Mini Project Report - Advanced Python

# Data Analysis and Visualization

L&T Technology Services

## Document History

| Ver. Rel. No. | Release Date | Prepared. By | Reviewed By | Approved By | Remarks/Revision Details |
|---|---|---|---|---|---|
| 1 | 20/7/2021 | 99004952<br>99004957<br>99004958<br>99004959 | | | |
| 2 | | | | | |
| 3 | | | | | |
| 4 | | | | | |
| 5 | | | | | |

## Table of contents

**GitHub Repository Folder Structure**

| Folder | Description |
|---|---|
| 1_Requirements | all requirements and research analysis work |
| 2_Architecture | all files with UML diagrams specifying design |
| 3_Implementation | all the code files |
| 4_TestPlan | details of test plan |
| 5_Report | documentation and report files |
| 6_Images_Videos | contains images and videos |
| 7_Others | information on standup calls |

## Features

| Feature | Description |
|---|---|
| F1 | Database for login |
| F2 | Implementation of Data Validation |
| F3 | Implementation of Data Analysis |
| F4 | Implementation of Data Visualization |

## Introduction

A learner's performance is measured in terms of his/her marks. It is one of the indicators of a learner's progress.This application helps to monitor the learning progress in a very clear and informative manner. The analysis and visualization of the data helps to understand the performance of all with respect to each course and overall in a better way. The application provides a detailed sheet containing the total, percentages, and grades of all learners. It also provides various options to users to get the deatils of the learners like getting details of learners with scope of improvement, to search for a particular learner and so on.

## Research

- Data consolidation is the process where all the datas from the organization is integrated in one place.
- The consolidation of data is an important step for data management process and integration. This makes the information of data to be available quickly and easily, also this increases productivity and efficiency by having all datas in one place.
- The consolidation of data allows us to gather the datas from different worksheets to a master worksheet and vice versa.
- The excel consolidate function allows us to select datas from the various locations and creating a table which summarizes the information for you.

## Requirements

Cost and Features

| Solution | Time | Cost | Feature | Difference |
|---|---|---|---|---|
| Manual calculations | Available since very long time | Time consumption | Can be done using only pen-paper | Our system is automatic, time efficient & accurate |
| Spreadsheet softwares (MS Excel, Libre Calc) | Available since long time | Paid/Free | Pleathora of features | For specific task, they are only semi-automatic while our system is automatic |
| Web-applications (doodu.io, data analysis) | Fairly latest | Paid/Free | High availability (on every platform) | Takes more time then our system. Not suitable for large data |
| Various template excel files | Came out after spreadsheet softwares | Mostly free | Pre-formatted, Automatically updates with data | Only pre-defined amount of data can be analyzed |
| Data analysis libraries for python (like pandas) | Since last decade | Free | Large number of analysis tools | Requires more learning to use, while our system requires almost no learning |

## Detail requirements

### High Level Requirements

| ID | Description | Category | Status |
|---|---|---|---|
| HLR01 | Data validation | Technical | Implemented |
| HLR02 | Working with excel files | Requirement | Implemented |
| HLR03 | Data analysis | Process | Implemented |
| HLR04 | Visualization of data | Requirement | Implemented |
| HLR05 | User interface | Requirement | Implemented |

**Low level Requirements**

| ID | Description | HLR ID | Status |
|---|---|---|---|
| LLR01 | Marks should only be number(no alphabet) | HLR01 | Implemented |
| LLR02 | A row in input sheet must contain data of all fields/columns | HLR01 | Implemented |
| LLR03 | Reading data from the excel sheet | HLR02 | Implemented |
| LLR04 | Writing data into the excel sheet | HLR02 | Implemented |
| LLR05 | Finding average for overall & each courses | HLR03 | Implemented |
| LLR06 | Finding Total & percentage for overall program | HLR03 | Implemented |
| LLR07 | Relative grading of the students based on performance (overall) | HLR03 | Implemented |
| LLR08 | Defining criteria for pass/fail | HLR03 | Implemented |
| LLR09 | Identifying & highlighting students with scope of improvements (overall & each courses) | HLR03 | Implemented |
| LLR10 | Displaying top performers for overall program & each courses | HLR03 | Implemented |
| LLR11 | Generating the bell curve based on data analysis (overall & each) | HLR04 | Implemented |
| LLR12 | Generating the pie chart based on grades (overall & each) | HLR04 | Implemented |
| LLR13 | Generating the bar chart based on percentage (overall) | HLR04 | Implemented |
| LLR14 | Options to search for info of a particular employee | HLR05 | Implemented |
| LLR15 | Functions to display user interface requirements | HLR05 | Implemented |

**4 W's & 1 H**

WHO?

- Any faculty member who wants to generate the performance reports of his/her students.
- Any faculty who wishes to analyse and visualize the performance of the students in an organized manner.
- Any member of the faculty who would like to generate an accurate grade for all students in any subject.

  - WHAT?

- A utility that automates the analysis and visualization of the marks/results obtained by students and generate the respective grade for different subjects , statistics and performance reports.

  - WHEN?

- Whenever a faculty/teacher would like to generate performance reports, grades and statistics of the students' performance automatically without any stress and wastage of time
- When manual generation of the report, feedback and grades becomes tedious

  - WHERE?
- This utility is computer based and is compatible with python environment and can be accessed anywhere through your laptop/desktop.

  - HOW?
- Reflecting on how tiresome and time consuming it can be to manually analyse all the students' data, generate grades and reports and how we can provide a platform that can automate this task and help our faculty members.

  - **SWOT analysis**

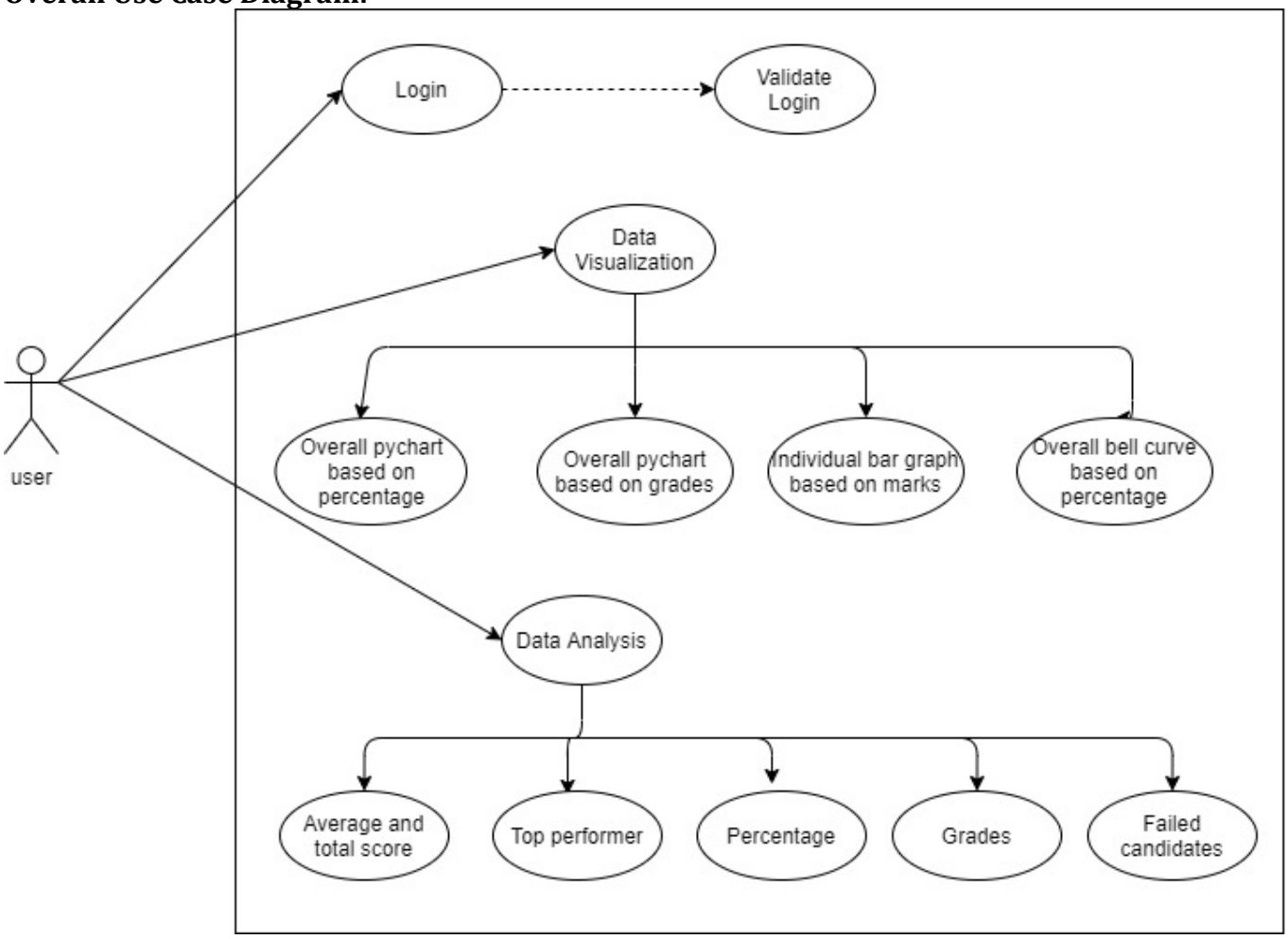
L&T Technology Services


| Strengths | Weaknesses | Opportunities | Threats |
|---|---|---|---|
| 1. Data analysis is automated | No Graphical user interface | Can include Graphical user interface | Not very visually pleasing since there is no GUI |
| 2. Data Visualization is automated | Processes limited number of excel sheets | Can be scaled to multiple sheets/subjects | Generates reports for fixed number of subjects |
| 3. Very time efficient | Does not provide individual reports for each student | Can generate report for each student | --------- |
| 4. Reduces faculty work load | ---------- | ---------- | --------- |
| 5. User Friendly | ---------- | ---------- | --------- |
| 6. Provides report for each subject | ---------- | ---------- | --------- |

