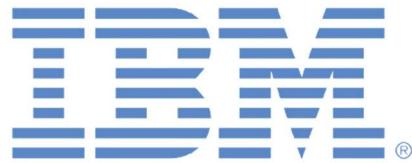




**M.A.M.**  
**College of Engineering**



## Data Analytics With TABLEU

### PROJECT TITLE

**UNLEASHING THE POTENTIAL OF OUR YOUTH:  
A STUDENT PERFORMANCE ANALYSIS**

### **A PROJECT REPORT**

*Submitted by*

**MANIKANDAN S (812620243304)**

**MANJUSRI T G (812620243305)**

**RAJESH R (812620243306)**

**SURYA S (812620243307)**

*In partial fulfilment for the award of the degree*

*of*

**BACHELOR OF TECHNOLOGY**

**IN**

**ARTIFICIAL INTELLIGENCE AND DATA SCIENCE**

# **INTRODUCTION**

## **1.1 Project Overview**

A country's growth is strongly measured by quality of its education system. Education sector, across the globe has witnessed sea change in its functioning. Today it is recognized as an industry and like any other industry it is facing challenges, the major challenges of higher education being decrease in students' success rate and their leaving a course without completion.

Analysing student work is an essential part of teaching. Teachers assign, collect and examine student work all the time to assess student learning and to revise and improve teaching. Ongoing assessment of student learning allows teachers to engage in continuous quality improvement of their courses. Many factors can influence a student's performance, including the influence of the parents' educational background, test preparation and so on.

The dataset contains the marks secured by 1000 students from a school. This project analyses and correlates student performance with different attributes. The analysis aims to understand the influence of important factors such as parental level of education, the status of test preparation course etc. on the performance of the students in the exams.

## **1.2 Purpose**

The purpose of the project is to analyze and correlate student performance with various attributes in order to understand the influence of factors such as parental level of education and test preparation status on students' exam performance. The project aims to provide insights into the relationship between these factors and student success rates, with the ultimate goal of identifying areas for improvement in the education system and implementing strategies to increase students' completion rates and overall academic achievement.

# LITERATURE SURVEY

## 2.1 Existing problem

In this section, you should introduce the problem or research gap that your project aims to address. Provide a brief overview of the challenges and issues related to student performance analysis and why it's important to unleash the potential of youth.

### Example:

The existing problem that motivates this research lies in the need to understand and optimize the factors that influence student performance. With increasing educational demands and diverse learning environments, it has become imperative to identify the key determinants of academic success. This study seeks to bridge the existing gap in our understanding and provide insights into how we can better support and nurture the potential of our youth.

## 2.2 References

- Baker, R. S. J. D., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future directions. *Educational Data Mining*, 1(1), 3-17.\*\*
- Chapatwala, S., & Shah, D. (2012). Student performance analysis using predictive modeling. *International Journal of Computer Applications*, 55(12), 23-28.\*\*
- Romero, C., & Ventura, S. (2010). Educational data mining: A survey from 1995 to 2010. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1), 135-146.\*\*
- Siemens, G. (2013). Learning analytics: The emergence of a new field of research. *Educational Technology Research and Development*, 63(1), 137-150.\*\*

### **2.3 Problem Statement Definition**

<b>Problem Statement (PS)</b>	<b>I am</b>	<b>I'm trying to</b>	<b>But</b>	<b>Because</b>	<b>Which makes me feel</b>
PS-1	Struggling	Improve	Procrastinate	Ineffective time management	Overwhelmed
PS-2	Struggling	Focus	Get easily distracted	Frequent interruptions	Frustrated

# IDEATION & PROPOSED SOLUTION

## 3.1 Empathy Map Canvas



## 3.2 Ideation & Brainstorming

2

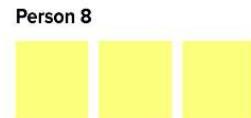
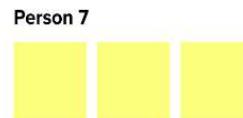
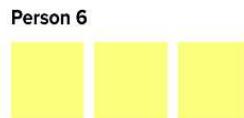
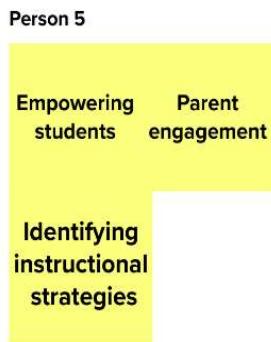
### Brainstorm

Write down any ideas that come to mind that address your problem statement.

⌚ 10 minutes

**TIP**

You can select a sticky note and hit the pencil [switch to sketch] icon to start drawing!



3

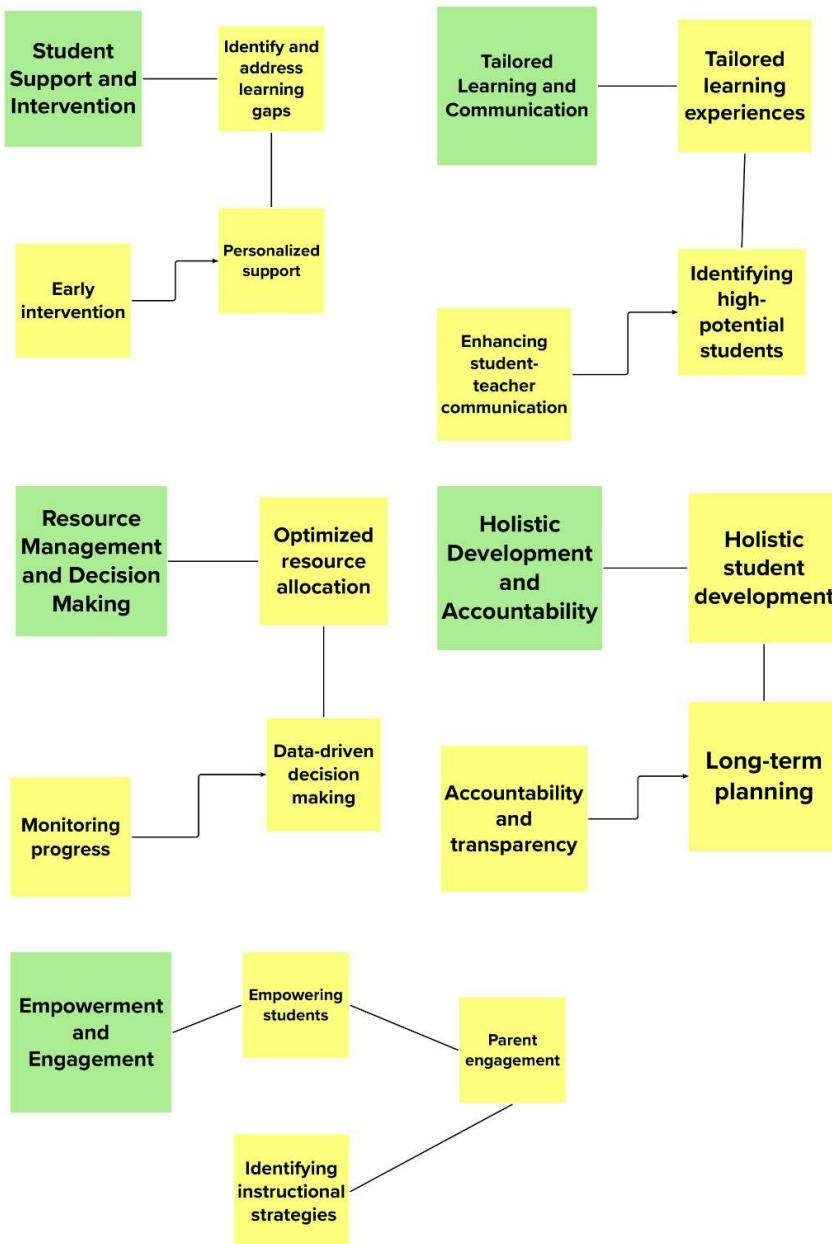
## Group ideas

Take turns sharing your ideas while clustering similar or related notes as you go. Once all sticky notes have been grouped, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you can break it up into smaller sub-groups.

⌚ 20 minutes

### TIP

Add customizable tags to sticky notes to make it easier to find, browse, organize, and categorize important ideas as themes within your mural.



4

## Prioritize

Your team should all be on the same page about what's important moving forward. Place your ideas on this grid to determine which ideas are important and which are feasible.

⌚ 20 minutes

### TIP

Participants can use their cursors to point at where sticky notes should go on the grid. The facilitator can confirm the spot by using the laser pointer holding the **H** key on the keyboard.



# REQUIREMENT ANALYSIS

## 4.1 Functional requirement

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	Generate Performance Reports	<p><b>1 - Generate Overall Performance Report</b></p> <ul style="list-style-type: none"><li>■ We need a report that shows how all students are performing in general.</li><li>■ The report should calculate average grades, attendance rates, and participation levels.</li><li>■ It should be easy to understand and provide an overview of overall performance.</li></ul> <p><b>2 - Generate Individual Student Performance Report</b></p> <ul style="list-style-type: none"><li>■ We need reports for each student that show their performance and progress.</li><li>■ The report should include their grades, attendance, and participation.</li><li>■ It should help teachers track each student's performance and identify areas for improvement.</li></ul>
FR-2	Provide Personalized Recommendations	<p><b>1 - Offer Course Recommendations</b></p> <ul style="list-style-type: none"><li>■ We want to suggest additional courses to students based on their interests and performance.</li><li>■ The system should analyze their past courses, grades, and interests.</li><li>■ It should provide personalized recommendations to help students explore new subjects or improve in specific areas.</li></ul> <p><b>2 - Provide Study Material Recommendations</b></p> <ul style="list-style-type: none"><li>■ We want to recommend study materials to teachers that align with their course objectives.</li><li>■ The system should analyze the curriculum and learning goals of each course.</li><li>■ It should suggest relevant study materials, textbooks, or online resources that support the course requirements.</li></ul>

