

GAYATHRI – MINI PROJECT REPORT

Name: GAYATHRI

Batch Code: TN_DA_ANB11

Contact No: 9344388514

Email ID: gayathri562001@gmail.com

Project Title: Student Performance Analysis

Project Domain: Education

Submission Date: 08 JAN 2025

Mentor: Kumaran M

Raw Dataset Link: [Mini Project/Gayathri Mini Project Raw Dataset.csv at main · gayathri520/Mini Project](#)

Cleaned Dataset Link : [Mini Project/Gayathri Miniproject Cleaned Dataset.xlsx at main · gayathri520/Mini Project](#)

Purpose of the project:

Purpose of the project is to analyze academic outcome, Identify trends in student scores, test preparation, grade across subjects. Provide teachers, administrators with evidence to allocate resources and improve teaching strategies. Compare performance across groups to ensure equal opportunities. Enable visualization through charts and dashboards and report so educator can easily interpret results. Track progress over time to measure the impact of educational programs and initiatives. Purpose is to convert raw data in valuable insights for accuracy.

Objectives of project:

- Evaluate overall academic performance and measure average score across subjects, Groups to identify general trends
- Compare the performance by gender, group subject to highlight strength and weakness.
- Analyze key performance such as percentage, pass and fail of the students in respective subjects with related to groups they belong and ranking distribution for actionable insights.
- Support data-driven decision-making. Visualize insights through charts and dashboards for easy interpretation.

Scope of the project:

- **Data Coverage:** Include student scores, test preparation, gender, groups, and performance rating.
- **Metric considered:** Key indicators such as percentage, Pass and fail of students, total marks, grades and ranking distribution will be calculated.
- **Tools and Techniques:** Data cleaning and imputation in Excel by using power query and some formulas used in excel for data cleaning like AVERAGE, AVERAGEIFS, COUNTIF, COUNTIFS, RANK.EQ, IF, ISBLANK.
- **Reporting format:** Finding will be presented in a structured report with tables and charts for visualization through charts and dashboard in Power BI.

Data cleaning on dataset :

Data cleaning was carried out to ensure the accuracy and reliability of the student dataset. Missing values were imputed using averages and group by function, duplicates were removed, and categorical inconsistencies were standardized. Error handling and validation rules were applied to maintain data integrity. As a result, the dataset was prepared for meaningful analysis and visualization.

1. Converting Range to Table:

The conversion of a data range into a structured table ensures consistency, easier referencing, and improved data management. Tables allow for automatic filtering, sorting, and structured formulas, making analysis more efficient.

Select Range → Insert Table [Insert → Table] → Confirm Headers
 [Checked “My table has headers”] → Apply Table Design.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Roll No	Gender	Race Ethnicity	Parental Level Of Education	Lunch	Test Prep	Math	Reading	Writing	Science	Total	Grade	Eligibility	Percentage	Rank	Number	Number of			
2	Std-8035	Female	Group D	Master's Degree	1	1	100	98	99	86	383 A	Eligible	95.75	1	1	0				
3	Std-8779	Female	Group C	Some High School	1	0	87	96	100	100	383 A	Eligible	95.75	1	2	0				
4	Std-7555	Female	Group D	Associate's Degree	0	0	89	100	99	94	382 A	Eligible	95.5	3	1	0				
5	Std-5908	Female	Group D	High School	1	0	79	79	100	99	378 A	Eligible	94.5	4	2	0				
6	Std-2709	Male	Group B	Some College	1	1	95	98	93	91	377 A	Eligible	94.25	5	0	0				
7	Std-4510	Male	Group A	Associate's Degree	1	0	98	99	88	92	377 A	Eligible	94.25	5	0	0				
8	Std-9182	Female	Group B	Bachelor's Degree	1	0	100	95	96	86	377 A	Eligible	94.25	5	1	0				
9	Std-1208	Male	Group C	High School	1	0	99	93	84	100	376 A	Eligible	94	8	1	0				
10	Std-2297	Male	Group E	Associate's Degree	0	1	83	100	92	100	375 A	Eligible	93.75	9	2	0				
11	Std-3246	Male	Group C	Associate's Degree	0	1	87	100	89	98	374 A	Eligible	93.5	10	1	0				
12	Std-8445	Female	Group C	Bachelor's Degree	1	0	100	84	100	90	374 A	Eligible	93.5	10	2	0				
13	Std-2821	Male	Group B	Associate's Degree	1	0	82	95	99	97	373 A	Eligible	93.25	12	0	0				
14	Std-3047	Male	Group E	Some College	1	1	76	99	98	100	373 A	Eligible	93.25	12	1	0				
15	Std-6843	Female	Group D	Some High School	1	0	87	98	95	93	373 A	Eligible	93.25	12	0	0				
16	Std-303	Male	Group D	Bachelor's Degree	1	0	75	96	100	100	371 A	Eligible	92.75	15	2	0				
17	Std-5047	Female	Group B	Some College	1	1	87	94	99	90	370 A	Eligible	92.5	16	0	0				
18	Std-66	Male	Group D	Associate's Degree	1	0	81	88	100	100	369 A	Eligible	92.25	17	2	0				
19	Std-6537	Female	Group D	Some College	1	1	87	97	84	100	368 A	Eligible	92	18	1	0				
20	Std-8964	Female	Group B	Associate's Degree	1	1	69	100	100	99	368 A	Eligible	92	18	2	0				
21	Std-8686	Female	Group E	Bachelor's Degree	1	0	89	100	86	93	368 A	Eligible	92	18	1	0				
22	Std-9214	Female	Group B	Some High School	1	1	100	76	100	92	368 A	Eligible	92	18	2	0				
23	Std-996	Male	Group B	Master's Degree	0	0	88	99	94	86	367 A	Eligible	91.75	22	0	0				
24	Std-2833	Male	Group D	Some College	1	0	100	95	72	100	367 A	Eligible	91.75	22	2	0				
25	Std-4190	Male	Group D	Some College	1	1	98	100	100	69	367 A	Eligible	91.75	22	2	0				
26	Std-4663	Male	Group B	Associate's Degree	1	0	75	100	100	92	367 A	Eligible	91.75	22	2	0				
27	Std-1465	Male	Group D	Some College	1	0	99	92	92	83	366 A	Eligible	91.5	26	0	0				

