



# Indian Institute of Technology Roorkee

## IBM-312 Project Report

### Analysis of performance of students in exams

Group 28: Vaishnavi Andhalkar (19116086),

Dipankar Chaudhary (19116021),

Karan (19116032)

#### 1 Motivation:

Performance of students during the exams is an important factor that contributes to the overall results and grades, performance depends on various factors such as IQ, nutrition, sleep, genetics and a lot of other features, in this report we have analysed the performance of the students and its correlation with such features and have tried to gain an insight on how performance is affected with these features.

#### 2 Snapshot of data:

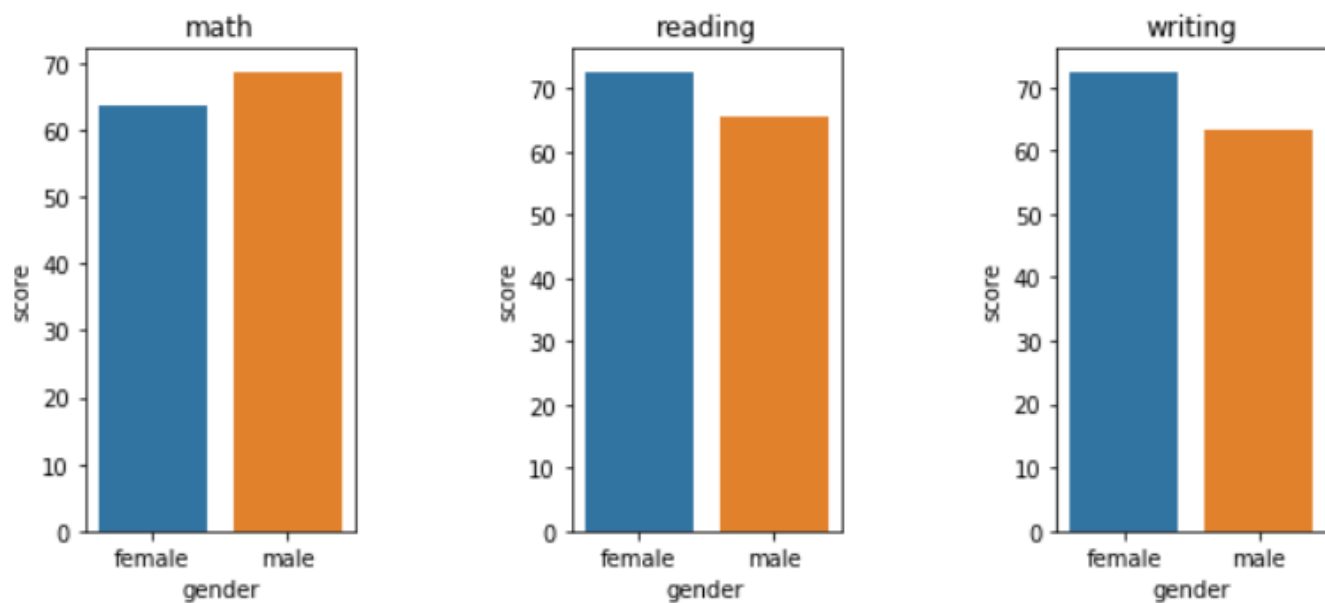
	A	B	C	D	E	F	G	H	I
1	gender	race/ethni	parental level of education	standard	test preparation cou	math score	reading score	writing score	
2	female	group B	bachelor's degree	standard	none	72	72	74	
3	female	group C	some college	standard	completed	69	90	88	
4	female	group B	master's degree	standard	none	90	95	93	
5	male	group A	associate's degree	free/reduced	none	47	57	44	
6	male	group C	some college	standard	none	76	78	75	
7	female	group B	associate's degree	standard	none	71	83	78	
8	female	group B	some college	standard	completed	88	95	92	
9	male	group B	some college	free/reduced	none	40	43	39	
10	male	group D	high school	free/reduced	completed	64	64	67	
11	female	group B	high school	free/reduced	none	38	60	50	
12	male	group C	associate's degree	standard	none	58	54	52	
13	male	group D	associate's degree	standard	none	40	52	43	
14	female	group B	high school	standard	none	65	81	73	
15	male	group A	some college	standard	completed	78	72	70	
16	female	group A	master's degree	standard	none	50	53	58	
17	female	group C	some high school	standard	none	69	75	78	
18	male	group C	high school	standard	none	88	89	86	
19	female	group B	some high school	free/reduced	none	18	32	28	
20	male	group C	master's degree	free/reduced	completed	46	42	46	
21	female	group C	associate's degree	free/reduced	none	54	58	61	
22	male	group D	high school	standard	none	66	69	63	
23	female	group B	some college	free/reduced	completed	65	75	70	
24	male	group D	some college	standard	none	44	54	53	
25	female	group C	some high school	standard	none	69	73	73	
26	male	group D	bachelor's degree	free/reduced	completed	74	71	80	
27	male	group A	master's degree	free/reduced	none	73	74	72	
28	male	group B	some college	standard	none	69	54	55	
29	female	group C	bachelor's degree	standard	none	67	69	75	

