Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 02, February-2019

Data Visualization on Movies Dataset using Python

Ms. Sushmita Roy
Post-Graduation Student
Department of Information Technology
Thakur College of Science and Commerce
Mumbai, India

***Abstract*— Data Visualization is the representation of data in the graphical format. It actually helps people to illustrate and understand the real significance of information by illustrating and presenting huge amounts of information in a simple and easy and to understand format and helps to communicate information clearly and effectively. In this paper, we consider using Python for understanding the data through different visualizations. When data visualization is applied on the Movies Dataset, it helps us to understand the data by providing various useful insights.**

Keywords:- Movie; python; data visualization.

I. INTRODUCTION

Human has a long history with basic data visualization, and data visualization is still a hot topic today. The history of visualization has to some extent been shaped by technology available and by the pressing needs of the time probably, including those of: Prehistoric paintings on clays, stones, maps on walls, images, table of numbers (with rows and columns), all these are a kind of data visualization - although we may not call them under this name at the time. Visualization is the graphical presentation of information in an attempt to give the viewer a qualitative knowledge of the actual content of the information. It is also the process of trying to transform objects, concepts and numbers into a visible and familiar form for human eyes. We can indeed refer to information, procedures, relationships or concepts also when we say "information".

II. WHAT IS DATA VISUALIZATION?

Data visualization is simply the representation of information in the form of charts, graphs, images, etc. It helps in understanding the patterns and trends out of the datasets.

Data visualization is all about understanding ratios and numerical relationships. Not to understand individual numbers, but to understand the patterns, trends and relationships in groups of numbers.

III. NEED FOR DATA VISUALIZATION

Seeing and understanding images is one of the basic instincts of human beings and understanding the numerical data requires training skills from the schools, and yet many people are still not good at numerical data.

It is much easier to identify trends, patterns and relationships from a well - drawn picture. Because graphical presentation of relevant information takes full advantage of the human eye's huge and often underutilized ability to measure picture and illustration information, the visualization of information shifts the load from numerical to visual. Gathering information from pictures probably saves somewhat more time than looking through original text and actual numbers - that is why many decision - makers probably prefer to have information provided to them in graphic form rather than in written or text form.

IV. DATA SOURCE

Data was collected using secondary sources. This data was then then to carry out the visualizations using Python language. The Movies Dataset obtained consists of 45,000 movies for films released on or before July 2017.

V. RESULTS

The following analysis has been performed using Python libraries on the movies data set. The results obtained help us to visualize the data easily in the form of graphs or charts.

We can obtain the different languages that have been used in the movies. Figure 1 shows the list of languages of the movies in the dataset. We check for the top 11 languages. The top most language is “English”, however, we have excluded the first language that is the top most language and display the graph as a bar chart for the top second to eleventh languages.

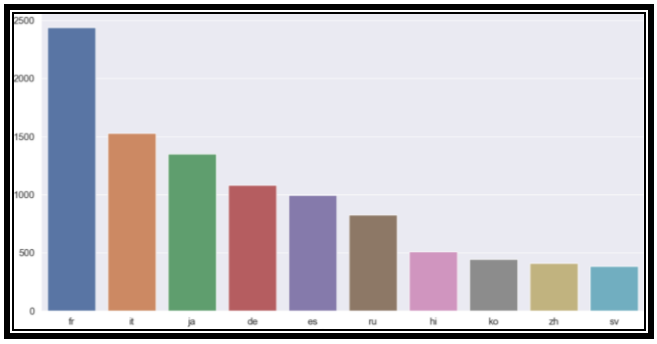

Fig. 1. Popular languages

We can also find the number of movies released in a particular month throughout the years. This gives a clear picture of the month which has the highest number of releases. Figure 2 shows the number of movies that have been released in a particular month. Thus, we see the results as January is the month in which most movies are being released, followed by September, October and so on. July, however, has the least number of movies released.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 02, February-2019