For data cleaning I uploaded the uncleaned data into power query for data cleaning.

In given dataset headers are in lower case so used “Capitalize Each Word” in power query

Data tab → From Table/Range.

2. Content given in the data set was also in lower case so used “Capitalize Each Word”

Before:

roll_no	gender	race_ethnicity
std-01	male	group D
std-02	male	group B
std-03	male	group C
std-04	male	group D
std-05	male	group C
std-06	male	group B
std-07		group C
std-08	male	group B
std-09	male	group C
std-10	male	group C

After:

Roll No	Gender	Race Ethinicity
Std-8035	Female	Group D
Std-8779	Female	Group C
Std-7555	Female	Group D
Std-9508	Female	Group D
Std-2709	Male	Group B
Std-4510	Male	Group A
Std-9182	Female	Group B
Std-1208	Male	Group C
Std-2297	Male	Group E
Std-3246	Male	Group C

3.Gender column:

Problem Identified: The dataset contained some inconsistencies.

Action Taken: By using filtering found there were inconsistencies like “Male” was written as “Boy”, “\Tmale”, “Girl” – “female”

gender	Gender
Boy	Male
Boy	Male
Boy	Male
Male	Male
Boy	Male
Boy	Male
Boy	Male

=> Male

gender	Gender
Girl	Female

=> Female

4.Race Ethnicity:

Problem Identified: The given dataset contains inconsistencies.

Action Taken: By using filtering option found “Group A, Group B, Group C, Group D, Group E” was written as “A,B,C,D,E” and Group C\N.

Race Ethnicity
group D
group B
group C
group D
group C
group B
group C
group B
group C
group C
group E
group A
group E
A

=> Group E

Missing Values:

1.Race Ethnicity:

Problem Identified: The dataset contains missing values which may cause inconsistency and lead to inaccuracy.

Steps Followed: To fill missing values in dataset used “Group by” function in power query. “null” is replaced with “Group D”.

race_ethnicity	Race Ethnicity
	Group D

2. Parental level of Education:

Problem Identified: The data were missing in dataset which may lead to inaccuracy while giving insights.

Steps followed: By using filtering option, founded the “null” values and by performing “Groupby” function identified missing values and filled with respective result.

“Null” is replaced with “Some college”.

parental_level_of_education	Parental Level Of Education
	Some College
=>	Some College

Lunch:

Problem Identified: The dataset contains many missing values in lunch which had been indicated in numerical for as "0"-who need lunch and "1" as those who don't want lunch.

Step involved: By filtering the lunch column able to find the "null" value and by performing "Groupby" function we can find the value which can be replace "null"

By groupby function "null" is replaced with "1".

Test Preparation of course:

Problem Identified: By analyzing dataset there were missing values which lead to inaccuracy while performing calculation and provides inaccurate insights.

Steps Taken: By filtering the test preparation of course able to find the null values in numerical form as “0” indicates students doesn’t finished his course

And “1” indicates student finished their preparation.

Null is replaced by “0” by group by function.

Performing Calculation for Valuable Insights:

1. Eligibility:

Created a new column named “Eligibility” to check the students whether students are eligible for higher studies. For identify the students performed some calculation to check the students eligibility criteria.

Formula Used: =IF(Q2>200,"Eligible","Not Eligible")

2. Percentage:

Created a new column “Percentage” to calculate the percentage of each students.

Formula Used: =Total Score/4

3. Rank:

Created a new column “Rank” to give the rank for each student and know their positions on classroom.

Formula Used: =RANK.EQ(Q2,\$Q\$2:\$Q\$10001,0)

4. Number of centum :

Created a new column “Number of centum” to know how many centum got by each student in each subject .

