Barbie vs. Oppenheimer: Answered by Data Science

A UVA Data Science Case Study by Maajid Husain, May 2024

Introduction

Imagine two of the most hyped and looked forward to movies of the summer with vastly different messages battling for the attention of Americans to get the most box office views. Public sentiment for the movies Barbie and Oppenheimer was a highly debated topic on social media and the internet prior to the premiers of each of the movies. Barbie, a vibrant and light-hearted adventure film, portrayed a drastically different message from the somber and dark message of the Manhattan project shown in Oppenheimer. What do these differences tell us about viewer preferences and expectations and how did the differences in production effect how people liked the movies in reviews.

Purpose: This case study provides you with the opportunity to explore the fascinating topic of sentiment analysis through real IMDb reviews. By querying and comparing the sentiments expressed in the reviews of these two blockbusters, you will gain insights into how different themes and storytelling techniques resonate with audiences. These are the two movies that were the hit of the academic year (2023-2024) but can be adapted and reshaped to see other two large competing movies of the time for upcoming years.

Objective

Mission: Your mission is to uncover how distinct movie genres influence viewer emotions and sentiments using data from IMDB, you will compare the sentiments from reviews of Barbie and Oppenheimer to decode patters and divergences in public perception.
Importance: Understanding these dynamics is a fun was to see what the American population prefers in their films and shows Hollywood what types of movies to produce more of.

What You Will Do

Tasks: You will conduct a comprehensive sentiment analysis using Python applying techniques to quantify and compare emotional responses from many reviewing following the ETL pipeline (Extract, Load, Transform).

Data: You will collect data by scraping the IMDb website using Python, under the audience reviews page as a reliable and easy way to get text data on movie reviews.

Deliverable: Produce a slideshow presentation that includes the motivation behind your project, the data preparation, the model plan, and your results and conclusions.