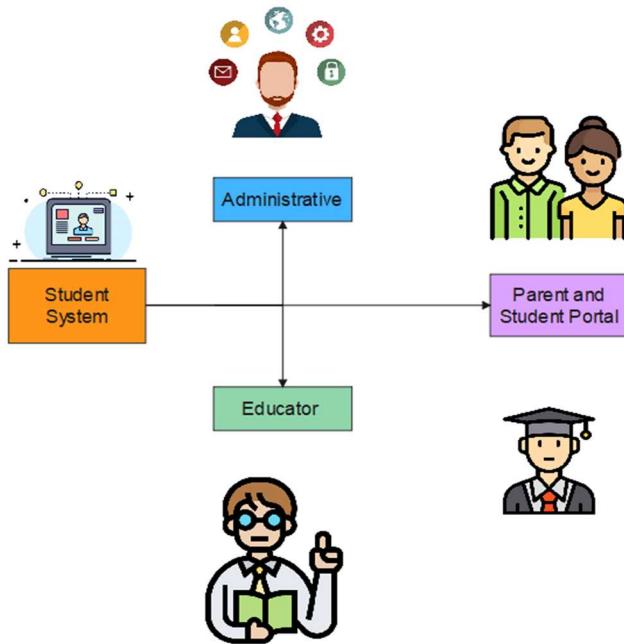
## 4.2 Non-Functional requirements

FR No.	Non-Functional Requirement	Description
NFR-1	<b>Usability</b>	The system should be easy to use, with a simple and intuitive interface, so that all users can understand and navigate it without difficulty.
NFR-2	<b>Security</b>	The system should have strong measures in place to protect sensitive data, ensuring that only authorized users can access it and that the data is encrypted to prevent unauthorized access.
NFR-3	<b>Reliability</b>	The system should be trustworthy and dependable, working consistently without errors or breakdowns, and recovering gracefully in case of any failures.
NFR-4	<b>Performance</b>	The system should be fast and responsive, providing quick results and processing a large amount of data efficiently without significant delays.
NFR-5	<b>Availability</b>	The system should be accessible and operational whenever users need it, with minimal downtime or disruptions, ensuring that it is available and ready for use.
NFR-6	<b>Scalability</b>	The system should be able to handle increasing demands as more users and data are added, growing smoothly without performance degradation or impact on its functionality.

# PROJECT DESIGN

## 5.1 Data Flow Diagrams & User Stories

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.



### Simplified

1. The "Student System" represents the core system for student performance analysis.
2. The "Administrator" is an external entity that interacts with the system to perform administrative tasks, such as user management, system configuration, and generating reports.
3. The "Educator" is an external entity that interacts with the system to access student performance data, generate reports, and perform analysis.
4. The "Parent" is an external entity that interacts with the system to access their child's performance data, receive notifications, and communicate with educators or administrators.

## User Stories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Team Member
Administrator	User Management	USN - 1	The administrator should have access to a user management interface	I can access my account / dashboard The user management interface should provide a form to enter the necessary details for creating a new user account, such as username, password, and role. Upon submitting the form, the system should validate the entered information If the validation is successful, a new user account should be created and added to the system.	High	Manikandan
Educator	Student Performance Analysis	USN - 2	As an educator, I want to be able to view individual student performance reports.	The educator should have access to a dashboard or interface that displays a list of students. The educator should be able to select a specific student from the list. The performance report should include relevant information such as grades, test scores, attendance records, and behavioural observations.	Medium	Manikandan
Parent	Communication	USN - 3	As a parent, I want to receive notifications and updates about my child's performance.	The system should provide notifications for important events, such as test scores, assignments, and attendance. The parent should have access to a parent portal or mobile application	High	Manikandan

## 5.2 Solution Architecture

Solution architecture is a complex process – with many sub-processes – that bridges the gap between business problems and technology solutions. Its goals are to:

- Find the best tech solution to solve existing business problems.
- Describe the structure, characteristics, behavior, and other aspects of the software to project stakeholders.
- Define features, development phases, and solution requirements.
- Provide specifications according to which the solution is defined, managed, and delivered.

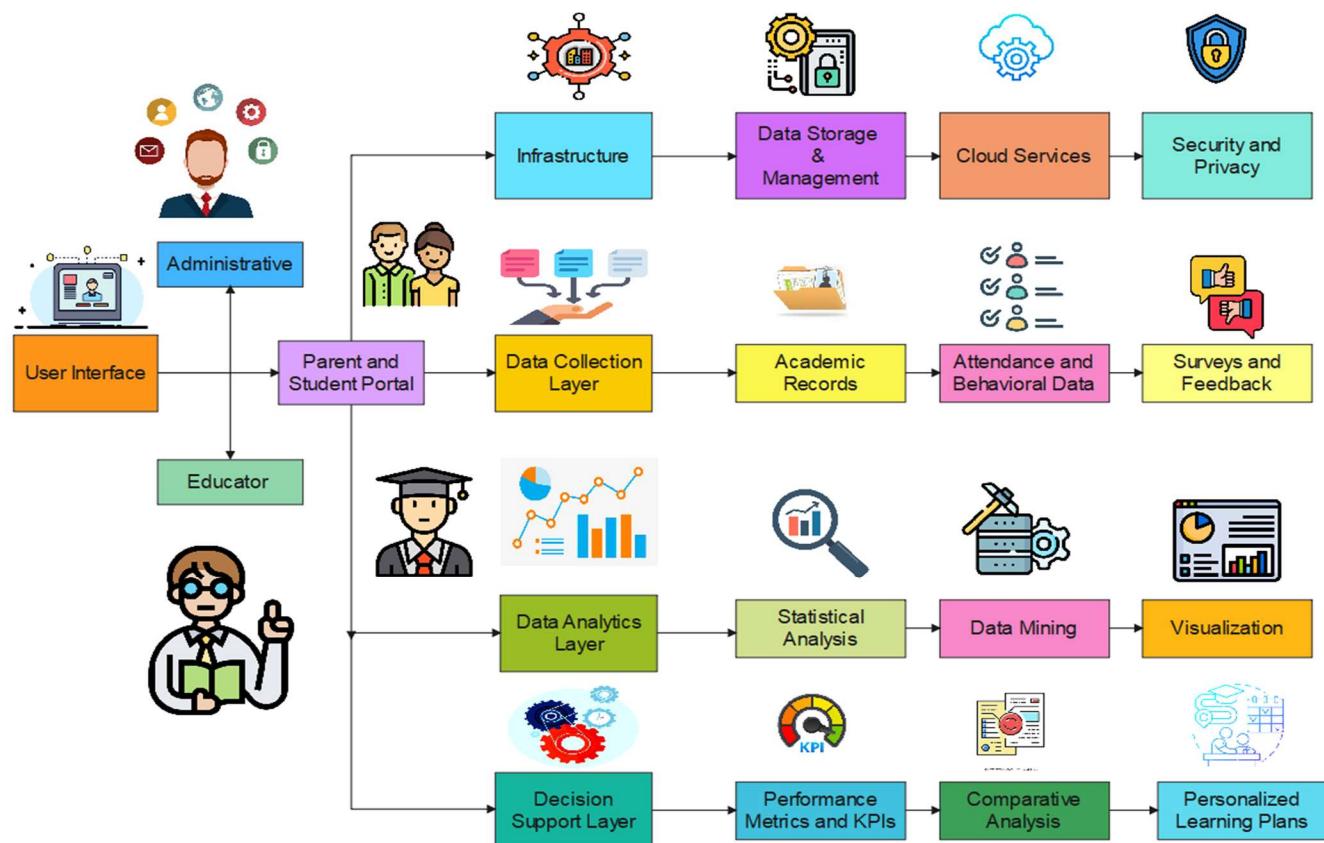
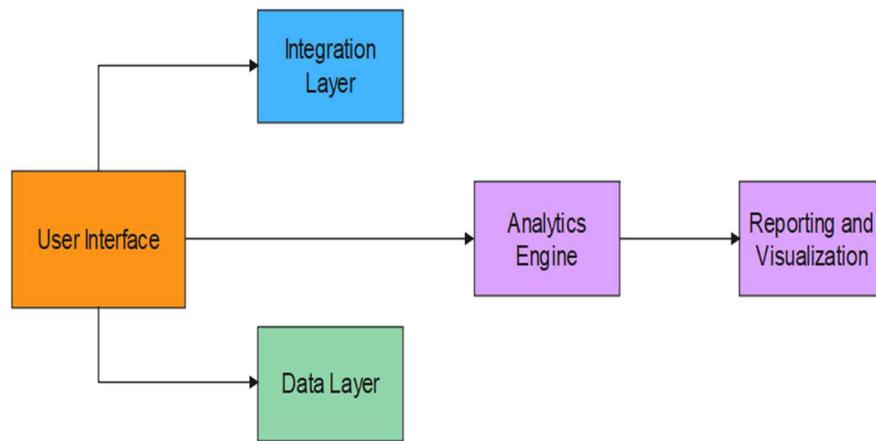


Figure : Architecture and data flow of the Student Performance Analysis

# PROJECT PLANNING & SCHEDULING

## 6.1 Technical Architecture

The Deliverable shall include the architectural diagram as below and the information as per the Table1 & Table 2.



### Guidelines:

1. **User Interfaces:** Design intuitive and user-friendly interfaces for easy system interaction.
2. **Integration Layer:** Develop robust APIs for seamless integration with external systems.
3. **Data Layer:** Implement a scalable and secure data storage solution.
4. **Analytics Engine:** Utilize advanced analytics techniques to derive actionable insights from data.
5. **Reporting and Visualization:** Create comprehensive and visually appealing reports for effective data presentation.

## 6.2 Sprint Planning & Estimation

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint 1	User Authentication	USN - 1	Allow users to register with email and password	5	High	Manikandan
Sprint 1	User Authentication	USN - 2	Implement social media login options	3	Medium	Surya
Sprint 1	Data Analysis	USN - 3	Create a data ingestion module	8	High	Manjusri
Sprint 2	Reporting	USN - 4	Generate basic user activity reports	5	High	Surya
Sprint 2	Reporting	USN - 5	Implement custom report generation	8	High	Rajesh
Sprint 2	User Authentication	USN - 6	Reset password functionality	3	Low	Manjusri

### 6.3 Sprint Delivery Schedule

Sprint	Total Story Points	Duration (in weeks)	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint 1	20	2	Aug 5, 2023	Aug 18, 2023	15	Aug 18, 2023
Sprint 2	15	2	September 19, 2023	October 2, 2023	14	October 2, 2023
Sprint 3	18	3	November 3, 2023	November 23, 2023	16	November 23, 2023
Sprint 4	22	3	December 24, 2023	December 13, 2023	21	December 13, 2023

## CODING & SOLUTIONING

### 7.1 Feature 1 - Overall Score Calculation

One useful feature to add is calculating the overall score for each student. The overall score can be obtained by summing up the math score, reading score, and writing score. This feature provides a consolidated measure of the student's performance across all subjects.