Formula Used: =COUNTIF(Table3[@[Math]:[Science]],"=100")

5. Number of students “<=35”:

Created a new column “Number of Students <=35”, for easy interpretation on failed students and focus on students for better marks on upcoming examination.

Formula Used: =COUNTIF(Table3[@[Math]:[Science]], "<=35")

Created a new excel sheet:

In MS Excel created a new sheet named “ Student Marksheets” for clear understanding

By copying the Roll No from cleaned dataset and pasted in new sheet where Roll No is unique value.

By using “VLOOKUP” function data from cleaned sheet is transferred to new sheet named “Student Marksheets”.

Formula Used: =VLOOKUP(\$A2,'Cleaned '!\$A\$2:\$U\$10001,2,0)

New Excel sheet:

Further another new excel sheet were created to summarize the overall marks of the student.

Initially created a new column named “Groups” from using Groups of the students belong we can summarize information related to Gender, Count of Centum in each Subject, Count of Fail mark in each Subject with related to Groups they belong by using formulas.

Count of Gender:

Male and Female are counted for each groups

Group A:

Formula Used:

For Male =COUNTIFS('Student
Marksheet'!\$B\$2:\$B\$10001,"Male",'Student
Marksheet'!\$C\$2:\$C\$10001,"Group A")

For Female =COUNTIFS('Student
Marksheet'!\$B\$2:\$B\$10001,"Female",'Student
Marksheet'!\$C\$2:\$C\$10001,"Group A")

Same for other groups by changing group name in criteria count of male and female can be calculated.

Count of Centum:

Counted the number of centum got by each group in each subject.

Formula Used: =COUNTIFS('Student
Marksheet'!\$G\$2:\$G\$10001,"100",'Student
Marksheet'!\$C\$2:\$C\$10001,"Group A")

This formula has been used for the rest of groups and subjects by changing column name and group name .

Count of Fail:

Counted the number of fail marks(<=35) in each group and each subjects.

Formula Used: =COUNTIFS('Student
Marksheet'!\$G\$2:\$G\$10001,"<=35",'Student
Marksheet'!\$C\$2:\$C\$10001,"Group A")

Final Cleaned Dataset Extraction

After completion of all data cleaning process cleaned dataset was selected and loaded in new sheet.

This helped the final dataset separate from the raw data and ensured a clean and structured report.

Roll No	Gender	Race Ethnicity	Parental Level Of Education	Lunch	Test Preparation Of Course	Math	Reading	Writing	Science	Total	Grade	Eligibility	Percentage	Rank	Number	Number of
Std-7712	Female	Group D	Some College	1	0	92	88	95	85	360	A	Eligible	90	43	0	0
Std-7973	Female	Group D	Some College	1	0	97	100	88	63	348	A	Eligible	87	122	1	0
Std-9379	Female	Group D	Some College	1	0	60	96	93	99	348	A	Eligible	87	122	0	0
Std-5189	Female	Group D	Some College	1	0	83	100	85	78	346	A	Eligible	86.5	154	1	0
Std-8438	Female	Group D	Some College	1	0	95	73	98	76	342	A	Eligible	85.5	217	0	0
Std-9156	Female	Group D	Some College	1	0	64	76	100	100	340	A	Eligible	85	255	2	0
Std-7137	Female	Group D	Some College	1	0	78	100	83	76	337	A	Eligible	84.25	312	1	0
Std-8648	Female	Group D	Some College	1	0	99	68	95	74	336	A	Eligible	84	333	0	0
Std-6356	Female	Group D	Some College	1	0	56	87	97	94	334	A	Eligible	83.5	390	0	0
Std-9135	Female	Group D	Some College	1	0	79	73	100	79	331	A	Eligible	82.75	466	1	0
Std-7977	Female	Group D	Some College	1	0	75	83	72	100	330	A	Eligible	82.5	494	1	0
Std-6561	Female	Group D	Some College	1	0	86	75	87	79	327	A	Eligible	81.75	590	0	0
Std-7649	Female	Group D	Some College	1	0	88	78	96	63	325	A	Eligible	81.25	658	0	0
Std-5732	Female	Group D	Some College	1	0	69	99	90	65	323	A	Eligible	80.75	714	0	0
Std-7759	Female	Group D	Some College	1	0	66	90	90	76	322	A	Eligible	80.5	754	0	0
Std-7812	Female	Group D	Some College	1	0	62	93	81	85	321	A	Eligible	80.25	806	0	0
Std-7094	Female	Group D	Some College	1	0	92	56	100	71	319	B	Eligible	79.75	905	1	0
Std-9961	Female	Group D	Some College	1	0	61	81	82	93	317	B	Eligible	79.25	995	0	0
Std-5396	Female	Group D	Some College	1	0	65	93	95	63	316	B	Eligible	79	1032	0	0
Std-5418	Female	Group D	Some College	1	0	79	57	92	88	316	B	Eligible	79	1032	0	0
Std-7039	Female	Group D	Some College	1	0	94	99	75	47	315	B	Eligible	78.75	1079	0	0
Std-5646	Female	Group D	Some College	1	0	90	94	50	80	314	B	Eligible	78.5	1135	0	0
Std-6689	Female	Group D	Some College	1	0	81	75	72	85	313	B	Eligible	78.25	1192	0	0
Std-6471	Female	Group D	Some College	1	0	88	34	96	93	311	B	Eligible	77.75	1306	0	1
Std-6512	Female	Group D	Some College	1	0	82	96	76	57	311	B	Eligible	77.75	1306	0	0
Std-9394	Female	Group D	Some College	1	0	83	54	100	73	310	B	Eligible	77.5	1359	1	0
Std-7457	Female	Group D	Some College	1	0	60	90	70	72	299	C	Eligible	77	1474	0	0

