Project Report

Project Title: Horror Movie Data Analysis

I. PROBLEM

[1] Franchise Analysis: The Franchise vs. Non-franchise analysis confronts the intricate challenge posed by the ever-evolving landscape of movie preferences, a challenge that vexes movie producers in their decision-making process. The analysis hinges on the exploration of three pivotal attributes: Collection Name, Popularity, and Release Date. Collection Name serves as the linchpin, categorizing movies into either franchise or non-franchise, while Popularity quantifies the audience's reception. Release Date, formatted as MM/DD/YYYY, facilitates the temporal grouping of movies into decades, providing a comprehensive historical context.

Within each decade, movies are further dissected into two subgroups—Franchise and Non-franchise—based on their Collection Name attribute. The crux of the analysis lies in the computation of the average Popularity for each subgroup, offering nuanced insights into shifting audience preferences over time. Visualizing these findings is achieved through a meticulously crafted clustered column chart. The X-axis portrays the temporal evolution in decades, marked by Release Dates, while the Y-axis vividly captures the ebb and flow of audience preferences. Noteworthy is the presence of two distinct bars for Franchise and Non-franchise movies within each decade, providing a visual narrative of how audience inclinations have evolved over successive periods.

This visual representation transcends mere aesthetics, evolving into a powerful decision-making tool for movie producers. It becomes a lens through which producers can discern the intricate patterns and trends in audience preferences, thereby making informed and strategic decisions in the intricate realm of franchise and non-franchise movie production. The analysis serves as a compass, guiding producers through the intricate terrain of cinematic creation, where the synthesis of historical context and contemporary trends informs decisions with a precision that aligns with the dynamic nature of audience preferences.

[2] Popular Genre Analysis: The Popular Horror-Genre Analysis addresses the intricate challenge faced by movie producers in navigating the ever-changing landscape of audience preferences within the horror genre. Employing a meticulous examination of key attributes, namely genre_names, title, and popularity, we embark on a nuanced exploration. The initial phase involves the creation of an array encompassing genre categories intricately linked with horror, spanning diverse spectrum such as Action, Adventure, Crime, Fantasy, Thriller, Drama, Mystery, Science Fiction, War, Comedy, Music, Romance, Documentary, Animation, Family, and Western.

Movies are meticulously grouped based on genre_names, resulting in distinct categories like Horror-Action, Horror-Adventure, and others, each encapsulating a unique blend of horror with its associated genre. The subsequent analysis focuses on identifying the top movie within each genre category, unraveling the intricate details of audience preferences. A pivotal aspect of this examination entails calculating the average popularity for each genre category, providing a quantitative measure of audience favor.

The findings are artistically presented through a pie chart, offering a visual representation of how audience preferences are distributed among various genre combinations. This comprehensive visualization serves as a valuable tool for movie producers, enabling them to discern patterns of audience engagement with specific genre blends. Additionally, the pie chart provides a detailed breakdown of the top popular movies within each genre category, arming producers with valuable insights for reference and strategic decision-making. Through this multifaceted analysis, producers gain a profound understanding of the nuanced dynamics within the horror genre, empowering them to navigate the intricate terrain of movie production with astute precision.

[3] Month Profit Trend Analysis: The Month-wise Profit Trend Analysis addresses a pivotal concern for film producers—identifying the optimal month for releasing movies to ensure profitability. This intricate examination delves into three fundamental attributes: Release Date, Budget, and Revenue. The Release Date is methodically parsed into distinct months, ranging from "January" to "December." The subsequent analysis involves calculating the profit for each movie by subtracting its budget from revenue and attributing it to the corresponding release month. This meticulous process aggregates profits for all movies released in a given month, shedding light on the most lucrative months for movie releases.

The presentation of these findings is elegantly executed through a bar chart, where the X-axis delineates the months and the Y-axis signifies the average profit. This visual representation allows producers to discern patterns in profitability across different months, facilitating strategic decision-making. By strategically selecting a release month based on historical profitability trends, producers can enhance their prospects of attaining financial success. This analysis provides a nuanced understanding of the intricate interplay between release timing and profitability, empowering producers to make informed choices in the competitive landscape of the film industry.

