Attendance Aggregator Project Documentation

Table of Contents

1. Introduction

1.1 Objective

The objective of this document is to define the functional and non-functional requirements for the Attendance Aggregator application. The application aggregates attendance data from multiple PDFs, resolves conflicts, and generates combined reports in either PDF or Excel format.

1.2 Scope

The application is a desktop utility developed using Python and its libraries like tkinter, pandas, PyPDF2, and reportlab. It processes PDF files containing attendance data and outputs a merged report with calculated attendance percentages. The generated report can be saved as either a PDF or Excel file.

2. Overall Description

2.1 Product Perspective

The Attendance Aggregator is a standalone desktop application that allows users to:

- Select multiple PDF files containing attendance data.

- Automatically extract attendance data from the selected files.

- Combine the data into a single comprehensive report.

- Resolve any data conflicts (e.g., duplicate roll numbers with different names).

- Generate the final attendance report in either PDF or Excel format.

2.2 Product Features

- **File Selection**: Users can select multiple PDF files for the attendance aggregation.

- **Data Extraction**: Attendance data is extracted from the PDF files, and roll numbers, names, total classes, and attendance are parsed into a DataFrame.

- **Conflict Resolution**: In case of conflicting data (e.g., same roll number with different names), the user can manually resolve these conflicts through a dialog box.

- **Report Generation**: The final aggregated data is output in either PDF or Excel format.

- **Calculation of Attendance Percentages**: For each student, the application calculates attendance percentage per subject, as well as total classes attended and overall attendance percentage.

- **Progress Indicator**: The application includes a progress bar to indicate the status of the report generation process.

2.3 User Characteristics

- Basic computer skills to operate a desktop application.

- Familiarity with PDF and Excel file formats.

- No programming knowledge is required.

2.4 Constraints

- The application requires that the attendance data in the PDF files follows a specific format (i.e., roll number, name, total classes, attendance).

- The input PDFs should contain clear tabular data to be processed correctly.

- Conflict resolution must be handled manually by the user in cases where discrepancies exist between names for the same roll number.

3. Functional Requirements

3.1 File Selection

Description: Users can select one or more PDF files from their local storage to be processed.

- **Input**: PDF files

- **Output**: List of file paths to be processed

- **Error Handling**: If no files are selected, an error message is shown.

3.2 Data Extraction

Description: The application extracts table data (roll number, name, total classes, and attendance) from the selected PDF files.

- **Input**: PDF file(s)

- **Output**: Parsed attendance data in a DataFrame

- **Error Handling**: If the file format is incorrect or unreadable, an error message is shown.

3.3 Data Aggregation

Description: Attendance data from multiple PDFs is aggregated into a single DataFrame, ensuring that each student's data (roll number and name) is consolidated correctly.

- **Input**: Multiple DataFrames

- **Output**: Combined DataFrame

- **Error Handling**: If data conflict occurs (e.g., same roll number with different names), the user is prompted to resolve the conflict manually.

3.4 Conflict Resolution

Description: When a conflict is detected (e.g., multiple names for the same roll number), the application prompts the user to select one of the conflicting names.

- **Input**: Conflict data (roll number, conflicting names)

- **Output**: Resolved name selection

- **Error Handling**: If the user cancels conflict resolution, the process is aborted.

3.5 Attendance Percentage Calculation

Description: The application calculates the attendance percentage for each student in each subject and overall.

- **Input**: DataFrame with attendance data

- **Output**: DataFrame with calculated percentages

- **Error Handling**: Attendance percentage is set to 0 if the total number of classes is 0.

3.6 Report Generation

Description: The aggregated attendance data is saved in either PDF or Excel format based on the user's choice.

- **Input**: DataFrame with aggregated data

- **Output**: PDF or Excel file

- **Error Handling**: If the file cannot be saved (e.g., due to permission issues), an error message is shown.

3.7 Progress Indicator

Description: A progress bar shows the status of the report generation process.

- **Input**: Progress percentage

- **Output**: Updated progress bar

- **Error Handling**: If an error occurs during the process, the progress is halted, and an error message is displayed.

4. Screenshots of Project

4.1 Input

S no.	Roll no.	Name	Total Classes	Classes Attended
1	0801CS221068	Jay Patel	10	2
2	0801it221065	Abhishek Kanade	10	
3	0801IT221079	Dhruv Patel	10	5
4	0801BM211058	Amesha Raghuvanshi	10	10
5	0801CS221011	Jagadish Kurkure	10	8
6	0801CE221447	Haridwar Kajwe	10	1
7	0801Ie221039	Sanjana Pal	10	0
8	0801CC221048	Ayush Pal	10	
9	0801ET221049	Manish Ashutosh	10	9
10	0801EE221002	Agrad Kumar	10	10
11	0801BM221051	Ashutosh Jayesh	10	2
12	0801EE221001	Jay	10	6

