DOUGLASCOLLEGE

Computing Studies and Information Systems

CSIS 4495 002 WINTER 2025

Applied Research Project

Developing a Python-Based Interactive Dashboard for Analyzing Movie Trends and Insights

Submitted to: Prof. Padmapriya Arasanipalai Kandhadai

Submitted by: Lan Dinh

Student number: 300383107

A. INTRODUCTION

The film industry has long captured the imagination of audiences worldwide, serving as a significant cultural force and an economic driver. As technology has advanced, the volume of data related to movies—such as box office performance, audience ratings, critical reviews, and genre trends—has grown exponentially. Harnessing this data to uncover actionable insights is critical for industry professionals, researchers, and movie enthusiasts alike.

While vast datasets on movies are readily available, many of them are presented in raw, unstructured forms, making it challenging to derive meaningful insights without technical expertise. Questions such as "What movies are currently trending?" or "How have audience preferences for certain genres changed over the years?" often remain unanswered due to the lack of accessible and user-friendly tools for exploration.

Stephen Few (2022) remarked, "Data visualization is just a tool. We could build houses before we had hammers and saws, the tools just let us do it better.". Similarly, N. Amar (2022) emphasized that a dashboard serves as a visual display of data, often pulling information from linked databases to provide a consolidated view of activities and trends. A dashboard is necessary of an organization for analyzing the activities and growth of the company. Additionally, Genevieve Hayes (2015) identified Python is by far the most relevant programming language for data science, making it an ideal choice for building dashboards. By applying these principles to movie analytics, a Python-based interactive dashboard has the potential to bridge existing gaps, making data-driven insights accessible to users without requiring technical expertise.

This research aims to develop such a Python-based interactive dashboard that integrates multiple movies datasets and empowers users to explore movie trends through dynamic visualizations and intuitive interfaces. The goal is to not only make the movies data more understandable but also to enhance public access to it. Ultimately, this research aims to contribute to the broader domain of data-driven decision-making in the film industry, providing a comprehensive, user-friendly platform for analyzing movie trends and insights.

B. PROPOSED RESEARCH PROJECT

1. Research Design and Objectives

The primary objective of this research is to develop an interactive Python-based dashboard that analyzes movie trends and insights, making complex datasets accessible to a diverse audience. This project will employ a combination of descriptive and exploratory research designs, focusing on identifying patterns, trends, and correlations within movie-related datasets. Drawing on insights from past studies on data visualization and dashboard development, as well as knowledge gained from coursework on data analytics, this project aims to bridge the gap between raw data and actionable insights. The dashboard will address questions such as trending movies, audience preferences, and key factors influencing movie success.

2. Methodology

This project focuses on creating simple and user-friendly dashboards that use refined information to present data patterns. The dashboard aims to make movie data comprehensible and actionable for a variety of audiences. Below is the step-by-step methodology:

2.1. Architecture

The project employs a streamlined architecture to process and visualize data effectively:

- The raw datasets are cleaned, processed, and stored in a structured database.
- Queries are applied to the database to generate refined, actionable datasets.
- Aggregation queries are used to summarize and process the data, which is then visualized in the form of analytics graphs and charts.
- The dashboard is designed to dynamically display these visualizations, ensuring simplicity and ease of navigation.

Figure 1. Architecture of entire methodology from where data is collected, then analyzed and results are interpreted.
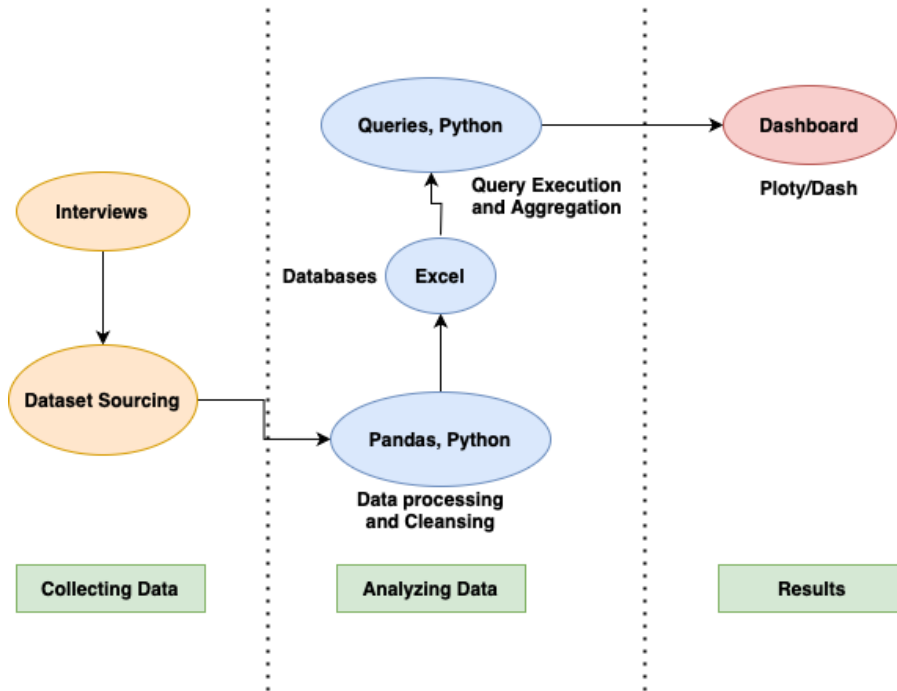
2.2. Data Collection Methods