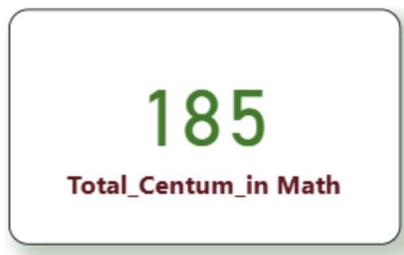
Visualization Using PowerBI

Measures:

Total Centum in Math: A new DAX measure was developed in Power BI to calculate the number of centum in math

Formula:

```
1 Total_Centum_in Math = SUM(Summary[Centum in Math])
```



Total Centum in Reading: A new DAX measure was developed in Power BI to calculate the number of centum in Reading.

Formula:

```
1 Total_Centum_in Reading = SUM(Summary[Centum in Reading])
```



Total Centum in Science: A new DAX measure was developed in Power BI to calculate the number of centum in Science.

Formula:

```
1 Total_centum_in_Science = SUM(Summary[Centum in Science])
```

319

Total_centum_in_Science

Total Centum in Writing: A new DAX measure was developed in Power BI to calculate the number of centum in writing.

Formula:

```
1 Total_Centum_in_Writing = SUM(Summary[Centum in Writing])
```

602

Total_Centum_in_Writing

Total Fail in Math: A new DAX measure was developed in Power BI to calculate the number of centum in math.

Formula Used:

```
1 Total_Fail_in_Math = sum(Summary[Fail in Math])
```

1887

Total_Fail_in_Math

Total Fail in Reading: A new DAX measure was developed in Power BI to calculate the number of centum in Reading.

Formula Used:

```
1 Total_Fail_in_Reading = SUM(Summary[Fail in Reading])
```



Total Fail in Science:

A new DAX measure was developed in Power BI to calculate the number of centum in Science.

Formula Used:

```
1 Total_Fail_in_Science = SUM(Summary[Fail in science])
```



Total Fail in Writing: A new DAX measure was developed in Power BI to calculate the number of centum in writing

Formula Used:

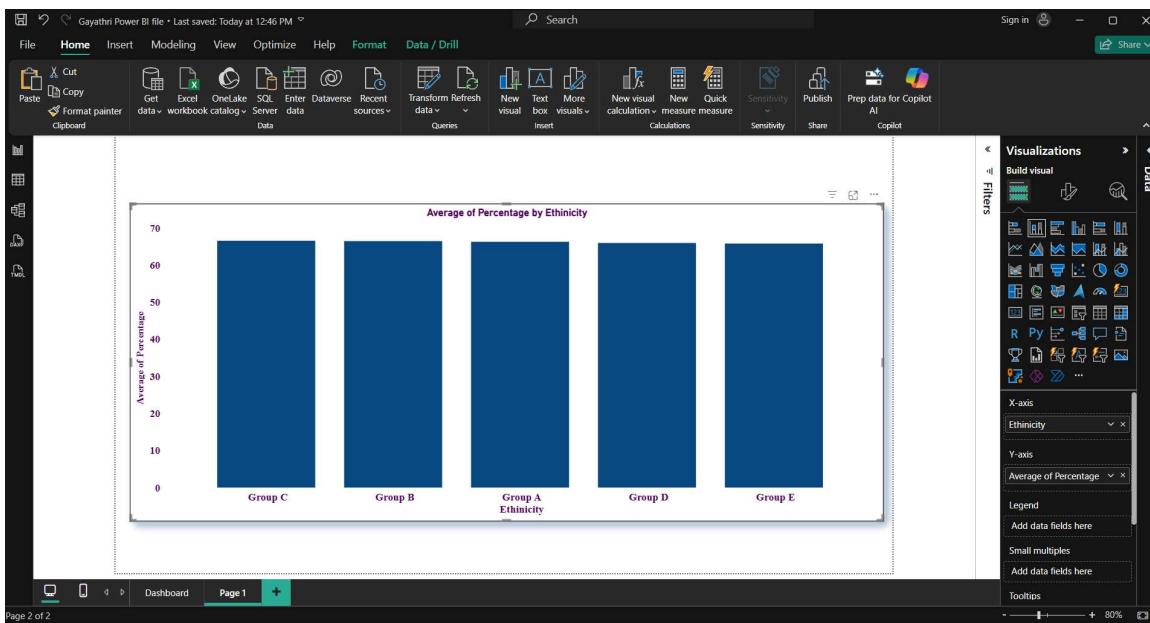
```
1 Total_Fail_in_Writing = SUM(Summary[ Fail in Writing])
```



