:- Project Documentation & Report :-

L0NG TERM INTERNSHIP ON DATA ANALYTICS

Tittle of project : *Unleashing The Potential Of Our Youth: A Student Performance Analysis*

Industry Mentor : *Indra*

Faculty Mentor : *K.Munisankar*

Team ID : LTVIP2023TMID08632

Team Size : 5

Team Leader : *S Jayaprakash*

Team Members :

K Vamsivardhanreddy

Baitinti Ganesh

K Sainathreddy

Chagalamarri Vijaykumar

1.INTRODUCTION

1.1 Overview

PROJECT DESCRIPTION

- In this project, we focus on the crucial role of the education sector in measuring a country's growth, emphasizing the quality of its education system. Over time, the functioning of the education sector has undergone significant transformation, now being recognized as an industry in its own right. However, similar to other industries, the education sector also faces challenges, with one of the major concerns being the decreasing success rate of students and their tendency to leave courses without completing them.
- To address these challenges, we delve into the analysis of student work, which plays a vital role in effective teaching. Teachers routinely assign, collect, and examine student work to assess their learning progress and continuously enhance their teaching methodologies. By consistently evaluating student learning, teachers can engage in continuous quality improvement, ultimately benefiting the students' educational experience.
- Our dataset comprises the marks secured by 1000 students from a specific school. The primary objective of this analysis is to identify and correlate student performance with various attributes. Notably, we aim to understand how factors like the educational background of parents, the status of test preparation courses, and other pertinent variables influence students' performance in their exams.
- By conducting this analysis, we aspire to gain valuable insights into the relationships between different factors and academic performance. These insights can aid educators, policymakers, and stakeholders in the education system in making informed decisions to improve student outcomes and enhance the overall quality of education. Ultimately, we hope that our findings will contribute to the ongoing efforts to address the challenges faced by the higher education sector and foster a more conducive learning environment for students.

1.2 PURPOSE

- The project we have described aims to analyze and correlate student performance with various attributes, such as parental level of education and test preparation status, using a dataset of marks secured by 1000 The project will help identify key factors that influence student performance in exams students from a school. By conducting this analysis, several important insights and benefits can be achieved:

- **Understanding Factors Affecting Student Performance**:. This understanding can be used to target specific areas for improvement in the education system.
- **Insights for Teachers and Educators**: Teachers can gain valuable insights into the factors that contribute to students' success or challenges. This knowledge can be used to tailor teaching strategies and interventions to meet the diverse needs of students.
- **Continuous Quality Improvement of Courses**: By engaging in ongoing assessment and analysis of student performance, educators can make informed decisions to enhance the quality and effectiveness of their courses.
- **Identifying At-Risk Students**: The analysis can help identify students who are at risk of not completing the course or struggling academically. Early intervention strategies can be put in place to support these students and improve their chances of success.
- **Policy and Resource Allocation**: Education policymakers can use the findings to make informed decisions about resource allocation and policy changes that address the challenges faced by the education system.
- **Influence Parental Involvement**: Understanding the impact of parental level of education on student performance can encourage parental involvement in their children's education and highlight the importance of education at home.
- **Designing Test Preparation Programs**: Insights into the influence of test preparation on student performance can help in designing more effective test preparation programs and strategies.

- **Data-Driven Decision Making**: The project promotes data-driven decision-making in education. By analyzing and using data effectively, educators and policymakers can base their actions on evidence and improve overall educational outcomes.
- **Benchmarking and Comparison**: The analysis can be used for benchmarking and comparison across different student groups or schools to identify best practices and areas for improvement.
- Overall, the project's use can lead to evidence-based interventions and improvements in the education system, contributing to increased student success rates, higher completion rates, and ultimately, the growth and development of the country's education sector.

2.LITERATURE SURVEY

2.1 Existing problem

- **Data Preprocessing:** The first step in any data analysis project is data preprocessing. This involves cleaning the data, handling missing values, and transforming the data into a suitable format for analysis. In this case, the dataset containing the marks of 1000 students needs to be thoroughly checked for any errors or inconsistencies, and missing values should be appropriately dealt with.
- **Descriptive Statistics:** One of the initial steps in analyzing student performance is to compute descriptive statistics. This includes calculating the mean, median, standard deviation, and other relevant metrics to get an overview of the distribution of student scores and identify any patterns or outliers.
- **Data Visualization:** Visualizing the data through graphs and charts can provide valuable insights into the relationship between different attributes and student performance. For example, creating scatter plots or box plots to observe the correlation between parental level of education and student scores, or bar graphs to compare the success rate of students who took a test preparation course versus those who didn't.
- **Correlation Analysis:** Perform correlation analysis to quantify the relationships between different attributes and the students' performance. This can be done using statistical methods such as Pearson correlation coefficient or Spearman rank correlation, helping to identify which factors are strongly correlated with student success.
- **Regression Analysis:** Utilize regression analysis to model the relationship between the dependent variable (student performance) and independent variables (parental education level, test preparation status, etc.). Multiple regression can be employed if there are multiple factors to consider simultaneously.
- **Machine Learning Models:** Employ machine learning algorithms for predictive analysis. You can split the dataset into training and testing sets and use algorithms like decision trees, random forests, or support vector machines to predict student success based on various attributes.
- **Feature Importance:** If using machine learning models, analyze the feature importance to understand which attributes have the most significant impact on student performance. This information can guide educational institutions in focusing on the crucial factors for improving student outcomes.