This research project employs both primary and secondary data collection methods to gather the necessary information for developing the movie analytics dashboard.

Primary Data Collection: Interviews

To design a dashboard that effectively meets the needs of its users, interviews were conducted with a diverse group of potential users. The goal of these interviews was to understand user expectations, and preferences related to data dashboards. Sample interview questions included:

- Q1: How often do you use dashboards for analyzing or exploring movie-related data?
- Q2: What key features or functionalities would you like to see in a movie analytics dashboard?
- Q3: What type of visualizations (e.g., bar charts, pie charts, heatmaps) do you find most useful for exploring data?
- Q4: What challenges have you faced while using existing dashboards or tools for analyzing movie datasets?

Secondary Data Collection: Dataset Sourcing

The datasets used in this study were sourced from the Kaggle platform, which is a well-known repository for data science and machine learning practitioners. These datasets included:

- The Movies Dataset: Includes information about the casts, the crews and movie_id
- Netflix Movies and TV Shows Dataset: Contains data on Netflix's catalog, including release year, types, title of movies, directors, ratings and etc.
- IMDB Movies Dataset: Features movie-related information such as genre, meta_score, overviews, directors, and etc.

2.3. Data Analysis Technique

This project will apply exploratory data analysis (EDA) techniques, utilizing a variety of visual representations such as bar charts, scatter plots, line graphs, and heatmaps. These visualizations will simplify the process of uncovering relationships between variables, such as correlations between movie genres and audience ratings. The visualizations will also help in detecting broader trends, such as the rising popularity of specific genres or shifts in audience preferences across different periods. This ensures that data-driven insights are obtained, enabling the development of a dashboard that presents meaningful and actionable information to users through clear and compelling visual exploration.

3. Technologies

3.1. Programming Language

This project will use Python as the primary programming language due to its efficiency, versatility, and ease of use. Python is widely utilized in web and application development, machine learning, and big data processing. Its simple and intuitive syntax makes it one of the most popular languages for developers and data analysts alike. (Anna, 2023)

Python provides a comprehensive ecosystem of libraries and tools for data analysis and visualization, such as:

- **Pandas** for data manipulation and cleaning.
- **NumPy/SciPy** for mathematical and scientific computations.
- **Plotly** for interactive and dynamic data visualization.

These libraries can be easily installed using Python's package manager, pip, making Python a convenient choice for building a scalable and data-driven application.

3.2. Data Storage

The dataset will be stored and managed in Microsoft Excel for simplicity and accessibility. Excel is a widely used tool that provides an intuitive interface for small to moderately sized datasets. It allows for easy sharing and manual updates during the development and testing phases.

While Excel is not suitable for handling large-scale datasets or concurrent access, its simplicity and flexibility make it a practical choice for this project, ensuring efficient data storage and management.

3.3. Back-end and Front-end Frameworks

The back end of the dashboard will be developed using Flask, a lightweight and versatile Python web framework. Flask is well-suited for managing server-side operations, including handling user requests, routing, and integrating with data sources.

The front end of the dashboard will be developed using either Dash or Streamlit, both of which are Python-based frameworks designed for creating interactive user interfaces and visualizations.

By combining Flask for the back end and Dash or Streamlit for the front end, the project ensures a clear separation of responsibilities while delivering a seamless, user-friendly experience.

4. Expected Results

This project is expected to enhance my technical and analytical skills, including Python programming, data manipulation, visualization, and dashboard development. By working through the process of integrating multiple technologies and managing datasets, I will gain practical experience in building interactive applications and applying data analysis techniques to real-world problems.