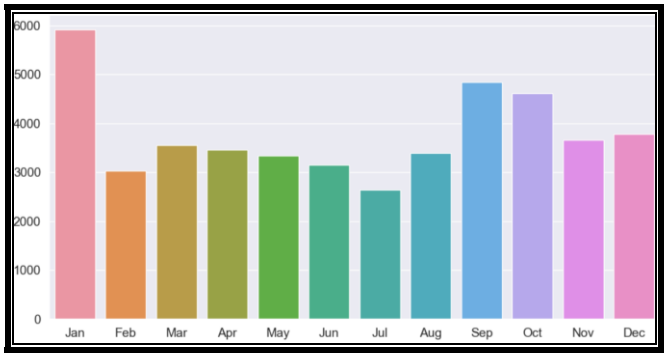

Fig. 2. Number of Movies Released in a particular month

We can also make an insight for finding the day which has the highest number of releases. Figure 3 shows the most popular day for movie release. As can be seen from the results, Friday is clearly the most popular day where the number of movies released is the maximum.

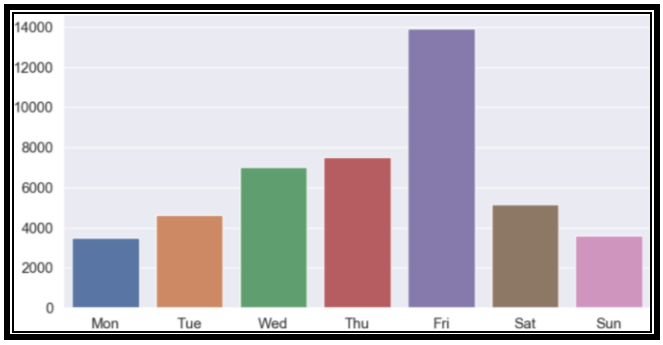

Fig. 3. Popular Day for Movie Releases.

Next, we can understand the popular genres. Figure 4 shows the top 15 most popular genres of movies. We can see from the results below that "Drama" is the most popular genre of movie while "Foreign" is the least popular form amongst these.

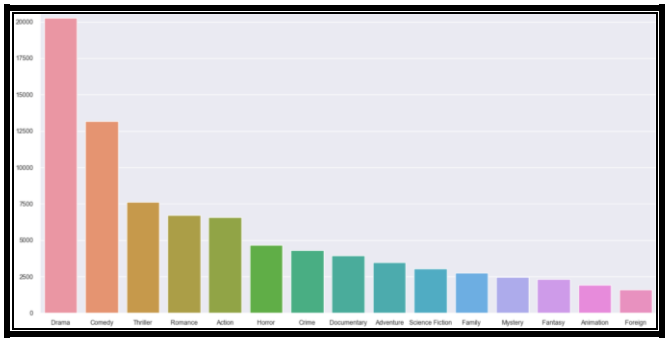

Fig. 4. Popular Genres.

VI. CONCLUSION

There is no limit to how many experiments can be performed for these visualizations. It actually depends on what is under analysis as well as the knowledge of packages such as Pandas and matplotlib. Using these packages, our job becomes easier.

The results obtained help us understand the data easily using these visualizations.

VII. FUTURE WORK

Apart from the results obtained, we can further use other forms of graphs available with Python to make various forms of graphs/charts.

Also, we can use other Python libraries to obtain different graphs and charts for various forms of data visualization.

ACKNOWLEDGMENT

I would like to express my acknowledgement to the college for providing me with the opportunity to accomplish and conduct this project.

I express my sincere gratitude to the Principal, Dr. (Mrs.) C.T. Chakraborty for helping with her kind and motivational words.

Also, I am thankful to the Department (IT) and Dr. Santosh Kumar Singh, Head of Department of Information Technology for the kind co-operation with providing the infrastructure for the completion of the project.

Special vote of thanks to my project guides, Dr. (Mr.) Santosh Kumar Singh and Prof. Mahendra Sharma who gave me the opportunity to do this and have guided me in it.

REFERENCES

[1] Zhao Kaidi, "Data Visualization" – https://www.cs.uic.edu/~kzhao/Papers/00_course_Data_visualization.pdf
[2] https://www.geeksforgeeks.org/data-visualization-different-charts-python/
[3] www.kaggle.com