UCD Michael Smurfit
Graduate Business School

M.Sc. Business Analytics

MIS41110 Programming for Analytics

Ukulele Tuesday Data Analyzer

(22nd November 2024)

Group 1: Team Members
Gourish Deshpande (24200847)
Jaee Jain (24206089)
Nidhi Sharma (24201811)
Staffin Thomas Shabu (24201287)

Individual Contributions

Team Member	Contribution
Gourish Deshpande	Contributed to Data Visualization and played a significant role in the Data Validation for the project.
Jaee Jain	Focused on Data Querying and Filtering, contributed to Data Visualization, and actively worked on the UI Development.
Nidhi Sharma	Actively worked on Data Querying and Filtering, contributed significantly to Data Visualization, and participated in UI Development.
Staffin Thomas Shabu	Worked on Data Visualization and made important contributions to the Data Validation.

User Manual To Run The Code

1. Introduction

Ukulele Tuesday Data Analysis – This program is targeted to Ukulele Tuesday group to allow analysis on its collection of songs. The application allows for querying and filtering song data and view trends in performances.

2. Installation and Setup

1. Download and install Python from python.org.
2. Install the required libraries – PyQT6, Pandas, Matplotlib
3. Download the program files in a location of your choice.
4. Locate your data files (in csv format) – Tabdb.csv, Playdb.csv, Requestdb.csv (your files can have any name)

3. Program flow

3.1 Loading Data Files

1. Open the program.
2. Upload files tabdb.csv, playdb.csv, and requestdb.csv.
3. The program will validate your data (as per section 5) for consistency, reporting (an error message window will pop up) any missing columns or errors in data format. You must correct your data in the files and re-upload to move forward.
4. Confirm successful data loading by reviewing the summary displayed.
5. Click on 'NEXT' button to go to Query Page.

3.2 Filtering and Querying Data

1. Navigate through the tabs to query data for each file.
2. Select the columns to display (e.g., title, artist, duration) and enter date range to narrow down the results.
3. Click 'APPLY FILTERS' to display the filtered results in a table/pie-chart.
4. Click on 'NEXT' button to see the graphs.

3.3 Generating Plots

1. Navigate through the tabs to see the plots.
2. The section contains 7 tabs with graphical data (based on entire data given by user) with each type of graph mentioned below.
   The program offers the following visualizations:
   a. **Histogram of Songs by Difficulty**: Visualizes the distribution of song difficulty levels.
   b. **Histogram of Songs by Duration**: Provides a view of song lengths.
   c. **Bar Chart of Songs by Language**: Shows language distribution.
   d. **Bar Chart of Songs by Source**: Compares songs from different sources.
   e. **Bar Chart of Songs by Decade**: Groups songs based on their release decade.
   f. **Cumulative Line Chart of Songs Played Over Time**: Shows the trend of song performances.
   g. **Pie Chart of Songs by Gender**: Visualizes the gender of lead vocalists in a pie chart format.
3. Click on 'DOWNLOAD ALL GRAPHS (PDF)' to get a local copy of all plots.