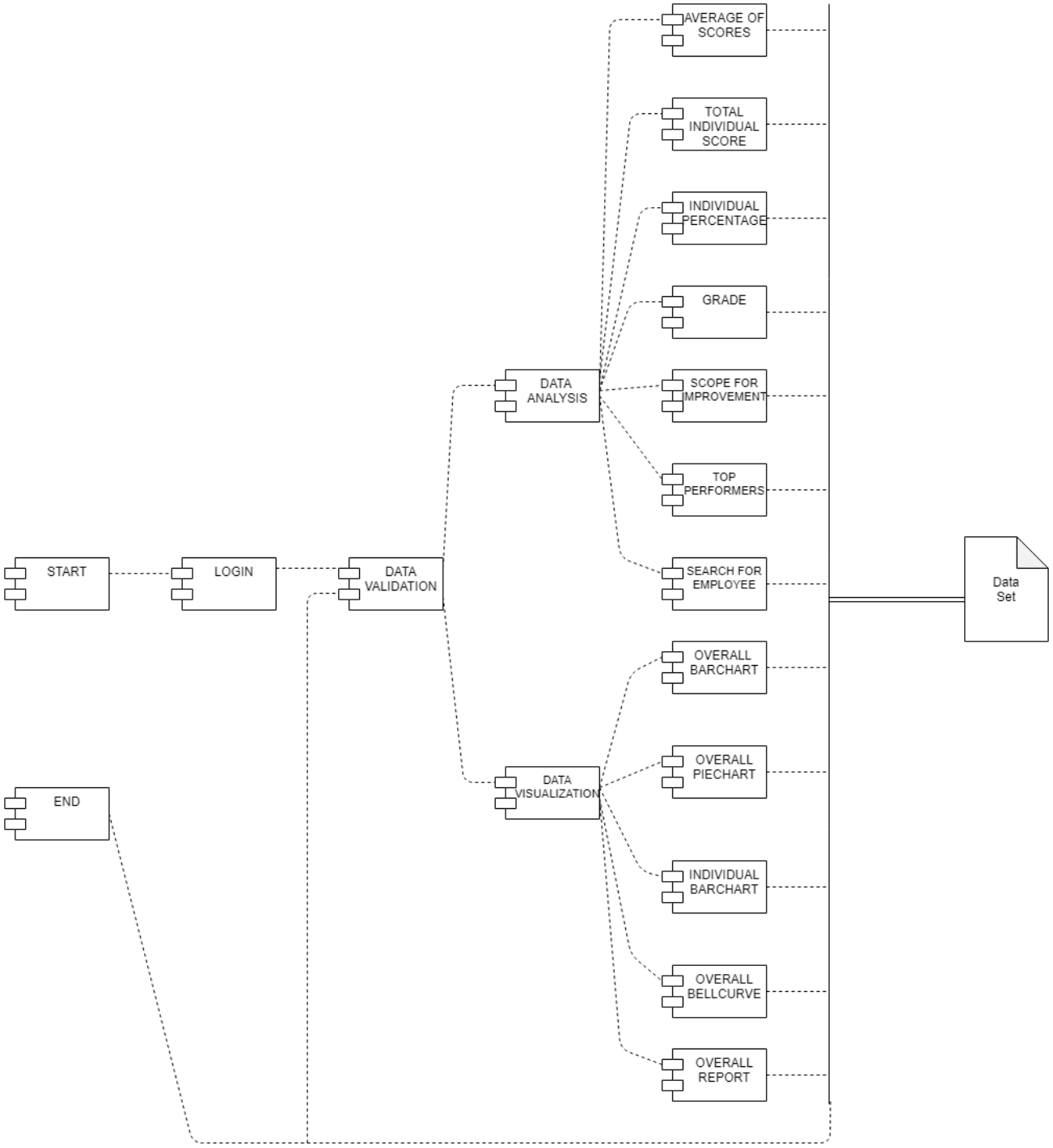
## ARCHITECTURE:

### HIGH LEVEL UML DIAGRAMS:

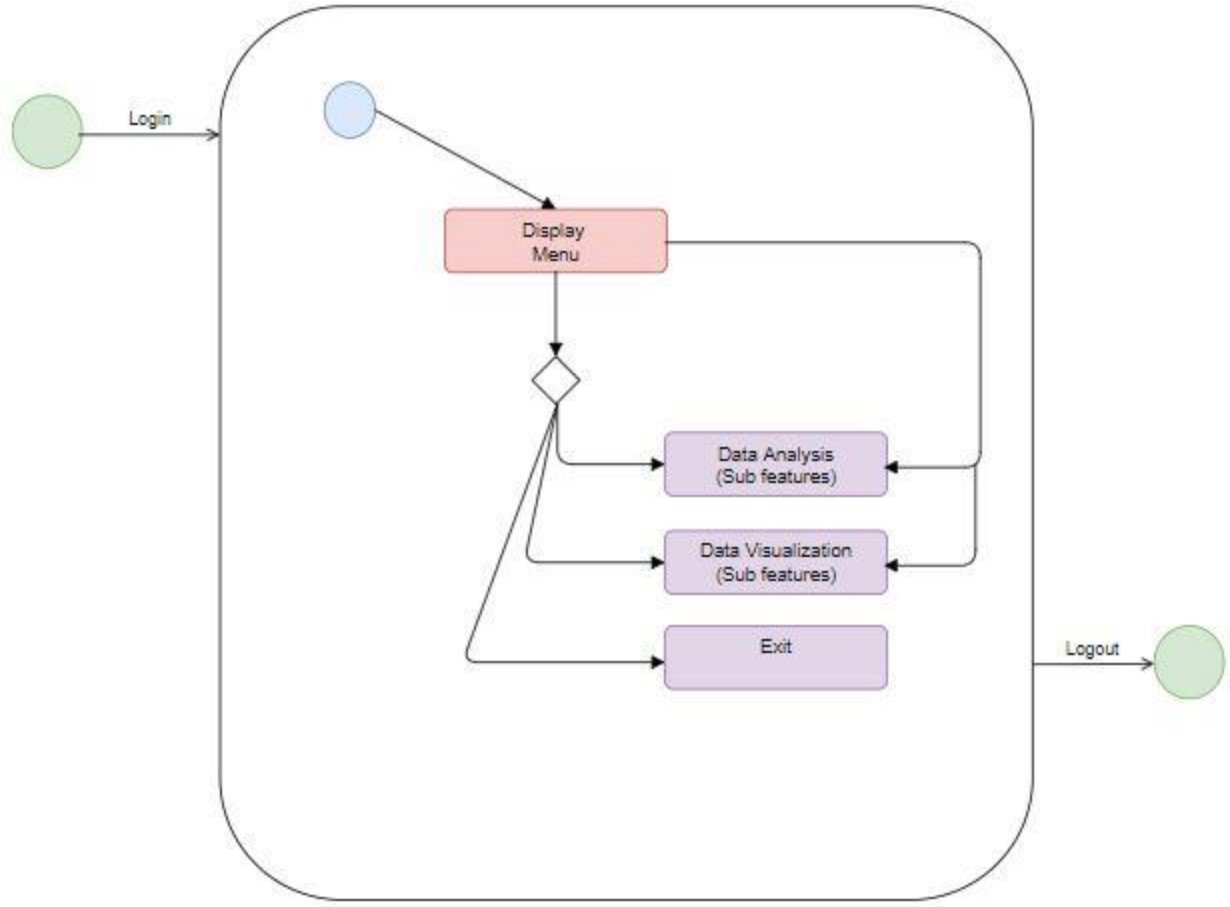
### Overall Use Case Diagram:



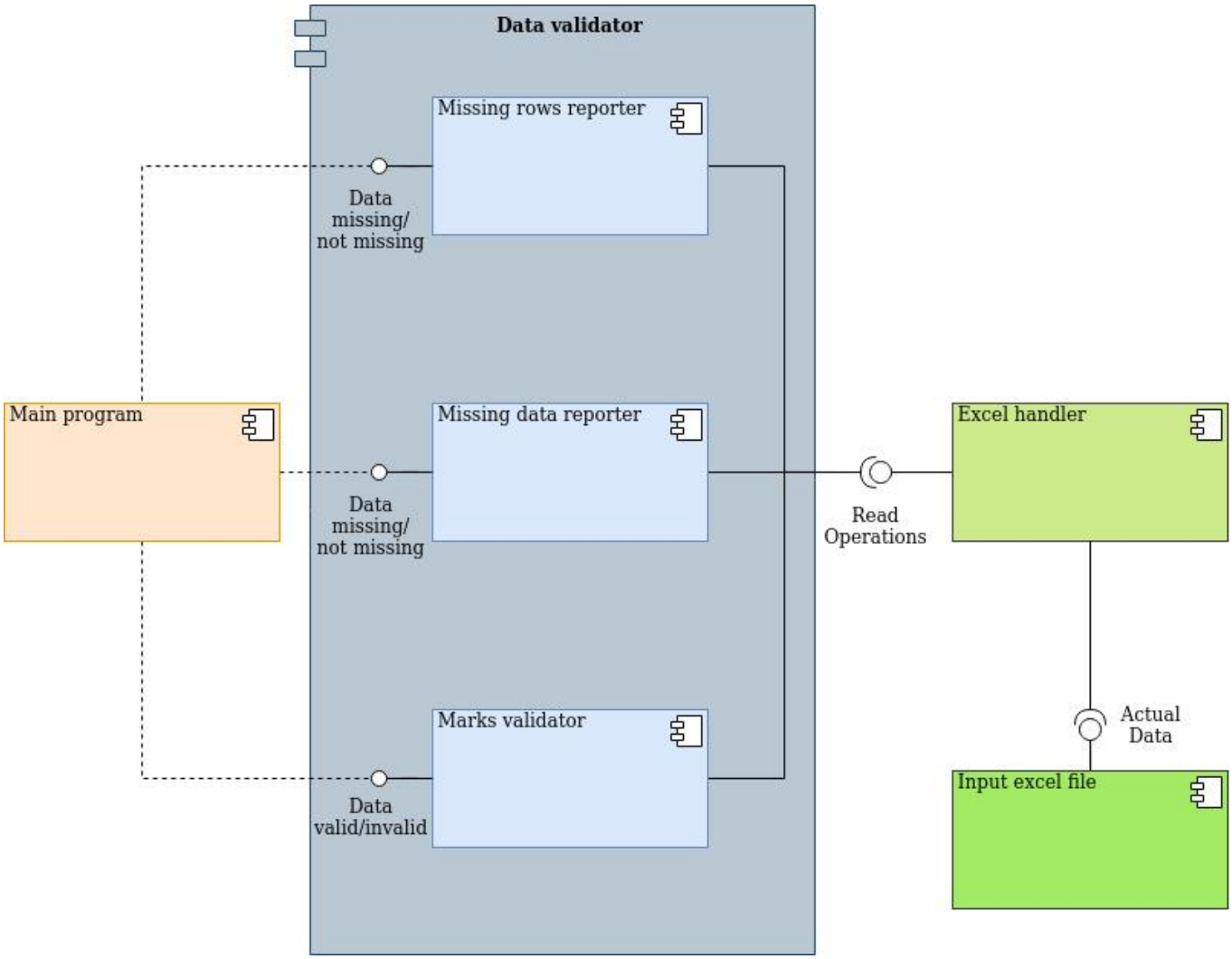
## Overall Component Diagram:



**Overall State Diagram:**



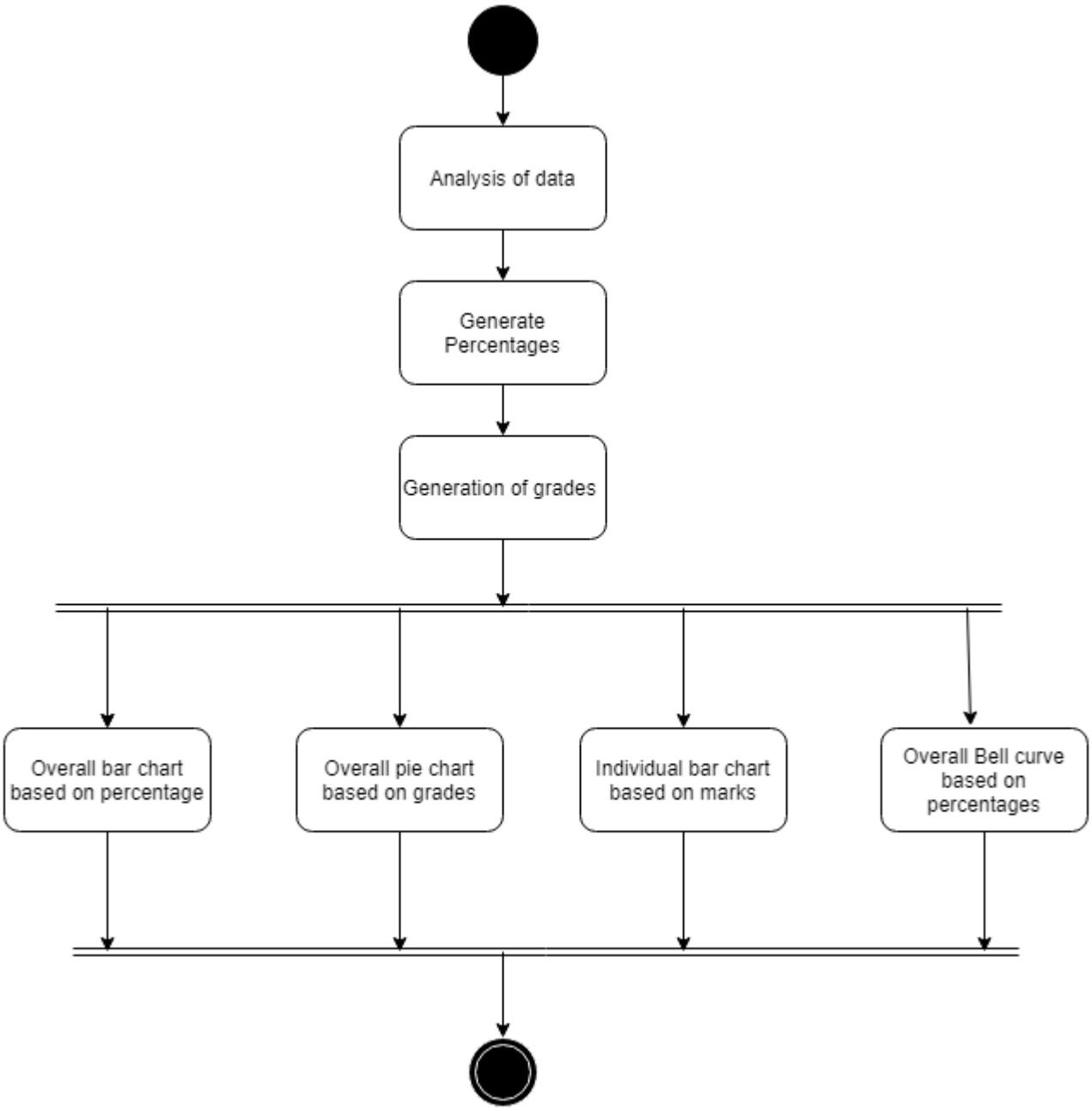
**Component Diagram (Data Validator) :**