[4] The Runtime Language Analysis: addresses the intricate challenge confronted by film producers aiming to devise an effective language strategy for the production of both short and long films. This analytical framework revolves around crucial attributes, namely original_language, runtime, and popularity, offering a nuanced exploration of the dynamics governing film production. The preliminary stage involves the meticulous selection of five languages—English, Japanese, Russian, Korean, and Espanol—as central points for the analysis.

The systematic grouping of movies is based on these chosen languages, with an additional subdivision within each language category based on runtime duration. This meticulous categorization segregates films into two distinct classes: those with a runtime below 120 minutes, designated as short films, and those exceeding 120 minutes, classified as long films. The rationale behind this segmentation is to unravel the nuanced audience reception and popularity dynamics associated with short and long films within each language category.

II. SOFTWARE DESIGN AND IMPLEMENTATION

A. Software Design and NoSQL-Database and Tools Used

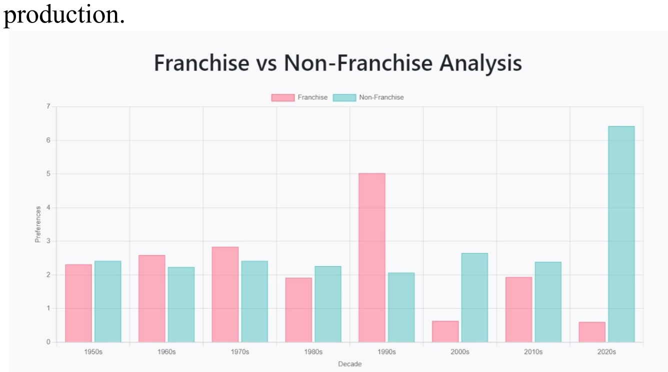
The back-end logic of the code has been implemented using the Python Flask framework, and MongoDB has been employed for database purposes through MongoDB Atlas. The front-end is developed using HTML, CSS, and JavaScript. This combination of Flask, MongoDB, PyMongo, along with front-end technologies, highlights a strategic and proficient methodology in software development, database handling, and the display of results.

To manage and manipulate a substantial dataset like the horror movies dataset, which consists of 21 attributes and spans over 32,540 values, this stack proves to be robust and efficient. Utilizing MongoDB's dynamic schema makes it a suitable choice for handling such a diverse and voluminous dataset, while Flask serves as a lightweight and flexible framework that can readily serve the data to the front-end.

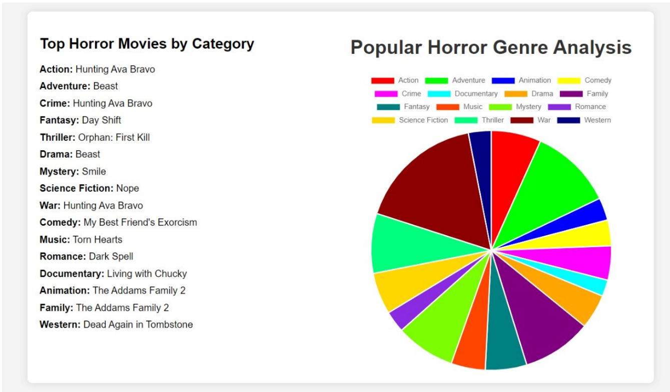
B. Parts that you have implemented

[1] Franchise Analysis: The intricate revelations of this analysis find their embodiment in a meticulously crafted clustered column chart. The temporal evolution across decades, delineated on the X-axis through Release Dates, converges with the dynamic shifts in audience preferences vividly captured on the Y-axis. This chart, marked by two distinct bars representing Franchise and Non-franchise movies within each decade, transcends mere aesthetics to become a potent decision-making instrument for movie producers. It serves as a discerning lens through which producers can unravel intricate patterns and evolving trends in audience inclinations, empowering them to make strategic decisions in the complex realm of franchise and non-franchise movie production. Beyond a visual spectacle, this analysis assumes the role of a compass, guiding producers through the intricate terrain of cinematic creation. The synthesis of historical context with contemporary trends equips producers with a

precision that resonates with the ever-changing dynamics of audience preferences, offering a comprehensive and insightful tool for navigating the multifaceted landscape of cinematic production.

