REPORT ON MINI PROJECT

Name : **Sowbarnika Sri U**

Batch Code : **TN_DA_FNB07**

Contact Number : **8489439449**

E-mail ID : **asurevthi@gmail.com**

Project Title : **Student Performance Analysis**

Project Domain : **Education**

Submission Date : **10 DEC 2025**

Mentor : **Kumaran M**

Raw Dataset link : Mini_Project/Sowbarnika Sri Mini Project Raw Dataset.csv at main · Sowbarnika-sri/Mini_Project

Cleaned Dataset link : Mini_Project/Sowbarnika Sri Miniproject Cleaned.xlsx at main · Sowbarnika-sri/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 a 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 groupby 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.

roll_no	gender	race_ethnicity	parental_level_of_education	lunch	test_preparation_course	math_score	reading_score	writing_score	science_score	total_score	grade
std-01	male	group D	some college	1	1	89	38	85	26	238	C
std-02	male	group B	high school	1	0	65	100	67	96	328	A
std-03	male	group C	master's degree	1	0	10	99	97	58	264	B
std-04	male	group D	some college	1	1	22	51	41	84	198	D
std-05	male	group C	some college	0	1	26	58	64	65	213	C
std-06	male	group B	high school	1	0	40	49	69	48	206	C
std-07		group C	some college	0	0	34	48	84	73	239	C
std-08	male	group B	associate's degree	1	1	25	80	34	17	156	D
std-09	male	group C		1	1	28	79	100	76	283	B
std-10	male	group C	some high school		0	71	22	85	39	217	C
std-11	male	group C	high school	1		55	64	87	60	266	B
std-12	male	group E	associate's degree	1	0		53	54	55	228	C
std-13	male	group A	associate's degree	1	0	29		76	91	293	B
std-14	male	group E	high school	0	1	18	70		45	162	D
std-15	male	A	associate's degree	1	1	72	97	52		289	B
std-16	male	group D	master's degree	0	0	23	96	34	54		C
std-17	male	group E	some high school	0	0	35	70	87	91	283	
std-18	male	group B	some college	0	1	38	82	82	73		B
std-19	Boy	group C	bachelor's degree	1	1	41	76	78		280	B
std-20	male	group B	some high school	1	1	22	73		75	224	C
std-21	male	group E	master's degree	1	1	44		50	47	208	C
std-22	male	group C	some high school	1	1		68	62	54	195	D
std-23	male	group B	bachelor's degree	1	1	39	71	49	78	237	C
std-24	male	group B	associate's degree		1	85	97	43	57	282	B
std-25	male	group A		1	0	30	63	63	9	165	D
std-26	male		some college	1	1	100	38	74	95	307	B

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.

Before:

Formula:

```
= Table.RenameColumns(#"Changed Type",{{"roll_no", "Roll_No"}, {"gender", "Gender"}, {"race_ethnicity", "Race_Ethnicity"},
    {"parental_level_of_education", "Parental_Level_Of_Education"}, {"lunch", "Lunch"}, {"test_preparation_course",
    "Test_Preparation_Course"}, {"math_score", "Math_score"}, {"reading_score", "Reading_score"}, {"writing_score", "Writing_score"},
    {"science_score", "Science_score"}, {"total_score", "Total_score"}, {"grade", "Grade"}})
```

After:

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

Before:

Roll_No	Gender	Race_Ethnicity	Parental_Level_Of_Education
std-01	male	group D	some college
std-02	male	group B	high school
std-03	male	group C	master's degree
std-04	male	group D	some college
std-05	male	group C	some college
std-06	male	group B	high school
std-07	*null*	group C	some college
std-08	male	group B	associate's degree
std-09	male	group C	*null*
std-10	male	group C	some high school
std-11	male	group C	high school
std-12	male	group E	associate's degree
std-13	male	group A	associate's degree
std-14	male	group E	high school
std-15	male	A	associate's degree
std-16	male	group D	master's degree
std-17	male	group E	some high school
std-18	male	group B	some college
std-19	Boy	group C	bachelor's degree
std-20	male	group B	some high school
std-21	male	group E	master's degree
std-22	male	group C	some high school
std-23	male	group B	bachelor's degree
std-24	male	group B	associate's degree
std-25	male	group A	*null*
std-26	male	*null*	some college

Formula:

```
= Table.TransformColumns(#"Renamed Columns",{{"Roll_No", Text.Proper, type text}, {"Gender", Text.Proper, type text}, {"Race_Ethnicity",
    Text.Proper, type text}, {"Parental_Level_Of_Education", Text.Proper, type text}})
```

After:

Roll_No	Gender	Race_Ethnicity	Parental_Level_Of_Education
Std-01	Male	Group D	Some College
Std-02	Male	Group B	High School
Std-03	Male	Group C	Master'S Degree
Std-04	Male	Group D	Some College
Std-05	Male	Group C	Some College
Std-06	Male	Group B	High School
Std-07	*null*	Group C	Some College
Std-08	Male	Group B	Associate'S Degree
Std-09	Male	Group C	*null*
Std-10	Male	Group C	Some High School
Std-11	Male	Group C	High School
Std-12	Male	Group E	Associate'S Degree
Std-13	Male	Group A	Associate'S Degree
Std-14	Male	Group E	High School
Std-15	Male	A	Associate'S Degree
Std-16	Male	Group D	Master'S Degree
Std-17	Male	Group E	Some High School
Std-18	Male	Group B	Some College
Std-19	Boy	Group C	Bachelor'S Degree
Std-20	Male	Group B	Some High School
Std-21	Male	Group E	Master'S Degree
Std-22	Male	Group C	Some High School
Std-23	Male	Group B	Bachelor'S Degree
Std-24	Male	Group B	Associate'S Degree
Std-25	Male	Group A	*null*

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”.

Steps Taken:

Formula:

```
= Table.ReplaceValue(#"Filtered Rows","Boy","Male",Replacer.ReplaceText,{"Gender"})
```

Before:

	Roll_No	Gender
1	Std-19	Boy
2	Std-978	Boy
3	Std-2840	Boy
4	Std-3659	Boy
5	Std-4190	Boy
6	Std-4512	Boy

After :

	Roll_No	Gender
1	Std-19	Male
2	Std-978	Male
3	Std-2840	Male
4	Std-3659	Male
5	Std-4190	Male
6	Std-4512	Male

Formula:

```
= Table.ReplaceValue(#"Filtered Rows","\Tmale","Male",Replacer.ReplaceText,{"Gender"})
```

Before:

Roll_No	Gender
Std-2872	\Tmale

After:

Roll_No	Gender
Std-2872	Male

Formula:

```
= Table.ReplaceValue(#"Filtered Rows","Girl","Female",Replacer.ReplaceText,{"Gender"})
```

Before:

After:

	Roll_No	Gender
1	Std-5120	Girl
2	Std-5251	Girl
3	Std-5973	Girl
4	Std-6412	Girl
5	Std-6581	Girl
6	Std-7303	Girl
7	Std-7742	Girl
8	Std-8063	Girl
9	Std-9241	Girl
10	Std-9697	Girl

	Roll_No	Gender
1	Std-5120	Female
2	Std-5251	Female
3	Std-5973	Female
4	Std-6412	Female
5	Std-6581	Female
6	Std-7303	Female
7	Std-7742	Female
8	Std-8063	Female
9	Std-9241	Female
10	Std-9697	Female

4.Race Ethnicity:

Problem Identified: The given dataset contains inconsitencies.