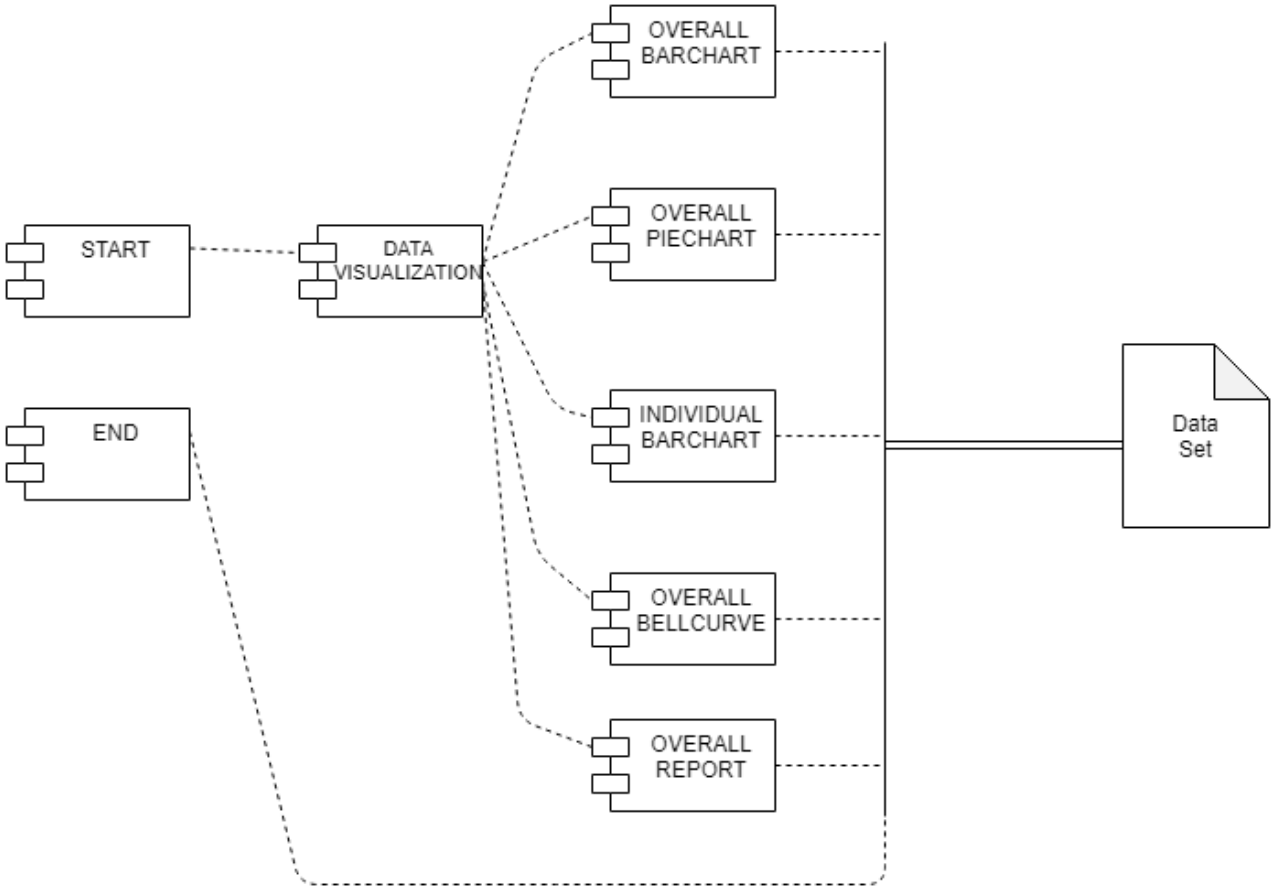
**Low Level UML Diagrams:**

## Data Visualization:

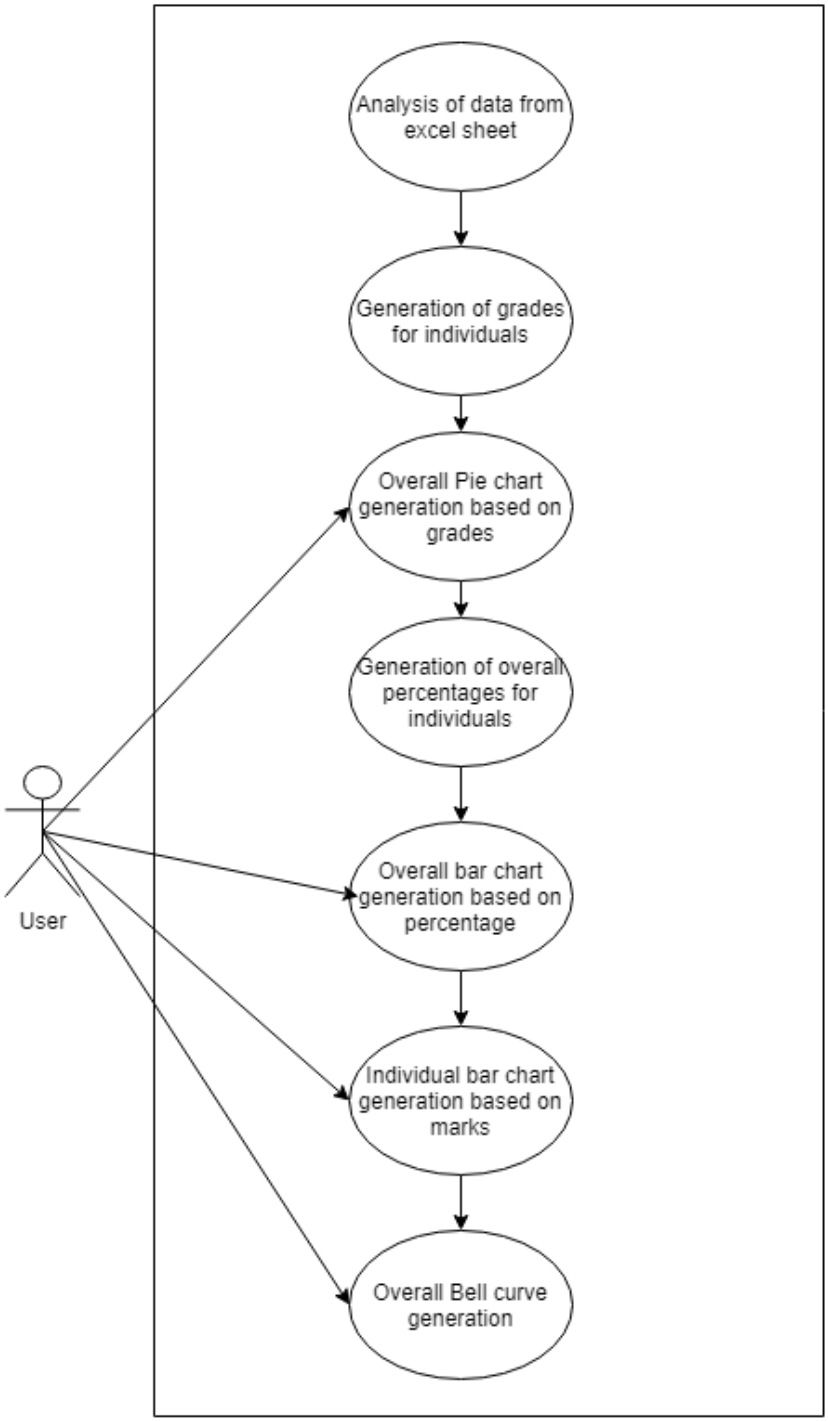
**Action Case Diagram:**



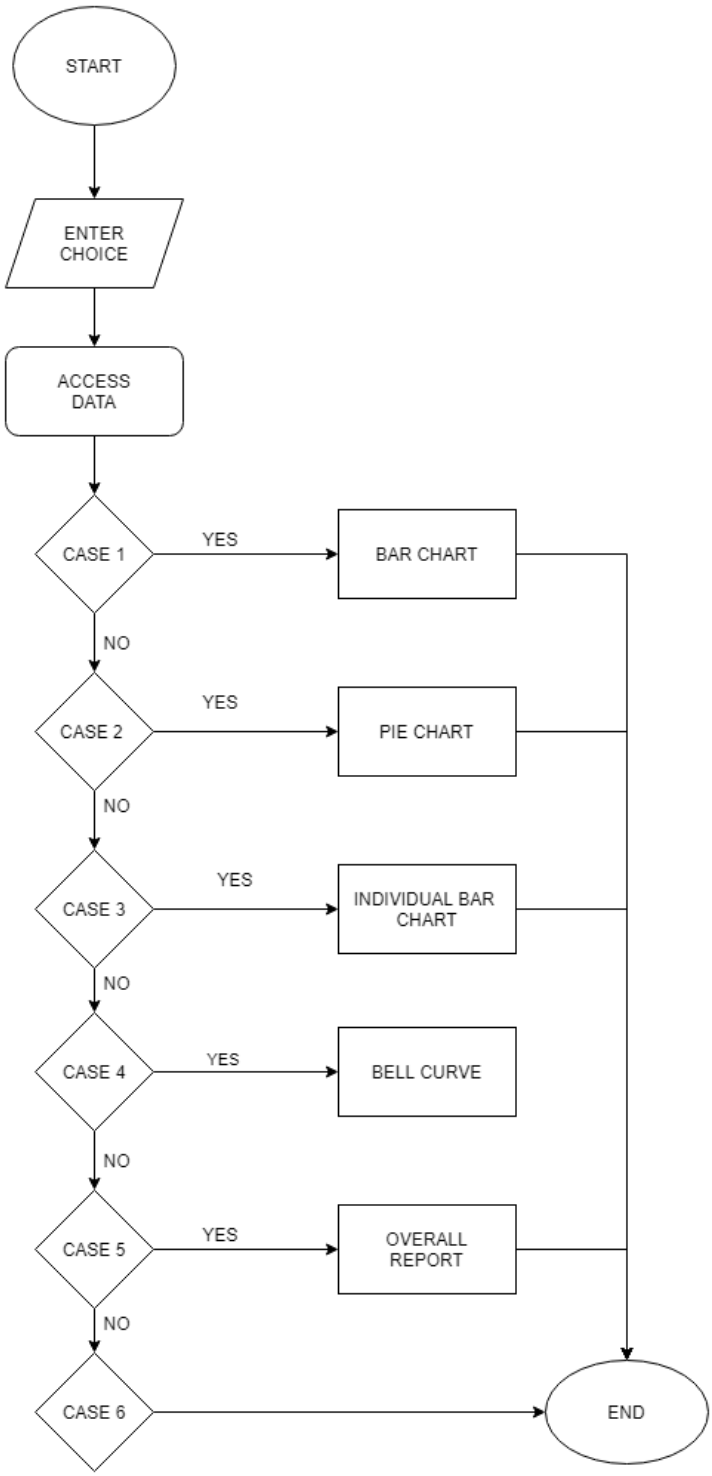