You can implement this feature using Python and the panda's library.

**Here's an example code,**

```
import pandas as pd

# Read the dataset
data = pd.read_csv('StudentsPerformance.csv')

# Calculate overall score
data['overall_score'] = data['math score'] + data['reading score'] + data['writing score']

# Display the updated dataset with the overall score
print(data.head())
```

This code reads the dataset from the CSV file, adds a new column called 'overall\_score,' and calculates it by summing up the scores from math, reading, and writing. The updated dataset is then displayed, showing the added overall score column.

## 7.2 Feature 2 - Success Rate Analysis

Another important feature is analyzing the success rate of students based on a certain threshold or criteria.

**For Example**, you may want to determine the percentage of students who scored above a specific mark in a particular subject or overall.

To calculate the success rate for students scoring above a certain threshold (e.g., 70) in the math subject:

**Here's an example code,**

```
import pandas as pd

# Read the dataset
data = pd.read_csv('StudentsPerformance.csv')

# Set the threshold for success
threshold = 70

# Calculate success rate for math subject
total_students = len(data)
success_students = len(data[data['math score'] > threshold])
success_rate = (success_students / total_students) * 100

# Display the success rate
print(f"Success rate in math subject (> {threshold}): {success_rate:.2f}%")
```

This code reads the dataset and sets a threshold value for success (e.g., 70 marks). It then calculates the number of students who scored above the threshold in the math subject, calculates the success rate as a percentage, and displays the result.

### **7.3 Database Schema (if Applicable)**

If you decide to store the dataset in a database for further analysis or application development, you can define a database schema to structure the data.

Here's an example of a possible database schema for the student performance dataset:

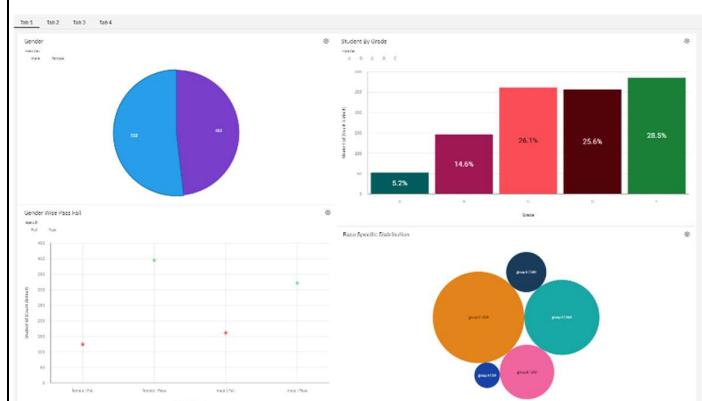
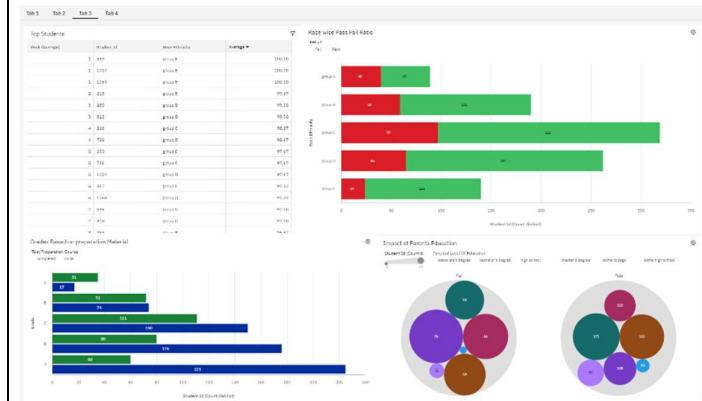
```
CREATE TABLE Students (
    student_id INT PRIMARY KEY,
    gender VARCHAR(10),
    race_ethnicity VARCHAR(50),
    parental_education VARCHAR(50),
    lunch VARCHAR(20),
    test_preparation_course VARCHAR(20),
    math_score INT,
    reading_score INT,
    writing_score INT
);
```

This schema creates a table called "Students" with columns corresponding to the dataset's attributes. The `student_id` column is set as the primary key to uniquely identify each student.

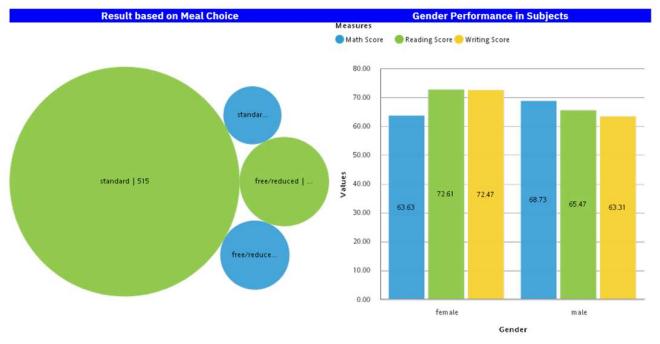
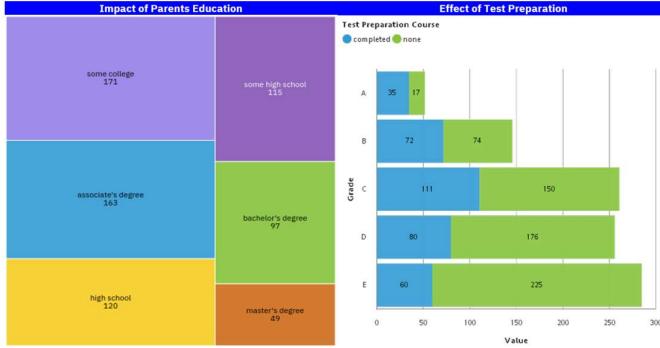
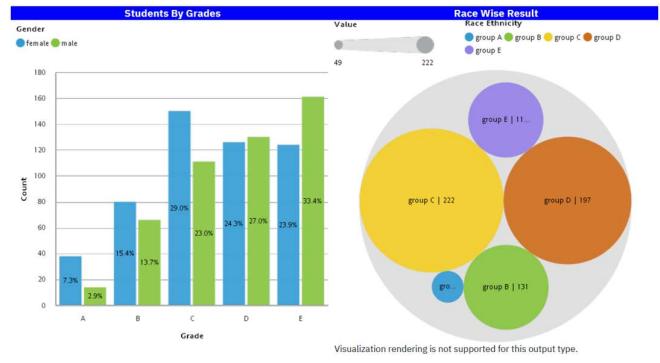
You can then load the dataset into the database table using appropriate SQL statements or utilize an ORM (Object-Relational Mapping) library in your preferred programming language to interact with the database.

# PERFORMANCE TESTING

## 8.1 Performance Metrics

S.No.	Parameter	Screenshot / Values
1.	Dashboard design	<p>No of Visualizations / Graphs -</p>  <p>The dashboard displays three tabs of data:</p> <ul style="list-style-type: none"> <li><b>Tab 1:</b> Student by Grade. A bar chart showing the percentage of students in each grade: 5.2% (K), 14.6% (1), 26.1% (2), 25.6% (3), and 28.5% (4).</li> <li><b>Tab 2:</b> Gender Wise Pass Ratio. A scatter plot showing the pass ratio for different gender groups: Male (High), Male (Med), Female (High), and Female (Med).</li> <li><b>Tab 3:</b> Grade Specific Distribution. A bubble chart showing the distribution of students across different grades.</li> </ul>  <p>The dashboard displays three tabs of data:</p> <ul style="list-style-type: none"> <li><b>Tab 1:</b> Parents Education Level. A treemap chart showing the distribution of parents' education levels: Primary School (20%), Secondary School (20%), Higher Secondary (30%), and Degree (30%).</li> <li><b>Tab 2:</b> Test Preparation Effectiveness. A stacked bar chart showing the effectiveness of test preparation: Moderate (22.94%) and High (77.05%).</li> <li><b>Tab 3:</b> Average Result of Maths and Physics. A grouped bar chart showing average results for Physics, Maths, and English across four categories: Male (High), Male (Med), Female (High), and Female (Med).</li> </ul>  <p>The dashboard displays three tabs of data:</p> <ul style="list-style-type: none"> <li><b>Tab 1:</b> Log Students. A table showing student log details with columns: student_id, name, and average.</li> <li><b>Tab 2:</b> Grade Wise Pass Ratio. A stacked bar chart showing the pass ratio for different grade groups: groupA (15.00), groupB (15.00), groupC (15.00), groupD (15.00), groupE (15.00), groupF (15.00), groupG (15.00), groupH (15.00), groupI (15.00), and groupJ (15.00).</li> <li><b>Tab 3:</b> Effect of Parents Education. A bubble chart showing the effect of parents' education level on student scores: primary school (100), secondary school (100), higher secondary (100), and degree (100).</li> </ul>