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.

Formula:

```
= Table.ReplaceValue(#"Filtered Rows","A","Group A",Replacer.ReplaceValue,{"Race_Ethnicity"})
```

Before:

Gender	Race_Ethnicity
Male	A
Female	A

After:

Male	Group A
Female	Group A

Formula:

```
= Table.ReplaceValue(#"Filtered Rows","B","Group B",Replacer.ReplaceValue,{"Race_Ethnicity"})
```

Before:

Gender	Race_Ethnicity
Female	B

After:

Gender	Race_Ethnicity
Female	Group B

Formula:

```
= Table.ReplaceValue(#"Filtered Rows","C","Group C",Replacer.ReplaceValue,{"Race_Ethnicity"})
```

Before:

Gender	Race_Ethnicity
Male	C
Female	C
Female	C

After:

Gender	Race_Ethnicity
Male	Group C
Female	Group C
Female	Group C

Formula:

```
= Table.ReplaceValue(#"Filtered Rows","D","Group D",Replacer.ReplaceValue,{"Race_Ethnicity"})
```

Before:

Gender	Race_Ethnicity
Male	D
Male	D
Male	D
Female	D
Female	D
Female	D

After:

Gender	Race_Ethnicity
Male	Group D
Male	Group D
Male	Group D
Female	Group D
Female	Group D
Female	Group D

Formula:

```
= Table.ReplaceValue(#"Filtered Rows","E","Group E",Replacer.ReplaceValue,{"Race_Ethnicity"})
```

Before:

Gender	Race_Ethnicity
Female	E
Female	E
Female	E
Female	E

After:

Gender	Race_Ethnicity
Female	Group E
Female	Group E
Female	Group E
Female	Group E

Formula:

```
= Table.ReplaceValue(#"Filtered Rows","Group C\N","Group C",Replacer.ReplaceText,{"Race_Ethnicity"})
```

Before:

Gender	Race_Ethnicity
Male	Group C\N
Male	Group C\N
Male	Group C\N
Male	Group C\N

After:

Gender	Race_Ethnicity
Male	Group C
Male	Group C
Male	Group C
Male	Group C

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”.

Groupby on Race ethnicity:

	Race_Ethnicity	Count
1	Group D	2572
2	Group B	1900
3	Group C	2928
4	Group E	1513
5	Group A	1064
6	null	23

Formula:

```
= Table.ReplaceValue(#"Filtered Rows",null,"Group D",Replacer.ReplaceValue,{"Race_Ethnicity"})
```

Before:

Race_Ethnicity
null
null
null
null
null
null
null
null
null
null
null
null
null
null
null
null
null
null
null
null
null

After:

Race_Ethnicity
Group D
Group D
Group D
Group D
Group D
Group D
Group D
Group D
Group D
Group D
Group D
Group D
Group D
Group D
Group D
Group D
Group D
Group D
Group D
Group D
Group D
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”.

Groupby of Parent level on Education:

	Parental_Level_Of_Education	Count
1	Some College	2272
2	High School	1987
3	Master's Degree	711
4	Associate's Degree	1905
5	null	22
6	Some High School	1742
7	Bachelor's Degree	1361

Before:

Parental_Level_Of_Education
null
null
null
null
null
null
null
null
null
null
null
null
null
null
null
null
null
null
null
null
null

After:

Parental_Level_C
Some College
Some College
Some College
Some College
Some College
Some College
Some College
Some College
Some College
Some College
Some College
Some College
Some College
Some College
Some College
Some College
Some College
Some College
Some College
Some College
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".

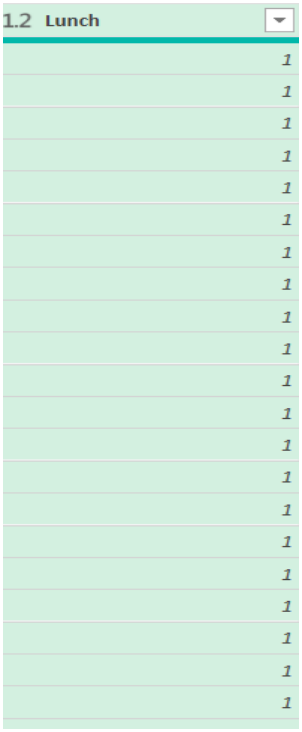
Group by on Lunch:

	Lunch	Count
1	1	6427
2	0	3549
3	null	24

Before:

Lunch
null
null
null
null
null
null
null
null
null
null
null
null
null
null
null
null
null
null
null
null
null

After:



1.2 Lunch
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
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.

Group by function on Test preparation on course:

	Test_Preparation_Course	Count
1	1	3878
2	0	6099
3	null	23

Before:

After:

The file is close and loaded to the MS Excel for further cleaning .

Math Column:

Problem Identified: In dataset some students math marks were missing . with missing value we cant perform calculation.

Step Taken: By filtering the column founded the missing values and perform some calculation to fill the missing values.

Formula used: =IF(ISBLANK(G2),AVERAGE($G$2:$G$10001),G2)

Average value of Math=57

Before:

Math_Score
85
78
36
68
97
38
38
78
64
60
60
38
69
87
83
42
72
52
65
27
46
12
51
32
75

After:

Math Score
85
78
36
68
97
38
38
57
78
64
60
60
38
69
87
83
42
72
57
52
65
27
46
12
51
32
75

Science Column:

Problem Identified: In dataset some students math marks were missing . with missing value we cant perform calculation.

Step Taken: By filtering the column founded the missing values and perform some calculation to fill the missing values.

Formula Used: **=IF(ISBLANK(J2),AVERAGE($J$2:$J$10001),J2)**

Average value of Science =66

Before:

Science_Score
55
91
45
54
91
73
75
47
54
78
57
9
95
44
46
82
78
81
9
31
65
70
57
71
93

After:

Science Score
55
91
45
66
54
91
73
66
75
47
54
78
57
9
95
44
46
82
78
81
9
31
65
70
57
71
93

Reading Column:

Problem Identified: In dataset some students math marks were missing . with missing value we cant perform calculation.