**Component Diagram:**



**Use Case Diagram:**

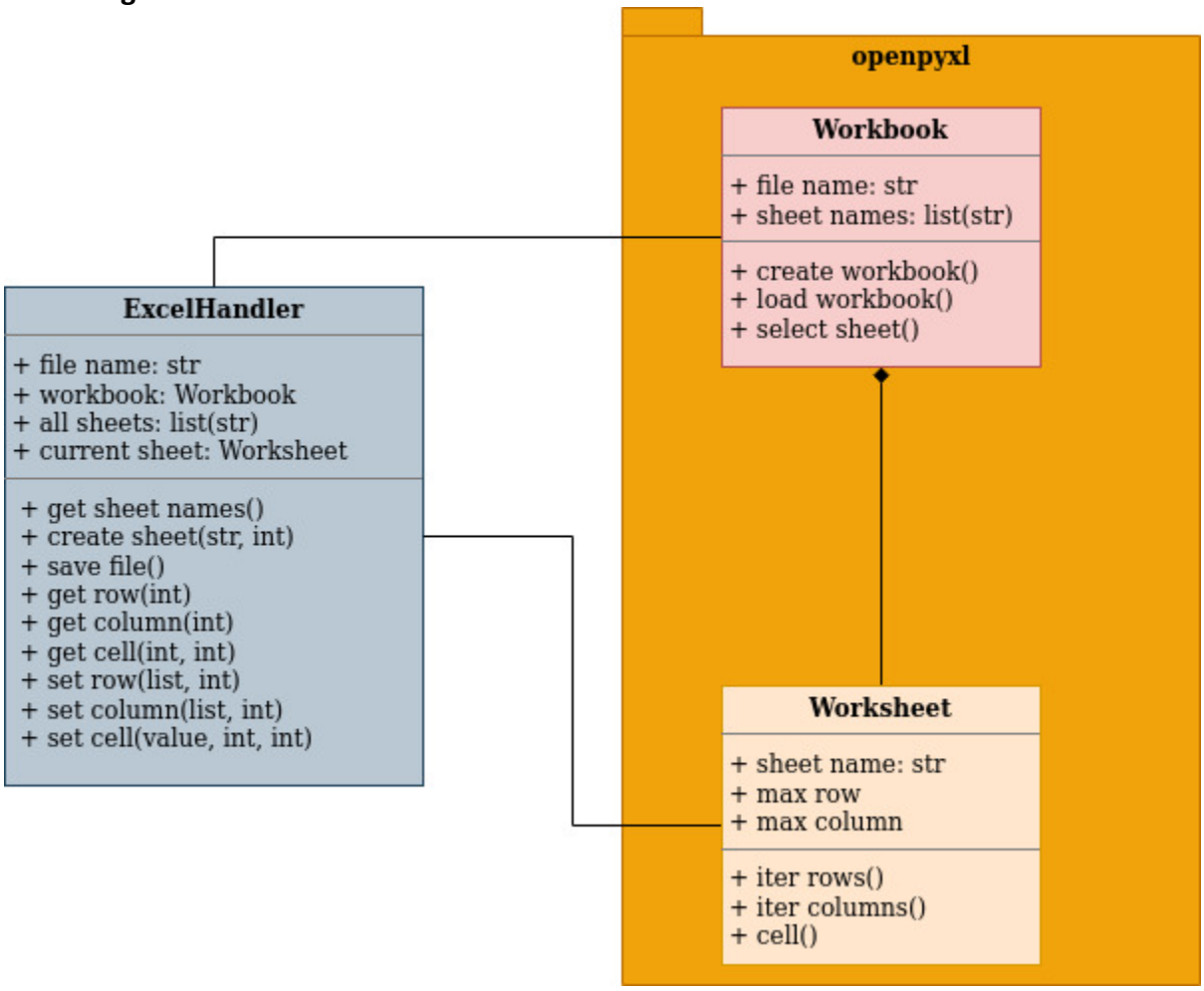


**Flow Chart:**

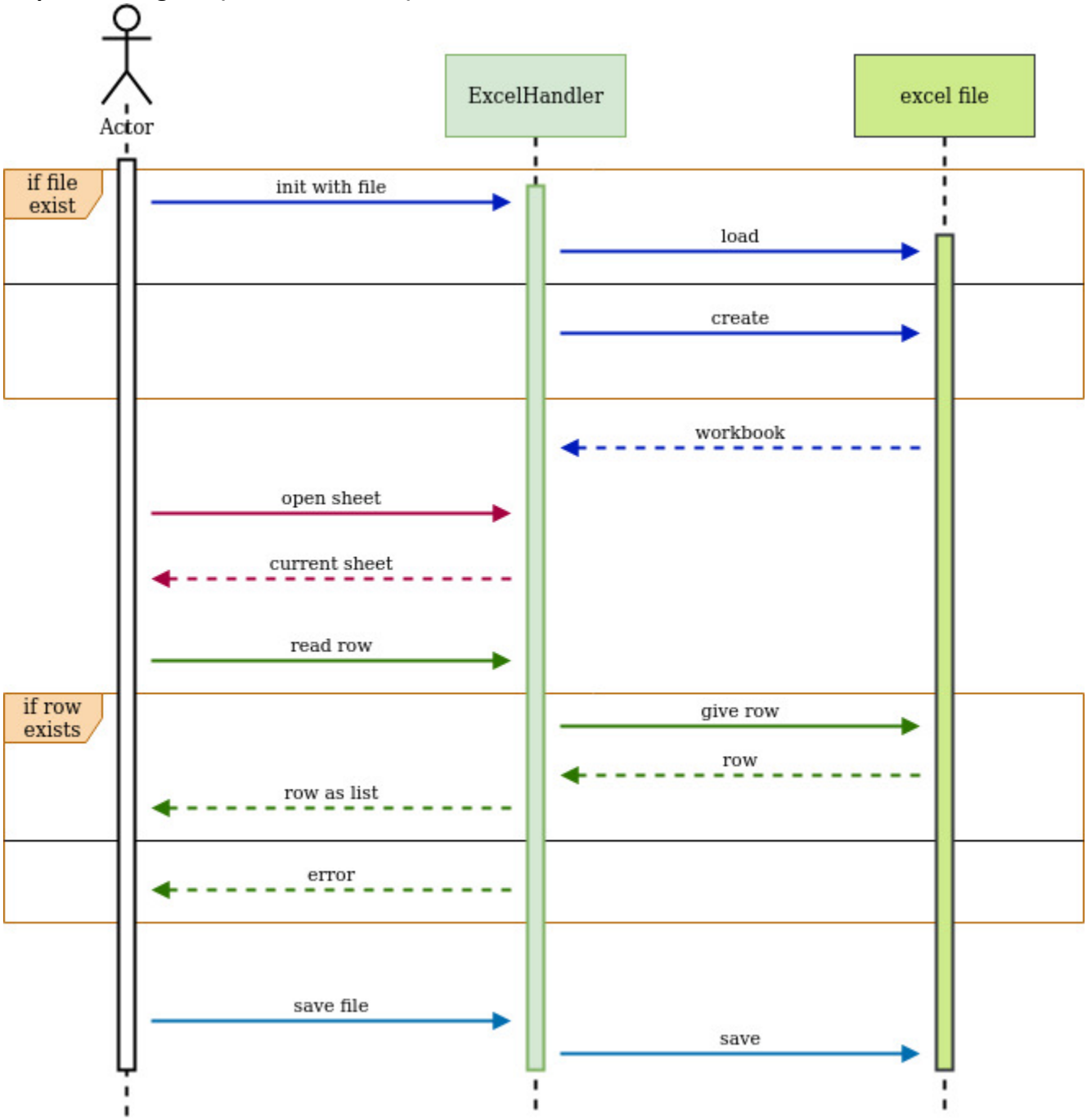

START
ENTER CHOICE
ACCESS DATA
CASE 1
YES
BAR CHART
NO
CASE 2
YES
PIE CHART
NO
CASE 3
YES
INDIVIDUAL BAR CHART
NO