This dataset contains various features which can affect the performance of students such as gender, ethnicity, education of parents, test preparation, and it also contains the scores of students in various subjects like reading, writing and maths which will help us analyse the overall performance in correlation with specific features.

### 3 Methodology analysis:

#### 3.1 Gender vs. score

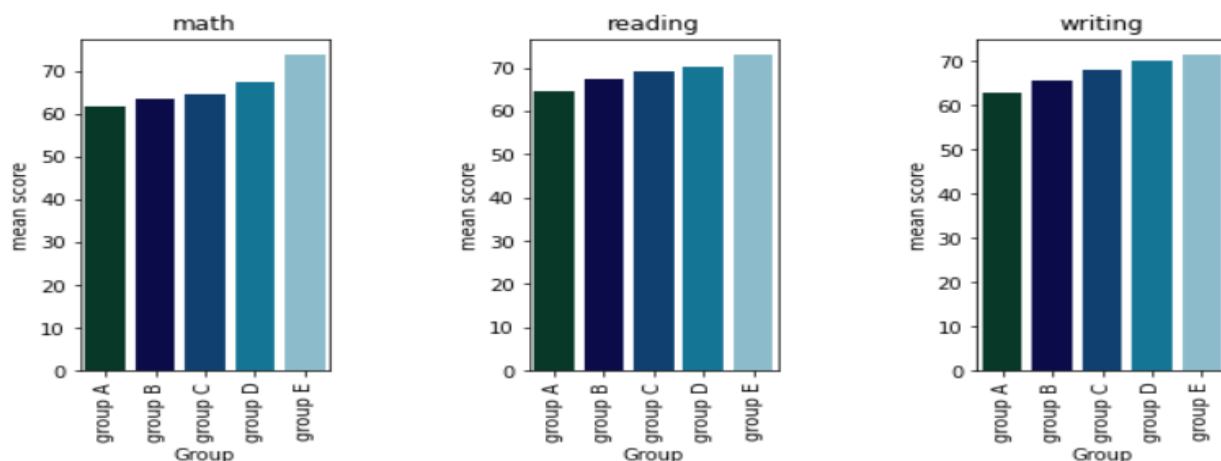
Firstly, we compared the scores of males and females in all the three subjects to seek the correlation between gender and scores.



From the results we can see that gender does not seem to affect the scores in a critical way, we can see that males perform better in maths, whereas females are better in reading and writing.