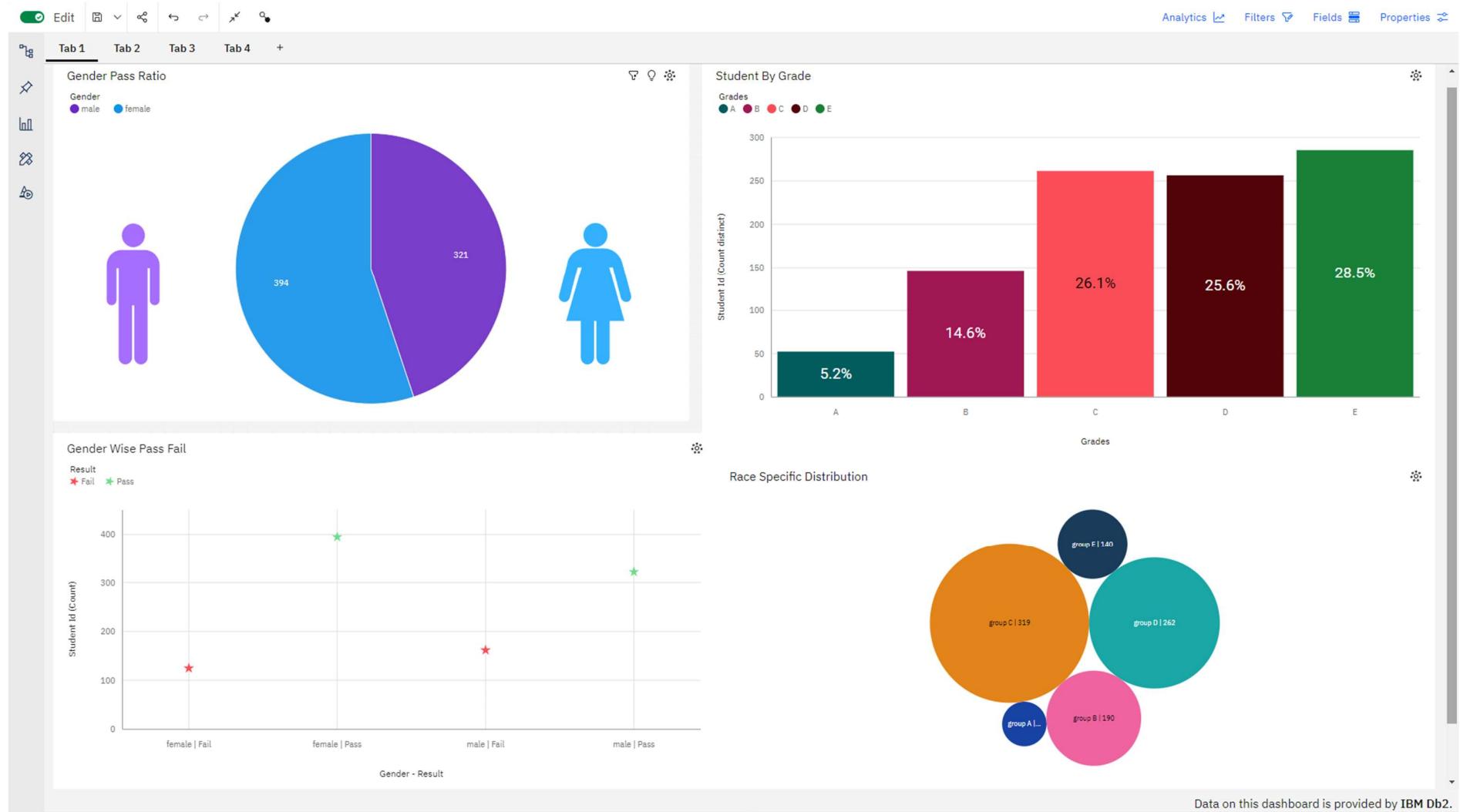
		<p><b>Student Result based on Lunch Choice</b></p> <table border="1"> <thead> <tr> <th>Lunch</th> <th>Result</th> <th>Count</th> </tr> </thead> <tbody> <tr> <td>free/reduced</td> <td>Fail</td> <td>155</td> </tr> <tr> <td>free/reduced</td> <td>Pass</td> <td>200</td> </tr> <tr> <td>standard</td> <td>Fail</td> <td>130</td> </tr> <tr> <td>standard</td> <td>Pass</td> <td>515</td> </tr> </tbody> </table>	Lunch	Result	Count	free/reduced	Fail	155	free/reduced	Pass	200	standard	Fail	130	standard	Pass	515	
Lunch	Result	Count																
free/reduced	Fail	155																
free/reduced	Pass	200																
standard	Fail	130																
standard	Pass	515																
2.	Data Responsiveness	The data is responsive, meaning that it can be updated and displayed in real time.																
3.	Amount Data to Rendered (DB2 Metrics)	The amount of data rendered is measured using DB2 metrics, and it is confirmed that the data is rendered from DB2 using IBM Cloud.																
4.	Utilization of Data Filters	Data filters were used in the data, which means that users can filter the data to view specific subsets of interest.																
5.	Effective User Story	The effective user story is "No of Scenes Added - 10", which means that users can track the number of scenes that have been added to the project.																
6.	Descriptive Reports	<p>No of Visualizations / Graphs</p> <p><b>Gender Distribution</b></p> <table border="1"> <thead> <tr> <th>Gender</th> <th>Percentage</th> </tr> </thead> <tbody> <tr> <td>female</td> <td>48.2%</td> </tr> <tr> <td>male</td> <td>51.8%</td> </tr> </tbody> </table> <p><b>Gender Wise Result</b></p> <table border="1"> <thead> <tr> <th>Gender - Result</th> <th>Count</th> </tr> </thead> <tbody> <tr> <td>female   Fail</td> <td>321</td> </tr> <tr> <td>male   Fail</td> <td>161</td> </tr> <tr> <td>female   Pass</td> <td>394</td> </tr> <tr> <td>male   Pass</td> <td>124</td> </tr> </tbody> </table>	Gender	Percentage	female	48.2%	male	51.8%	Gender - Result	Count	female   Fail	321	male   Fail	161	female   Pass	394	male   Pass	124
Gender	Percentage																	
female	48.2%																	
male	51.8%																	
Gender - Result	Count																	
female   Fail	321																	
male   Fail	161																	
female   Pass	394																	
male   Pass	124																	