CASE 4
YES
BELL CURVE
NO
CASE 5
YES
OVERALL REPORT
NO
CASE 6
END

## Data Validation:

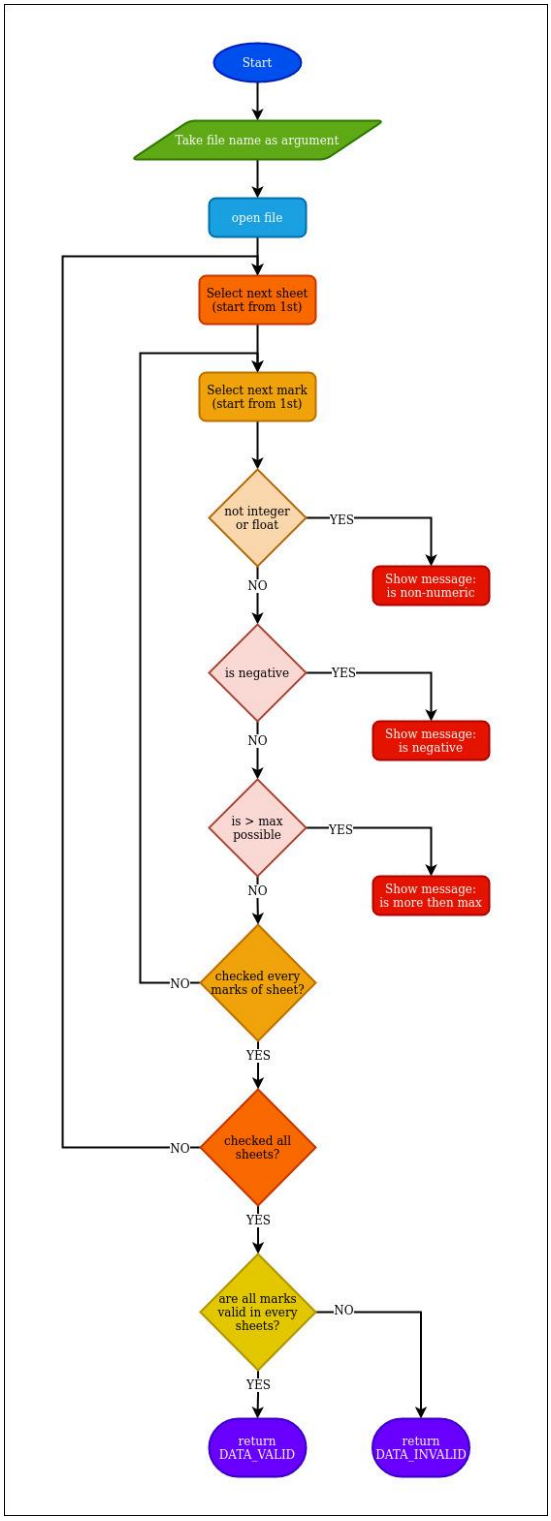
### Class Diagram:



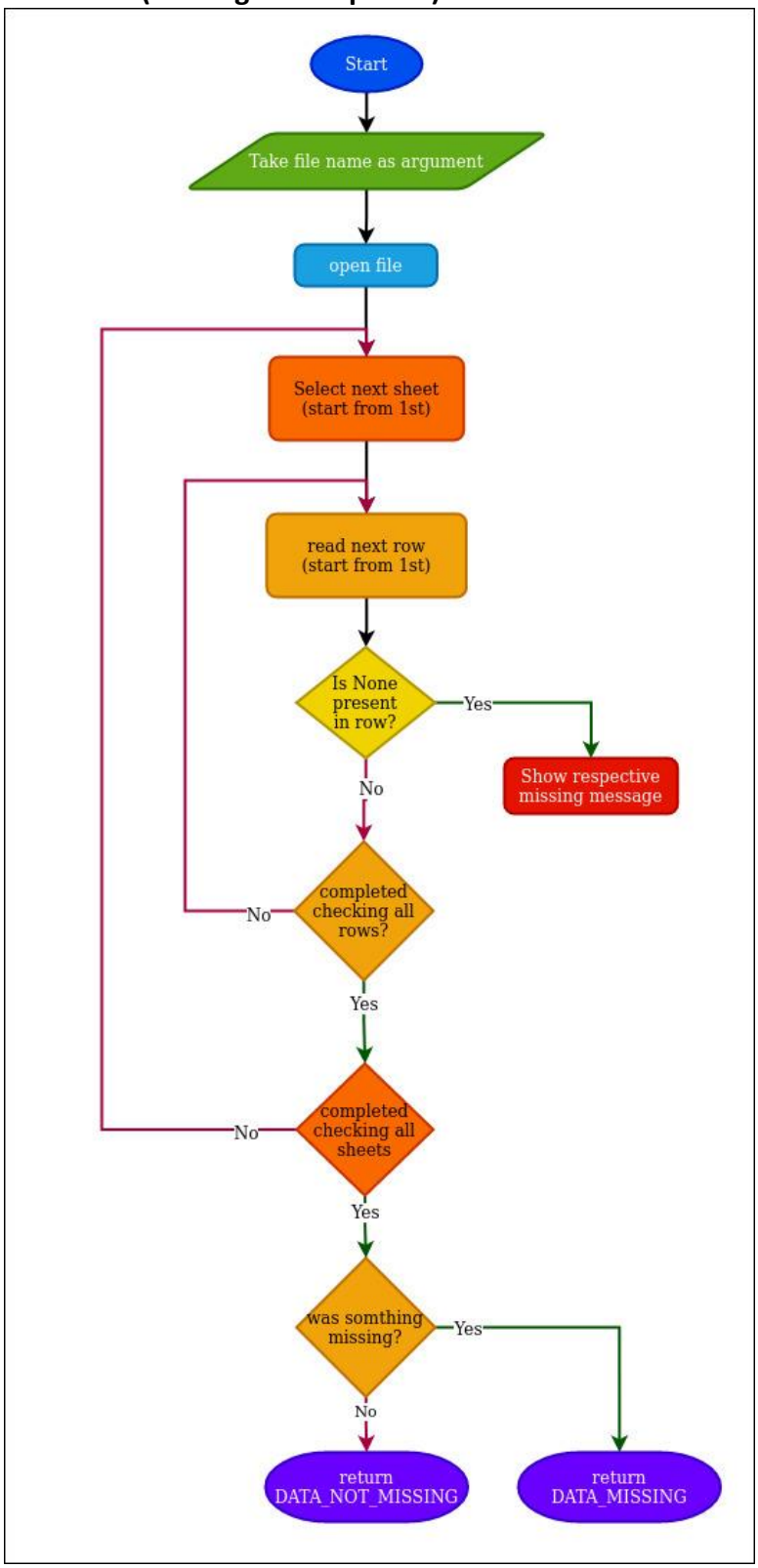
**Sequence Diagram (Excel read/write) :**



