Viet Nam National University - HCMC

UNIVERSITY OF SCIENCE

FACULTY OF INFORMATION TECHNOLOGY

PLAN REPORT ON FINAL PROJECT

Topic: RESEARCH ON MUSIC TASTE WORLDWIDELY

Course: Introduction to Data Science

Student on Project: Group 8
21127528 - Nguyen Thi Minh Minh
21127063 - Nguyen Van Dang Huynh
21127423 - Tran Hieu Tam

Instructor:
Teacher: Nguyen Bao Long
Teacher: Le Ngoc Thanh
Teacher: Pham Trong Nghia
Teacher: Nguyen Ngoc Thao

December 14, 2023

Contents

1   Member Information

STT	Full name	Student ID	Email
1	Nguyen Thi Minh Minh	21127528	ntmminh21@clc.fitus.edu.vn
2	Nguyen Van Dang Huynh	21127063	nvdhuynh21@clc.fitus.edu.vn
3	Tran Hieu Tam	21127423	thtam21@clc.fitus.edu.vn

Table 1: Member Information

2   Project Information

2.1   Project Title

Research on music taste wordwidely

2.2   Project Description

This is a project officially called "Researching on music taste worldwide", carried out from November to December 2023 at the Faculty of Information Technology, University of Science, VNU-HCM. Currently, digital media in general and music in cyberspace in particular are developing very strongly, creating a digital community where people can easily share and get rich from it. According to statistics from Youtube, their Total videos reach Over 800 million videos (2023), with Total users reaching 2.70 billion of 100+ countries worldwide (2023), bringing the platform's revenue in 2022 to 29.24 billion USD (2022). At the same time, according to Spotify's statistics, their Total songs reach Over 100 million songs (2023), with Total users reaching 551 million of 184 countries worldwide (2023), and have revenue in 2022 reaching 12.33 billion USD (2022). . Through our project, the team aims to understand and provide accurate analysis of the market and today's music listening tastes of the world's netizens. From there, there are appropriate suggestions for authors in the process of publishing their works, creating many development opportunities for everyone with ability. Information is taken from sources Kworb.net, Youtube.com and Spotify.com. We took a list of the top 2,500 videos with the highest views on Youtube, from which we have a basis to analyze and learn valuable experiences from the above videos. The above project will still have many shortcomings. Therefore, we look forward to receiving everyone's candid comments to help our project be more objective, accurate and effective.

2.3   Plan management tool

- Trello: `https://trello.com/b/GvZXYaQX/youtube-project-ds`

- Github: `https://github.com/danghuynh-bit/youtube_project_DS.git`

3 Project Plan

3.1 Project Schedule: Phase 1

STT	Task	Content	Member	Due date
1	First meeting	Meeting and decide topic.	All	27/11/2023
2	Making plan on Trello	Making plan on Trello	21127423	27/11/2023
3	Crawl Data (1st Crawl)	Crawl data from web kworb.net	21127528	27/11/2023 - 29/11/2023
4	Crawl Data (2nd Crawl)	Crawl data from web kworb.net	21127423	27/11/2023 - 29/11/2023
5	Crawl Data (3rd Crawl)	Crawl data from Youtube using API	21127063	27/11/2023 - 29/11/2023
6	Crawl Data (4th Crawl)	Crawl data from Spotify using API	21127528	29/11/2023 - 31/11/2023
7	Write Introduction Notebook	Write introduction notebook	21127423	29/11/2023 - 31/11/2023
8	Second Meeting	Gather all data collected and plan for next phase	All	31/11/2023

Table 2: Project Schedule: Phase 1

3.2 Project Schedule: Phase 2

STT	Task	Content	Member	Due date
1	Data Processing (1st process)	Process data on the first sample	21127063	31/11/2023 - 02/12/2023
2	Data Processing (2nd process)	Process data on the first sample	21127423	31/11/2023 - 02/12/2023
3	Data Processing (3rd process)	Process data on the second sample	21127528	31/11/2023 - 02/11/2023
4	Data Exporation (Basic Level)	Doing overview on the first and second sample	21127063	02/12/2023 - 03/12/2023
5	Data Exporation (Basic Level)	Doing overview on the first and second sample	21127528	02/12/2023 - 03/12/2023
6	Data Exporation (Advanced Level)	Propose 2 questions and answer with visualization	21127423	02/12/2023 - 04/12/2023
7	Data Exporation (Advanced Level)	Propose 3 questions and answer with visualization	21127063	02/12/2023 - 04/12/2023
8	Data Exporation (Advanced Level)	Propose 2 questions and answer with visualization	21127528	02/12/2023 - 04/12/2023
9	Third Meeting	Gather all data exporation sample and plan for next phase	All	04/12/2023

Table 3: Project Schedule: Phase 2

3.3 Project Schedule: Phase 3

STT	Task	Content	Member	Due date
1	Data Modeling Preparation	Data Preparation	21127528	04/12/2023 - 05/12/2023
2	Data Modeling (1st)	Problem statement and Data Preparation	21127063	04/12/2023 - 05/12/2023
3	Data Modeling (2nd)	Build Model 1: Linear Regression	21127063	05/12/2023 - 09/12/2023
4	Data Modeling (3rd)	Build Model 2: Polynomial Regression	21127063	05/11/2023 - 09/12/2023
5	Data Modeling (4th)	Build Model 3: Neural Network	21127528	05/11/2023 - 09/12/2023
6	Data Modeling Evaluation	Evaluate Models	21127423	09/12/2023 - 10/12/2023

Table 4: Project Schedule: Phase 3

3.4 Project Schedule: Phase 4

STT	Task	Content	Member	Due date
1	Final Meeting	Gather all data modeling sample and start final report	All	11/12/2023
2	Write Plan Report	Write plan report	21127528	12/12/2023
3	Slide for presentation	Slide for presentation	21127423	11/12/2023 - 13/12/2023
4	Reorganize Source code and Report	Reorganize Source code and Report	21127063	11/12/2023 - 13/12/2023

Table 5: Project Schedule: Phase 4