Charts:

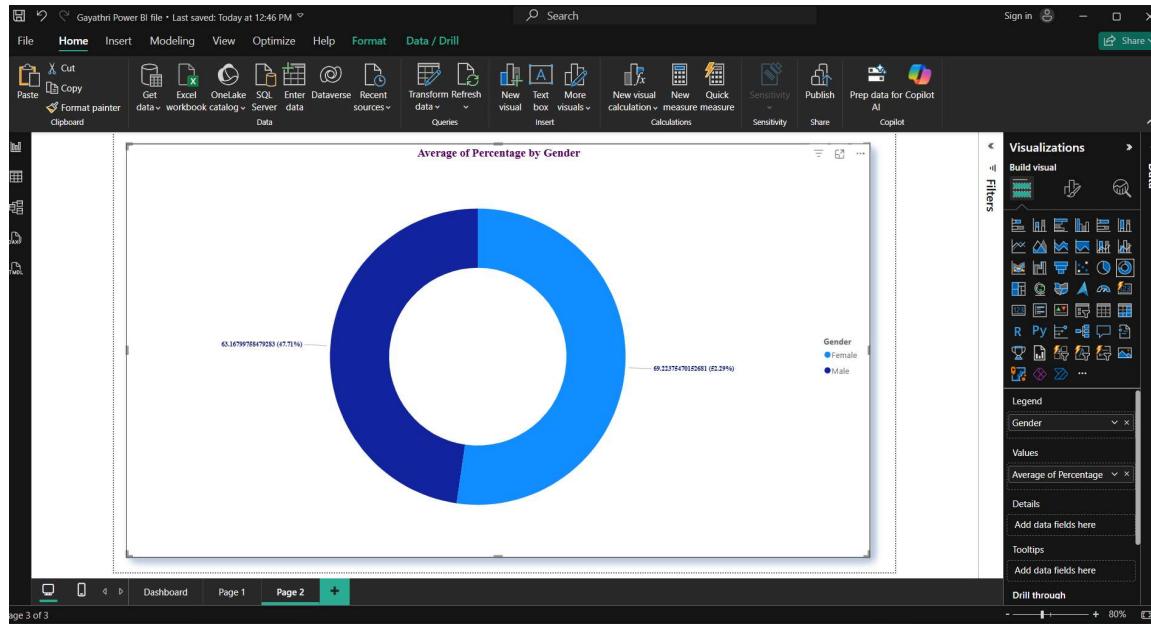
- Stacked Column chart:** Chart is created for analysis taking X axis as Ethinicity and in Y axis as Average of Percentage

Findings: Stacked column chart indicates “Group C” has highest percentage while “Group E” is lowest



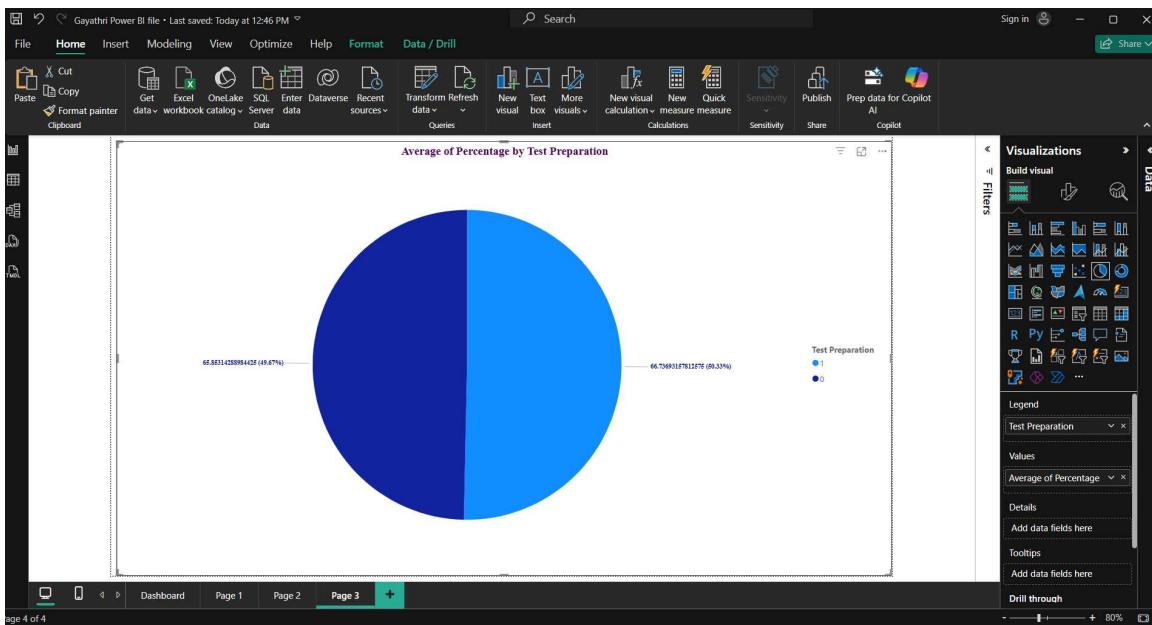
- Donut Chart:** Chart is created for analysis between gender and average percentage .