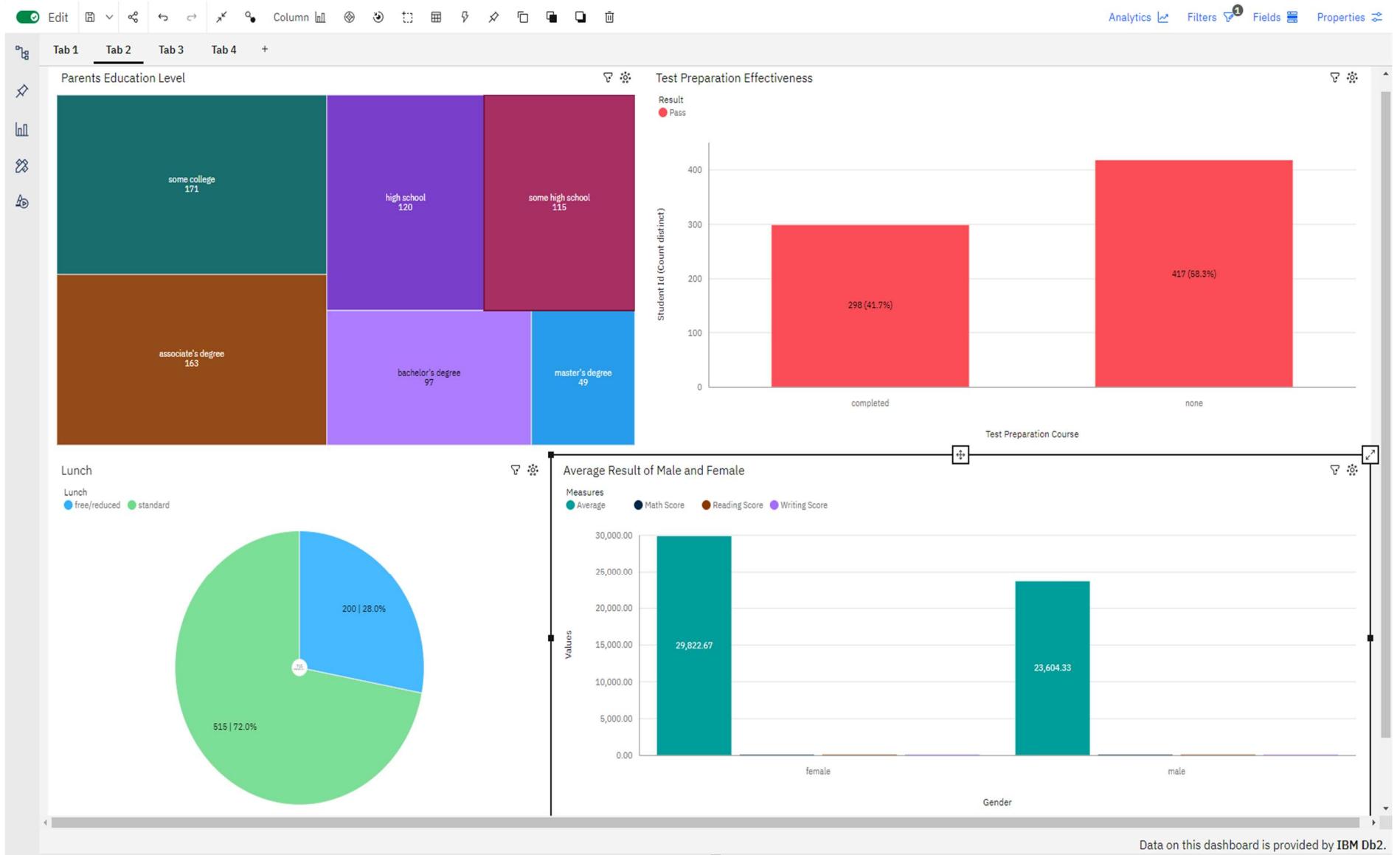


# RESULTS

## 9.1 Output Screenshots

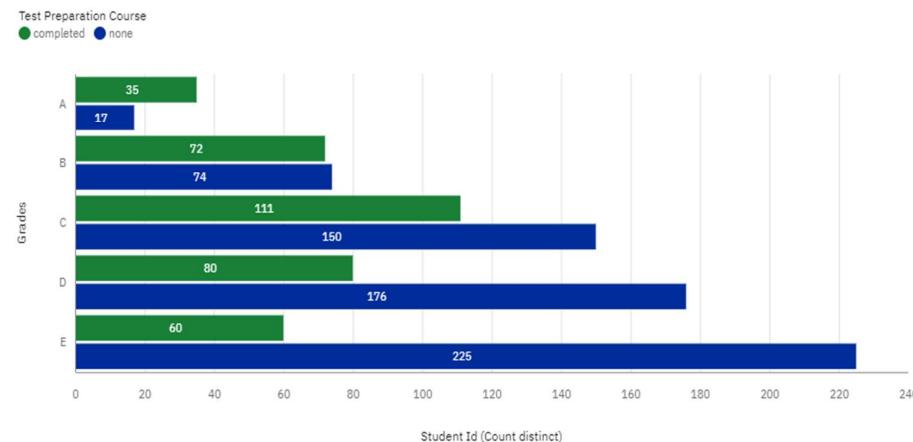
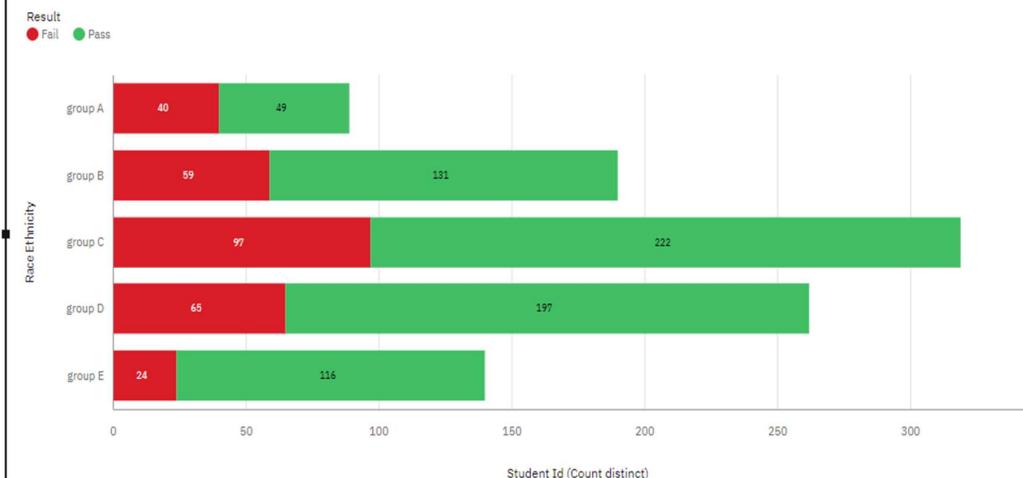
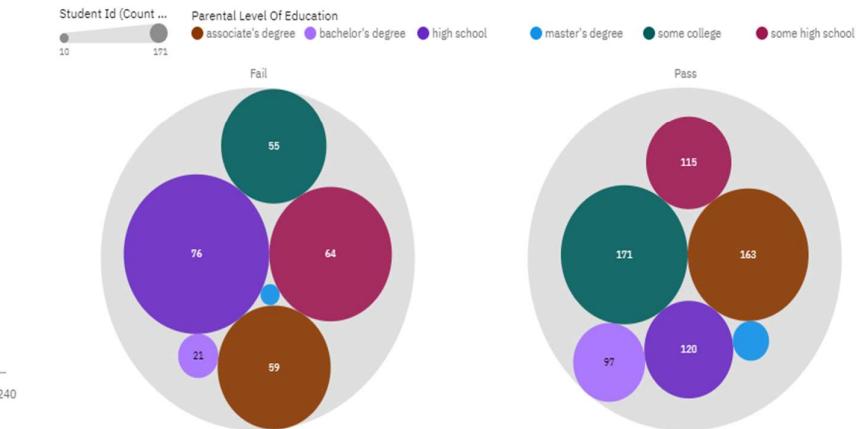
### 1. Dashboard design



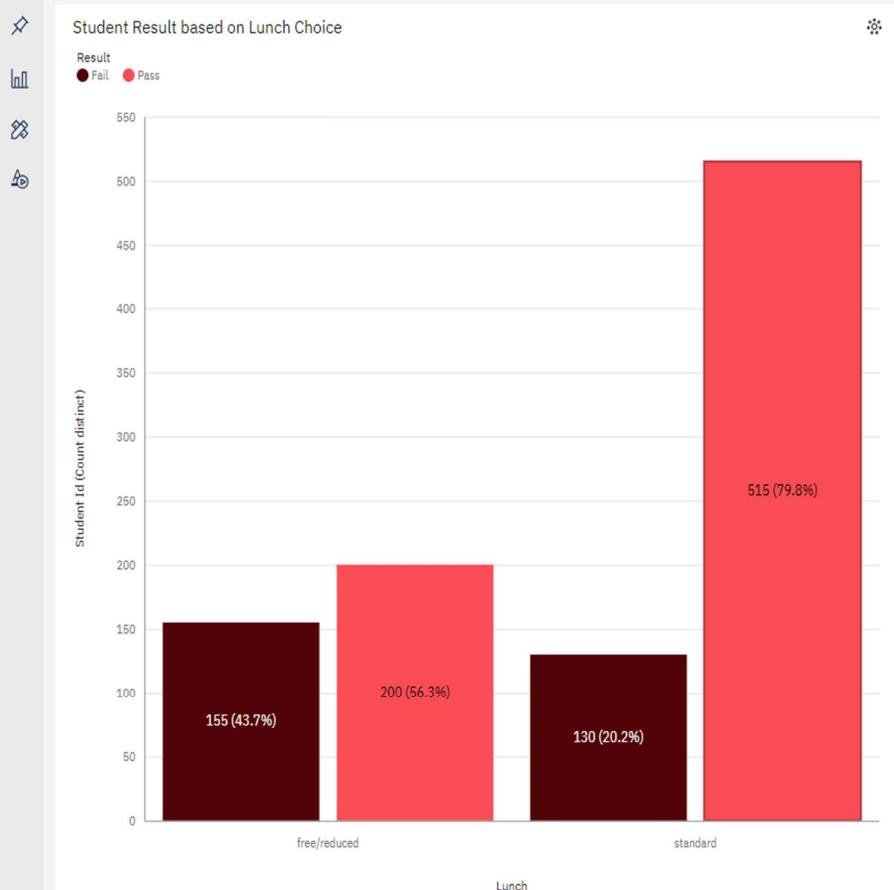


Tab 1 Tab 2 Tab 3 Tab 4 +
**Top Students**

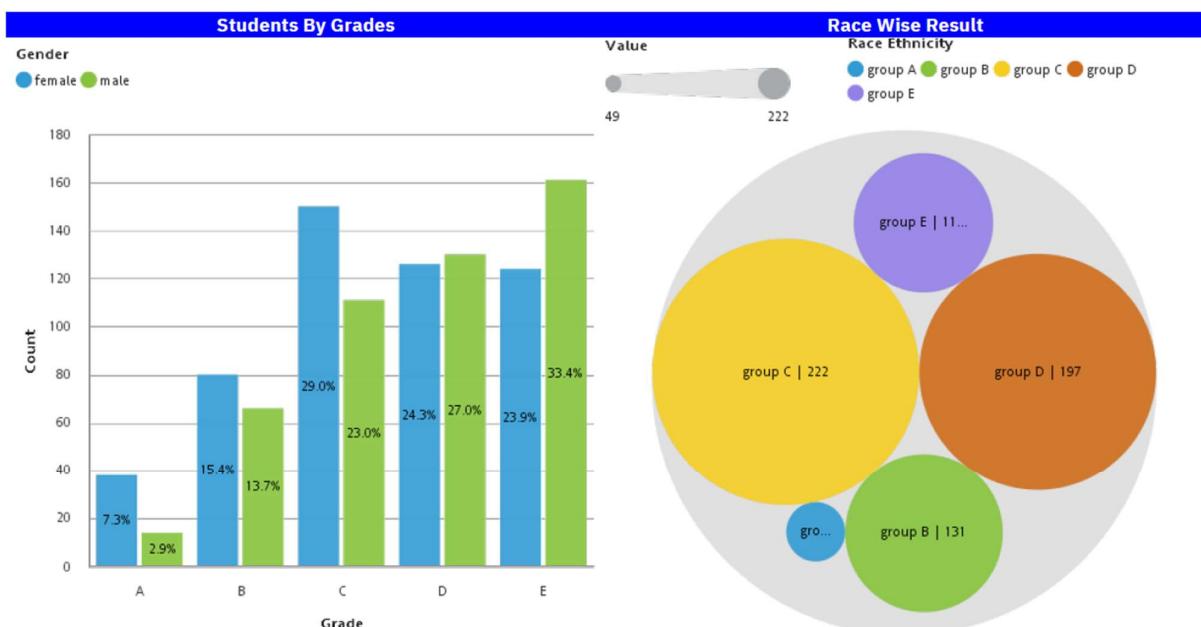
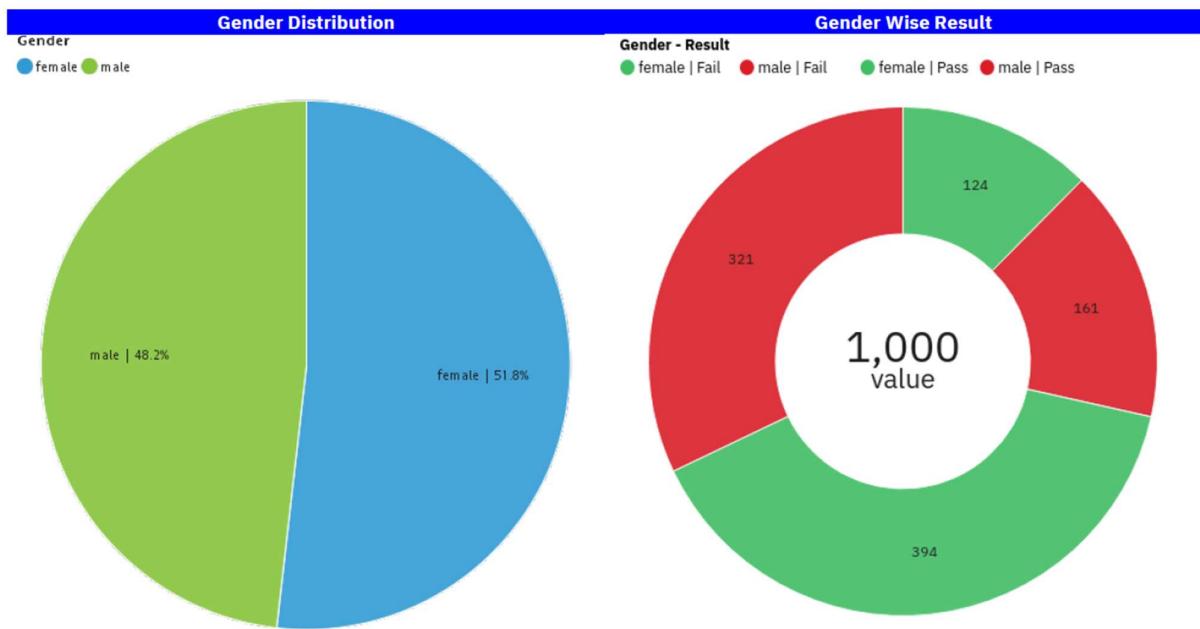
Student Id	Race Ethnicity	Average
101	group B	72.67
Summary		72.67
102	group C	82.33
Summary		82.33
103	group B	92.67
Summary		92.67
104	group A	49.33
Summary		49.33
105	group C	76.33
Summary		76.33
106	group B	77.33
Summary		77.33
107	group B	91.67

