**PROJECT TITLE**

**Basic Student Performance Analysis**

**ORGANIZATION/ DEPARTMENT NAME & ADDRESS**

**WAYSPIRE PVT.LTD**

**SUBMITTED BY**

**Ms. Janhavi Sachin Khare**

**M.Sc. (Applied Statistics)**

**PRN:24060641047**

**ACADEMIC YEAR**

**2024**

**-**

**2**

**6**

**Under the guidance of**

**Name of Project Guide / Mentor**

**Designation:**

**Email Id:**

**xxxxx@domain.com**

**Mobile: xxxxxxxxxx**



**Contents**

1. **Executive Summary** ………………………………………………………………………

………………………………………………………………………………………………

………………………………………………………………………………………………

1. **Study Background** ………………………………………………………………………..

………………………………………………………………………………………………

………………………………………………………………………………………………

1. **Literature Review (If any)** ………………………………………………………………..

………………………………………………………………………………………………

………………………………………………………………………………………………

1. **Aims & Objectives** ………………………………………………………………………..

………………………………………………………………………………………………

………………………………………………………………………………………………

1. **Methodology** ………………………………………………………………………………

………………………………………………………………………………………………

………………………………………………………………………………………………

1. **Results** ……………………………………………………………………………………..

………………………………………………………………………………………………

1. **Discussion and Conclusion** ……………………………………………………………….

………………………………………………………………………………………………

1. **Recommendation** ………………………………………………………………………….

………………………………………………………………………………………………

1. **Acknowledgement** ………………………………………………………………………...

………………………………………………………………………………………………

1. **References (including Bibliography and web references)** ……………………………...

………………………………………………………………………………………………

1. **Annexure – (Questionnaire & other details)** ……………………………………………

………………………………………………………………………………………………

1. **Executive Summary**

This project aimed to analyse student performance metrics using Python programming fundamentals. The analysis included calculating average scores, classifying students as Pass/Fail, and identifying top/low performers using the dataset. Key Python concepts like operators, loops, functions, and conditional logic were implemented to derive meaningful insights from the data. The project successfully demonstrated how basic programming techniques can be applied to educational datasets.

1. **Study Background**

Educational institutions require data-driven approaches to evaluate student performance. This project analysed exam scores (math, reading, writing) alongside demographic factors to understand performance patterns. The analysis serves as a foundation for more advanced educational data mining applications.

1. **Literature Review**

While no formal literature review was conducted, the project aligns with established educational research practices that use performance data to:

* Identify at-risk students
* Evaluate teaching effectiveness
* Understand demographic influences on academic achievement

1. **Aims & Objectives**

* Implement Python programming concepts for data analysis
* Calculate and analyse average performance scores
* Develop classification systems for student performance
* Identify exceptional cases (top/weak performers)
* Demonstrate core programming techniques

1. **Methodology**

* Data processing

Used Pandas to load and clean the dataset

Created new metrics (average\_score)

Formatted categorical data (race/ethnicity)

* Analysis Technique

Arithmetic operations for score calculations

String operations for data standardization

Conditional statements for Pass/Fail classification

Loop structures for iterative processing

Custom functions for reusable logic

* Key functions

def classify\_student(avg): # Pass/Fail classifier

def get\_topper(df): # Identifies highest scorer

def get\_weakest(df): # Identifies lowest scorer

1. **Results**

* **Overall performance metrics**

|  |  |
| --- | --- |
| Metric | Value |
| Total Students | 1000 |
| Average Maths Score | 66.09 |
| Average Reading Score | 69.17 |
| Average Writing Score | 68.05 |
| Overall Pass rate | 87.4% |

* Passing threshold: 50/100
* Topper: Female student with 100 average score
* Weakest: Female student with 9 average score
* **Top/Bottom 5 Performers**

**Top 5 Students**:

1. Female (Race: Group E) – Avg: **100.0**
2. Female (Race: Group D) – Avg: **99.67**
3. Female (Race: Group C) – Avg: **99.0**
4. Male (Race: Group E) – Avg: **97.67**
5. Female (Race: Group D) – Avg: **96.33**

**Bottom 5 Students**:

1. Female (Race: Group B) – Avg: **9.0**
2. Male (Race: Group E) – Avg: **23.33**
3. Female (Race: Group C) – Avg: **26.0**
4. Male (Race: Group B) – Avg: **30.0**
5. Male (Race: Group E) – Avg: **32.33**

* **Performance Distribution**

Majority of students scored above passing threshold

Significant score variance observed across demographics

1. **Discussion and Conclusion**

The analysis revealed:

* Clear performance differences among student groups
* Effectiveness of basic programming for educational analytics
* Potential for expansion with more sophisticated techniques

1. **Recommendation**

* Enhanced Analysis

Incorporate statistical tests for significance

Add visualization (histograms, box plots)

* Technical Improvements

Implement error handling for data quality issues

Optimize code for larger datasets

* Practical Applications

Develop early warning systems for at-risk students

Create performance benchmarking tools

1. **Acknowledgement**

I would like to express my gratitude to my mentor and the institute for providing guidance and resources to complete this project.

1. **References**

* Pandas Documentation: <https://pandas.pydata.org/>
* Kaggle Dataset: <https://www.kaggle.com/datasets/spscientist/students-performance-in-exams>

1. **Annexure**

* Dataset: <https://drive.google.com/file/d/1Ws17mDDao3eEglhNrgwiv1JKLXcgwyxJ/view?usp=sharing>
* Python notebook :

<https://colab.research.google.com/drive/1OgzbMSBYUU0UeB5zL93uMMZkc3PmwhqJ?usp=sharing>