Step Taken: By filtering the column founded the missing values and perform some calculation to fill the missing values.

Formula used: =IF(ISBLANK(H2),AVERAGE($H$2:$H$10001),H2)

Average of Reading = 70

Before:

Reading_Score
76
82
70
66
89
97
92
38
81
43
56
90
98
63
56
64
100
98
66
55
91
83
73
100
86

After:

Reading Score
76
82
70
66
89
97
92
38
70
81
43
56
90
98
63
56
64
70
100
98
66
55
91
83
73
100
86

Writing Column:

Problem Identified: In dataset some students math marks were missing . with missing value we cant perform calculation.

Step Taken: By filtering the column founded the missing values and perform some calculation to fill the missing values

Formula Used: **=IF(ISBLANK(I2),AVERAGE($I$2:$I$10001),I2)**

Average of Writing=71

Before:

Writing_Score
95
54
81
33
78
53
98
48
57
69
59
73
60
100
30
74
100
77
73
82
63
80
100
32
61

After:

Writing Score
95
54
81
71
33
78
53
98
48
71
57
69
59
73
60
100
30
74
100
77
73
82
63
80
100
32
61

Total Score:

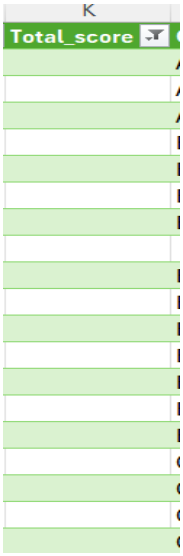
Problem Identified: Total marks of the student were missing which lead to create inaccurate insights.

Steps Taken: By filtering columns identified missing values in rows and performed some calculation to fill the missing values.

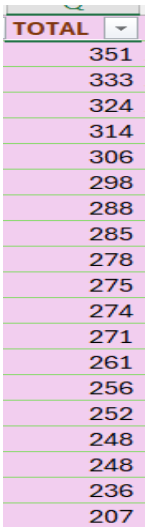
Formula used:

=[@MATHS]+[@READING]+[@WRITING]+[@SCIENCE]

Before:



After:



Grade Column:

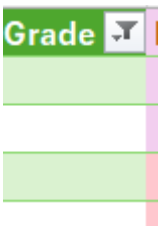
Problem Identified: Grade of the student were missing which lead to create inaccurate insights.

Steps Taken: By filtering columns identified missing values in rows and performed some calculation to fill the missing values.

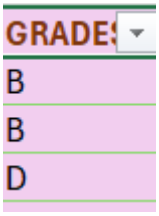
Formula Used:

=IF(Q2<150,"E",IF(Q2<200,"D",IF(Q2<250,"C",IF(Q2<320,"B","A"))))

Before:



After:



Gender:

Problem Identified: Some rows were missed in gender column.

Steps Taken: By analyzing gender column from “Std-01” to “Std-5000” were only male and from “Std-5001” to “Std-10000” were female.

So, by manual filled the missing values in Gender with related to their Roll no.

Manual filling on Gender

1	Roll_No	Gender	Race_Ethnicity	Parental_L
8	Std-07	Male	Group C	Some Colle
28	Std-27	Male	Group C	Some High
39	Std-38	Male	Group C	Associate'S
60	Std-59	Male	Group C	Associate'S
252	Std-251	Male	Group D	Some High
408	Std-407	Male	Group C	High Schoo
537	Std-536	Male	Group C	Associate'S
775	Std-774	Male	Group B	High Schoo
1326	Std-1325	Male	Group E	High Schoo
2107	Std-2106	Male	Group A	Some High
4503	Std-4502	Male	Group B	Some Colle
5762	Std-5761	Female	Group E	Some Colle
6170	Std-6169	Female	Group C	Bachelor'S
8037	Std-8036	Female	Group E	Some High
9064	Std-9063	Female	Group D	Some Colle
9174	Std-9173	Female	Group C	Associate'S
9287	Std-9286	Female	Group E	Master'S D
9642	Std-9641	Female	Group C	Some High

Performing Calculation for Valuable Insights:

1.Eligibility:

Created a new column named “Eligibility” to check the students weather 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")

Conditional Formatting:

Used conditional formatting on each subject to know the fail marks of students which is easy to grab the insights from given dataset.

19	45	58	47
38	55	67	9
16	39	48	66
39	65	39	26
33	56	43	37
24	49	59	36
32	67	32	37
48	46	49	25
48	51	16	53
23	77	27	41
49	58	49	12
41	50	32	45
28	33	84	23
33	61	51	22
19	70	54	24
24	76	54	13
10	43	67	47
19	17	62	69
24	52	76	15
20	76	27	43

Created a new excel sheet:

In MS Excel created a new sheet named " Student Marksheet" 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 transferd to new sheet named "Student Marksheet".

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 datset separate from the raw data and ensured a clean and structured report.

