**Data Analysis and Visualization**

**Project Report**

**Project Partners: Dawood Tanvir 21I-1665 and Umm-e-Hani 21I-1715**

**Introduction**

We have created two Text Visualizations

1. Word Cloud
2. Word Tree

**Information about Word Cloud Visualization**

For the word cloud visualization, we used Movies Plot dataset. A random word cloud will be created of a movie plot and the most occurring word will tell about the movie type. First, we did the dataset cleaning by removing special characters, lowering the case and removing stop words and then applied NLP technique Lemmatization to reduce words to its base form and saved my processed data in to a CSV file. After this, we used Python flask where we read the csv file and sent the data to the HTML file for visualizing the data. We found the idea of visualizing word cloud for movie plots interesting because the most occurring words will tell us about the type of movie and if you are a movie geek, you can even guess the movie by read the most occurring words. In addition, the most occurring words will have bigger size. Moreover, the cloud is in circular shape.

For Interactivity we added zooming and highlighting as whenever you hover on a words its size increases. Interactivity is important as it makes visualization interesting.

**Information about the Word Tree Visualization**

For the word tree visualization, we used a movies lines dataset. Through the word tree, we are showing phrasing, relationships between words. Our goal is to show keyword-in-context. First, we did cleaning of dataset in python by removing the special characters and lowering the cases. We did not remove the stop words because it would have caused problem in showing relationship between different words. For NLP we used Lemmatization to reduce words to its base form. For us we found word tree a good approach because it is a good way to visualize the relation between words and phrases.

For interactivity, we added colors as whenever you hover on a text its color changes and the colors are random, it also increases the size of text, and you can select the node. Interactivity makes the visualization interesting.

**Conclusion**

In conclusion, our exploration into text visualizations using Word Cloud and Word Tree techniques has provided valuable insights into the analysis of movie-related datasets.

The Word Cloud visualization, based on the Movie’s Plot dataset, offered a creative and intuitive way to represent the essence of movie plots. By preprocessing the data through techniques such as removing special characters, lowering case, eliminating stop words, and applying Lemmatization, we aimed to extract meaningful information. The resulting visual representation, a circular cloud where the size of words indicates their frequency, allows enthusiasts to discern the predominant themes of movies. The ability to potentially identify a movie by analyzing the most occurring words adds an element of engagement and entertainment to the visualization.

On the other hand, the Word Tree visualization, using a dataset of movie lines, focused on displaying phrasing and relationships between words. The decision not to remove stop words was deliberate, as it aimed to preserve the context and demonstrate connections between different words. The Word Tree approach provides a comprehensive view of keyword-in-context, revealing the intricacies of language usage within the dataset. This technique proves valuable for those interested in understanding the nuanced relationships between words and phrases.

Both visualizations contribute to the broader field of Natural Language Processing (NLP) and data visualization. They offer diverse perspectives on exploring textual data and extracting meaningful patterns. While the Word Cloud emphasizes the prevalence of certain words to infer movie genres, the Word Tree delves deeper into the relationships and contexts surrounding the words within movie lines.

In essence, our dual approach demonstrates the versatility of text visualizations in extracting valuable information from diverse datasets. These visualizations not only serve analytical purposes but also cater to the curiosity and engagement of users, making them accessible and enjoyable tools for exploring textual data in the realm of movies.

**That is All Folks!**