Findings: The donut chart indicates that percentage of female is more when compared to male .



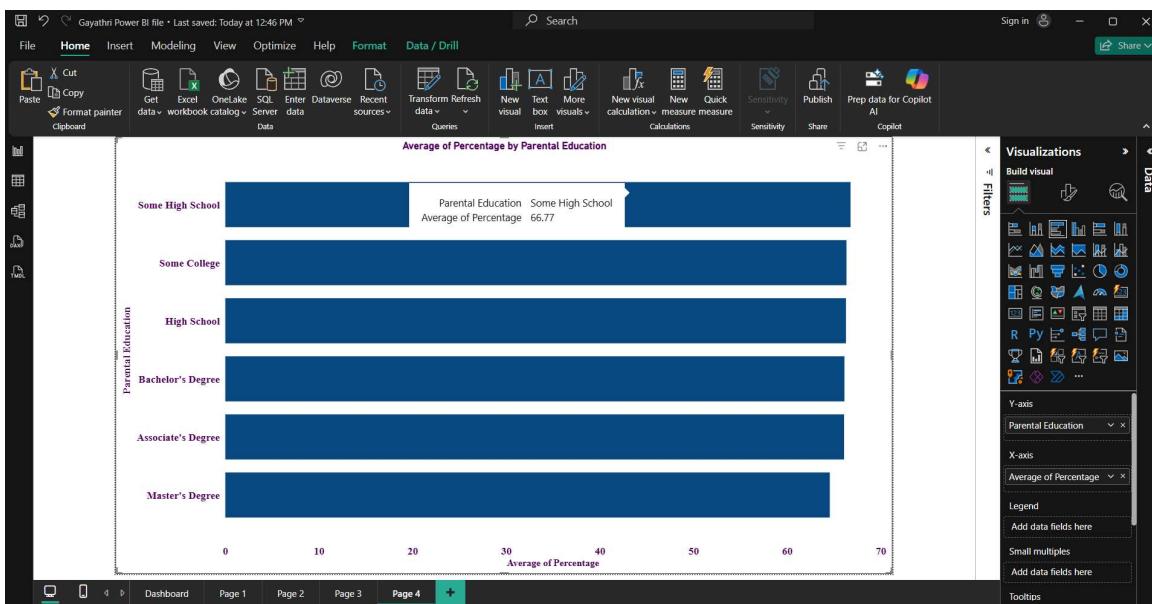
3.Pie Chart: Chart is created for doing analysis among students who prepared for test with average percentage.

Findings: Pie chart indicates student are well prepared for examination.



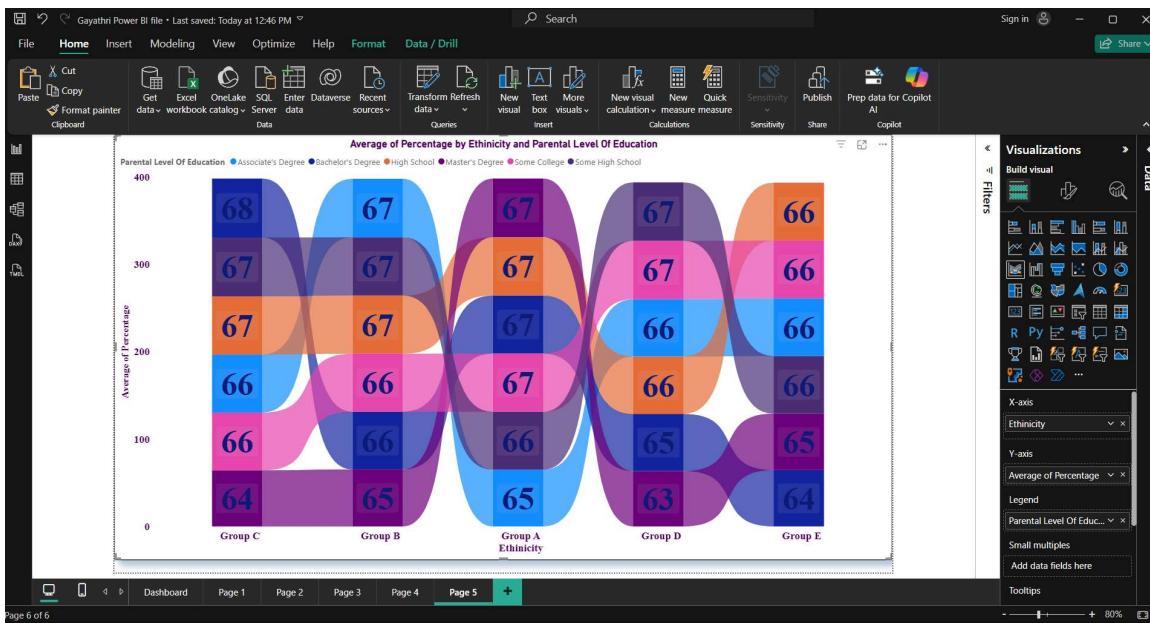
4. Clustered Bar Chart: Chart is created for analysis with parental education and average percentage.