**Grades Based on preparation Material**

**Race wise Pass Fail Ratio**

**Impact of Parents Education**


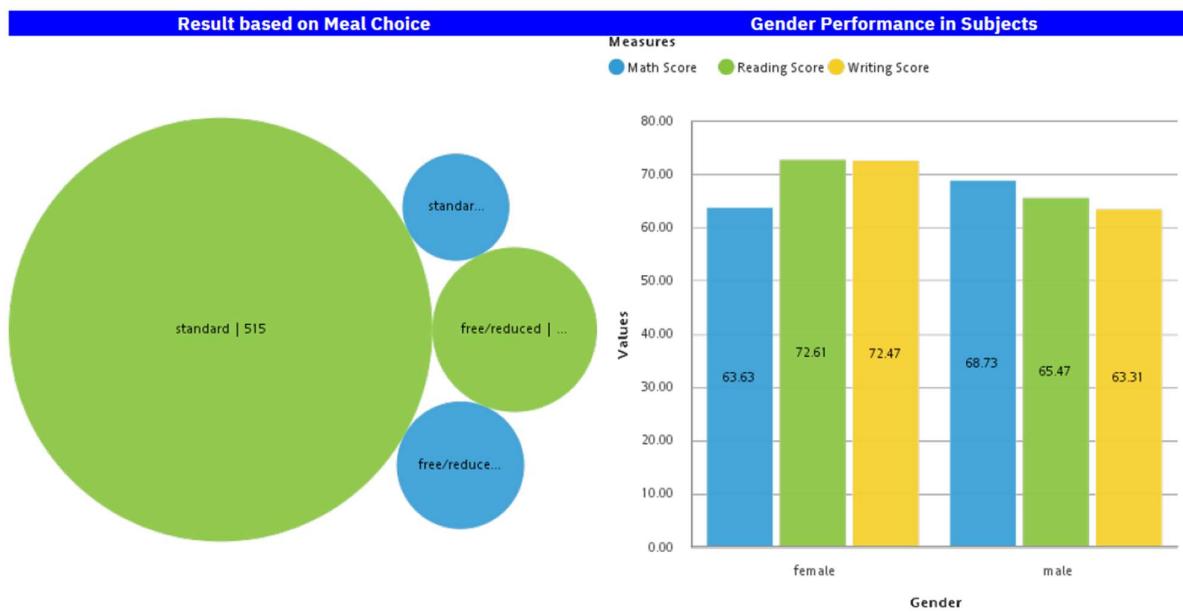
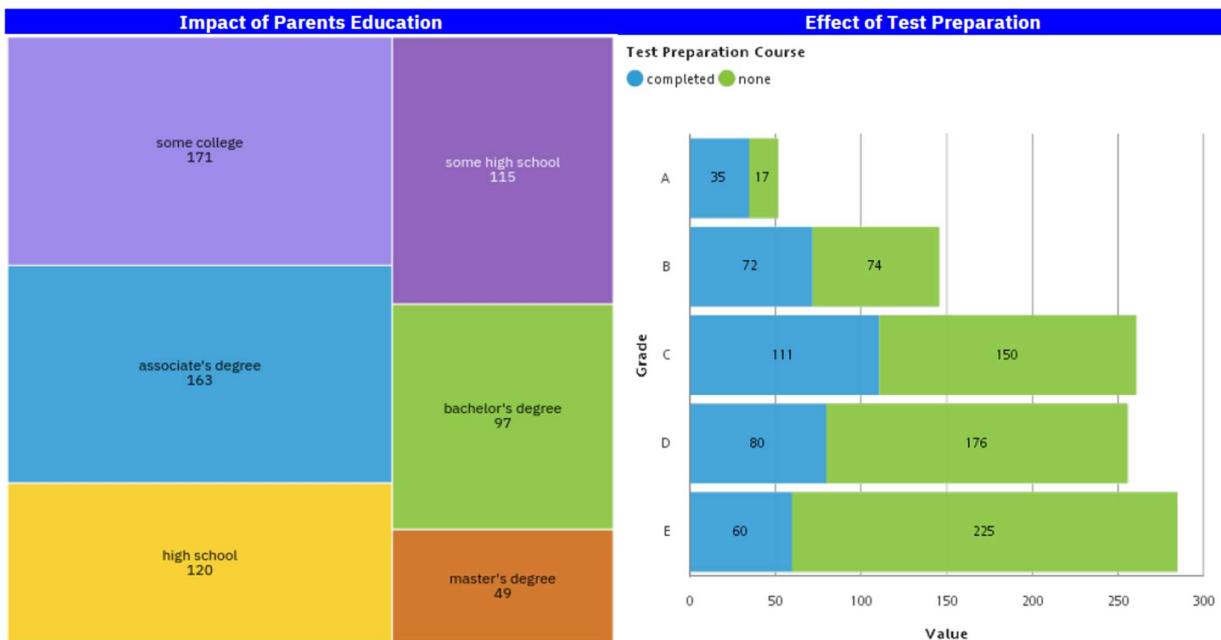
Tab 1 Tab 2 Tab 3 Tab 4 +



## 2. Descriptive Reports



Visualization rendering is not supported for this output type.



### 3. Story

A vibrant illustration set against a white background. In the center is a large laptop screen displaying a user interface with a purple header bar. On the screen, a man in an orange jacket and blue pants stands next to a purple graduation cap and three stars. A large yellow trophy is visible on the right side of the screen. A purple rocket ship is launching from the top left corner of the screen, leaving a trail of purple smoke. The background features abstract purple and blue shapes resembling leaves or clouds. In the top right corner of the slide, there is a grey box labeled "Analysis". At the bottom of the slide, the text "Unleash the True Potential of Students," is displayed in a large, bold, purple font.