Roll No	Gender	Race Ethinicity	Parental Level Of Education	Lunch	Test Preparation Of Course	Math	Reading	Writing	Science	Total	Grade	Eligibility	Percentage	Rank	Number of centum
Std-8035	Female	Group D	Master's Degree	1	1	100	98	99	86	383	A	Eligible	95.75	1	1
Std-8779	Female	Group C	Some High School	1	0	87	96	100	100	383	A	Eligible	95.75	1	2
Std-7555	Female	Group D	Associate's Degree	0	0	89	100	99	94	382	A	Eligible	95.5	3	1
Std-9508	Female	Group D	High School	1	0	79	100	100	99	378	A	Eligible	94.5	4	2
Std-2709	Male	Group B	Some College	1	1	95	98	93	91	377	A	Eligible	94.25	5	0
Std-4510	Male	Group A	Associate's Degree	1	0	98	99	88	92	377	A	Eligible	94.25	5	0
Std-9182	Female	Group B	Bachelor's Degree	1	0	100	95	96	86	377	A	Eligible	94.25	5	1
Std-1208	Male	Group C	High School	1	0	99	93	84	100	376	A	Eligible	94	8	1
Std-2297	Male	Group E	Associate's Degree	0	1	83	100	92	100	375	A	Eligible	93.75	9	2
Std-3246	Male	Group C	Associate's Degree	0	1	87	100	89	98	374	A	Eligible	93.5	10	1
Std-8445	Female	Group C	Bachelor's Degree	1	0	100	84	100	90	374	A	Eligible	93.5	10	2
Std-2821	Male	Group B	Associate's Degree	1	0	82	95	99	97	373	A	Eligible	93.25	12	0
Std-3047	Male	Group E	Some College	1	1	76	99	98	100	373	A	Eligible	93.25	12	1
Std-6643	Female	Group D	Some High School	1	0	87	98	95	93	373	A	Eligible	93.25	12	0
Std-303	Male	Group D	Bachelor's Degree	1	0	75	96	100	100	371	A	Eligible	92.75	15	2
Std-5047	Female	Group B	Some College	1	1	87	94	99	90	370	A	Eligible	92.5	16	0
Std-66	Male	Group D	Associate's Degree	1	0	81	88	100	100	369	A	Eligible	92.25	17	2
Std-6637	Female	Group D	Some College	1	1	87	97	84	100	368	A	Eligible	92	18	1
Std-6964	Female	Group B	Associate's Degree	1	1	69	100	100	99	368	A	Eligible	92	18	2
Std-8868	Female	Group E	Bachelor's Degree	1	0	89	100	86	93	368	A	Eligible	92	18	1
Std-9214	Female	Group B	Some High School	1	1	100	76	100	92	368	A	Eligible	92	18	2
Std-996	Male	Group B	Master's Degree	0	0	88	99	94	86	367	A	Eligible	91.75	22	0
Std-2833	Male	Group D	Some College	1	0	100	95	72	100	367	A	Eligible	91.75	22	2
Std-4190	Male	Group D	Some College	1	1	98	100	100	69	367	A	Eligible	91.75	22	2
Std-4663	Male	Group B	Associate's Degree	1	0	75	100	100	92	367	A	Eligible	91.75	22	2
Std-1465	Male	Group D	Some College	1	0	99	92	92	83	366	A	Eligible	91.5	26	0

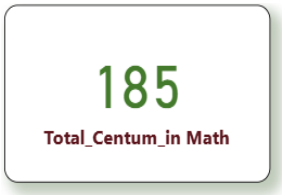
Visualization Using PowerBI

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])
```

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

Formula:

```
Total_Centum_in_Writing = SUM(Summary[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])
```

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

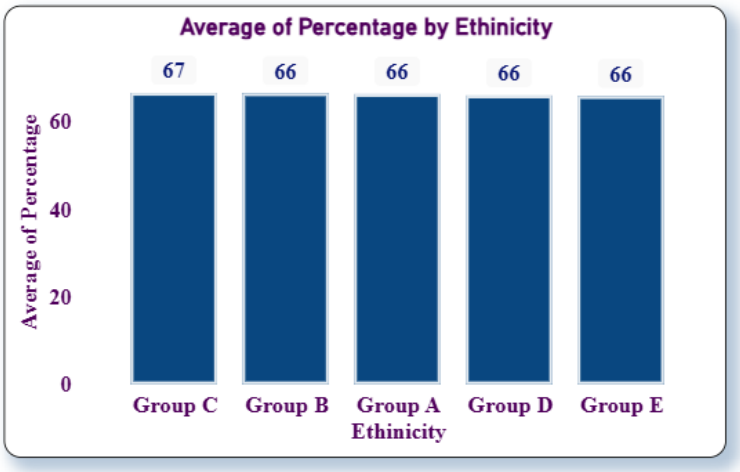
Formula Used:

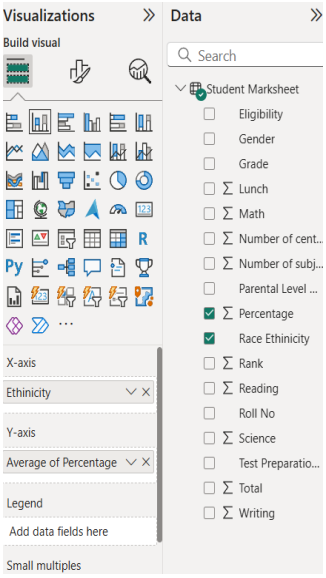
```
1  Total_Fail_in_Writing = SUM(Summary[ Fail in Writing])
```

Charts:

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

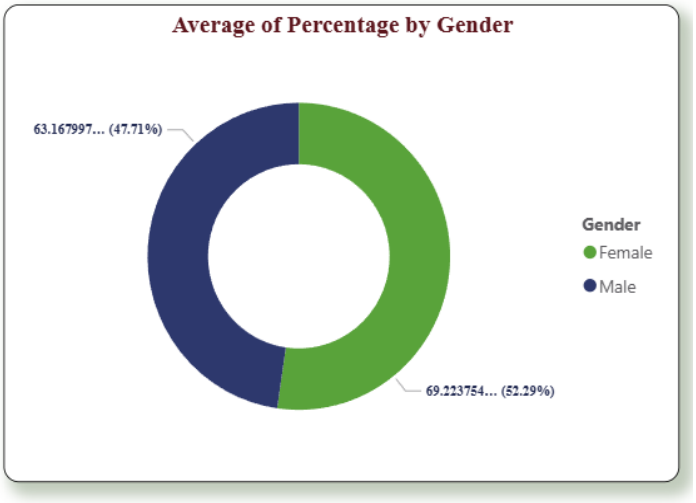
Findings: Stacked column chart indicates “Group C” has highest percentage while “Group E” is lowest.

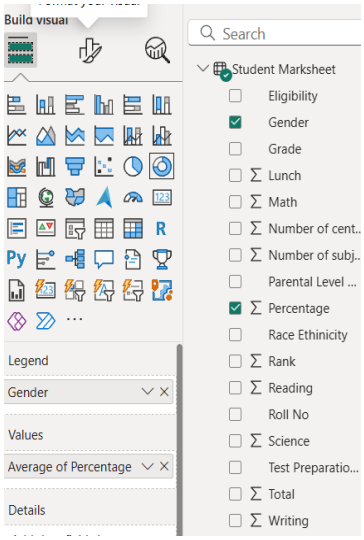




2. **Donut Chart**: Chart is created for analysis between gender and average percentage .

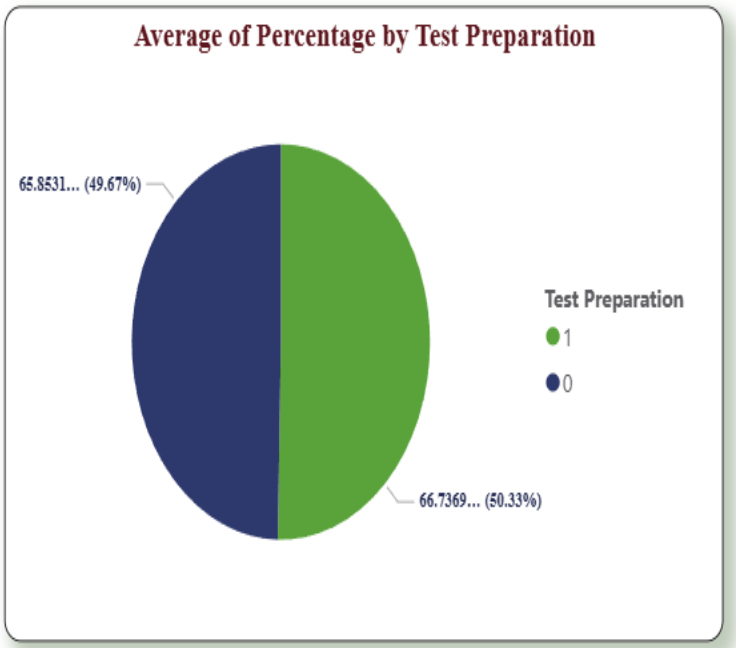
   Findings: The donut chart indicates that percentage of female is more when compared to male .