Findings: Clustered chart indicates that parents who had their education in high school their children have high percentage in score when compared with parent have master's degree.



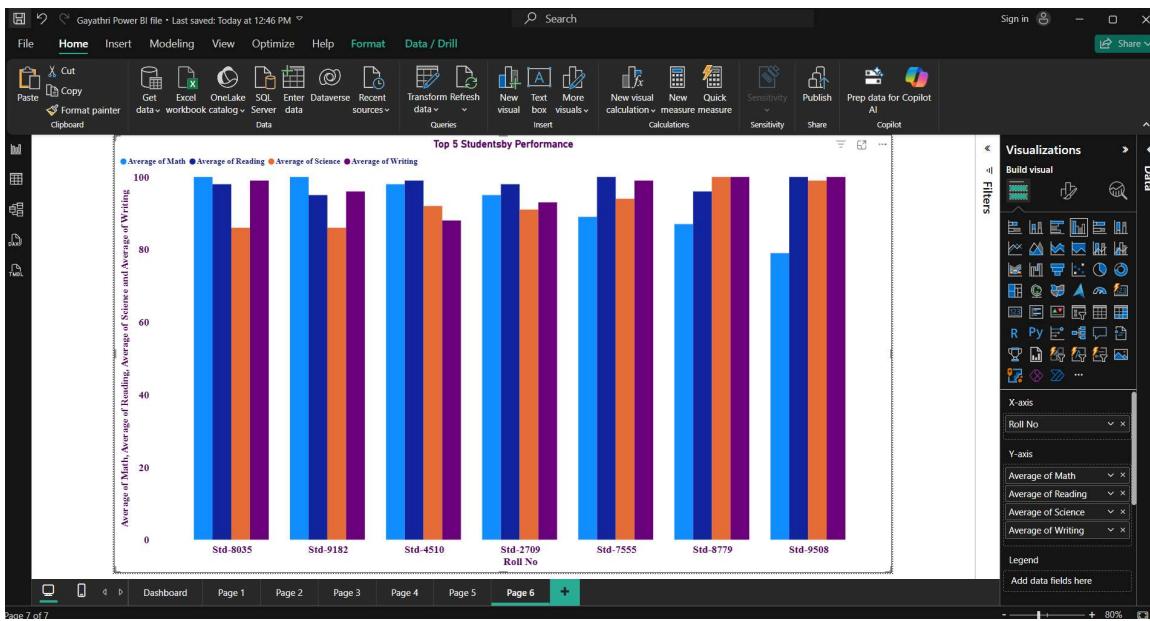
5.Ribbon Chart: Chart is created for analysis with ethnicity ,parental education, average percentage.

Finding: Ribbon chart indicates that the average percentage scores across all ethnic groups and parental education levels are tightly clustered between 64 and 67, indicating minimal variation.



1. **Clustered Column Chart:** Chart is created for analysis with roll no and average of each subject to find the top 5 students by using filter.

Findings: All seven students (Std-8035 to Std-9508) show consistently high performance across all subjects, with scores generally above 80.

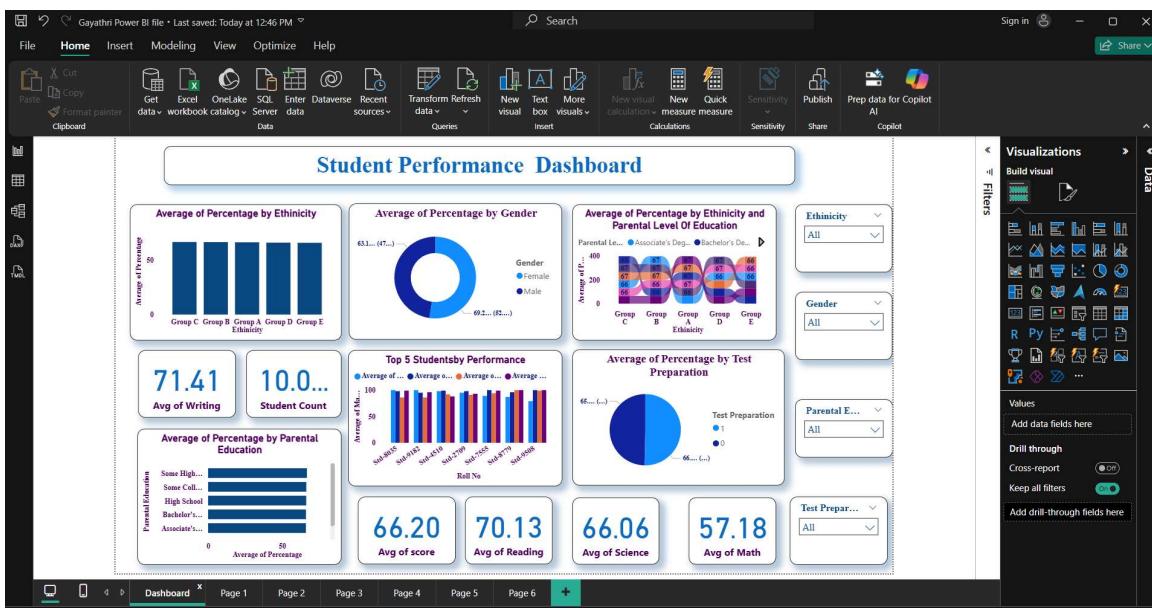


7.Add Slicers :