Press Esc to exit full screen

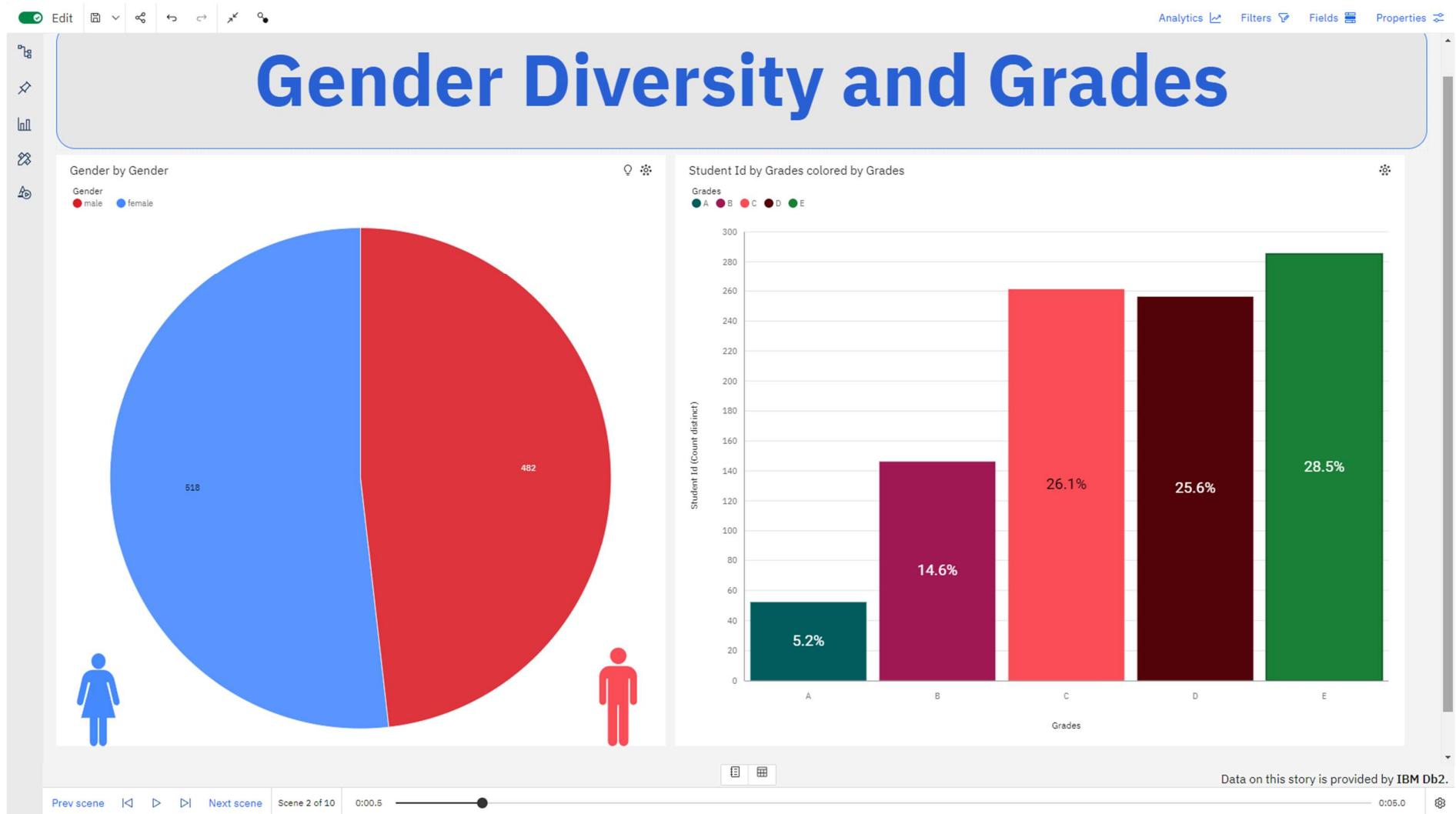
Analytics Filters

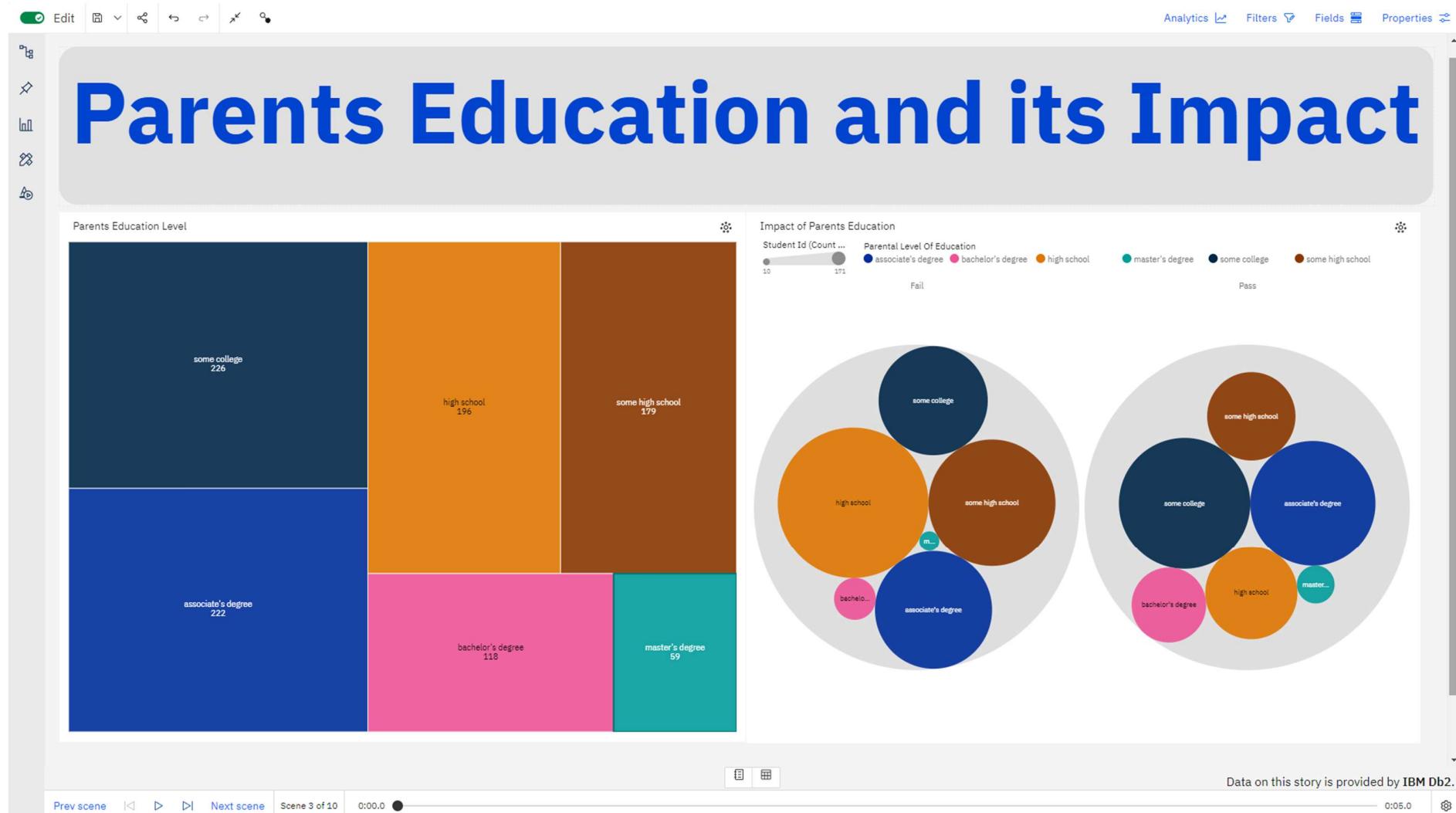
Analysis

Unleash the True Potential of Students,

Prev scene ⏪ ⏩ Next scene Scene 1 of 10 0:00.0 0:05.0

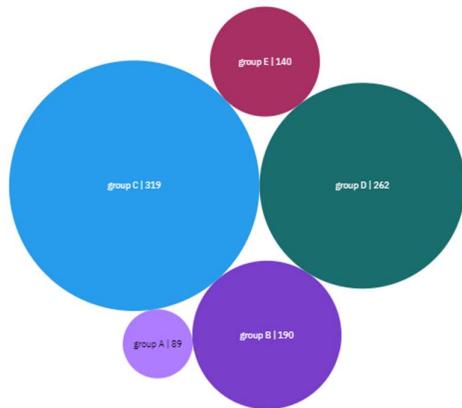
Data on this story is provided by IBM Db2.



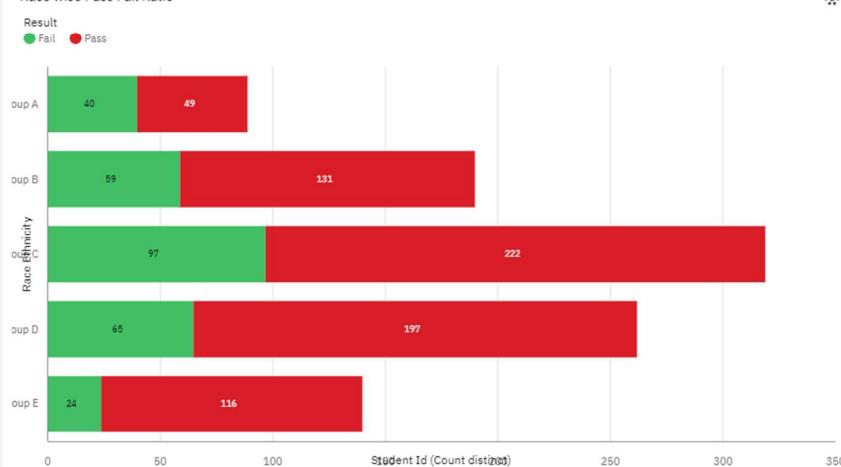


# Race Diversity

Race Specific Distribution

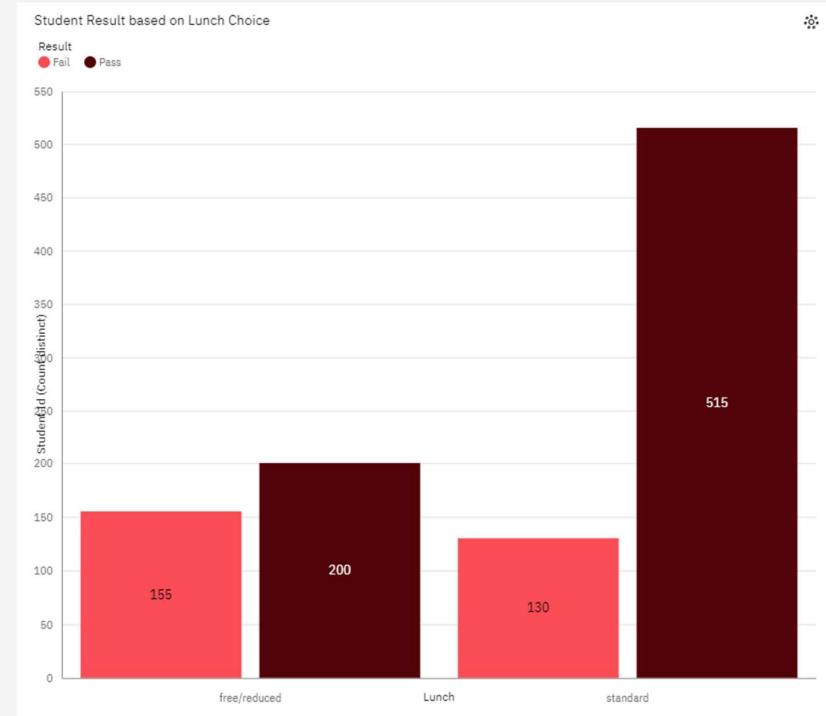
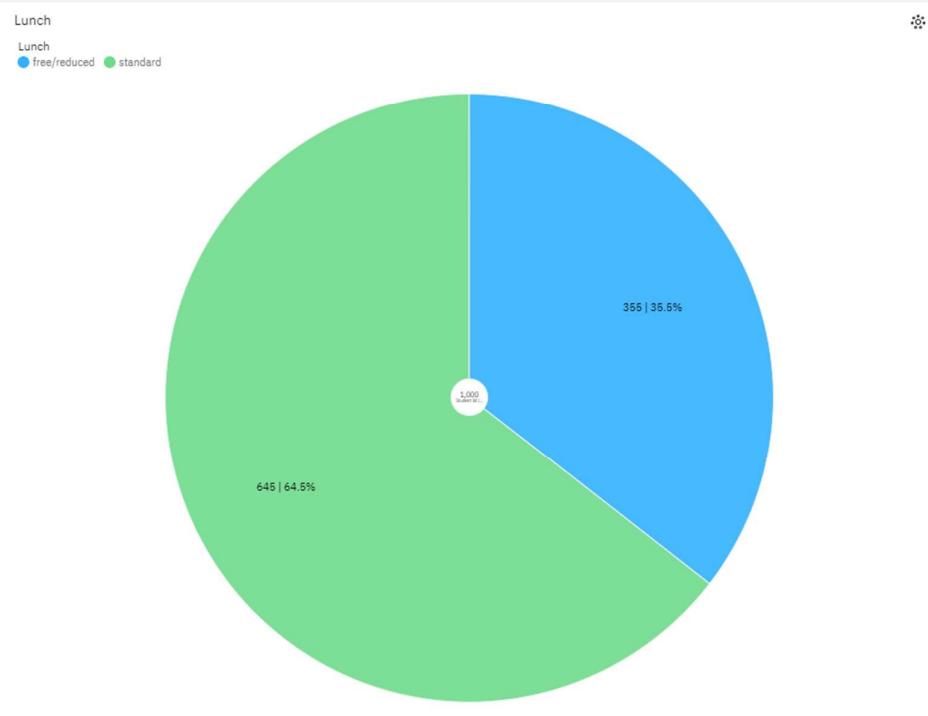