#### 3.2 Ethnicity vs. Score

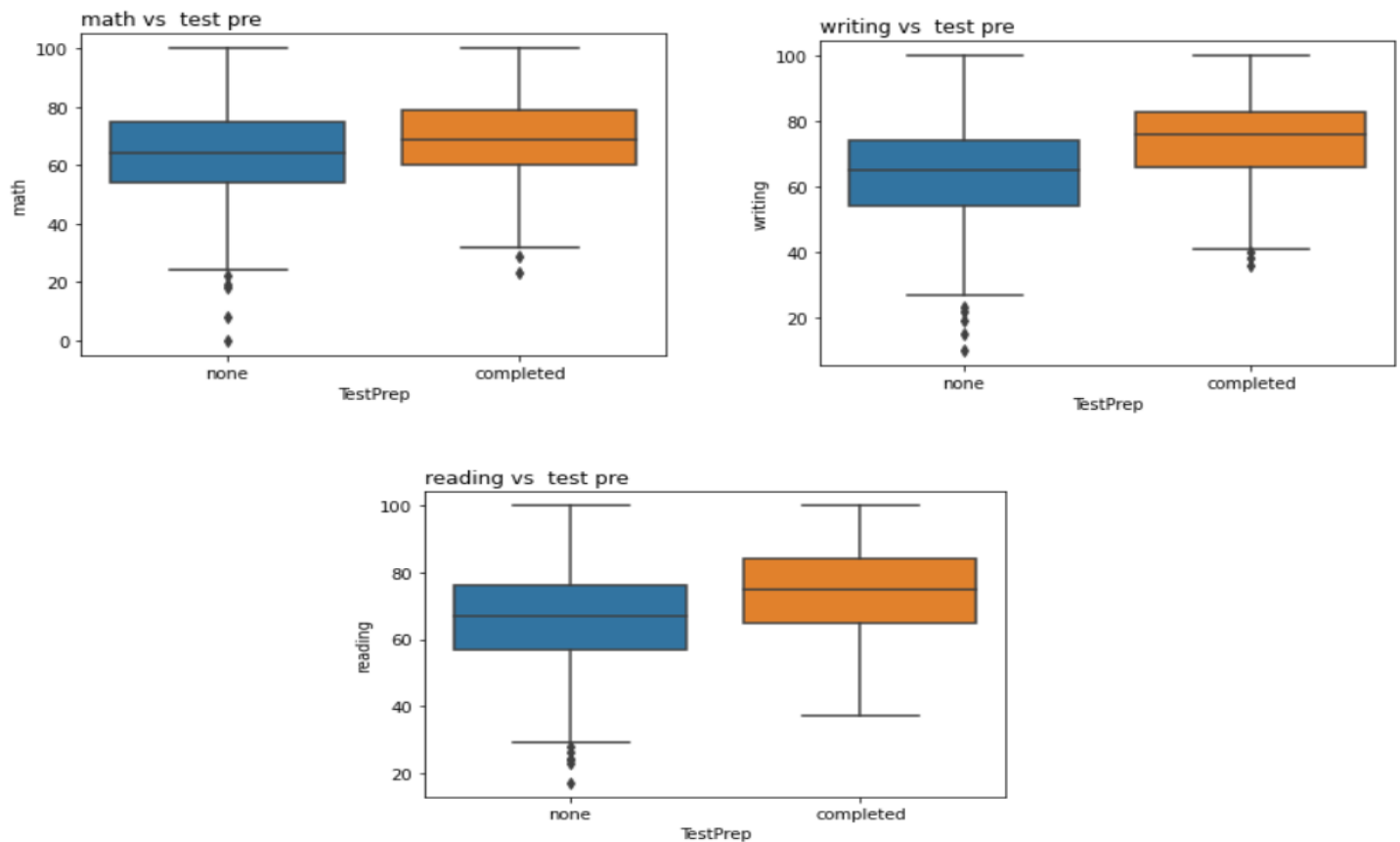
Then, we compared the scores for different ethnic groups, from the results we found out that group E performs the best and group A performs the worst in all the subjects.



From the results we can infer that ethnicity has some role to play in the performance.

### 3.3 Role of preparation before test

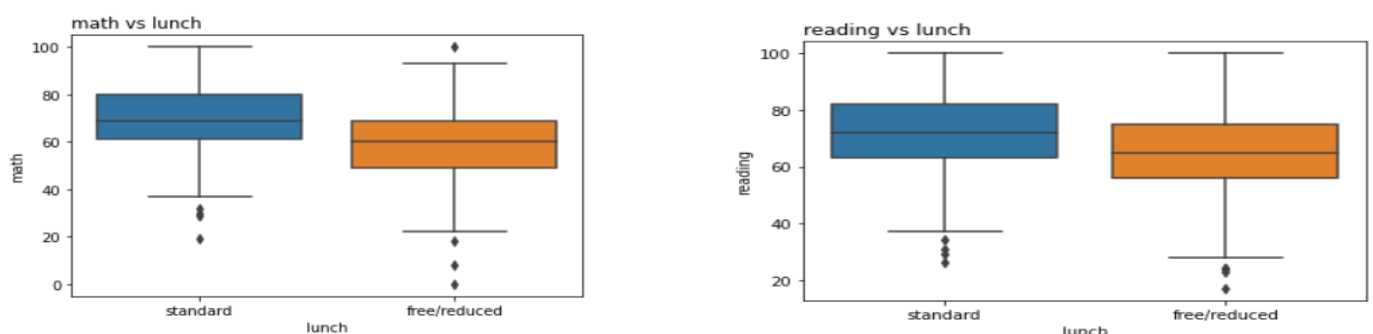
Now, we separated the students who have prepared for the exams and those who have not, and we plotted their marks in all three subjects.