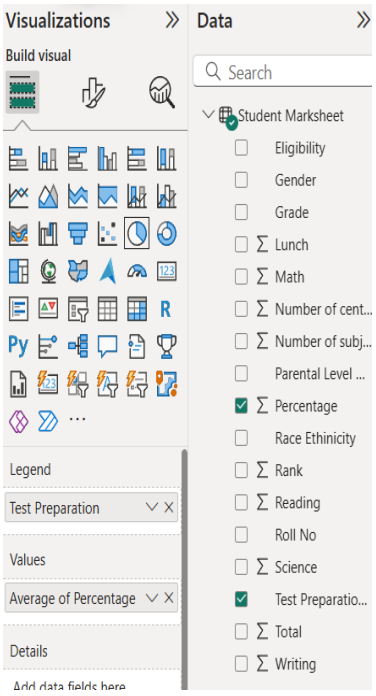




3. **Pie Chart:** Chart is created for doing analysis among students who prepared for test with average percentage.

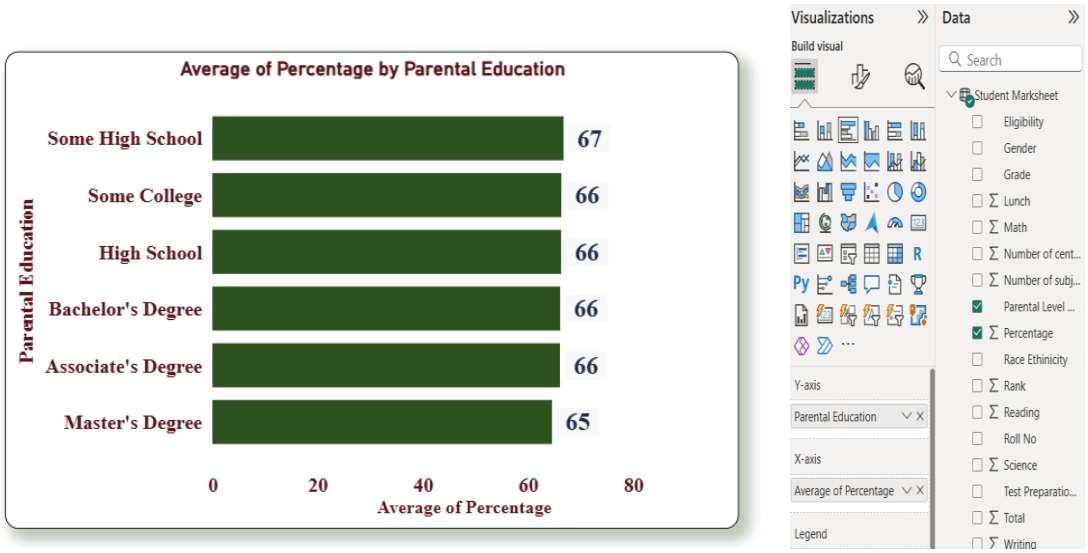
   Findings: Pie chart indicates student are well prepared for examination.





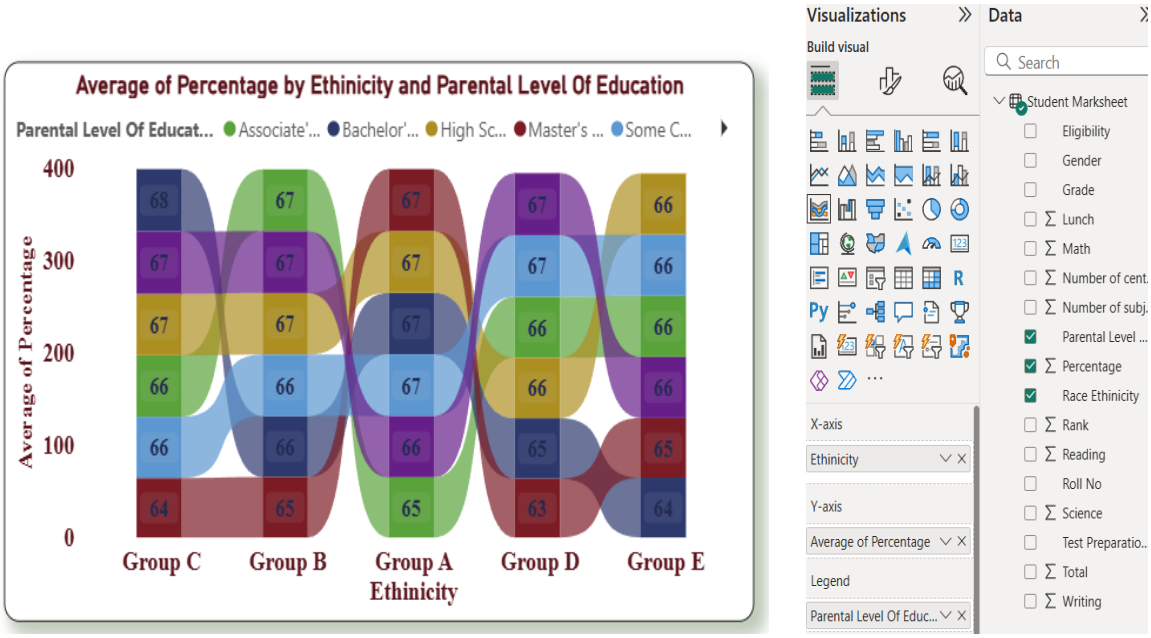
4. **Clustered Bar Chart**: Chart is created for analysis with parental education and average percentage.

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.



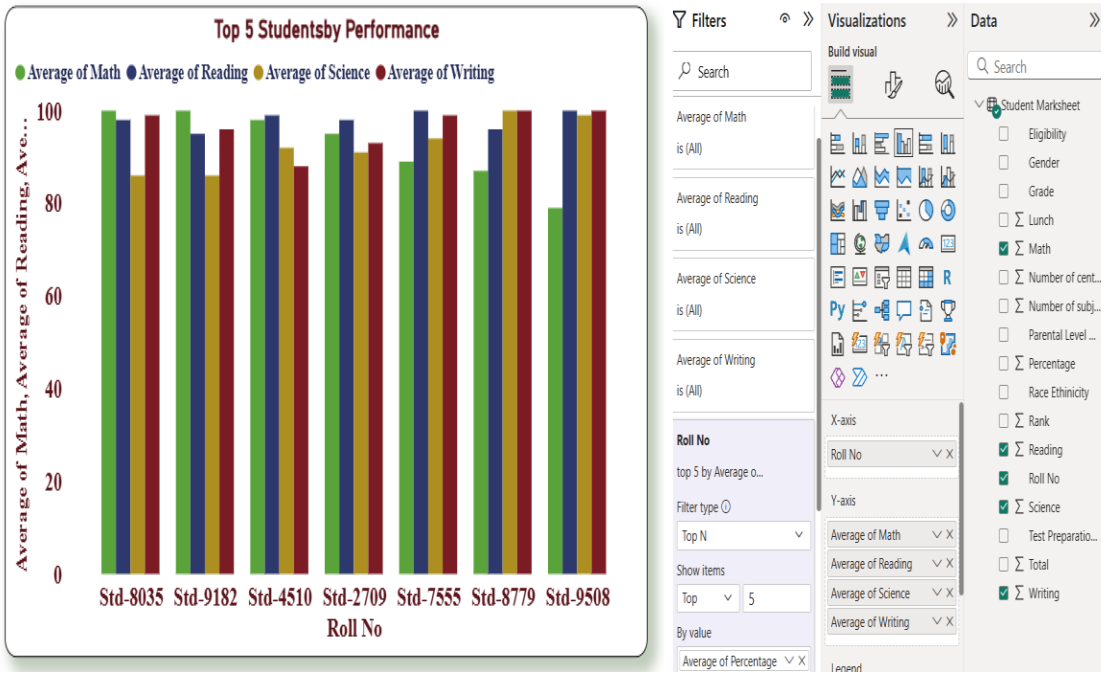
5. **Ribbon Chart:** Chart is created for analysis with ethnicity ,parental education, average percentage.

   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.



6. **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.

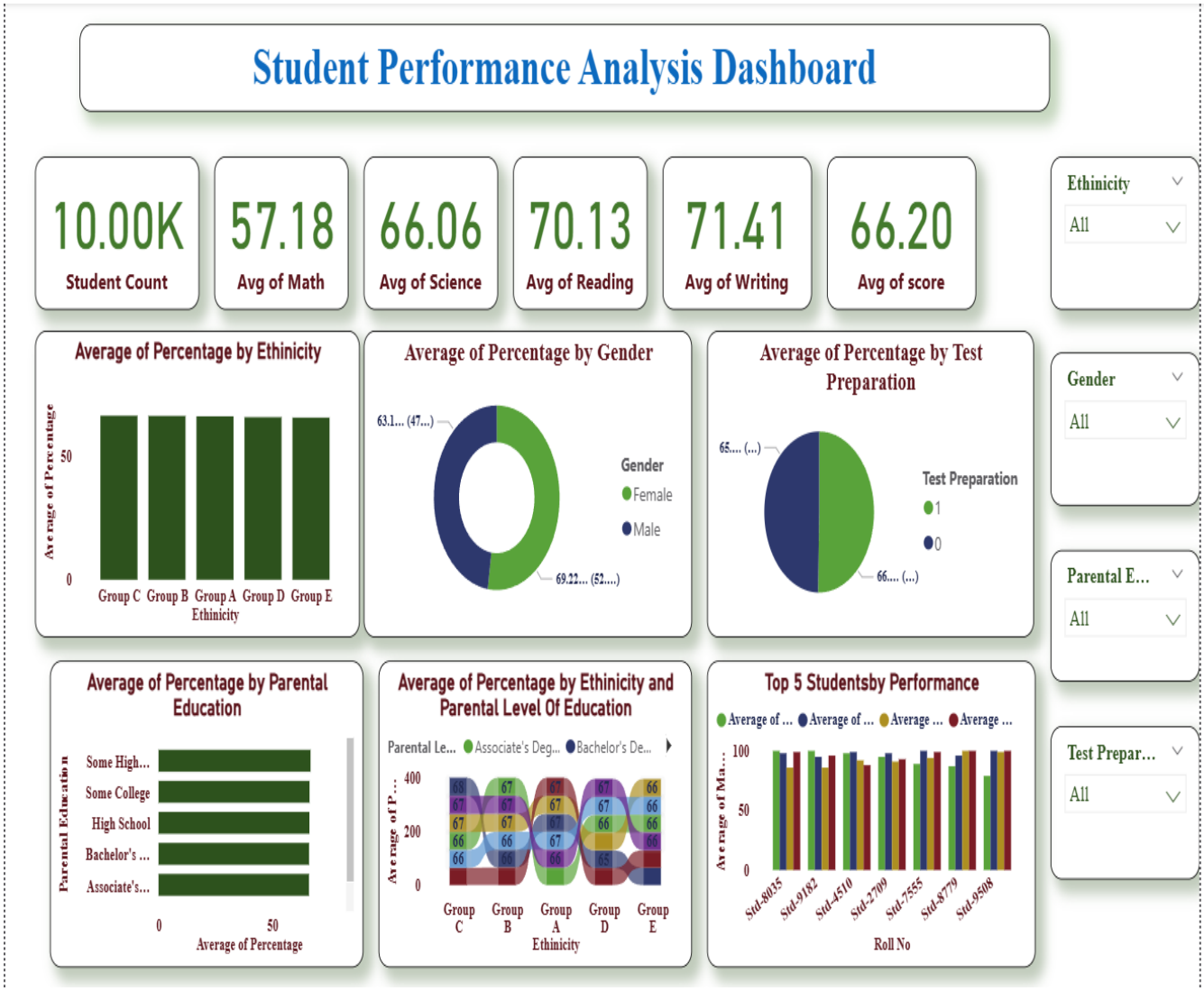
