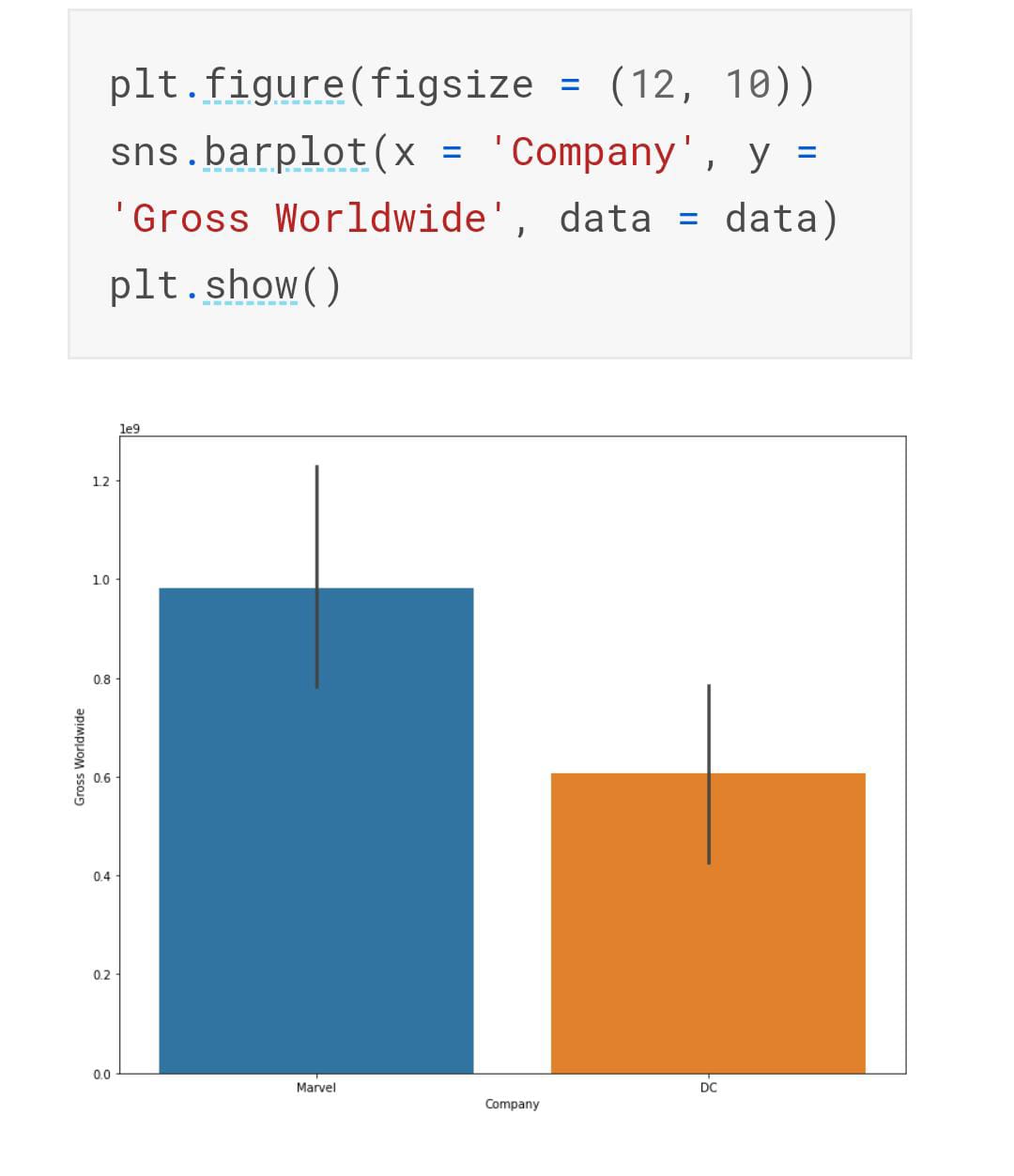


The critics also favor Marvel as the metascore of Marvel movies are also greater than DC.

The average runtime of both the Marvel and DC movies is almost equal. But there is a huge difference there in highest runtime movies.



The Budget of Marvel is slightly higher than Dc.



In terms of gross, few Marvel movies are far away from DC Movies.

Most of the Marvel movie has IMDB ratings lies between 6.7 to 8.2.

DC movies ratings are evenly distributed across the graph.

DC movies are performing well on IMDB gross but if you compare it with Marvel then they fall short.

We cannot compare both the Movie Making production house because who knows the future. In the future, DC may overshadow Marvel. But the best part is that both these productions houses are making good movies and entertained the audience for the past few decades.

# Conclusion

* Avg. rating of Marvel is greater than DC.
* Avg. metascore of Marvel is greater than DC.
* Avg. duration of DC movies is greater than Marvel.
* DC movies were released earlier than birth of Marvel, but since 2010 people tended to like Marvel more over the years.
* Gradually duration of movies of Marvel increased and whereas of DC decreased.
* Moreover, we can see that the avg. budget for Marvel was more than DC, but the outcome was also good as the Gross World was also high. So, it was worth it!!!

### **Well, in my opinion as per stats Marvel won, but let's see how future places these two in the tough competition.**

Well, that’s it for this article.

If this article sounds informative to you, make sure to follow and share it with your geek community.

**References**

[*https://www.kaggle.com/ikarosalpha/marvel-v-s-dc*](https://www.kaggle.com/ikarosalpha/marvel-v-s-dc)

[*https://medium.com/geekculture/marvel-vs-dc-data-analysis-in-python-e561cac72358*](https://medium.com/geekculture/marvel-vs-dc-data-analysis-in-python-e561cac72358)

[*https://betterprogramming.pub/how-to-perform-exploratory-data-analysis-with-marvel-vs-dc-comics-data-ec75f457ac60*](https://betterprogramming.pub/how-to-perform-exploratory-data-analysis-with-marvel-vs-dc-comics-data-ec75f457ac60)