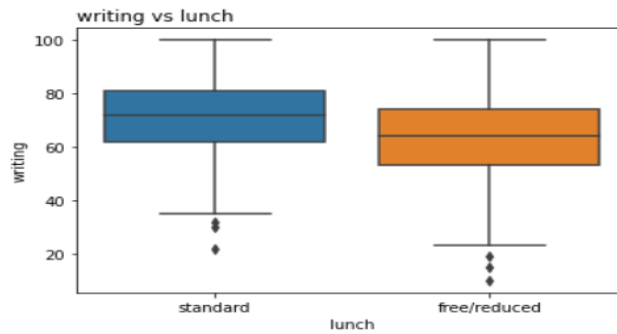


We found out that students who prepared for the exams performed significantly better, the average score increased, and the distribution got narrower which signifies that there is a lot of inconsistency in the marks of the students who did not prepare for the exams i.e., their marks were spread out in a larger range, but those who prepared for the exams performed with better consistency.

### 3.4 Role of Nutrition

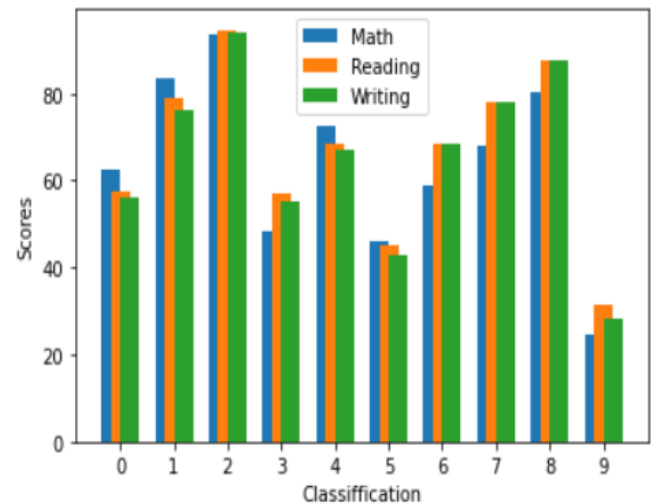
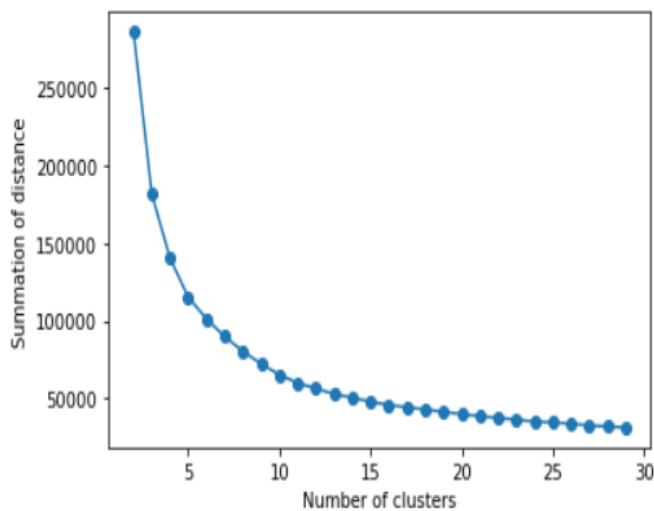
We also compared the performance of students who had standard amount of lunch with those who had insufficient amount of lunch and found out some interesting insights. Students who had standard lunch performed significantly better than those who did not had lunch (or had reduced lunch). The average marks of students who had lunch were much higher than those who did not had lunch.