The research is expected to produce a fully functional, Python-based interactive dashboard that allows users to dynamically analyze and visualize movie trends. The dashboard will enable users to explore insights such as trending genres, and audience preferences across different time periods.

The dashboard is designed to be user-friendly, ensuring that data analysis becomes accessible to a broad audience. Its interactive features, such as filters and customizable visualizations, will provide an engaging experience, empowering users to derive meaningful insights without requiring advanced technical skills.

5. Contribution to Practical Applications

The proposed dashboard addresses existing gaps in movie analytics by offering an integrated, interactive, and accessible tool. It will empower users to derive meaningful insights without requiring advanced technical skills, contributing to data-driven decision-making in the film industry. This research also demonstrates the practicality of Python web frameworks for developing robust, scalable, and user-friendly data-driven applications, providing a template for future innovations in other domains.

C. PROJECT PLANNING AND TIMELINE

1. Project Planning

Phase 1: Research Planning and Proposal Development (Week 1 and 2)

Milestones:

- Define the research objectives for the Python-based interactive dashboard (e.g., purpose, target users, and features) (Week 1).
- Identify the tools and frameworks to be used.
- Submit the research proposal and secure supervisor approval (End of January 26, 2025).

Deliverables:

- Approved research proposal.
- Initial requirements document detailing features and functionality.

Phase 2: Design and Prototype Development (Week 3 to 6)

Milestones:

- Create wireframes or mockups for the dashboard layout and features.
- Develop a basic prototype using Python and key libraries.

Deliverables:

- Dashboard wireframes or mockups.
- Prototype with basic functionality (e.g., loading data, basic visualizations).

Phase 3: Implementation and Feature Development (Week 7 and 8)

Milestones:

- Implement core dashboard features, such as data visualization, filtering, and interactivity.
- Add advanced features.

Deliverables:

- Fully functional dashboard with implemented features.
- Progress report highlighting the development process and milestones achieved. (For the midterm progress)

Phase 4: Testing and Optimization (Week 9 and 10)

Milestones:

- Conduct usability testing with client to identify issues and gather feedback.

- Optimize the dashboard for performance, usability, and responsiveness.

Deliverables:

- Usability testing report.

- Optimized and bug-free dashboard ready for deployment.

Phase 5: Documentation and Final Report Writing (Week 12)

Phase 6: Presentation and Submission (Week 13)

2. Worklog

WORK LOG FOR THE PROJECT						
NAME	**CONTACT**					
Lan Dinh	672-866-0515					
DEPARTMENT	**SUPERVISOR**				**TOTAL TIME ON SCHEDULE**	
ACTIVITY	**DATE**	**START TIME**	**END TIME**	**TOTAL TIME**	**STATUS**	**NOTES**
Doing Interest and Expertise Survey to find out what skills and technologies want to work on	January 12, 2025	5:00 PM	6:00 PM	1:00	Complete	
Finding the topic for the research project, looking for paper researches and previous works	January 14, 2025	5:00 PM	7:00 PM	2:00	Complete	

Creating and adding the Client in the Repo		9:00 AM	9:00 AM	0:00	Not Started	
Writing the draft proposal for the Client	January 20, 2025	9:00 AM	3:00 PM	6:00	In Progress	
Discussing to the Client about the choosing topic is Activity Reminder Mobile App for Douglas College students	January 23, 2025	10:00 AM	10:30 AM	0:30	Complete	
Changing the topic to the Dashboard Development by Python for analyzing movie trends and insights	January 24, 2025	10:00 AM	10:30 AM	0:30	Needs Review	
Rewriting the proposal	January 24, 2025	1:00 PM	8:00 PM	7:00	In Progress	

2. Project Contract

DASHBOARD DEVELOPMENT AGREEMENT #1

This Dashboard Development Agreement (the **"Agreement"**) is entered into 01/26/2024 (the **"Effective Date"**), by and between Padmapriya Arasanipalai Kandhadai (the **"Client"**) and Lan Dinh (the "**Developer**"), collectively referred to as "the **Parties.**"

1. **Project Description.**

Client wishes to hire Designer to create a Python-based Interactive Dashboard for Analyzing Movie Trends and Insights. The specific requirements and details as stated by Client are as follows:

- Design and develop a single-page interactive dashboard to visualize movie data, including trends, ratings, and movies performances.
- Integrate datasets sourced from Kaggle, such as "The Movies Dataset," "Netflix Movies and TV Shows Dataset," and "IMDB Movies Dataset."

- Enable dynamic visualizations, including bar charts, scatter plots, heatmaps, and line graphs.
- Include interactive features such as filters for genres, release years, and ratings, as well as a search function.
- Ensure the dashboard is user-friendly and responsive, with support for multiple devices.