   Findings: All seven students (Std-8035 to Std-9508) show consistently high performance across all subjects, with scores generally above 80.



7. **Add Slicers for:**

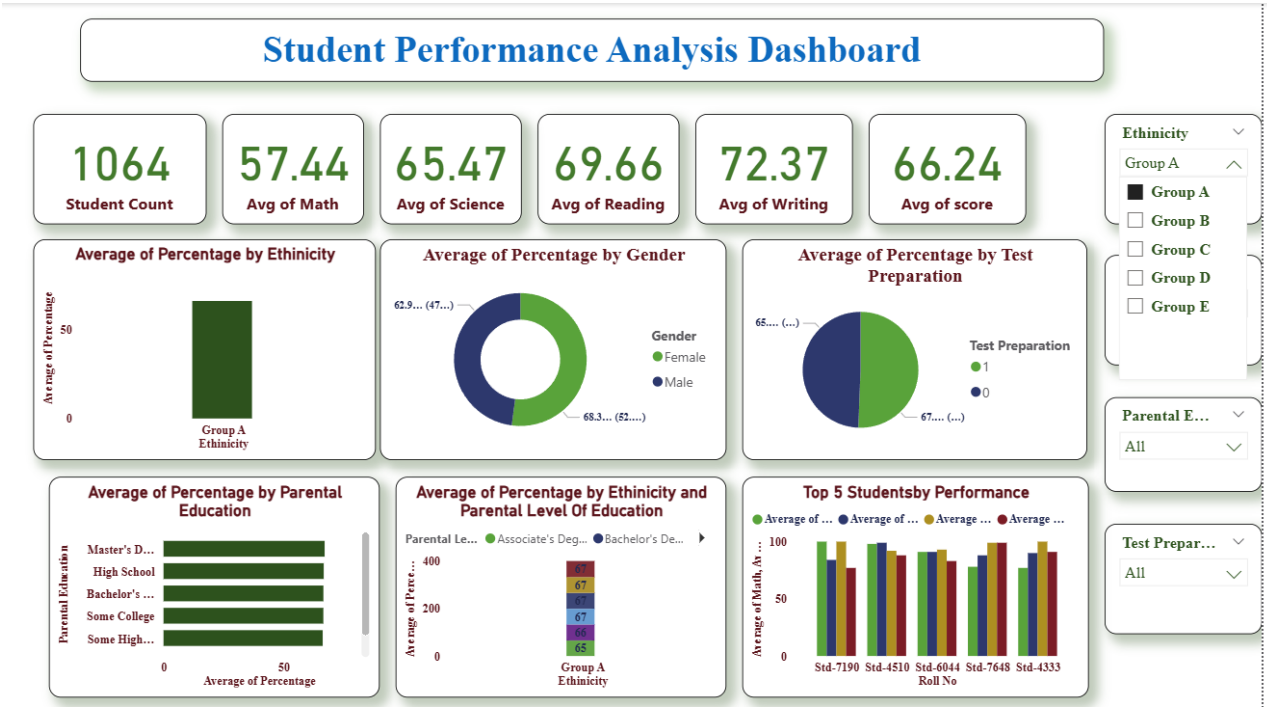
- Ethnicity
- Gender
- Parental Education
- Test Preparation

Dashboard

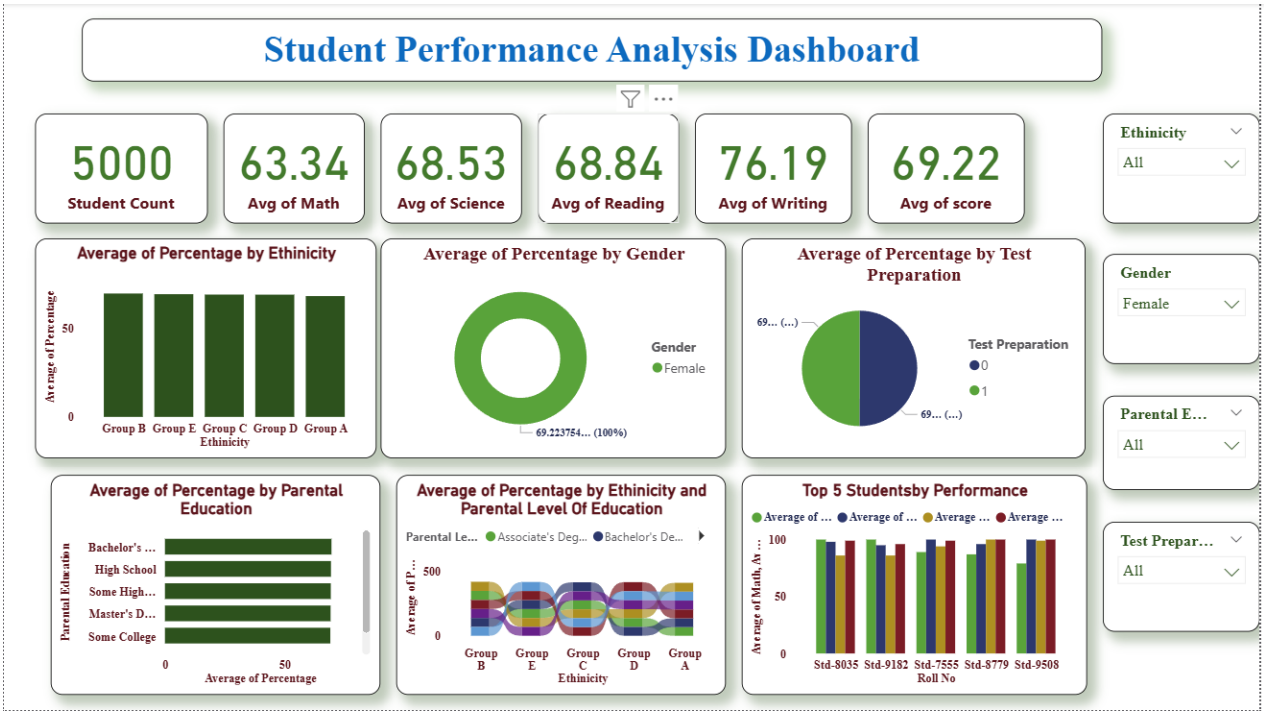

Student Performance Analysis Dashboard
10.00K
Student Count
57.18
Avg of Math
66.06
Avg of Science
70.13
Avg of Reading
71.41
Avg of Writing
66.20
Avg of score
Ethinicity
All
Average of Percentage by Ethinicity
Average of Percentage
50
0
Group C
Group B
Group A
Group D
Group E
Ethinicity
Average of Percentage by Gender
63.1... (47...)
69.22... (52....)
Gender
Female
Male
Gender
All
Average of Percentage by Test Preparation
65.... (...)
66.... (...)
Test Preparation
1
0
Parental E...
All
Average of Percentage by Parental Education
