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

February 9, 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	January 29, 2025	0:00	In Progress	
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	In Progress	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		0:00	In Progress	Still finding the potential participant and scheduling. The deadline is Feb 10
Doing the interview with Participant B		0:00	In Progress	Still finding the potential participant and scheduling. The deadline is Feb 12
Analyzing interview results		0:00	On Hold	Analyze feedback from Participants A and B to refine dashboard features.
Developing data visualizations for the dashboard and testing the dashboard functionality		0:00	On Hold	Test the dashboard for bugs, usability, and performance.

2. Description of work done

This week, I completed several key tasks to advance the project. I started by conducting an Interest and Expertise Survey to identify the skills and technologies I wanted to focus on. This process led me to explore Mobile Development, as I was more familiar with it. Initially, I proposed developing an Activity Reminder Mobile App for Douglas College students. However, after consulting with the client, I decided to challenge myself by learning new technologies and shifted the project focus to Dashboard Development using Python to analyze movie trends and insights.

To align with the new direction, I rewrote the proposal, which was subsequently approved. I selected the Netflix dataset from Kaggle and cleaned it using Jupyter Notebook, as it provided a clear and interactive way to process and inspect the data.

Additionally, I began drafting the Progress Report and started coding the dashboard in VS Code Studio, focusing on data visualization and user interface development. I also prepared interview questions and outlined initial ideas for the dashboard layout.

3. Challenges and Solutions

One of the key challenges I faced was the need to change the project topic at the last minute. The transition from mobile development to dashboard analytics required a quick adjustment in approach and research. I had to rewrite the proposal, redefine the project scope, and adapt to new technologies, all under tight time constraints. Despite the added pressure, I successfully revised the proposal, and it was ultimately approved.

Another challenge is the time required to clean the database, as the dataset was large and required careful handling to ensure accuracy. To address this, I used Python libraries like Pandas and NumPy to streamline the process. Moving forward, I plan to complete the dashboard coding, conduct interviews, and analyze the results to refine the dashboard features.

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. Research folder (added: February 1, 2025):
   Contains initial research on potential topics, aiding in the selection of the final project direction.
b. Reports and Documents folder:
- Approved Proposal (Added: February 6, 2025) – Reflects the updated project scope after the topic change.
- Interview Questions (Added: February 1, 2025) – Draft questions prepared for user research.
c. Misc folder:

- Work Logs (CSV Format) (Added: February 7, 2025) – Logs all project activities from January 12, 2025, onward, providing a detailed record of progress and task completion.
- Jupyter Notebook for Database Cleaning (Uploaded: February 6, 2025) – Includes scripts used to preprocess and clean the dataset for analysis.

d. Implementation folder:

- Netflix Database (Uploaded: February 6, 2025) – The primary dataset for analysis.
- Dashboard Code (Added: February 6, 2025) – Initial Python scripts for dashboard development, including data visualization components.

5. Dashboard Code Explanation

a. Importing Libraries and Setting Up the Dashboard

The code begins by importing essential 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.

The dashboard is configured with a title, layout, and custom styling using HTML and CSS. A Netflix image is displayed at the top to enhance the visual appeal.

b. Loading and Cleaning the Data

The Netflix dataset (netflix_titles.csv) is loaded into a Pandas DataFrame. The data is cleaned to handle missing values:

- Rows with missing `date_added` or `duration `values are dropped. The `date_added` column is converted to a datetime format, and the `duration` column is extracted and converted to a numeric format for analysis.
- Missing values in `rating`, `duration`, `country`, `cast`, and `director` columns are filled with appropriate defaults (e.g., mode for categorical columns and "Unknown" for text columns).

c. Creating Interactive Filters

The dashboard includes a sidebar with interactive filters:

- *Content Type:* Users can filter by movie or TV show.
- *Rating:* Users can select specific content ratings.
- *Release Year:* Users can choose specific years.
- *Country:* Users can select one or more countries.

d. Visualizations

The dashboard features four key visualizations:

Line Chart (Number of Titles Added Over Time by Type):

- Shows the trend of movies and TV shows added to Netflix over time. The date picker allows users to select a specific time range for analysis.
- The x-axis represents the date, and the y-axis represents the number of titles.
- Different colors distinguish between two content types are movies and TV shows.

Bar Chart (Countries by Duration):

- Displays the top 10 countries with the highest total duration of content.
- The x-axis represents countries, and the y-axis represents the total duration.
- A color gradient highlights countries with higher durations.

Filled Map (Number of Titles by Country):

- A choropleth map showing the number of titles available by country.
- Darker shades indicate countries with more titles.

Content Recommendations:

- Based on the selected country, the dashboard recommends titles from the most popular genre in that country.
- Recommendations are displayed in a clean, formatted table.