2. **Revisions.** Client shall be entitled to 3 revisions. Any revisions beyond this number shall be chargeable at a rate of $50 per revision.

3. **Payment.** The Parties agree to the following Payment and Payment Terms:

   Total Fee for Services: $1,500
   Upfront Fee (Due Before Project Start Date): $500
   Remaining Balance Due: $1,000 upon project completion

4. **Confidentiality.** During the course of this Agreement, it may be necessary for Client to share proprietary information, including trade secrets, industry knowledge, and other confidential information, to Developer in order for Developer to complete the Dashboard in its final form. Developer will not share any of this proprietary information at any time, even after the Agreement is fulfilled. Developer also will not use any of this proprietary information for her personal benefit at any time, even after the Agreement is fulfilled.

5. **Ownership Rights.** Client continues to own any and all proprietary information it shares with Developer during the term of this Agreement for the purposes of the Project. Developer has no rights to this proprietary information and may not use it except to complete the Project. Upon completion of the Agreement, Client will own the final dashboard design.

   While Developer will customize Client's Dashboard to Client's specifications, Client recognizes that websites generally have a common structure and basis. Designer continues to own any and all template designs it may have created prior to this Agreement. Developer will further own any template designs it may create as a result of this Agreement.

6. **Representations and Warranties.**

   Developer. Developer represents and warrants that he/she has the right to enter into and perform this Agreement. Developer further represents and warrants that she has the right to utilize and distribute the designs created for Client and that such designs are not owned by anyone else to Developer's knowledge. In the event that Developer does not have these rights, Developer will repay any associated damages Client may experience or will take responsibility so that Client does not experience any damages.

Client. Client represents and warrants that is has the rights to use any proprietary information, including, but not limited to trade secrets, trademarks, logos, copyrights, images, data, figures, content, and the like that it may provide to Developer to be included in this Dashboard. In the event that Client does not have these rights, Client will repay any associated damages Developer may experience or will take responsibility so that Developer does not experience any damages.

7. **Disclaimer of Warranties.** Designer shall create a Website for Client's purposes and to Client's specifications. DEVELOPER DOES NOT REPRESENT OR WARRANT THAT SAID DASHBOARD WILL CREATE ANY ADDITIONAL PROFITS, SALES, EXPOSURE, BRAND RECOGNITION, OR THE LIKE. DEVELOPER HAS NO RESPONSIBILITY TO CLIENT IF THE DASHBOARD DOES NOT LEAD TO CLIENT'S DESIRED RESULT(S).

8. **Limitation of Liability.** UNDER NO CIRCUMSTANCES SHALL EITHER PARTY BE LIABILE TO THE OTHER PARTY OR ANY THIRD PARTY FOR ANY DAMAGES RESULTING FROM ANY PART OF THIS AGREEMENT SUCH AS, BUT NOT LIMITED TO, LOSS OF REVENUE OR ANTICIPATED PROFIT OR LOST BUSINESS, COSTS OF DELAY OR FAILURE OF DELIVERY.

9. **Legal Fees.** In the event of a dispute resulting in legal action, the successful party will be entitled to its legal fees, including, but not limited to its attorneys' fees.

10. **Legal and Binding Agreement.** This Agreement is legal and binding between the Parties as stated above. This Agreement may be entered into and is legal and binding both in the United States and throughout Europe. The Parties each represent that they have the authority to enter into this Agreement.

The Parties agree to the terms and conditions set forth above as demonstrated by their signatures as follows:

"CLIENT"

Signed: _____

Date: _____

"DEVELOPER"

Signed: __Lan Dinh_____

Date: __01/26/2024_____

D. REFERENCES

Amar, N., & Baliga, S (2022). Design and Development of Analytical Dashboard. *Journal of Current Research in Engineering and Science*.

Kuznetsova, A. (2023). Dashboard solutions for building analysis optimization: data analysis of environmental conditions and smart controls performance of building data and comparison of existing dash-boarding possibilities for data analysis. *Electrical and Automation Engineering*

Few, S. (2022). *Stephen Few on Data Visualization: 8 Core Principles*. Tableau Software. Retrieved from Tableau Software Web site: https://www.tableau.com/blog/stephen-few-data-visualization

Hayes, G. (2019). Which Programming Language Should Data Scientists Learn First?. Towards Data Science. Retrieved from Towards Data Science: https://towardsdatascience.com/which-programming-language-should-data-scientists-learn-first-aac4d3fd3038