Parental Education
Some High...
Some College
High School
Bachelor's ...
Associate's...
0
50
Average of Percentage
Average of Percentage by Ethinicity and Parental Level Of Education
Parental Le...
Associate's Deg...
Bachelor's De...
Average of P...
400
200
0
Group C
Group B
Group A
Group D
Group E
Ethinicity
Top 5 Studentsby Performance
Average of ...
Average of ...
Average ...
Average ...
Average of Ma...
100
50
0
Std-8035
Std-9182
Std-4510
Std-2709
Std-7555
Std-8779
Std-9508
Roll No
Test Prepar...
All

Dashboard After Using Slicer

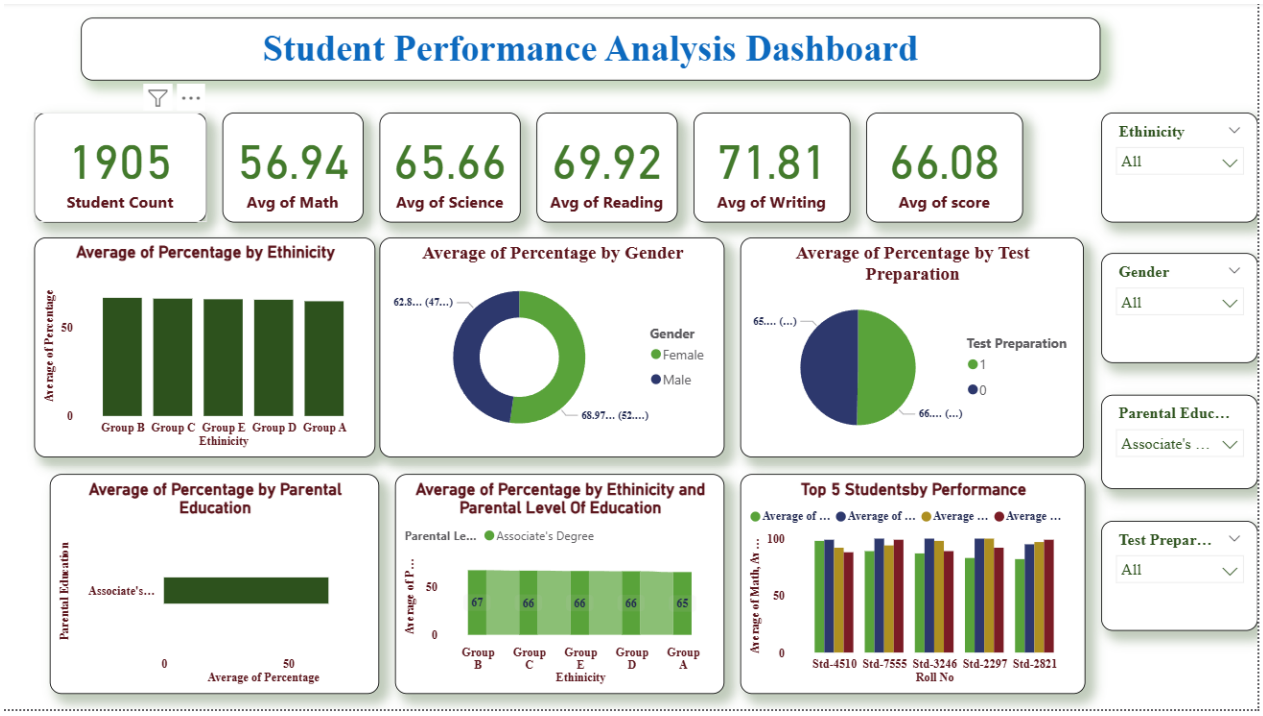
Applied slicer on Ethnicity “Group A”.



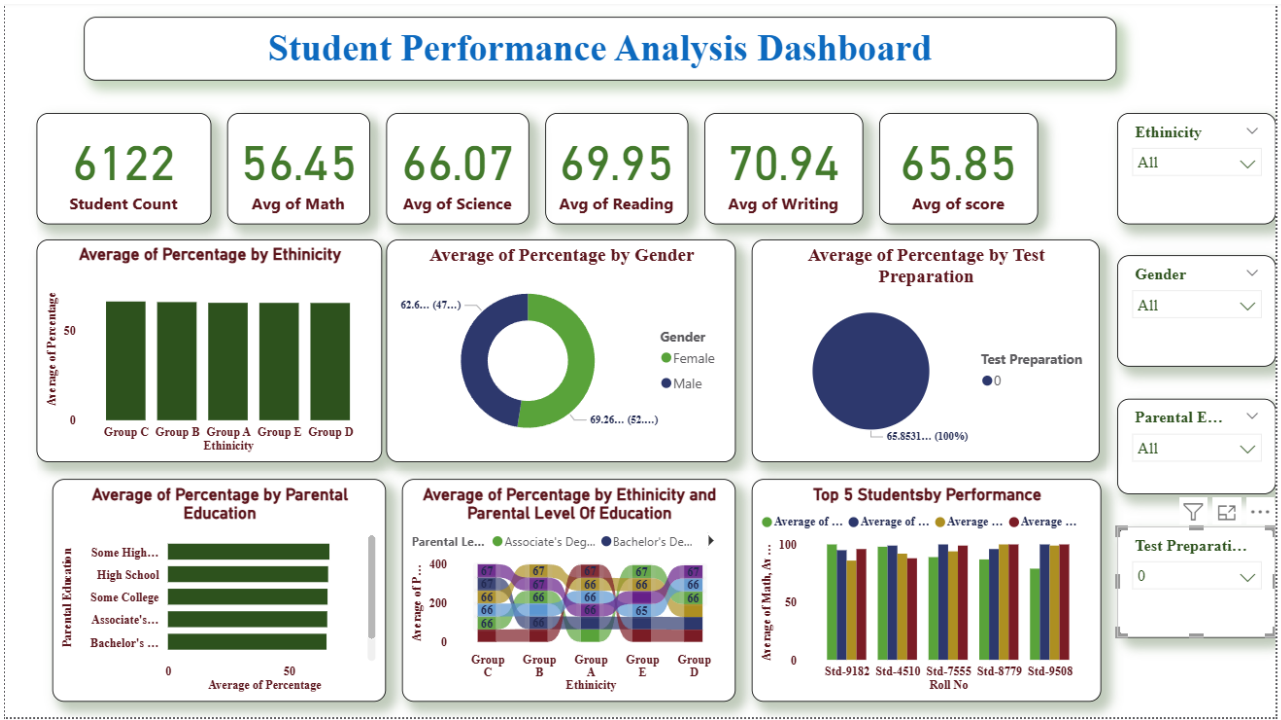
Applied slicer on Gender “Female”



Applied slicer on Parental Education



Applied slicer on Test Preparation



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.