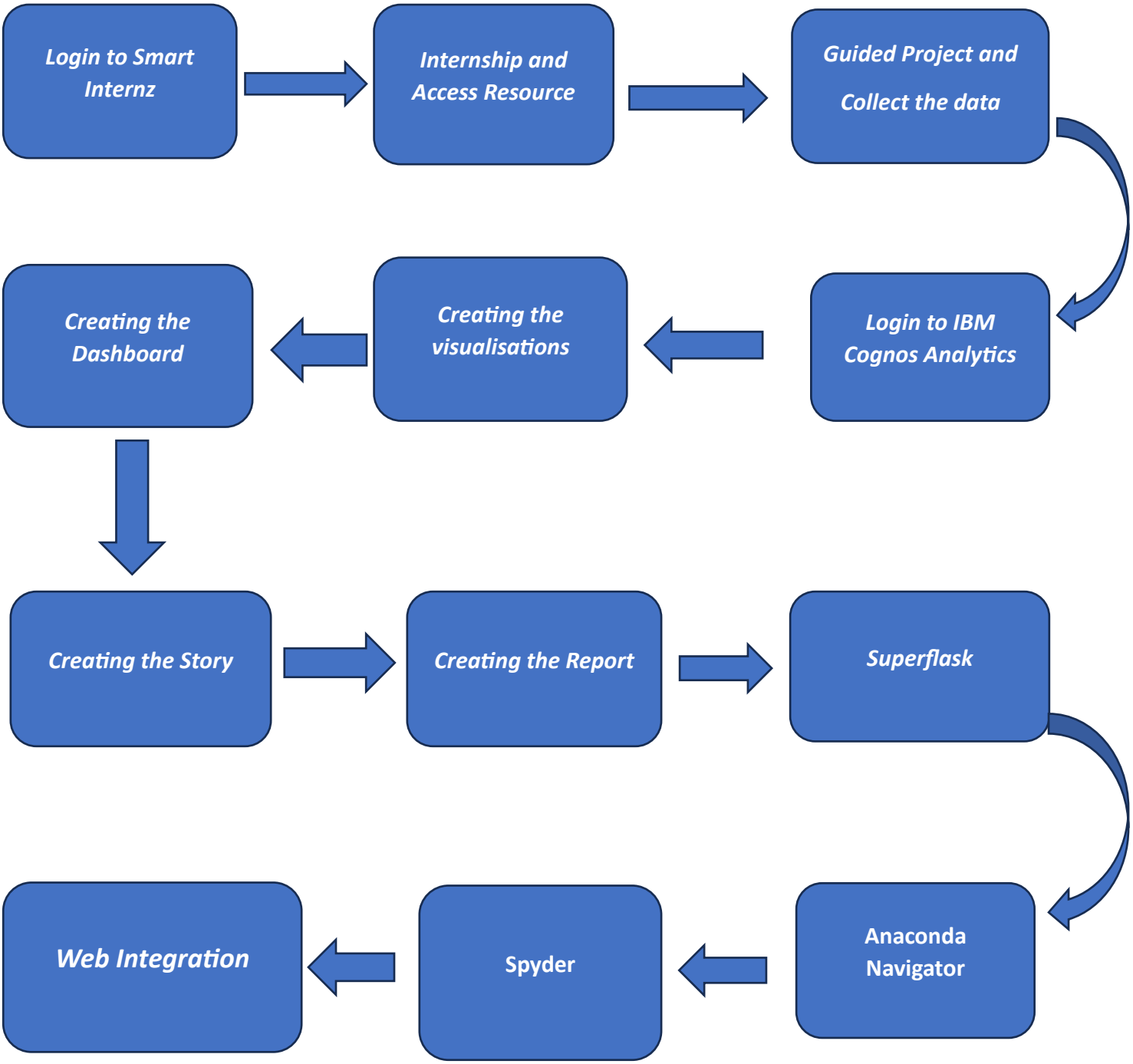
2.2 PROPOSED SOLUTION

- **Data Cleaning and Preparation:**
    - Begin by cleaning and preprocessing the dataset containing the marks secured by 1000 students. Handle missing values, outliers, and any inconsistencies in the data.

    - Convert categorical variables into numerical representations (e.g., one-hot encoding) to make them usable for analysis.
- **Exploratory Data Analysis (EDA):**
    - Perform exploratory data analysis to gain insights into the dataset. Visualize the distributions of various attributes, analyze correlations between variables, and identify patterns.
    - Explore the relationship between student performance and attributes such as parental level of education, test preparation status, etc., using graphs and statistical measures.
- **Predictive Modeling:**
    - Build a predictive model to forecast students' performance based on various attributes. Use machine learning algorithms such as regression, decision trees, or random forests, depending on the nature of the problem and the data.
    - Split the dataset into training and testing sets to evaluate the model's performance and avoid overfitting.
    - Assess the relative importance of different attributes in predicting student performance using feature importance techniques.
- **Recommendations and Insights:**
    - Summarize the findings of the analysis and highlight the significant factors that influence student performance the most.
    - Provide actionable recommendations to stakeholders (e.g., educators, policymakers) on how to improve student success rates and course completion. These recommendations may include targeted interventions for students with specific characteristics, improvements in test preparation programs, or initiatives to support students from diverse educational backgrounds.
- **Visualization and Reporting:**
    - Present the results of the analysis in a visually appealing and understandable manner. Create graphs, charts, and reports to effectively communicate the findings to a broader audience.
    - Use interactive visualizations to allow stakeholders to explore the data and findings in a more engaging way.
- **Continuous Improvement:**
    - Emphasize the importance of ongoing assessment and continuous improvement in the education system. Encourage the use of data-driven insights to make informed decisions and adapt teaching methodologies to better meet students' needs.

3 THEORITICAL ANALYSIS

3.1 BLOCK DIAGRAM



3.2 HARDWARE/SOFTWARE DESIGNING

Hardware Requirements:

- **Computer/Server:** A capable computer or server with sufficient processing power and memory to handle the dataset and perform the analysis.
- **Storage:** Adequate storage space to store the dataset and any intermediate or final results generated during the analysis.
- **Operating System:** The operating system should be compatible with the chosen data analysis tools and software.
- **Processor:** A multi-core processor (e.g., Quad-core or higher) is recommended to expedite data processing and analysis.
- **RAM:** A minimum of 8 GB RAM is recommended to handle large datasets and avoid performance bottlenecks.
- **Graphics Processing Unit (GPU) :** If you plan to use machine learning algorithms that support GPU acceleration, having a compatible GPU can significantly speed up certain computations.

Software Requirements:

- **Python:** Python is a popular programming language for data analysis, and numerous libraries and frameworks are available to facilitate the analysis. Ensure that Python is installed on the system.
- **Integrated Development Environment (IDE):** Choose a suitable IDE for Python development, such as Jupyter Notebook, JupyterLab. The IDE should provide an interactive environment to run code and visualize data.
- **Data Analysis Libraries:** Install essential Python libraries for data analysis, including but not limited to:
    - Pandas: For data manipulation and analysis.
    - NumPy: For numerical operations on data.
    - Matplotlib and Seaborn: For data visualization.
- **Machine Learning Libraries (Optional):** If you plan to build predictive models or perform machine learning tasks, consider installing relevant libraries, such as Scikit-learn or TensorFlow.
- **Database Management System (DBMS):** Depending on the size and structure of the dataset, you may need a DBMS like MySQL or SQLite to store and manage the data efficiently.
- **Text Editor:** A text editor can be helpful for making any changes to code or configuration files.
- **Spreadsheet Software:** You may need spreadsheet software like Microsoft Excel or Google Sheets to view, clean, or manipulate the data before importing it into Python.
- **Web Browser:** For accessing online resources, documentation, and any web-based tools related to the project.

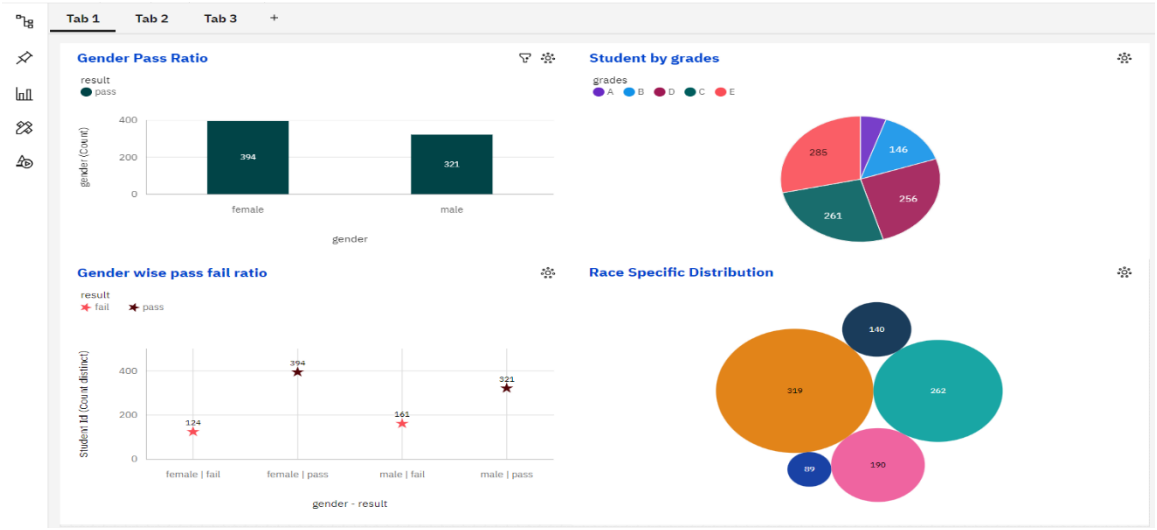
Internet Connectivity:

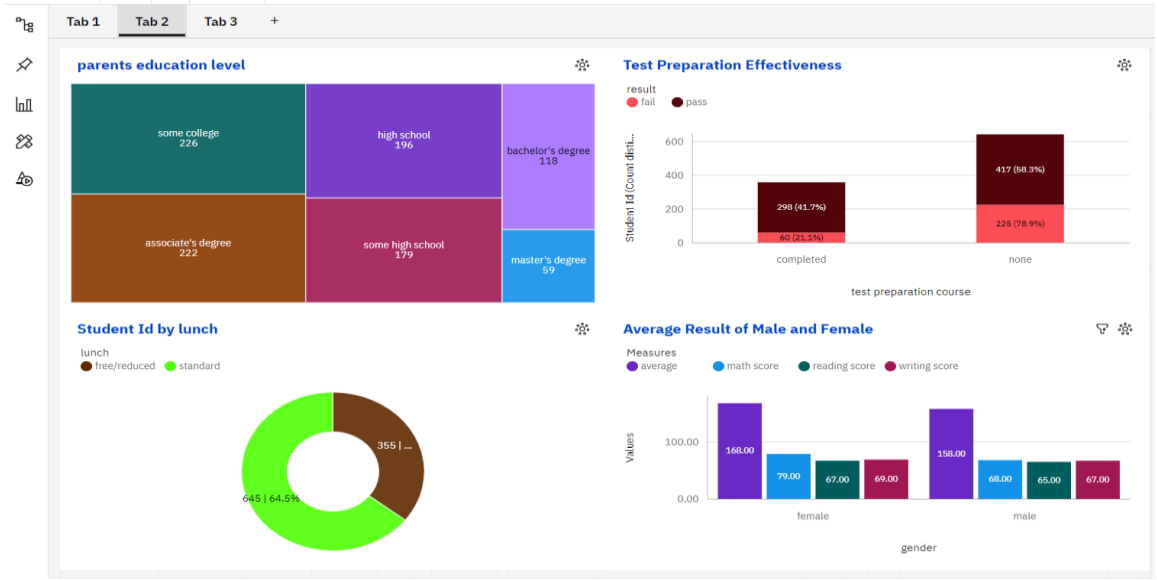
- A stable internet connection may be required if you need to download libraries, access external datasets, or refer to online documentation or resources.

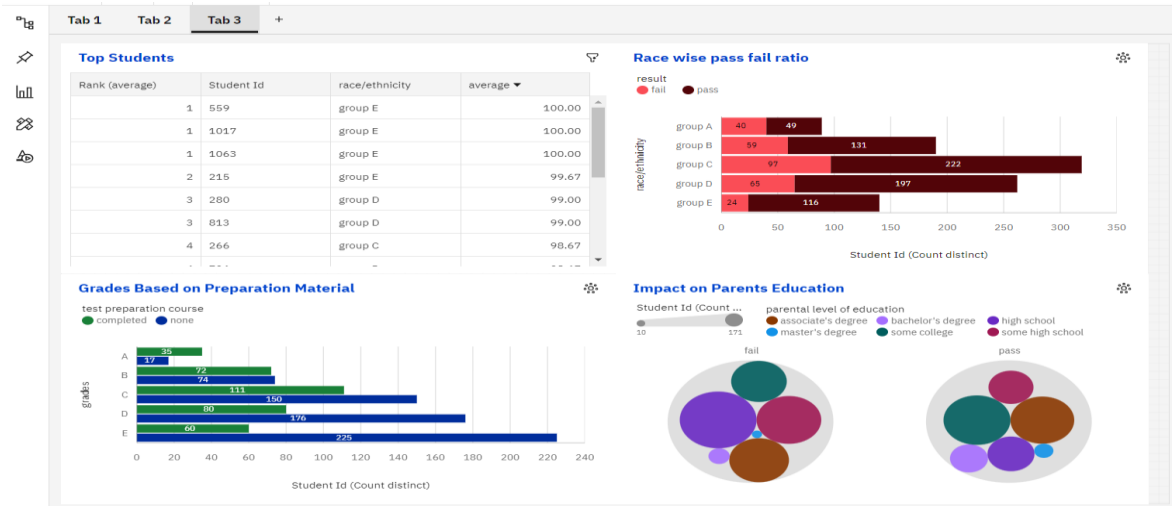
- Before starting the analysis, ensure that all the necessary software and libraries are installed and the hardware specifications meet the project's requirements. With the appropriate setup, you'll be ready to unleash the potential of our youth through this student performance analysis.

4 RESULT

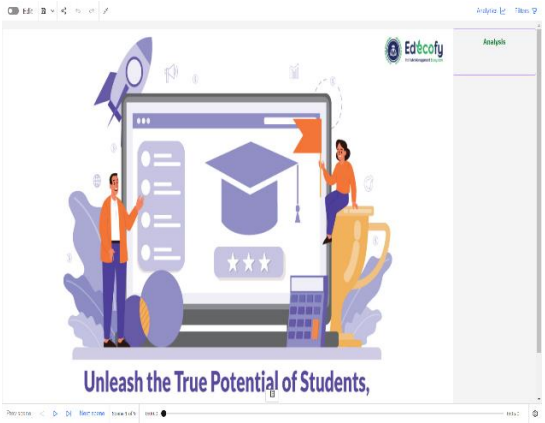
Dashboard:

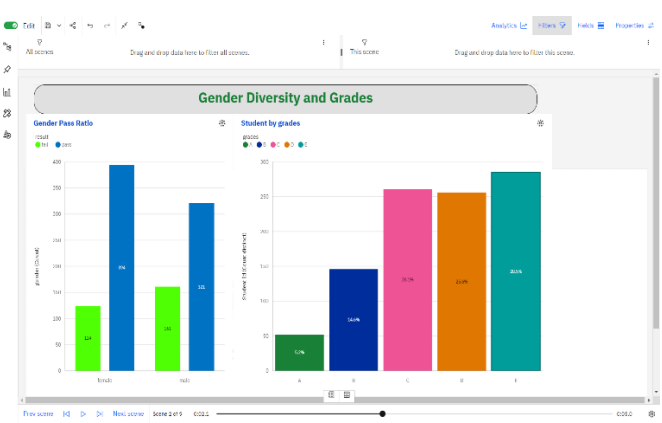


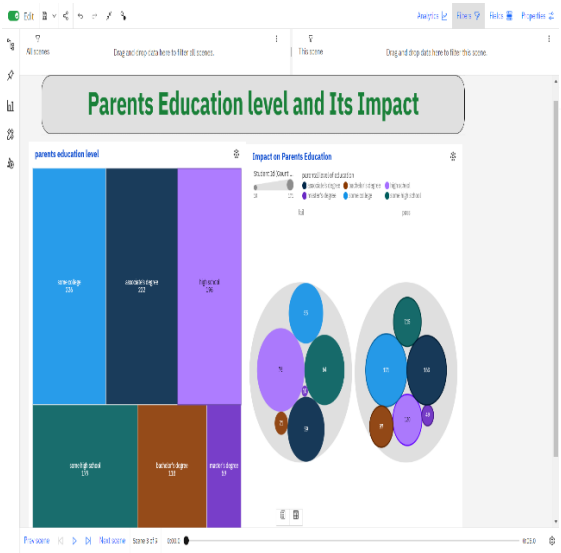


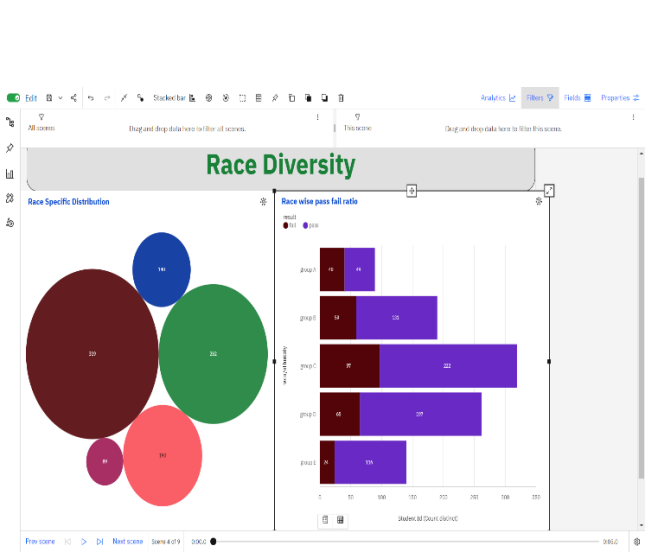

Tab 1
Tab 2
Tab 3
Top Students
Rank (average) | Student Id | race/ethnicity | average
1 | 559 | group E | 100.00
1 | 1017 | group E | 100.00
1 | 1063 | group E | 100.00
2 | 215 | group E | 99.67
3 | 280 | group D | 99.00
3 | 813 | group D | 99.00
4 | 266 | group C | 98.67
Race wise pass fail ratio
result
fail
pass
race/ethnicity
group A
group B
group C
group D
group E
Student Id (Count distinct)
Grades Based on Preparation Material
test preparation course
completed
none
grades
Student Id (Count distinct)
Impact on Parents Education
Student Id (Count ...
parental level of education
associate's degree
bachelor's degree
high school
master's degree
some college
some high school
fail
pass

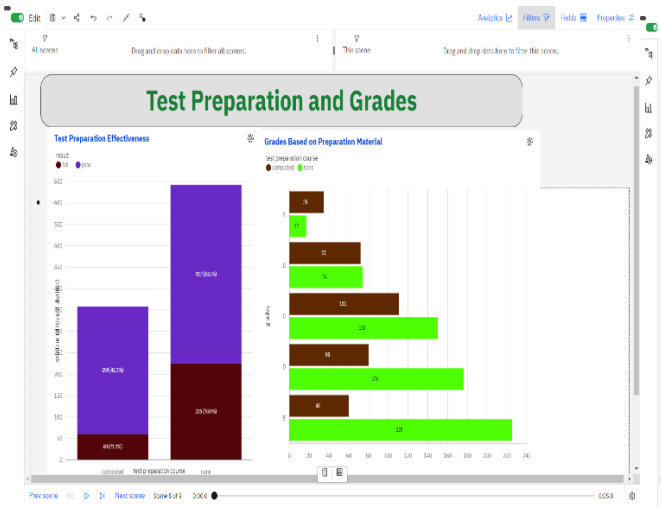
Story:

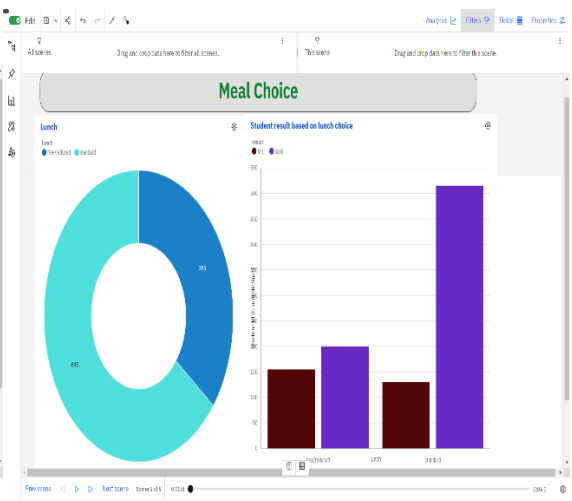

Edecofy
Analysis
Unleash the True Potential of Students,

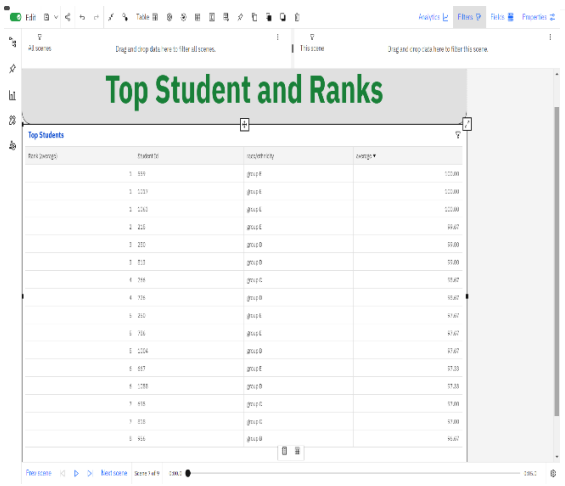

Gender Diversity and Grades
Gender Pass Ratio
Student by grades

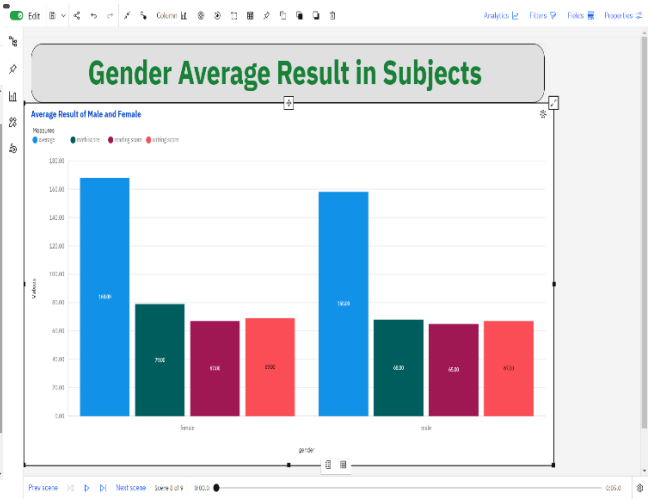

Parents Education level and Its Impact
parents education level
Impact on Parents Education


Race Diversity
Race Specific Distribution
Race wise pass fail ratio

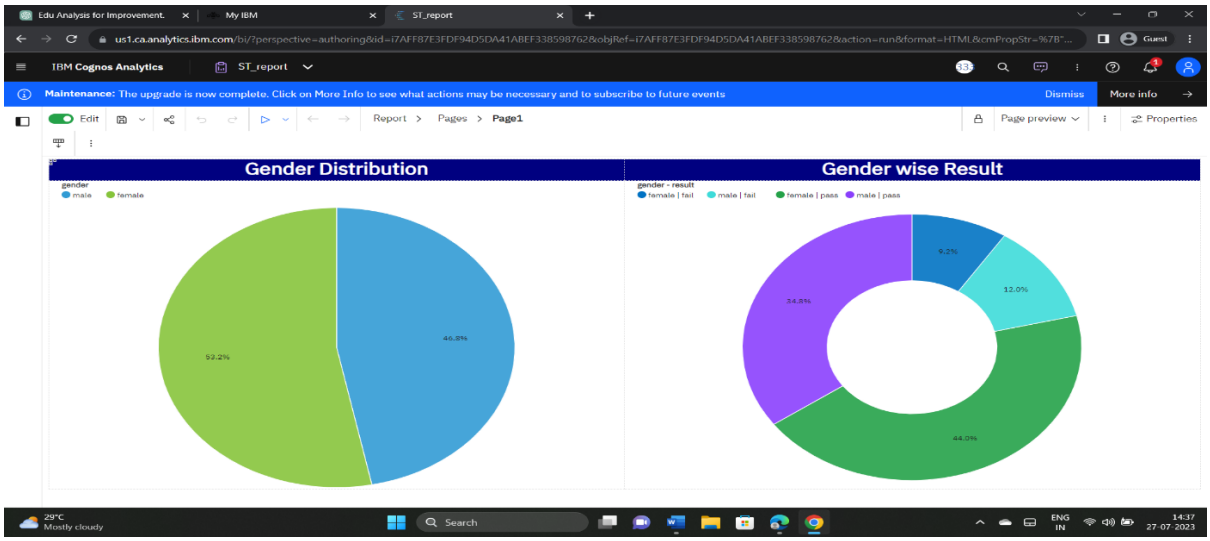

Test Preparation and Grades
Test Preparation Effectiveness
Grades Based on Preparation Material