- Ethnicity
- Gender
- Parental Education
- Test Preparation

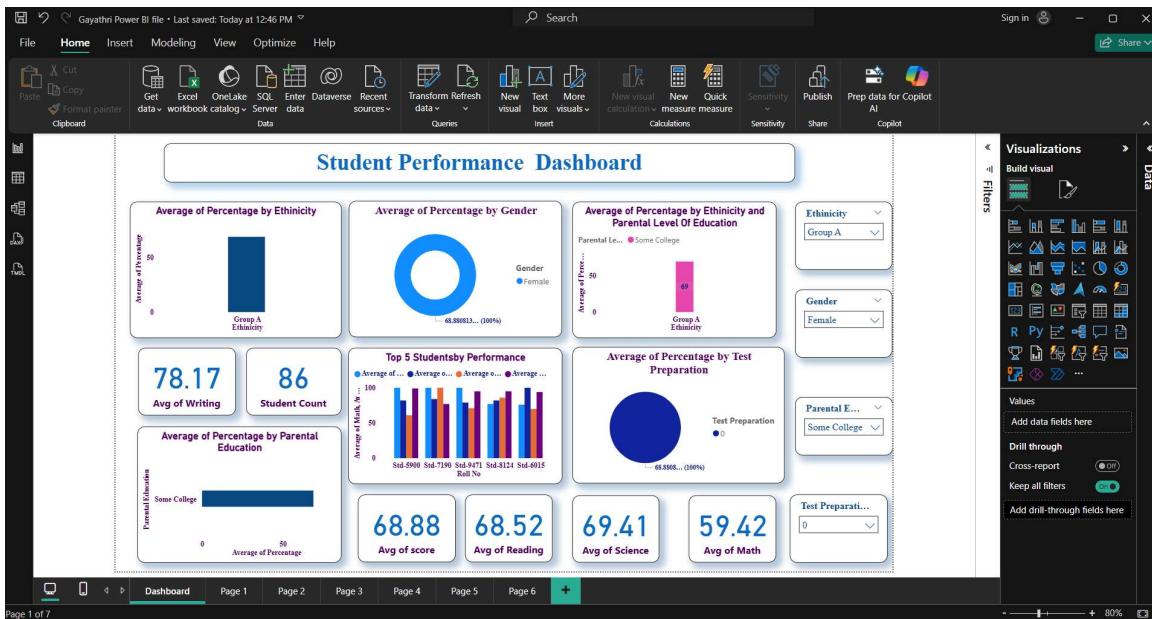
STUDENT PERFORMANCE ANALYSIS DASHBOARD:

(WITHOUT APPLYING SLICER)



STUDENT PERFORMANCE ANALYSIS DASHBOARD:

(APPLYING SLICER)



Insights:

1. Total Students Analyzed: 10,000

Average Scores by Subject:

- Math: 57.18
- Science: 66.06
- Reading: 70.13
- Writing: 71.41

Overall Average Score: 66.20

Insight: Writing and Reading are the strongest subjects overall, while Math shows the lowest average, indicating a potential area for academic support.

Ethnicity

- Group A leads with the highest average percentage.
- Groups D and E show slightly lower performance across metrics.

Insight: Ethnic disparities are minimal but consistent, suggesting further exploration into contextual factors like access or support.

Gender

- Female students average ~69.22%
- Male students average ~63.1%

Insight: Female students outperform males across all subjects, especially in Reading and Writing.

Parental Education Level

- Students whose parents hold Bachelor's or Associate's degrees tend to score higher.
- Those with Some High School or High School backgrounds show slightly lower averages.

Insight: Higher parental education correlates with better student performance, though the gap is modest.

Test Preparation

- Students who completed test preparation scored ~66%
- Those who didn't scored ~65%

Insight: Test preparation has a small but positive impact on performance, suggesting it may be underutilized or unevenly implemented.

Top Performers

- Highlighted students: Std-4903, Std-4192, Std-4510, Std-7555, Std-8779, Std-5960
- These students consistently score high across all subjects, especially in Science and Writing.

Insight: These profiles can be used to model successful academic behaviors or identify best practices.

Conclusion:

The dashboard provides a clear view of student performance, attendance and study habits drive stronger performance. Subject-level strengths and weaknesses highlight areas for targeted support. Background factors influence outcomes, stressing the need for equitable resources. Data-driven insights help identify at-risk students early and guide personalized interventions.