## Flow Chart (Marks validator) :

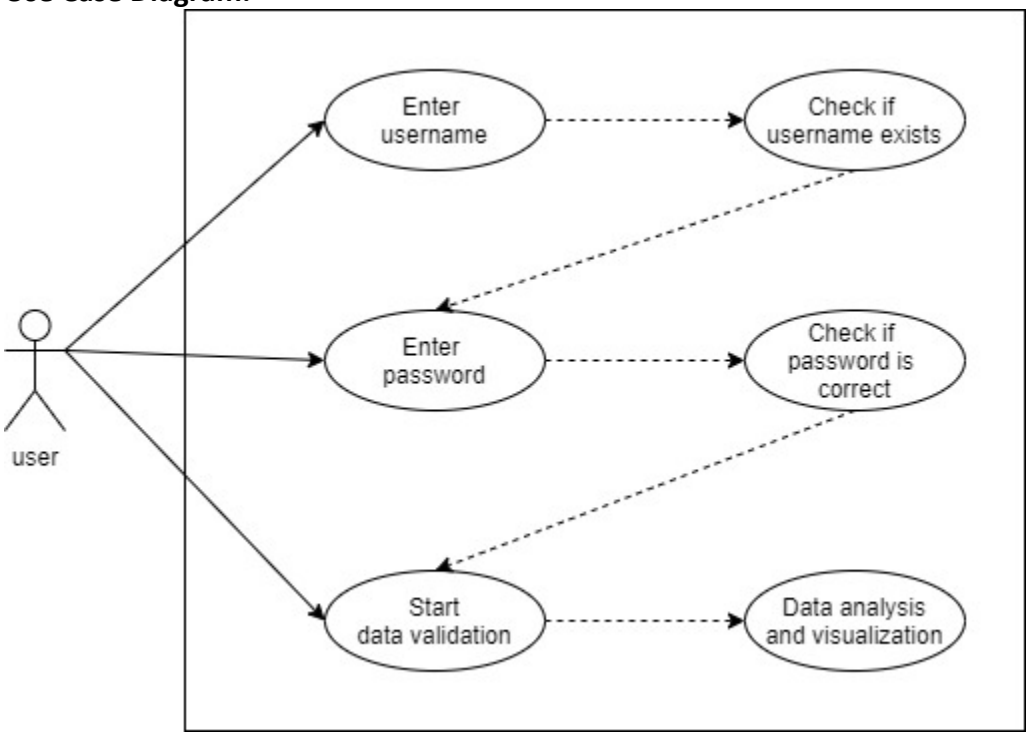


**Flowchart (Missing data reporter) :**

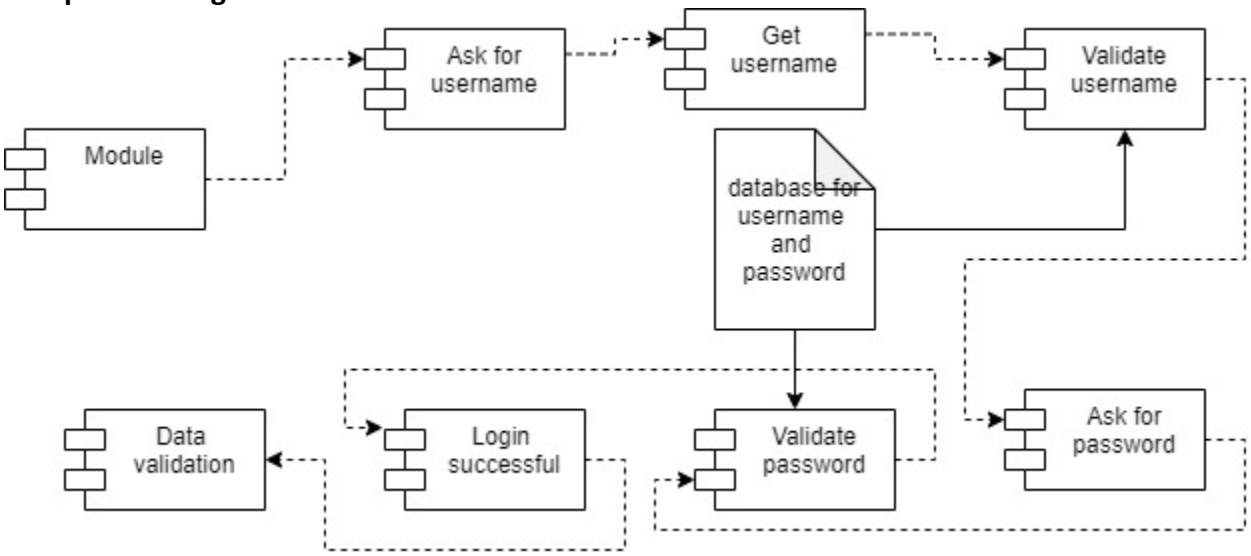


## Login Implementation:

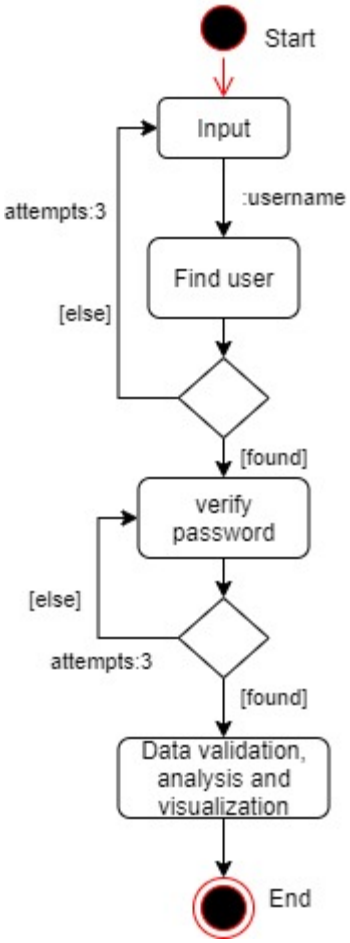
**Use Case Diagram:**



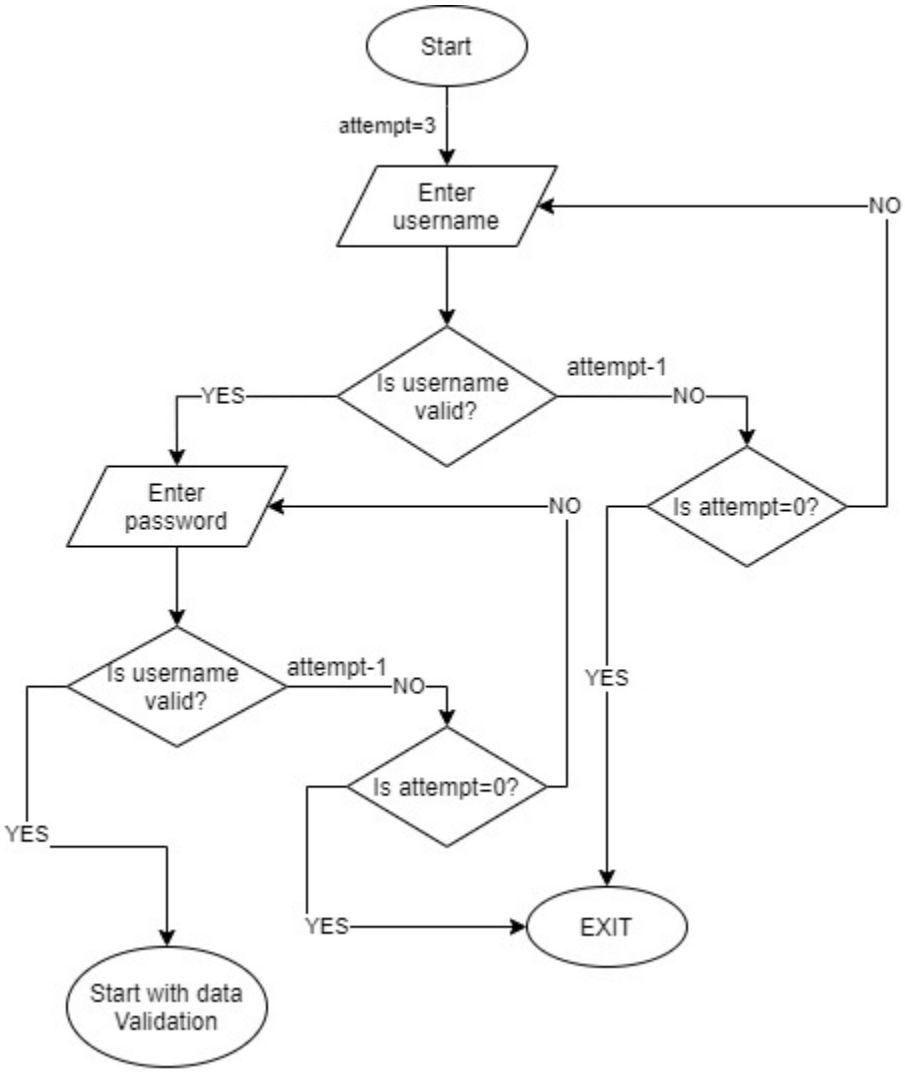
**Component Diagram:**



**Action Case Diagram:**

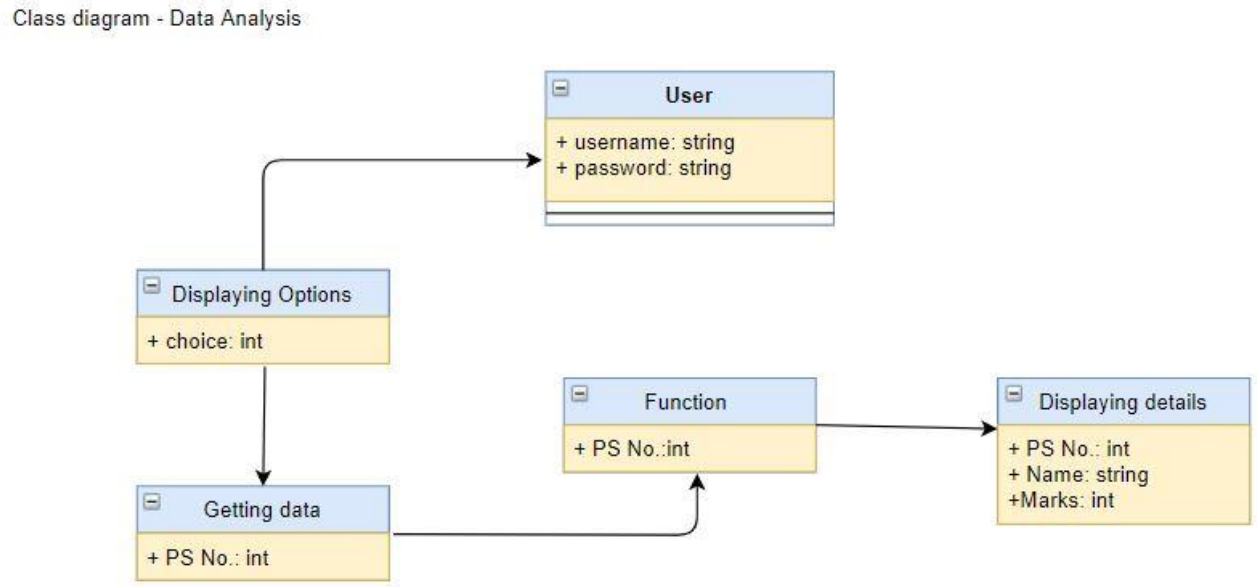


**Flow Chart:**

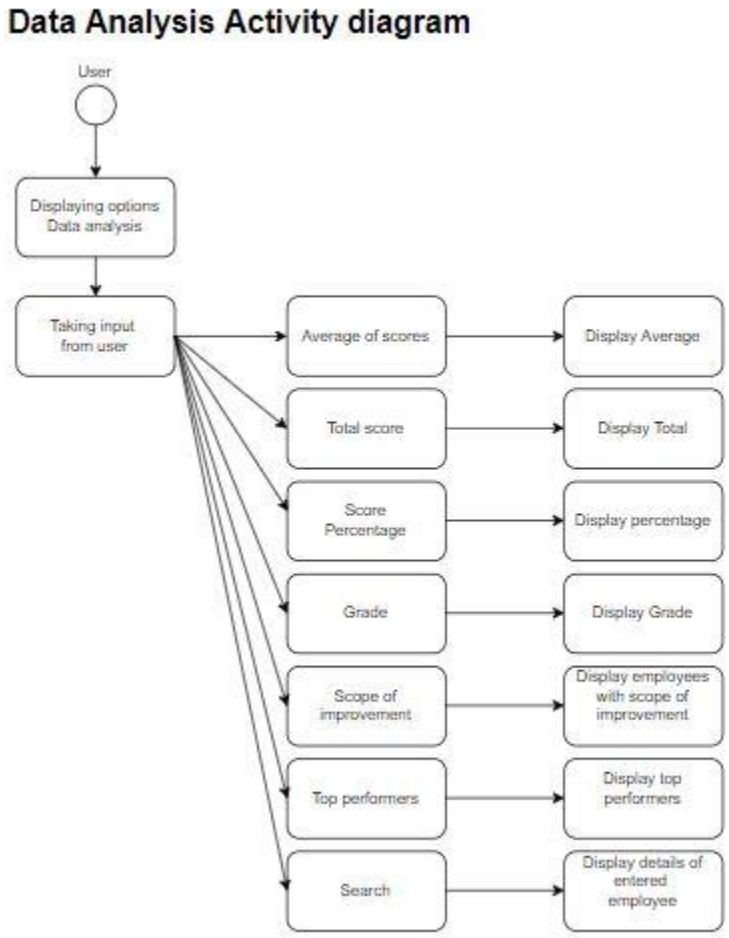

Start
attempt=3
Enter username
NO
Is username valid?
attempt-1
YES
NO
Enter password
NO
Is attempt=0?
Is username valid?
attempt-1
NO
YES
Is attempt=0?
YES
YES
EXIT
Start with data Validation

## Data Analysis:

### Class Diagram:



**Activity Diagram:**



## TEST PLAN:

### High level test plan

| Test ID | Description | Exp I/P | Exp O/P | Actual Out | Result | Type Of Test |
|---|---|---|---|---|---|---|
| HLT01 | Check the functionality of data_validator module | different excel files | respective identifications | identified & reported properly | PASS | Functionality |
| HLT02 | Check the functionality of data_visualization module | different excel files | respective identifications | identified & reported properly | PASS | Functionality |
| HLT03 | Check the functionality of data analysis module | Lists of data | Output of analysis functions | Proper output | PASS | Functionality |