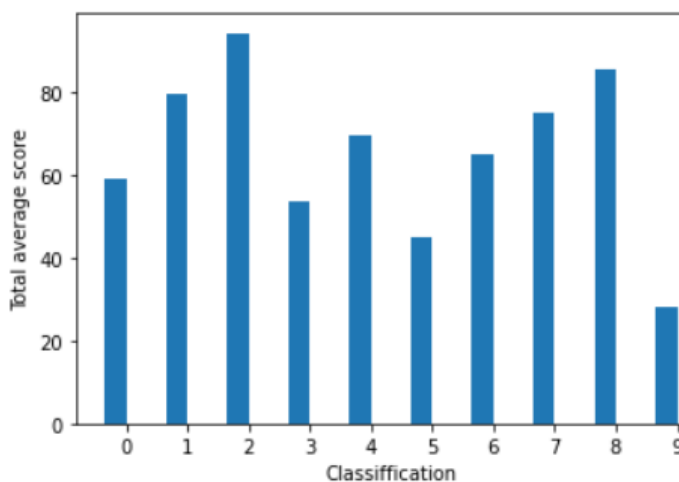


### 3.5 K-Means Clustering

Now, we encoded the education level of parents into 5 groups numbered from 0-4 and encoded the values of lunch and test preparation having values either 0 or 1. Further we used K-means clustering algorithm to classify the data. We found the optimal value of the clusters to be 10 (elbow point). We plotted the scores of different clusters in all subjects and found out that cluster 2 performed the best and cluster 9 performed the worst.



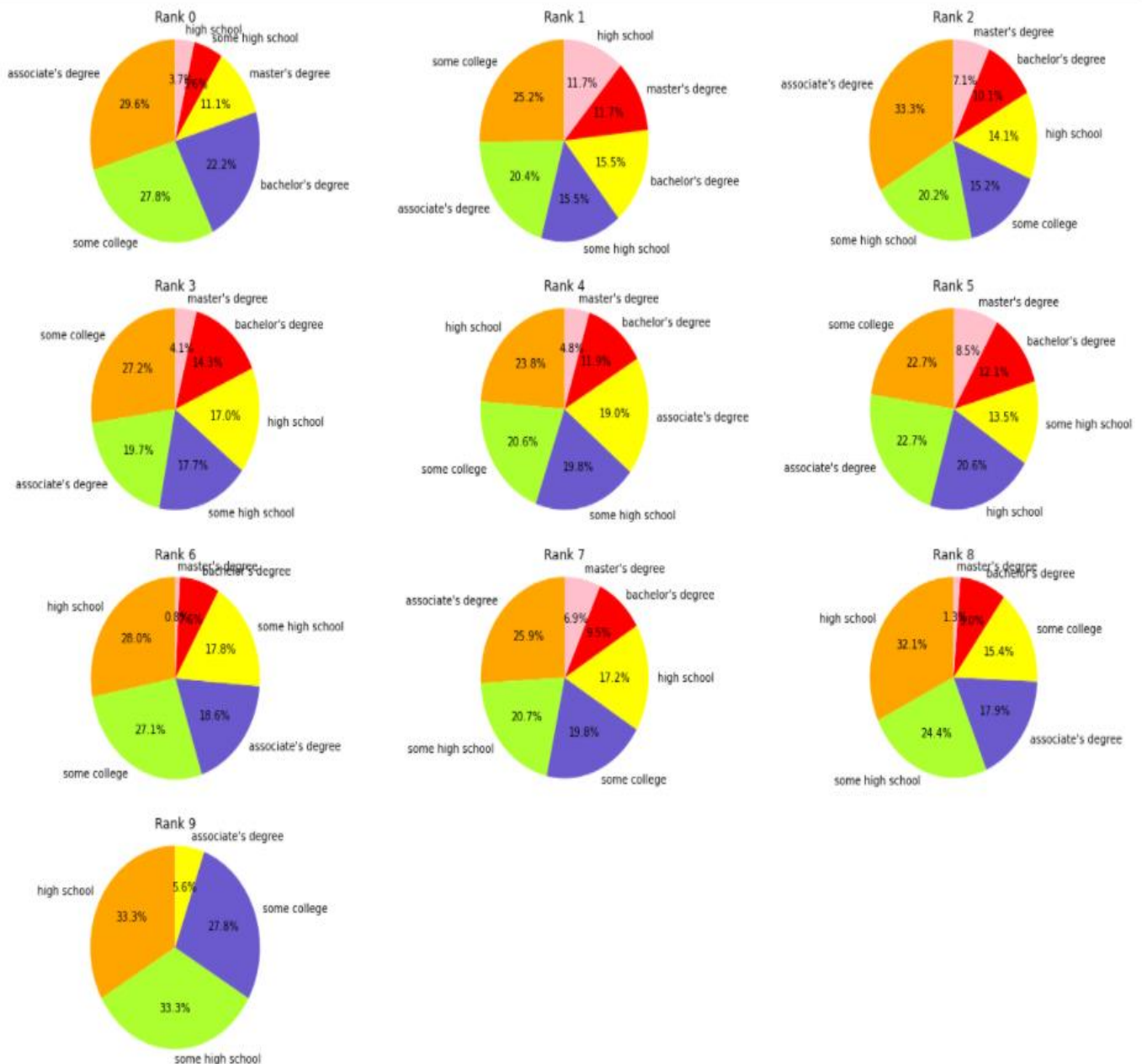
To get more insights on total performance, and rankings of all clusters we plotted the overall performance of all clusters.



```
In [55]: rank
Out[55]: classification
2      94.098765
8      85.271845
1      79.501684
7      74.678005
4      69.444444
6      65.066194
0      58.810734
3      53.402299
5      44.799145
9      28.055556
Name: total_ave_score, dtype: float64
```

