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

March 16, 2025

PROGRESS REPORT #1

1. Work Date/Hours Logs

ACTIVITY	DATE	NUMBER OF HOURS	STATUS	NOTES
Doing Interest and Expertise Survey to find out what skills and technologies want to work on	January 12, 2025	1:00	Complete	Survey completed successfully. Identified interest in Mobile Developments and Building Dashboards by Tableau and PowerBI
Finding the topic for the research project, looking for paper research and previous works	January 14, 2025	2:00	Complete	Explored multiple topics. Narrowed down to Activity Reminder Mobile App.
Creating and adding the Client in the Repo	February 1, 2025	0:00	Complete	Client added to the repository. No time spent as it was instantaneous.
Writing the draft proposal for the Client	January 20, 2025	6:00	Complete	
Discussing to the Client about the choosing topic is Activity Reminder Mobile App for Douglas College students	January 23, 2025	0:30	Complete	Client provided feedback. Topic was approved initially but later changed.
Changing the topic to the Dashboard Development by Python for analyzing movie trends and insights	January 24, 2025	0:30	Approved	Topic changed based on client's suggestion. New focus on Python dashboard developments.
Rewriting the proposal	January 24, 2025	7:00	Complete	Proposal rewritten to reflect the new topic. Submitted for approval. And it was approved

Cleaning database (Netflix database) by using JupyterNotebook	January 26, 2025	2:00	Complete	Data cleaned and prepared for analysis.
Writing the Progress Report #1	January 29, 2025	5:00	Complete	
Writing questions for an interview	February 1, 2025	1:00	Complete	Questions are being developed for Participant A and B.
Drafting the ideas for the dashboard	February 2, 2025	1:00	Complete	Initial ideas for dashboard layout and functionality drafted.
Coding for the dashboard by Python	February 6, 2025	10:00	In Progress	Coding in progress. Focused on data visualization and user interface.
Doing the interview with Participant A	February 8, 2025	0:30	Complete	Still finding the potential participant and scheduling. The deadline is Feb 10
Doing the interview with Participant B	February 10, 2025	0:30	Complete	Still finding the potential participant and scheduling. The deadline is Feb 12
Analyzing interview results	February 14, 2025	2:00	Complete	Analyze feedback from Participants A and B to refine dashboard features.
Writing the Interview Questions and Reponses	February 14, 2025	2:00	Complete	Interview questions and responses documented and uploaded to GitHub.
Cleaning database (IMDB database) by using Jupyter Notebook	February 16, 2025	1:00	Complete	Data cleaned and prepared for analysis.
Adding IMDb dataset (imdb_top_1000.csv) to the repository	February 16, 2025	0:00	In Progress	Primary dataset for "Gross Earnings Explorer" added to repo. No time spent as it was instantaneous.
Coding the "Movie Trends Dashboard" welcome page	February 18, 2025	1:00	Complete	Initial script (Welcome.py) completed with layout and styling

Coding the "Explore Global Trends" dashboard	February 20, 2025	5:00	Complete	Initial script (GlobalTrend.py) completed with Netflix-based visualizations and filters. Enhanced with custom styling.
Drafting the midterm report	February 24, 2025	3:00	Complete	Midterm report drafted to summarise progress, including interviews, and initial coding efforts.
Coding the "Gross Earnings Explorer" dashboard	March 1, 2025	7:00	Complete	Initial script (GrossEarnings.py) completed with data loading, cleaning, and filters.
Enhancing dashboards with custom styling and informative elements	March 5, 2025	8:30	Complete	Improved visual appeal and usability of Welcome page, Explore Global Trends, and Gross Earnings Explorer dashboards.
Drafting the progress report #2	March 9, 2025	4:00	Complete	Progress report #2 drafted to detail dashboard development, interview insights, and visualization progress.
Creating user authentication for the dashboards	March 14, 2025		In Progress	Began implementing user authentication for dashboards to enhance security and personalization
Creating history watchlist for each user			In Progress	

2. Description of work done

This week, I made some progress on the Movie Trends Dashboard project. I finalized the interview document, which includes user research insights from two participants (a casual moviegoer and a data analyst), aimed at shaping the dashboard's design and functionality.

Following this, I selected IMDB dataset from Kaggle and cleaned it using Jupyter Notebook. With the cleaned dataset, I began coding the initial dashboard prototypes using Streamlit and Python,

focusing on creating an intuitive interface and integrating basic data visualization features. The work included setting up the dashboard structure, importing necessary libraries, and designing layouts for two key sections: the "Movie Trends Dashboard" welcome page and the "Gross Earnings Explorer" dashboard.

Additionally, I enhanced three dashboards: Welcome page, Global Movie Trends, Gross Earnings Explore with custom styling and informative elements to improve their visual appeal and usability.

3. Challenges and Solutions

A primary challenge this week was initiating the coding process with limited prior experience in Streamlit. To address this, I consulted online tutorials and adapted sample code, enabling me to successfully implement a functional dashboard layout complete with custom styling and interactive filters.