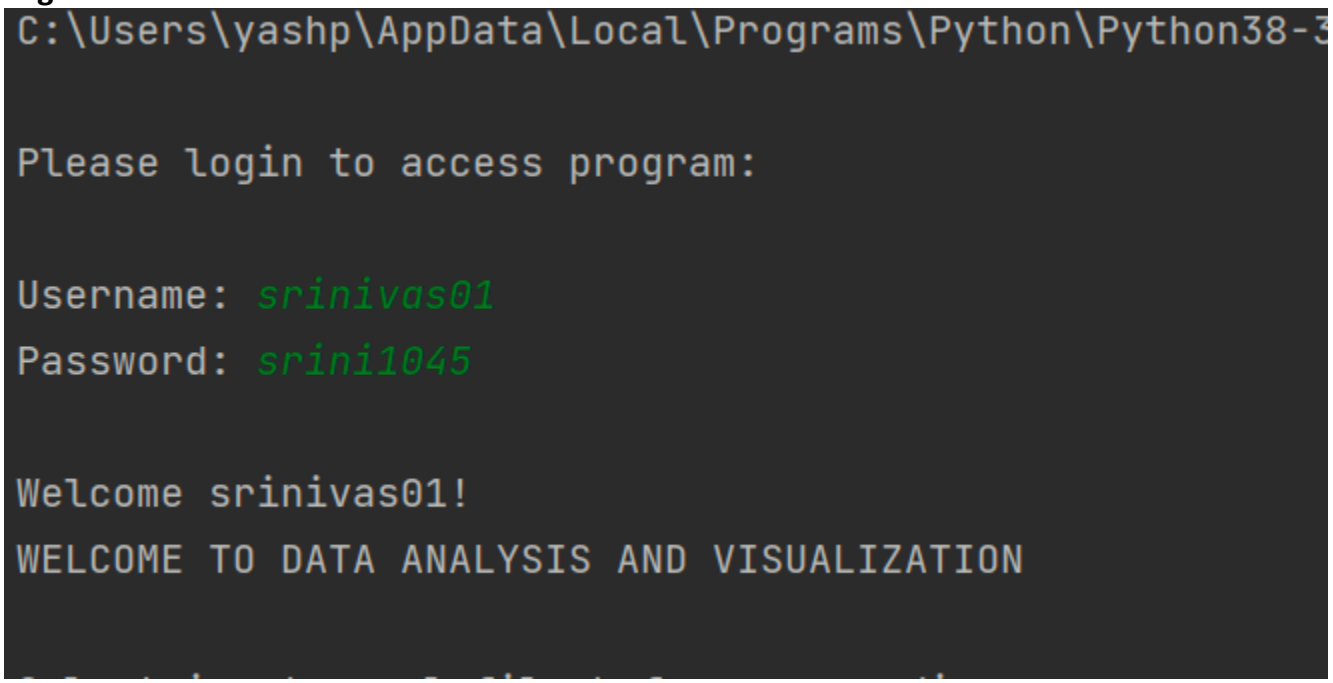
### Low level test plan

| Test ID | HLT ID | Description | Exp IN | Exp OUT | Actual Out | Result | Type Of Test |
|---|---|---|---|---|---|---|---|
| LLT01 | HLT01 | missing_row_reporter() --> Able to check if all rows are present | Test_Input_File1.xlsx | DATA_NOT_MISSING | DATA_NOT_MISSING | PASS | Requirement |
| LLT02 | HLT01 | missing_row_reporter() --> Able to report missing rows | Test_Input_File2.xlsx | DATA_MISSING | DATA_MISSING | PASS | Requirement |
| LLT03 | HLT01 | missing_data_reporter() --> Able to check if all data is present | Test_Input_File1.xlsx | DATA_NOT_MISSING | DATA_NOT_MISSING | PASS | Requirement |
| LLT04 | HLT01 | missing_data_reporter() --> Able to report missing SR No | Test_Input_File3.xlsx | DATA_MISSING | DATA_MISSING | PASS | Requirement |
| LLT05 | HLT01 | missing_data_reporter() --> Able to report missing PS No | Test_Input_File3.xlsx | DATA_MISSING | DATA_MISSING | PASS | Requirement |
| LLT06 | HLT01 | missing_data_reporter() --> Able to report missing Names | Test_Input_File3.xlsx | DATA_MISSING | DATA_MISSING | PASS | Requirement |
| LLT07 | HLT01 | missing_data_reporter() --> Able to report missing Marks | Test_Input_File3.xlsx | DATA_MISSING | DATA_MISSING | PASS | Requirement |
| LLT08 | HLT01 | marks_validator() --> Able to check if all marks are valid | Test_Input_File1.xlsx | DATA_VALID | DATA_VALID | PASS | Requirement |
| LLT09 | HLT01 | marks_validator() --> Able to check if marks are non-numeric | Test_Input_File4.xlsx | DATA_INVALID | DATA_INVALID | PASS | Requirement |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| LLT10 | HLT01 | marks_validator() --> Able to check if marks are negative number | Test_Input_File4.xlsx | DATA_INVALID | DATA_INVALID | PASS | Requirement |
| LLT11 | HLT01 | marks_validator() --> Able to check if marks are more then defined maximum | Test_Input_File4.xlsx | DATA_INVALID | DATA_INVALID | PASS | Requirement |
| LLT12 | HLT02 | Checks if the BarChart.xlsl file is saved in the folder | marks for 3 subjects and corresponding ps Number | PASS | PASS | PASS | Requirement |
| LLT13 | HLT02 | Checks if the PieChart.xlsl file is saved in the folder | marks for 3 subjects and corresponding ps Number | PASS | PASS | PASS | Requirement |
| LLT14 | HLT02 | Checks if the Bell_curve.xlsl file is saved in the folder | marks for 3 subjects and corresponding ps Number | PASS | PASS | PASS | Requirement |
| LLT15 | HLT02 | Checks if the Overall_Report.xlsl file is saved in the folder | marks for 3 subjects and corresponding ps Number | PASS | PASS | PASS | Requirement |
| LLT16 | HLT03 | Checking functionality of average function | list of marks | Average of marks | PASS | PASS | PASS |
| LLT17 | HLT03 | Checking functionality of percentage function | list of marks | Percentage of marks | PASS | PASS | PASS |
| LLT18 | HLT03 | Checking functionality of grades function | list of marks | Grades of learners | PASS | PASS | PASS |
| LLT19 | HLT03 | Checking functionality of fail_canidates function | list of marks | list of failed learners | PASS | PASS | PASS |
| LLT20 | HLT03 | Checking functionality of top_performer function | list of marks | list of top learners | PASS | PASS | PASS |

## Implementation Screenshots

### Login with credentials



**Data Validation**

```
Select input excel file before proceeding
0 :   PieChart.xlsx
1 :  BarChart.xlsx
2 :  Bell_curve.xlsx
3 :  IndividualBarChart.xlsx
4 :  Mark_Sheet.xlsx
5 :  Overall_Report.xlsx
Enter selection number: 4

File: Mark_Sheet.xlsx is selected

Every sheets have equal number of rows
All the data is present
Every marks are valid

Input file is complete and has valid data
```

**Main menu**

```
MAIN MENU
1. Data Analysis
2. Data Visualization
3. Exit
Enter selection number: 1
```

## Data Analysis

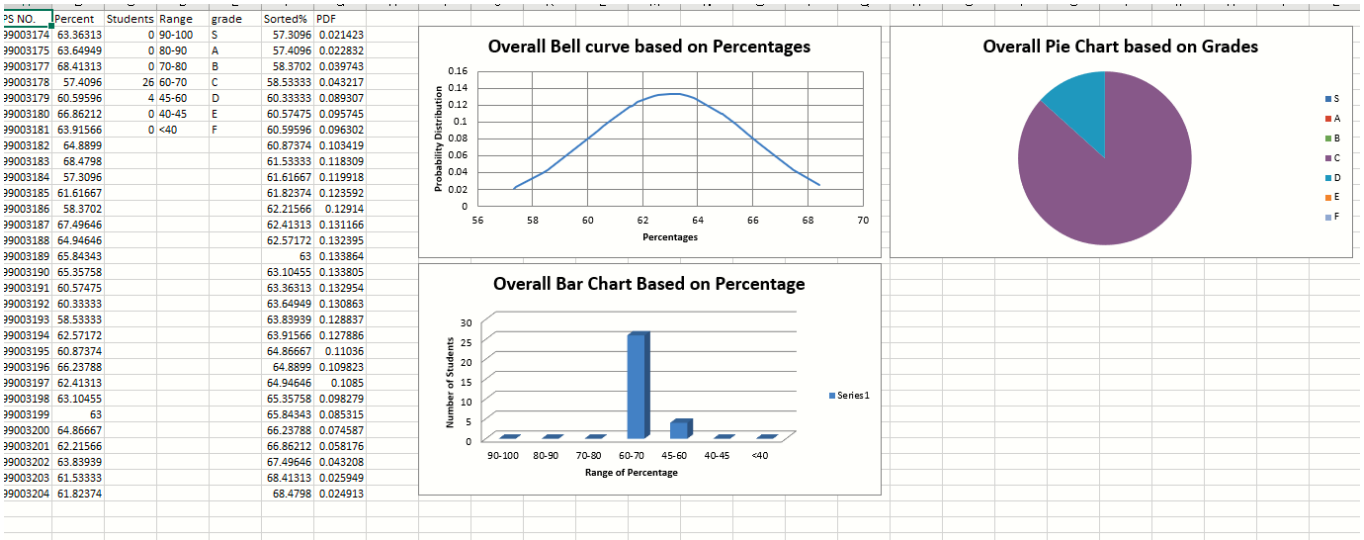
```
Data analysis options
1. Average of scores
2. Total score of employee
3. Score percentage of employee
4. Grade based on performance
5. List of employee with scope of improvement
6. Top performers
7. Search for employee
8. Save report to excel file
9. Exit
Enter selection number: 1

Courses list:
1. Applied SDLC
2. Advanced Python
3. MBSE
Enter selection number: 2
Average marks in Advanced Python  :66.36666666666666
```

## Data Visualization

```
Enter selection number: 2

Data visualization options
1. Overall Bar chart based on percentage
2. Overall Pie Chart based on grade
3. Individual Bar chart based on marks
4. Overall Bell Curve
5. Overall performance report
6. Exit
Enter selection number: 5
Please check the Overall_Report.xlsl file
```



| PS NO. | Percent | Students | Range | grade | Sorted% | PDF |
|---|---|---|---|---|---|---|
| 99003174 | 63.36313 | 0 | 90-100 | S | 57.3096 | 0.021423 |
| 99003175 | 63.64949 | 0 | 80-90 | A | 57.4096 | 0.022832 |
| 99003177 | 68.41313 | 0 | 70-80 | B | 58.3702 | 0.039743 |
| 99003178 | 57.4096 | 26 | 60-70 | C | 58.53333 | 0.043217 |
| 99003179 | 60.59596 | 4 | 45-60 | D | 60.33333 | 0.089307 |
| 99003180 | 66.86212 | 0 | 40-45 | E | 60.57475 | 0.095745 |
| 99003181 | 63.91566 | 0 | <40 | F | 60.59596 | 0.096302 |
| 99003182 | 64.8899 | | | | 60.87374 | 0.103419 |
| 99003183 | 68.4798 | | | | 61.53333 | 0.118309 |
| 99003184 | 57.3096 | | | | 61.61667 | 0.119918 |
| 99003185 | 61.61667 | | | | 61.82374 | 0.123592 |
| 99003186 | 58.3702 | | | | 62.21566 | 0.12914 |
| 99003187 | 67.49646 | | | | 62.41313 | 0.131166 |
| 99003188 | 64.94646 | | | | 62.57172 | 0.132395 |
| 99003189 | 65.84343 | | | | 63 | 0.133864 |
| 99003190 | 65.35758 | | | | 63.10455 | 0.133805 |
| 99003191 | 60.57475 | | | | 63.36313 | 0.132954 |
| 99003192 | 60.33333 | | | | 63.64949 | 0.130863 |
| 99003193 | 58.53333 | | | | 63.83939 | 0.128837 |
| 99003194 | 62.57172 | | | | 63.91566 | 0.127886 |
| 99003195 | 60.87374 | | | | 64.86667 | 0.11036 |
| 99003196 | 66.23788 | | | | 64.8899 | 0.109823 |
| 99003197 | 62.41313 | | | | 64.94646 | 0.1085 |
| 99003198 | 63.10455 | | | | 65.35758 | 0.098279 |
| 99003199 | 63 | | | | 65.84343 | 0.085315 |
| 99003200 | 64.86667 | | | | 66.23788 | 0.074587 |
| 99003201 | 62.21566 | | | | 66.86212 | 0.058176 |
| 99003202 | 63.83939 | | | | 67.49646 | 0.043208 |
| 99003203 | 61.53333 | | | | 68.41313 | 0.025949 |
| 99003204 | 61.82374 | | | | 68.4798 | 0.024913 |

# Pylint score

## Scrum Standup Call

16/7/21

1. What have we done
2. What are we Doing
3. Challeges faced

@vishal99004952

1. Completed identifying features part of requirements, figured out the HLR and LLRs for data validation , completed the cost and ageing
2. Update HLR and LLRs to github, Work on the UML diagrams and Make some test plans
3. Finding good resources for UML

@pavanmulimani

1. Completed the Introduction part of requirement. Identified features of data analysis
2. Figure out HLR and LLR for Mind reader and draw UML for the same
3. classifying HLR and LLR

@arunmaurya070

1. Identified threats in SWOT analysis, Discussed 4W1H of Application, Did cost and feature analysis
2. Figure out the approach for login and identify the HLR and LLR
3. Figuring out relevant requiremnent and UML diagram

@Sonia110

1. Identified the features and functions for data visualization and Discussed SWOT and 4W1H and updated
2. Identify HLR and LLR, design UML Diagrams and test plan
3. finding good resources to UML and figuring out how to design them

5/7/21

@pavanmulimani

1. UML diagrams and coding of Data analysis

2. Test plan and code integration
3. Testing

@arunmaurya070

1. Finished all the UML diagrams and started coding for the login part.
2. Finish the test plan and coding.
3. Unit testing
4.

@Sonia110

1. Completed the UML diagrams
2. Finish coding the data visualization
3. Testing

@vishal99004952

1. Completed UML diagrams
2. Finish Test plan and implementation of data validation
3. Testing