S no.	Roll no.	Name	Total Classes	Classes Attended
1	0801CS221068	Jay Patel	30	15
2	0801it221065	Abhishek Kanade	30	8
3	0801IT221079	Dhruv Patel	30	
4	0801BM211058	Amesha Raghuvanshi	30	19
5	0801CS221011	Jagadish Kurkure	30	27
6	0801CE221447	Haridwar Kajwe	30	9
7	0801Ie221039	Sanjana Pal	30	1
8	0801CC221048	Ayush Pal	30	0
9	0801ET221049	Manish Ashutosh	30	
10	0801EE221002	Agrad Kumar	30	10
11	0801BM221051	Ashutosh Jayesh	30	11
12	0801EE221001	Jay	30	18

Roll no.	Name	Total Classes	Classes Attended
0801CS221068	Jay Patel	22	20
0801it221065	Abhishek Kanade	22	
0801IT221079	Dhruv Patel	22	9
0801BM211058	Amesha Raghuvanshi	22	19
0801CS221011	Jagadish Kurkure	22	20
0801CE221447	Haridwar Kajwe	22	7
0801Ie221039	Sanjana Pal	22	0
0801CC221048	Ayush Pal	22	12
0801ET221049	Manish Ashutosh	22	
0801EE221002	Jay Narayan	22	9
0801BM221051	Ashutosh Jayesh	22	11
0801EE221001	Jay	22	14

4.2 Output

Name Conflict

Roll No 0801EE221002 has multiple names: Agrad Kumar, Jay Narayan. Please choose one name:

Agrad Kumar

OK   Cancel

Success

The combined attendance report has been generated as a Excel file!

OK

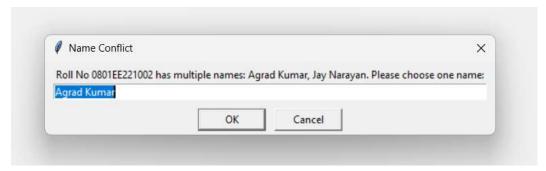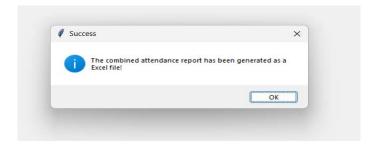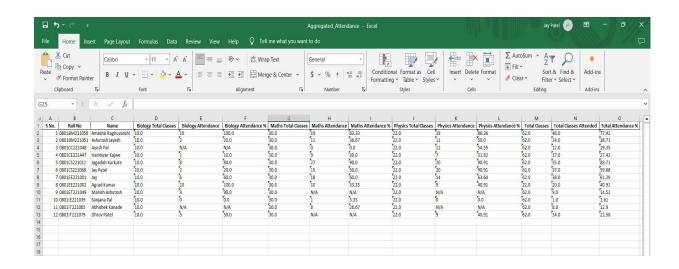| S No.	Roll No	Name	Biology Total Classes	Biology Attendance	Biology Attendance %	Maths Total Classes	Maths Attendance	Maths Attendance %	Physics Total Classes	Physics Attendance	Physics Attendance %	Total Classes	Total Classes Attended	Total Attendance %
1	0801BM211058	Amesha Raghuvanshi	10.0	10	100.0	30.0	19	63.33	22.0	19	86.36	62.0	48.0	77.42
2	0801BM221051	Ashutosh Jayesh	10.0	2	20.0	30.0	11	36.67	22.0	11	50.0	62.0	24.0	38.71
3	0801CC221048	Ayush Pal	10.0	N/A	N/A	30.0	0	0.0	22.0	12	54.55	62.0	12.0	19.35
4	0801CE221447	Haridwar Kajwe	10.0	1	10.0	30.0	9	30.0	22.0	7	31.82	62.0	17.0	27.42
5	0801CS221011	Jagadish Kurkure	10.0	8	80.0	30.0	27	90.0	22.0	20	90.91	62.0	55.0	88.71
6	0801CS221068	Jay Patel	10.0	2	20.0	30.0	15	50.0	22.0	20	90.91	62.0	37.0	59.68
7	0801EE221001	Jay	10.0	6	60.0	30.0	18	60.0	22.0	14	63.64	62.0	38.0	61.29
8	0801EE221002	Agrad Kumar	10.0	10	100.0	30.0	10	33.33	22.0	9	40.91	62.0	20.0	40.91
9	0801ET221049	Manish Ashutosh	10.0	9	90.0	30.0	N/A	N/A	22.0	N/A	N/A	62.0	9.0	14.52
10	0801IE221039	Sanjana Pal	10.0	0	0.0	30.0	1	3.33	22.0	0	0.0	62.0	1.0	1.61
11	0801IT221065	Abhishek Kanade	10.0	N/A	N/A	30.0	8	26.67	22.0	N/A	N/A	62.0	8.0	12.9
12	0801IT221079	Dhruv Patel	10.0	5	50.0	30.0	N/A	N/A	22.0	9	40.91	62.0	14.0	22.58

5. References

- Python 3.x Documentation : https://docs.python.org/3/

- Tkinter Documentation : https://docs.python.org/3/library/tk.html

- ReportLab User Guide : https://www.reportlab.com/docs/reportlab-userguide/

- Pandas Documentation : https://pandas.pydata.org/docs/