4. Error Handling and Troubleshooting

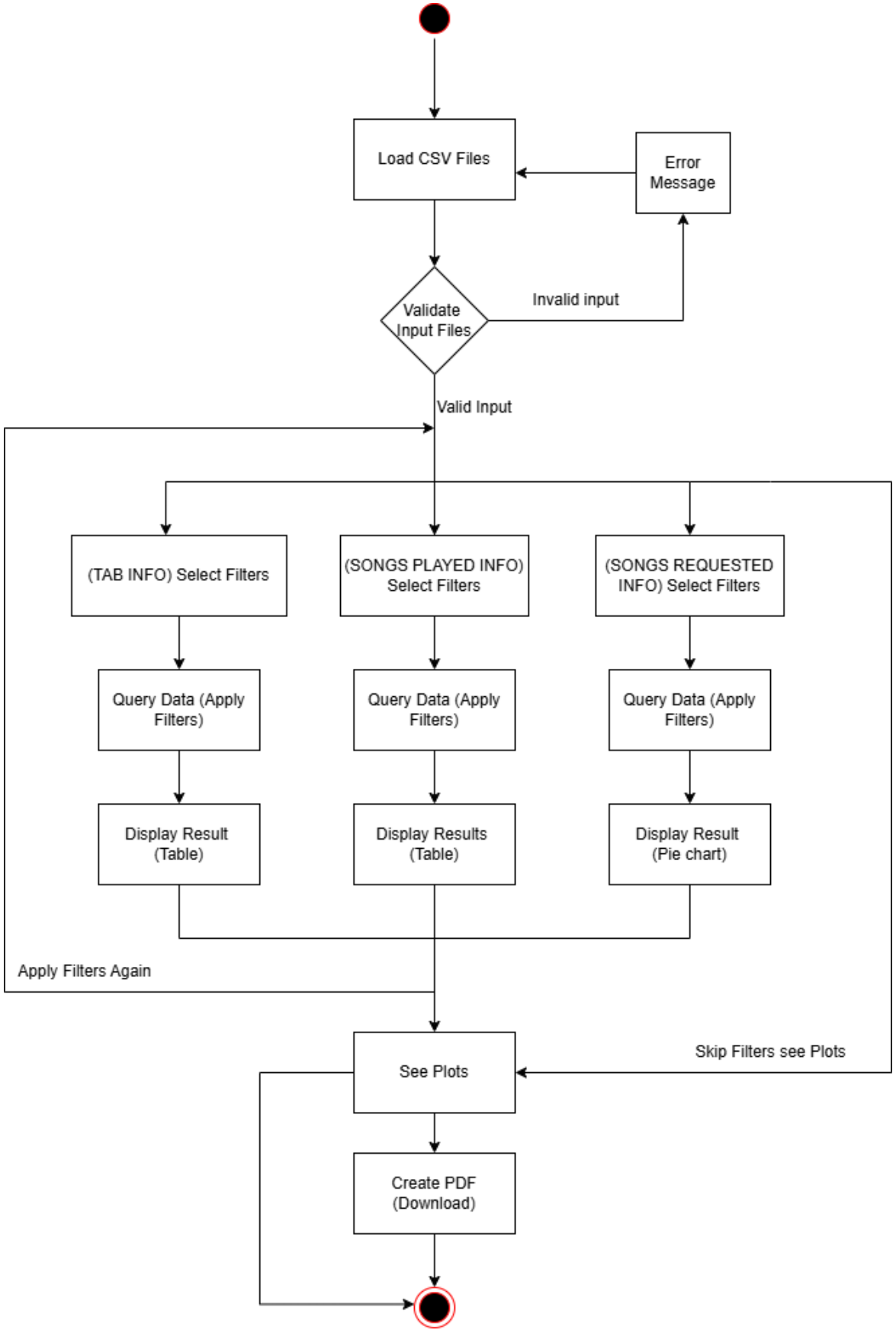
Refer below tables for validation of files.

Validations For PlayDB File	
All columns should be present in file	
Column	Valid Entries
Song	Numeric, Aplhanumric,Strings
Artists	Any value is accepted considered liberty in naming
Dates	Format - YYMMDD
Values in date column	Positive values including decimals

Validations For RequestDB File	
All columns should be present in file	
Column	Valid Entries
Song	Numeric, Aplhanumric,Strings
Artists	Any value is accepted considered liberty in naming.
Dates	Format - YYMMDD
Values in date column	G,A,? and blank values.

Validations For TabDB File	
All columns should be present in file	
Column	Valid Entries
Song	Numeric, Aplhanumric,Strings
Artists	Any value is accepted considered liberty in naming.
Year	Year should be > 1800
Type	Person,Group
Gender	Male,Female, Duet, Ensemble,Intrumental
Duration	Float Values
Language	Language liberty
Tabber	Tabber entry is not shown because of privacy concern.
Source	old,new,off
Date	Format YYMMDD
Difficulty	Values from 1 to 5
Speicalbooks	Liberty considerd for specialbooks.

UML Diagram


Load CSV Files
Error Message
Validate Input Files
Invalid input
Valid Input
(TAB INFO) Select Filters
(SONGS PLAYED INFO) Select Filters
(SONGS REQUESTED INFO) Select Filters
Query Data (Apply Filters)
Query Data (Apply Filters)
Query Data (Apply Filters)
Display Result (Table)
Display Results (Table)
Display Result (Pie chart)
Apply Filters Again
Skip Filters see Plots
See Plots
Create PDF (Download)

References

- Shaw, Z. 2017, Learn Python 3 the hard way: a very simple introduction to the terrifyingly beautiful world of computers and code, Addison-Wesley, London;Boston;https://learnpythonthehardway.org/
- Python Software Foundation (2024) Python 3.12.6 documentation. Available at: https://docs.python.org/3.12/
- The pandas development team (2024) pandas documentation. Available at: https://pandas.pydata.org/docs/
- The Matplotlib Development Team (2024) Matplotlib documentation. Available at: https://matplotlib.org/stable/contents.html
- Stack Overflow (n.d.) Stack Overflow: Where Developers Learn, Share, & Build Careers. Available at: https://stackoverflow.com/
- OpenAI (2024) ChatGPT (November 2024 version). Available at: https://chat.openai.com/