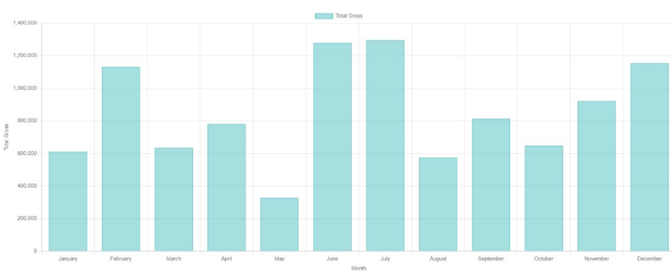


[2] Popular Genre Analysis: The discoveries are eloquently depicted using a pie chart, presenting a visual portrayal of the distribution of audience preferences across diverse genre combinations. This all-encompassing visualization serves as a valuable instrument for film producers, allowing them to decipher the intricacies of audience engagement with specific genre amalgamations. Moreover, the pie chart furnishes an elaborate breakdown of the highest-rated movies within each genre category, providing producers with insightful references for strategic decision-making. Through this multifaceted analysis, producers attain a profound comprehension of the subtle dynamics inherent in the horror genre, equipping them with the acumen to navigate the intricate landscape of movie production with discerning precision.



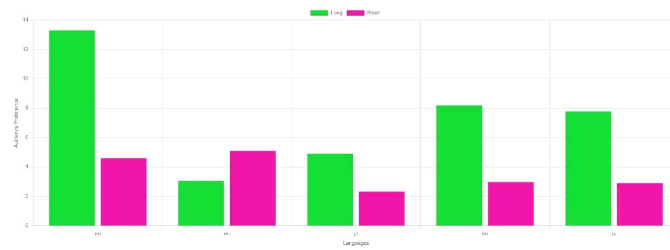
[3] Month Profit Trend Analysis: In this Month-wise Profit Trend Analysis emerges as a sophisticated tool, equipping producers with actionable insights derived from the intricate relationships between release months, budget, revenue, and profitability. Through this comprehensive examination, producers can optimize their decision-making processes, enhancing their ability to navigate the multifaceted dynamics of the film release schedule.

Month-wise Profit Trend Analysis



[4] The Runtime Language Analysis: An integral facet of the analysis entails the calculation of the mean popularity for each of these subgroups, offering a quantitative gauge of audience interaction for both short and long films across various languages. The synthesis of these insights is eloquently showcased through a Clustered Column Chart, where the X-axis delineates the selected languages, each manifested by two distinctive bars representing short and long films. Concurrently, the Y-axis visually captures the popularity metric. This refined analytical framework furnishes film producers with invaluable perspectives, enabling them to judiciously opt for a language congruent with their intended film duration—be it short or long. The nuanced comprehension of audience preferences within specific language and runtime categories empowers producers to make well-informed decisions, optimizing the probability of success in the dynamic landscape of film production. This analysis bestows upon producers a profound understanding of the intricate interplay between language, film duration, and audience popularity, acting as a guiding beacon toward strategic decisions in the ever-evolving domain of cinematic creation.

Runtime Language Analysis



III. PROJECT OUTCOME

The analyses explore different aspects of movie production and audience preferences. The Franchise Analysis uses a clustered column chart to reveal shifts in audience favor towards Franchise and Non-franchise movies over decades. In the Popular Genre Analysis, a pie chart visually represents audience preferences across various horror-genre combinations, providing insights into top-rated movies in each category. The Month Profit Trend Analysis offers a sophisticated tool for producers to optimize release schedules based on the relationships between release months, budget, revenue, and profitability. Finally, the Runtime Language Analysis employs a Clustered Column Chart to showcase audience preferences for short and long films in different languages, aiding producers in making informed decisions about film duration and language. These analyses collectively empower producers with valuable insights for navigating the dynamic landscape of movie production.

REFERENCES

[1]MongoDB Documentation
https://www.mongodb.com/docs/

[2]Flask Documentation (1.1.x),
https://flask.palletsprojects.com/_
/downloads/en/1.1.x/pdf/

[3] PyMongo 4.6.1 Documentation
https://pymongo.readthedocs.io/en/stable/atlas.html#

[4] Kaggledataset
https://www.kaggle.com/datasets/evangower/horror-movi es