4. Repo Check in of Implementation

The following files and folders have been checked into the project repository, categorized based on their purpose and contribution to the project:

a. Reports and Documents folder:
- Interview Questions and Responses (Added: February 22, 2025) – The completed document detailing user preferences and expectations for the dashboard.

b. Misc folder:
- Work Logs (CSV Format) (Updated: March 13, 2025) – Logs all project activities from January 12, 2025, onward, providing a detailed record of progress and task completion.
- IMDB movies dataset.ipynb (Added: February 20, 2025) – Jupyter Notebook with scripts to preprocess and clean the IMDb dataset.

c. Implementation folder:
- imdb_top_1000.csv (Added: February 20, 2025) – The primary dataset used for the "Gross Earnings Explorer" dashboard.
- Welcome.py (Added: March 13, 2025) – Initial Python script for the "Movie Trends Dashboard" welcome page, including layout and styling.

- gross_earnings_explorer.py (Added: March 13, 2025) – Initial script for the "Gross Earnings Explorer" dashboard, featuring data loading, cleaning, and interactive filters using the IMDb Top 1000 dataset.

5. Dashboard Code Explanation

This week's coding efforts produced three dashboard components: the "Movie Trends Dashboard" welcome page, the "Gross Earnings Explorer" dashboard (IMDb-based), and the "Explore Global Trends" dashboard (Netflix-based). Below is an explanation of the code developed so far:

a. Importing Libraries and Setting Up the Dashboard

The code for all dashboards begins by importing key libraries:

- *Streamlit (st):* Used to create the web interface.
- *Pandas (pd):* For data manipulation and analysis.
- *NumPy (np):* For numerical operations.
- *Seaborn (sns) and Matplotlib (plt):* For data visualization.
- *Plotly Express (px):* For creating interactive charts.
- *Datetime:* For handling date-related operations.

"Movie Trends Dashboard": Features a wide layout, a blue-themed title, and a movie-themed background image.

"Explore Global Trends": Includes a red-themed title and a global connectivity background, emphasizing worldwide content trends.

"Gross Earnings Explorer": Uses an orange-themed title and a film-related background, with a focus on earnings analysis.

b. Loading and Cleaning the Data

IMDb dataset (imdb_top_1000.csv) is loaded into a Pandas DataFrame. The data is cleaned by filling missing Certificate with "Unrated," Meta_score with the mode, converting Gross to millions, Released_Year to integers, and Runtime to minutes. Genres were split for multi-select filtering.

c. Creating Interactive Filters

- Gross Earnings Explorer: Sidebar filters include sliders for IMDb rating, Meta score, and release year, plus a multi-select genre dropdown.

- Explore Global Trends: Sidebar filters feature a release year slider, multi-select options for content type (Movie/TV Show), rating category, and country. A date picker on the main page filters by date_added.

These filters align with Participant A's request for genre and rating filters and Participant B's need for analytical flexibility.

d. Visualizations

Welcome page: The welcome page currently lacks charts, serving as an introductory interface with a navigation sidebar and planned features.

Explore Global Trends: The "Explore Global Trends" dashboard organizes its visualizations into three tabs—"Duration Trends," "Number of Titles" and "Recommendation". Each chart is designed to be interactive and responsive to sidebar filters.

Tab 1: Duration Trends – Top Countries by Duration (hours)

- Shows the total duration (in hours) of Netflix content for the top 10 countries, split by content type (Movies vs. TV Shows), to identify major contributors to watch time.
- Six metrics (e.g., total duration, avg. duration/title, Movies/TV Shows share, top country) are styled in black boxes with red labels.

Tab 2: Number of Titles – Global Map of Title Counts by Country

- Visualizes the number of unique Netflix titles per country globally, highlighting content availability distribution.
- Clicking a country triggers a donut chart (below) for genre breakdown. It will show the top 10 genres (plus "Other") by percentage of unique titles in a clicked country from the map, revealing local content preferences.

Tab 3: Recommendations – Personalized Recommendations

- Suggests up to five Netflix titles based on a user-selected genre and filtered countries, aiding content discovery.
- A dropdown lists available genres, updating the table with five random matching titles (or fewer if limited). The CSV button exports the table.

Gross Earnings Explorer: The "Gross Earnings Explorer" dashboard organizes its visualizations into three tabs—"Gross Trends" "Directors & Actors" and "Gross Impactor". Each chart is designed to be interactive and responsive to sidebar filters.

Metrics Display (Dataset Snapshot)

- Provides a quick summary of the filtered dataset before diving into detailed tabs, showing total movies, total gross earnings, average IMDb rating, and average Meta score.
- Updates dynamically with sidebar filter changes, offering an immediate snapshot without user input beyond filtering.

Tab 1: Gross Trends - Total Gross Earnings (M) Over Time: Peak Performers

- Displays the total gross earnings (in millions) of movies aggregated by release year, highlighting trends in box office performance over time and identifying peak earning years.
- Hovering activates a unified tooltip showing earnings across all years, while clicking a point reveals detailed information about the highest-grossing movie for that year (e.g., title, gross, IMDb rating, Meta score, runtime, genre, certificate, overview, and poster image).

Tab 2: Directors & Actors - Directors of Blockbusters: Top 10 by Gross and Stars of the Box Office: Top 10 by Gross

- Identifies the top 10 directors and actors by total gross earnings (in millions).
- Text labels above each bar display earnings (e.g., "$1,234.56M"), and two metrics below— mean and median gross—are presented in Streamlit metric widgets with a delta for comparisons.

Tab 3: Gross Impactor – Gross Impact Explorer

- Explores the relationship between gross earnings (in millions) and a user-selected numeric metric (e.g., IMDb rating, Meta score, runtime, release year), helping identify factors influencing box office success.
- An expander offers a second identical chart with hover data (title and metric value) and a white template for contrast.