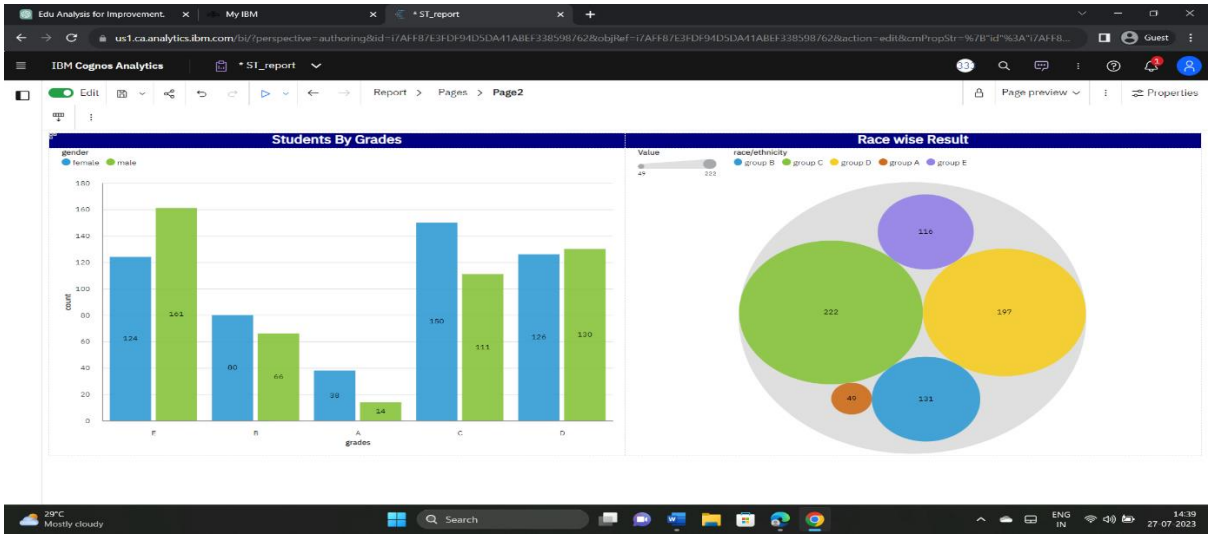

Meal Choice
Lunch
Student result based on lunch choice

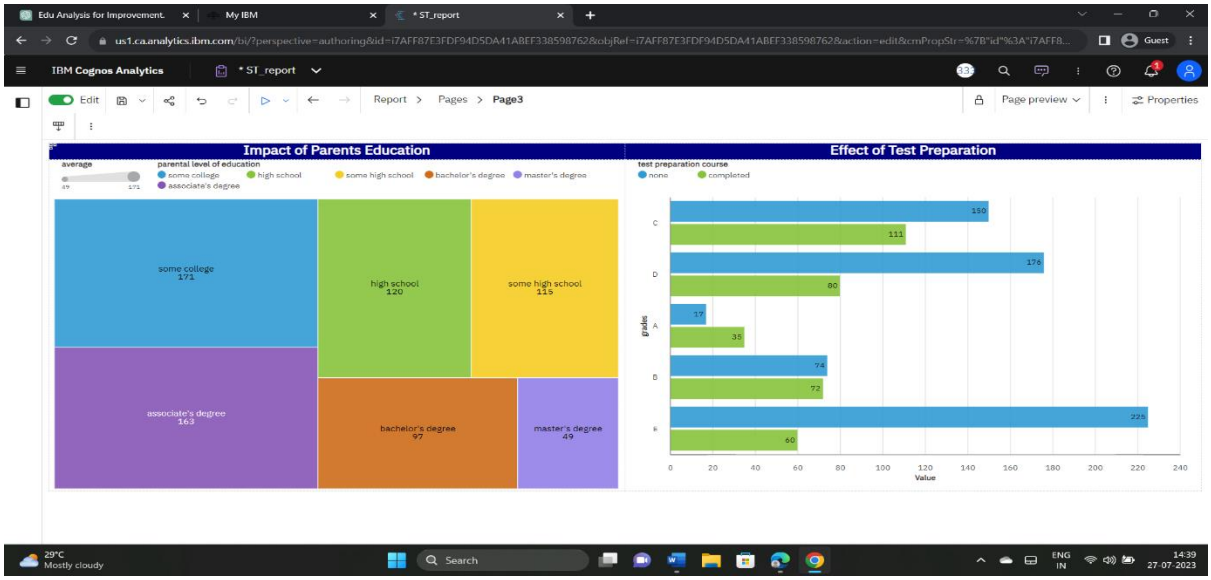

Top Student and Ranks
Top Students

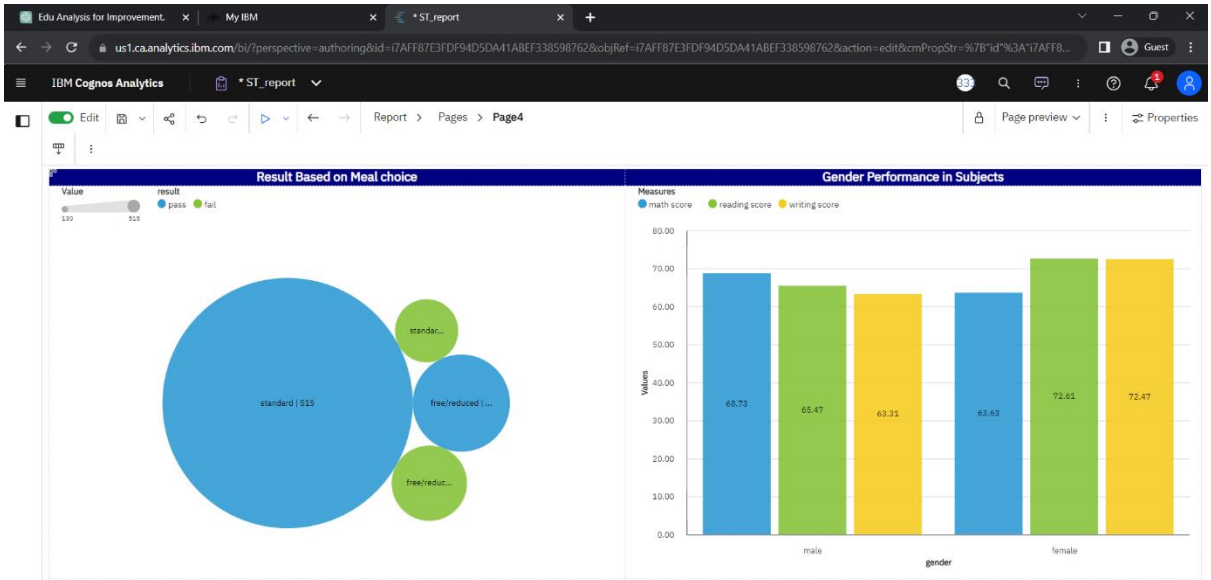

Gender Average Result in Subjects
Average Result of Male and Female

Report:

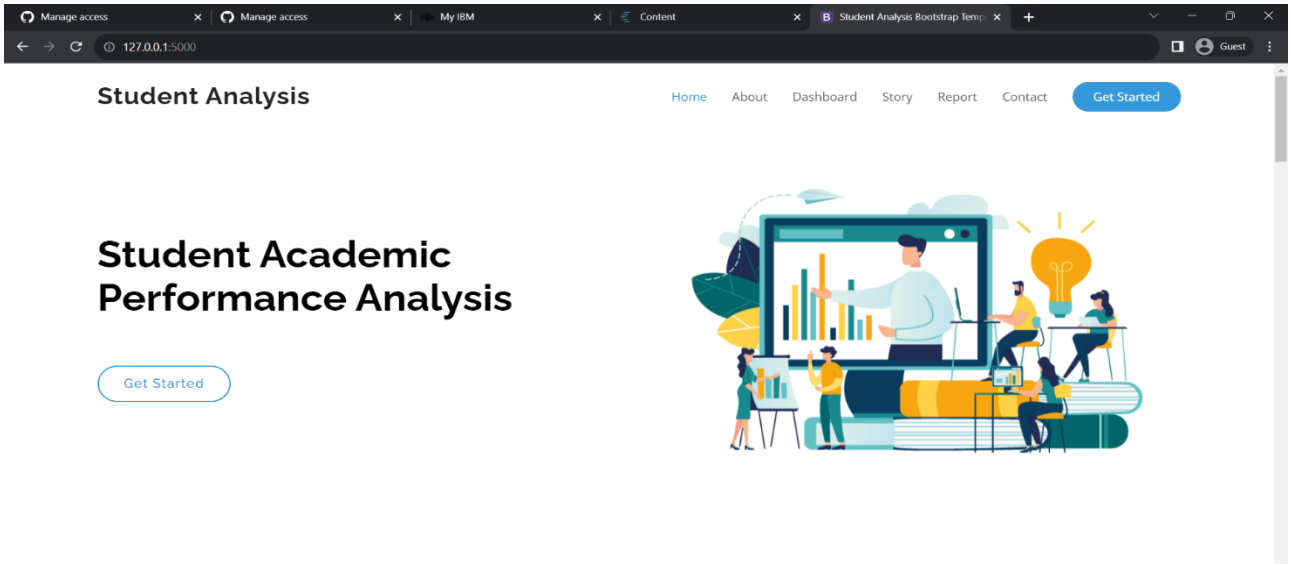

Edu Analysis for Improvement
My IBM
ST_report
us1.ca.analytics.ibm.com
IBM Cognos Analytics
ST_report
Maintenance: The upgrade is now complete. Click on More Info to see what actions may be necessary and to subscribe to future events
Dismiss
More info
Edit
Report > Pages > Page1
Page preview
Properties
Gender Distribution
gender
male
female
46.8%
53.2%
Gender wise Result
gender - result
female | fail
male | fail
female | pass
male | pass
9.2%
12.0%
34.8%
44.0%
29°C
Mostly cloudy
Search
ENG
IN
14:37
27-07-2023

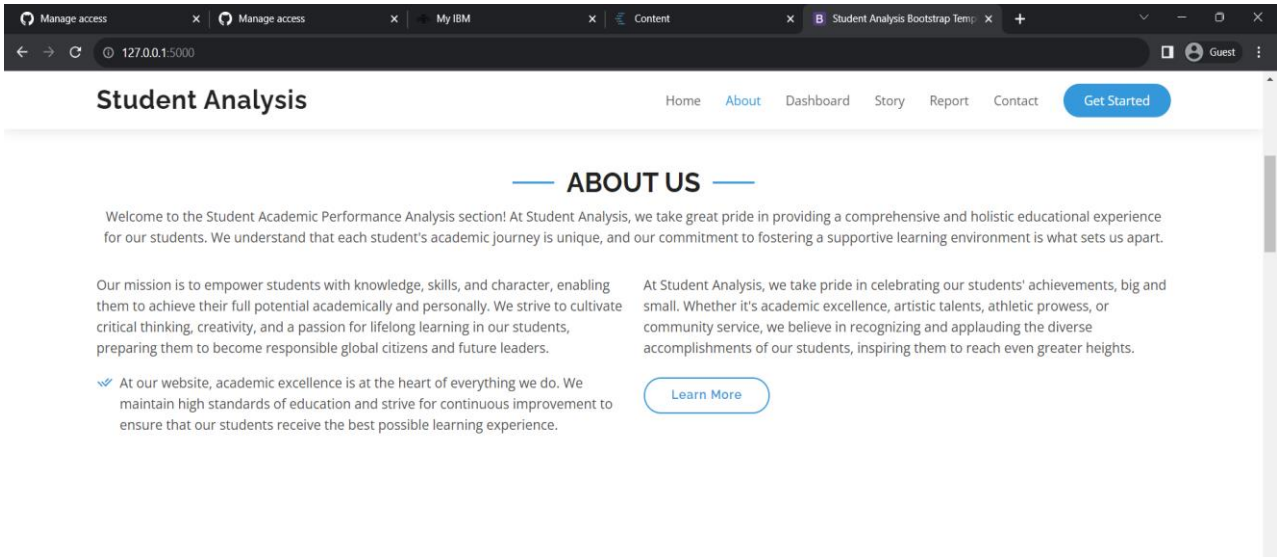

Edu Analysis for Improvement.
My IBM
* ST_report
IBM Cognos Analytics
Edit
Report > Pages > Page2
Page preview
Properties
Students By Grades
gender
female
male
count
grades
124
161
60
66
38
14
150
111
126
130
Race wise Result
Value
race/ethnicity
group B
group C
group D
group A
group E
116
222
197
49
131