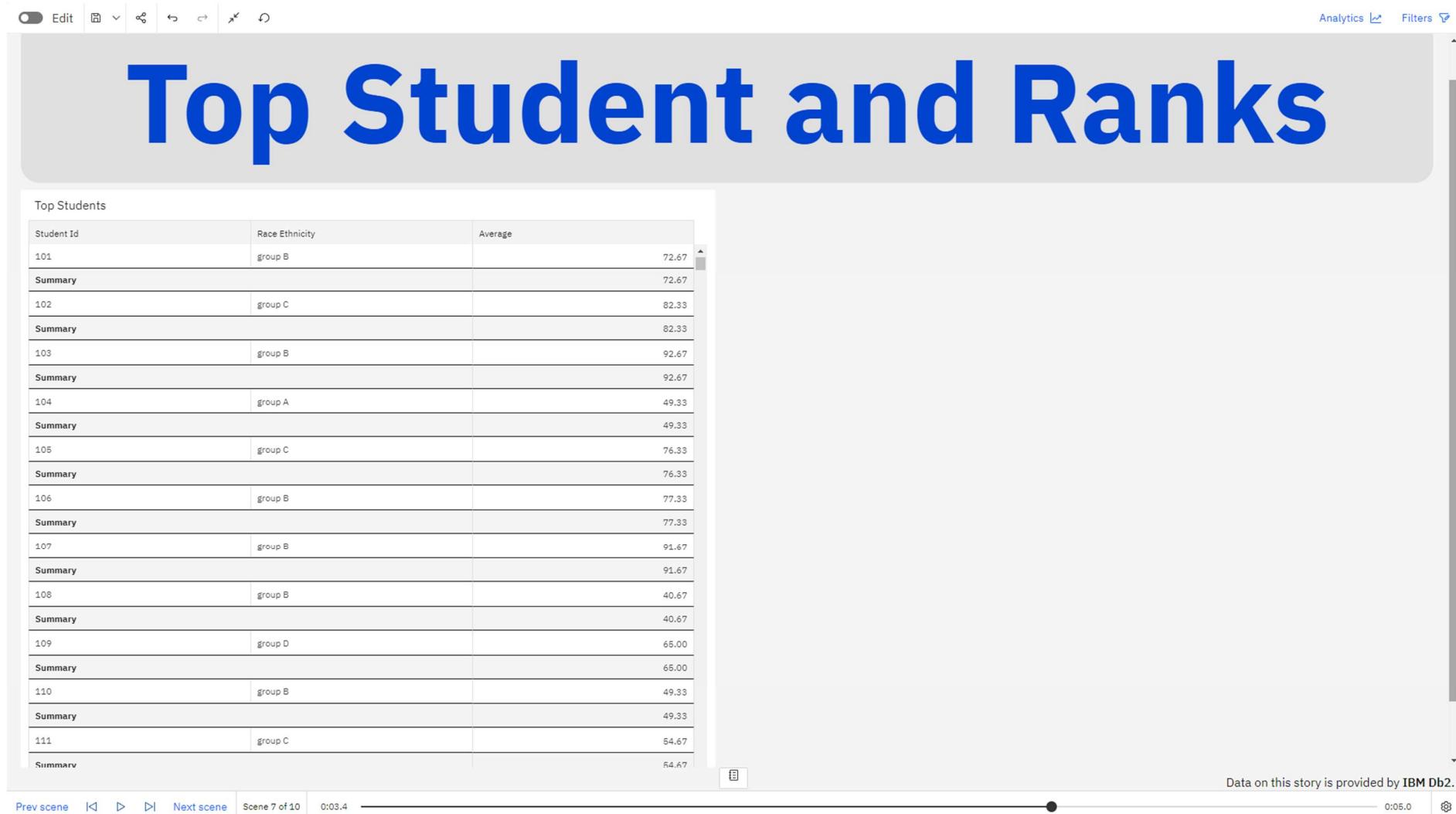
Race wise Pass Fail Ratio

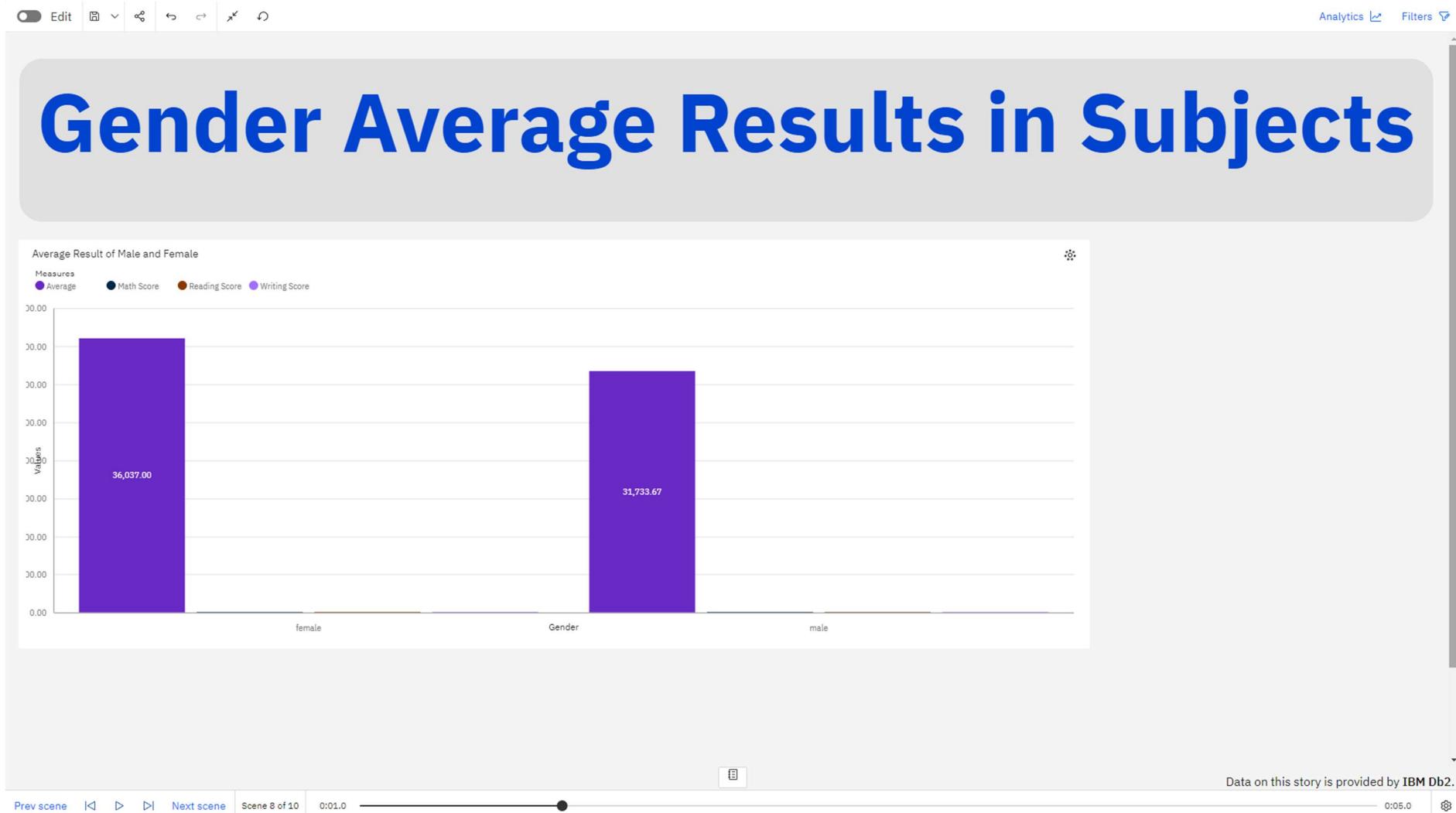




# Meal Choice







Edit |

Analytics Filters

Press to exit full screen

# Gender Wise Pass Fail Counts

Gender Wise Pass Fail

Result  
★ Fail ★ Pass



Prev scene Next scene Scene 9 of 10 0:02.0



Data on this story is provided by IBM Db2.

0:05.0



Edit |

Analytics Filters

# THANK YOU

Prev scene Next scene

Scene 10 of 10

0:01.4



Data on this story is provided by IBM Db2.

0:05.0



## **ADVANTAGES & DISADVANTAGES**

### **ADVANTAGES**

#### **Valuable Insights**

Analyzing the student performance data helps us understand why some students do well and others struggle. We can identify what factors, like parents' education and test preparation, contribute to student success.

#### **Informed Decision Making**

By using data to make decisions, we can choose the best ways to help students succeed. We can make changes to teaching methods, curriculum, or resources based on what the data tells us.

#### **Spotting Patterns**

Looking at the data helps us find patterns and trends. We can see if there are connections between different things and figure out how to improve student performance.

#### **Improving Teaching**

By analyzing student work and data, teachers can see where students are having trouble. This helps them change how they teach to better support students' learning.

## **DISADVANTAGES**

### **Limited Information**

The dataset we have might not include everything that affects student performance. Things like a student's home life or personal situation could be important but aren't in the dataset.

### **Biased Data**

The dataset could have biases, like not representing all types of students or schools. Biased data can lead to biased conclusions, so we need to be aware of that.

### **Missing Context**

The dataset alone might not give us all the information we need. We might need more details from surveys or interviews to fully understand why students perform the way they do.

### **Data Quality Issues**

The data might not be completely accurate or could have missing information. This can affect the analysis and make it hard to draw reliable conclusions.

### **Ethical Considerations**

When using student data, it's important to respect privacy and follow ethical guidelines. We need to protect students' information and make sure we're using the data responsibly.

## **CONCLUSION**

In conclusion, analyzing student performance data and incorporating features into the project provides valuable insights into the factors influencing student success rates. By examining factors such as parental education and test preparation, we can better understand how these attributes impact student performance. This analysis enables data-driven decision making, allowing for targeted interventions, curriculum improvements, and resource allocation to enhance student outcomes.

The project helps identify patterns and trends in student performance, aiding in the identification of strategies and interventions to improve academic achievement. Ongoing assessment of student learning supports continuous improvement in teaching practices and course content, ultimately enhancing the quality of education.

However, it is important to acknowledge the limitations of the dataset, such as potential biases, limited scope, and the need for additional contextual information. Data quality and completeness should also be carefully considered to ensure accurate and reliable analysis.

Ethical considerations regarding data privacy and protection must be addressed when dealing with student data. Respecting confidentiality and anonymity is paramount to ensure the project's integrity and safeguard student privacy.

By recognizing the advantages and disadvantages of the project, we can leverage the insights gained to drive positive change in the education system, supporting students' success and overall academic achievement.

## FUTURE SCOPE

- **Long-term Analysis**

Analyze data over multiple years to track student progress and identify trends.

- **Predictive Modeling**

Use machine learning to predict student success and provide targeted interventions.

- **Include External Data**

Incorporate additional data sources to gain a comprehensive understanding of student performance.

- **Compare Performance**

Compare performance across schools or regions to identify successful practices.

- **Visualize and Report**

Create visualizations and dashboards for easier understanding and decision making.

- **Combine Qualitative Research**

Use surveys and interviews to gain deeper insights into student performance.

- **Evaluate Interventions**

Implement and assess the effectiveness of interventions to improve outcomes.

- **Promote Collaboration**

Foster collaboration among educators, researchers, and policymakers for knowledge sharing and improvement.

The future scope revolves around leveraging data, technology, and collaboration to enhance student outcomes and advance the education sector.

## **APPENDIX**

- Source Code

<https://github.com/manime666/Student-Performance-Analysis--NM2023TMID05094/blob/main/Final%20Deliverables/HTML%20Code/Web%20Integration.rar>

- GitHub

<https://github.com/manime666/Student-Performance-Analysis--NM2023TMID05094>

- Project video demo link

<https://drive.google.com/file/d/1XVCe-K4QCHbWmdx3YkYx-oh61cfm26Rb/view?usp=sharing>