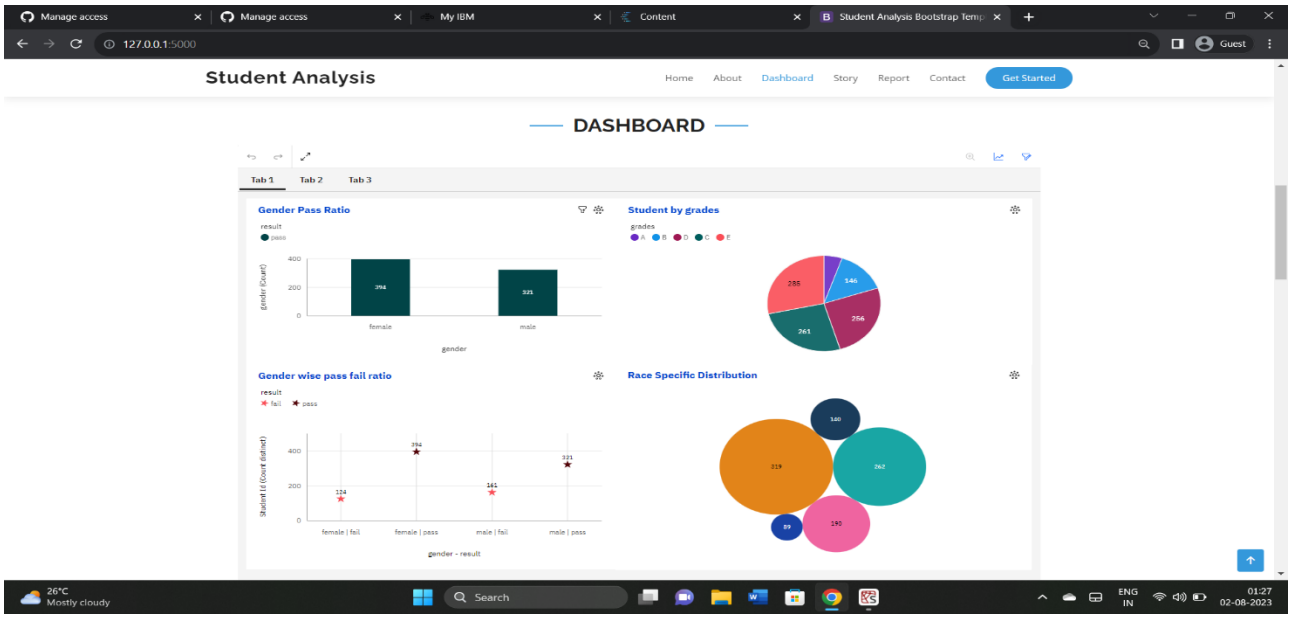

Edu Analysis for Improvement.
My IBM
* ST_report
IBM Cognos Analytics
Edit
Report > Pages > Page3
Page preview
Properties
Impact of Parents Education
average
parental level of education
some college
high school
some high school
bachelor's degree
master's degree
associate's degree
some college
171
high school
120
some high school
115
associate's degree
163
bachelor's degree
97
master's degree
49
Effect of Test Preparation
test preparation course
none
completed
grades
Value
150
111
176
80
17
35
74
72
223
60

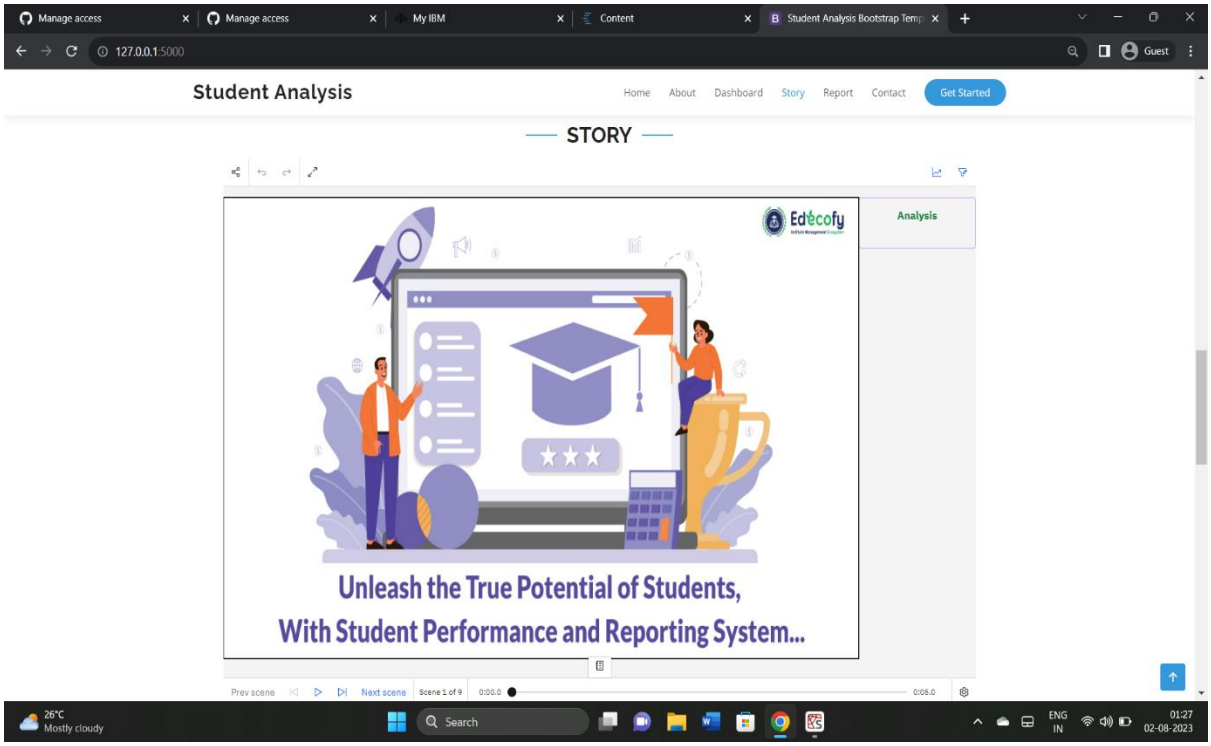

Edu Analysis for Improvement.
My IBM
* ST_report
IBM Cognos Analytics
Edit
Report > Pages > Page4
Page preview
Properties
Result Based on Meal choice
Value
result
pass
fail
standard | 515
free/reduced | ...
Gender Performance in Subjects
Measures
math score
reading score
writing score
Values
gender
68.73
65.47
63.31
63.63
72.61
72.47
male
female

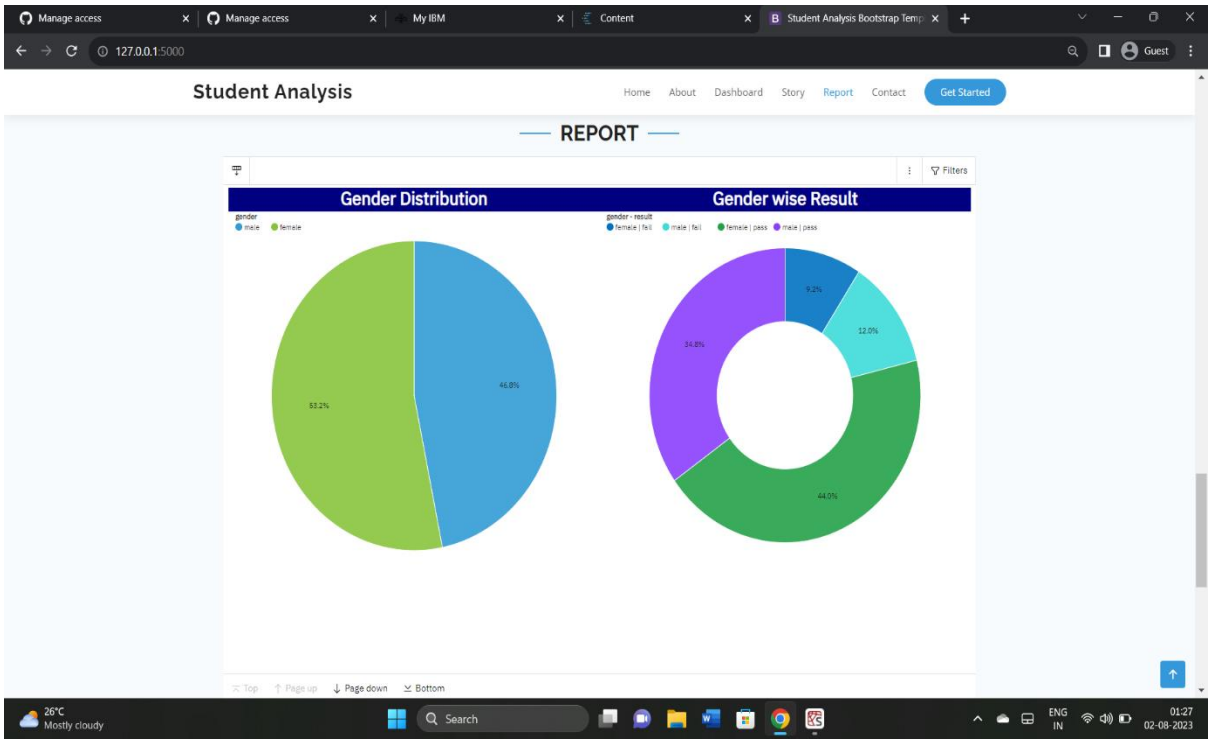
Web Integration:








Manage access
Manage access
My IBM
Content
Student Analysis Bootstrap Temp
127.0.0.1:5000
Guest
Student Analysis
Home
About
Dashboard
Story
Report
Contact
Get Started
STORY
Edecofy
Analysis
Unleash the True Potential of Students,
With Student Performance and Reporting System...
Prev scene
Next scene
Scene 1 of 9
26°C
Mostly cloudy
Search
ENG
IN
01:27
02-08-2023


Manage access
Manage access
My IBM
Content
Student Analysis Bootstrap Temp
127.0.0.1:5000
Guest
Student Analysis
Home
About
Dashboard
Story
Report
Contact
Get Started
REPORT
Filters
Gender Distribution
Gender wise Result
gender
male
female
gender - result
female | fail
male | fail
female | pass
male | pass
46.8%
53.2%
9.2%
12.0%
34.8%
44.0%
Top
Page up
Page down
Bottom
26°C
Mostly cloudy
Search
ENG
IN
01:27
02-08-2023

5 ADVANTAGES&DISADVANTAGES

Advantages of the Proposed Solution:

- **Data-Driven Insights**: The project is based on analyzing a dataset containing students' marks, which allows for data-driven insights into the factors influencing student performance. This can lead to more objective and evidence-based decision-making.
- **Understanding Factors Influencing Performance:** By analyzing factors like parental level of education and test preparation status, the project can provide valuable insights into the variables that have the most significant impact on student success. This understanding can help educators tailor interventions and support systems to address specific challenges.
- **Continuous Quality Improvement:** The ongoing assessment of student learning can enable teachers to engage in continuous quality improvement of their courses. Identifying areas where students struggle the most can lead to curriculum refinements and teaching strategies to enhance overall learning outcomes.
- **Evidence-Based Decision Making:** The project's findings can serve as a basis for educational policymakers to make evidence-based decisions. By understanding the factors influencing student performance, policymakers can design targeted interventions and policies to enhance the overall quality of education.
- **Student-Centric Approach:** The analysis aims to understand the factors affecting student success, which promotes a student-centric approach to education. This focus on individual needs and challenges can lead to more personalized support and academic growth.

Disadvantages of the Proposed Solution:

- **Limited Scope:** While analyzing the provided dataset can offer valuable insights, it may have limitations in representing the entire student population's diversity. Factors not present in the dataset could be equally significant but not considered in the analysis.
- **Causation vs. Correlation:** The analysis might identify correlations between factors and student performance, but it may not establish causation. For example, a correlation between test preparation and higher marks does not necessarily prove that test preparation directly caused better performance.
- **Generalizability:** The dataset may represent the performance of a specific school or region, making it challenging to generalize the findings to different educational contexts or countries. The project's conclusions should be approached with caution before applying them broadly.

- **Ethical Considerations:** Data analysis involving student performance raises ethical concerns, such as maintaining student privacy and ensuring that the analysis is not used to stigmatize or discriminate against specific student groups based on their backgrounds.
- **Addressing Underlying Issues:** While the project identifies factors affecting student performance, it may not directly address underlying systemic issues or societal disparities

that influence education quality. A holistic approach to education reform may be necessary to tackle these broader challenges.

6 APPLICATIONS

Potential Areas of Application:

- **Educational Policy Formulation:** The insights gained from this analysis can be used by educational policymakers to identify factors that significantly impact student performance. This information can be utilized to formulate targeted policies and interventions aimed at improving the overall quality of the education system.
- **Curriculum Improvement:** Understanding the correlation between different attributes and student performance can help educators and curriculum designers identify areas where the curriculum can be enhanced or modified. This analysis can lead to a more tailored and effective learning experience for students.
- **Student Support Programs:** Schools and educational institutions can utilize the findings to develop support programs for students at risk of underperforming or dropping out. By addressing the specific challenges faced by students, such as those from disadvantaged backgrounds, institutions can enhance student success rates.
- **Parental Engagement:** Recognizing the influence of parental education on student performance can encourage efforts to increase parental involvement in their children's education. Schools can collaborate with parents and provide resources to enhance their support in their child's academic journey.
- **Test Preparation Strategies:** Analyzing the impact of test preparation courses on student performance can assist educators in refining and improving such programs. Institutions can design more effective test preparation courses to boost student achievement.
- **Personalized Learning:** The analysis can help identify individual student needs and strengths, enabling the implementation of personalized learning approaches. Tailoring education to suit each student's requirements can lead to better engagement and improved outcomes.

- **School Ranking and Accreditation:** Educational authorities can use the insights from this analysis to assess and rank schools based on their ability to improve student performance. It can also aid in accrediting institutions that demonstrate effective teaching methodologies.
- **Research and Data-Driven Decision Making:** The project's findings can contribute to the existing body of research on educational outcomes and serve as a basis for data-driven decision making in the education sector.

7 CONCLUSION

The Analysis Process:

- Data Collection: Marks secured by 1000 students are gathered from the school dataset.
- Correlation of Student Performance: The project examines the relationship between students' academic performance and various attributes, such as parental educational background and participation in test preparation courses.
- Identifying Key Influencers: The analysis aims to pinpoint the most influential factors affecting students' academic success.
- Continuous Quality Improvement: Ongoing assessment of student learning is essential for teachers to refine and enhance their teaching methods continually.

Findings:

The project uncovers valuable insights into the factors that contribute significantly to student performance:

- Parental Education: It is observed that students with parents having higher educational backgrounds tend to perform better academically. This highlights the positive influence of parents' education on their children's learning outcomes.

- Test Preparation Courses: Students who have participated in test preparation courses demonstrate improved performance, suggesting that such courses play a crucial role in enhancing student learning and exam outcomes.

Conclusion:

In conclusion, the project "Unleashing the Potential of Our Youth: A Student Performance Analysis" emphasizes the vital role of education in a country's growth. By analyzing the dataset and correlating student performance with various attributes, the project sheds light on the significance of parental education and the impact of test preparation courses on student success. These findings underscore the need for continuous quality improvement in the education system and encourage educators to consider tailored interventions to support students in their academic journey. To foster the potential of our youth and ensure a prosperous future, it is essential to address the identified factors and provide a conducive learning environment that empowers every student to excel.

8 FUTURE SCOPE

Future Enhancements:

In the future, the project could be enhanced in several ways to further explore and improve the education system. Some potential enhancements include:

- Longitudinal Analysis: Instead of focusing solely on a snapshot of student performance, conducting a longitudinal analysis could provide valuable insights into student progress over time. This approach could help identify trends, patterns, and areas of improvement for continuous support and development.
- Qualitative Research: In addition to quantitative data analysis, incorporating qualitative research methods like interviews or surveys with students, parents, and educators could

provide deeper insights into the underlying reasons behind certain trends in student performance.

- Intervention Strategies: Building on the analysis, the project could propose and test various intervention strategies to address the identified challenges. These strategies could focus on improving parental involvement, optimizing test preparation programs, or implementing personalized learning approaches to cater to individual student needs.
- Comparison with Other Schools: Comparing the performance analysis with data from other schools or educational institutions could provide a broader perspective on the challenges faced by the education sector and highlight successful practices that can be adopted.
- Technology Integration: Exploring the role of technology in education and how it can be leveraged to enhance student learning and engagement could be a valuable addition to the project. This could include investigating the impact of e-learning platforms, digital resources, or educational apps on student performance.
- Policy Recommendations: Based on the project's findings, proposing policy recommendations to educational authorities could lead to systemic changes aimed at improving the overall quality of education and addressing the identified challenges.

- By continuously improving and expanding the project's scope, stakeholders in the education sector can gain valuable insights to foster a more effective and supportive learning environment for students, ultimately contributing to the growth and development of the country.