### 3.6 Insights on role of education level of parents

By clustering the data, we can now get a better idea of the correlation between the performance and features. Now we plotted the pie charts for all the different clusters to find whether education of parents affects the overall performance.



From the pie charts we can see that different clusters have different composition of parents with a certain education level. The cluster with rank 0 (best) has 33.3% students whose parents have an associates degree. Whereas rank 9 (worst) cluster has the majority of students whose parents have high school level of education. But as we can see the graphs above, rank 8 cluster performs better than rank 7 although the majority of parents in rank 8 have just high school education whereas rank 7 has majority of parents with associates degree but they still underperform. Also, we can see that many of the clusters which have higher number of parents with master's degree are outperformed those which have majority of parents with bachelor's degree.

So, we can infer that education level of parents is important to a certain extent, but its role isn't crucial to the performance of students in the exams.

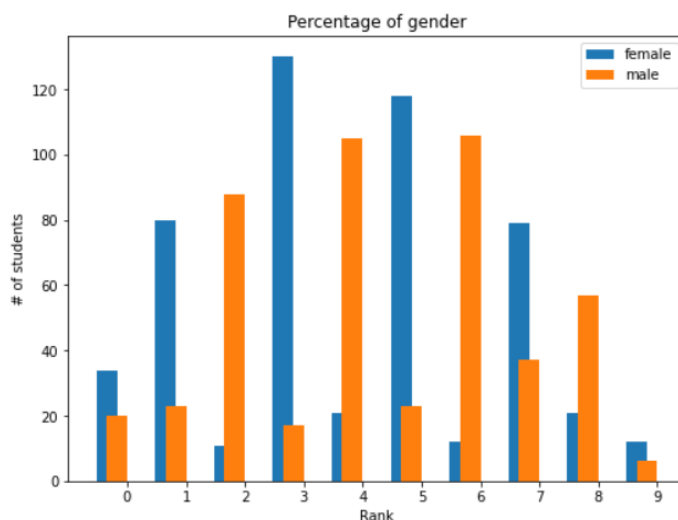
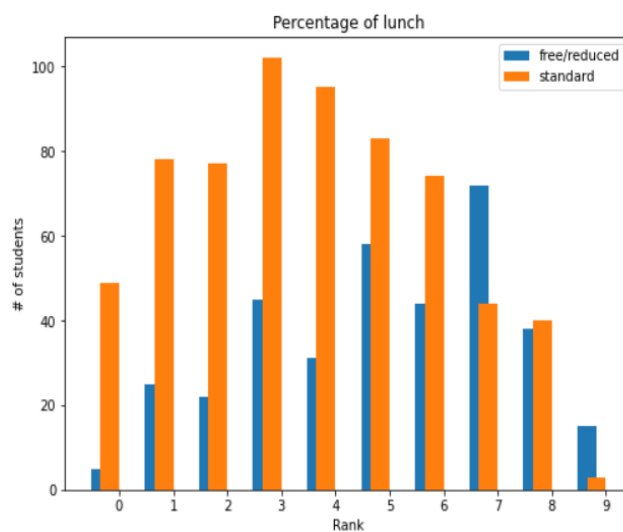
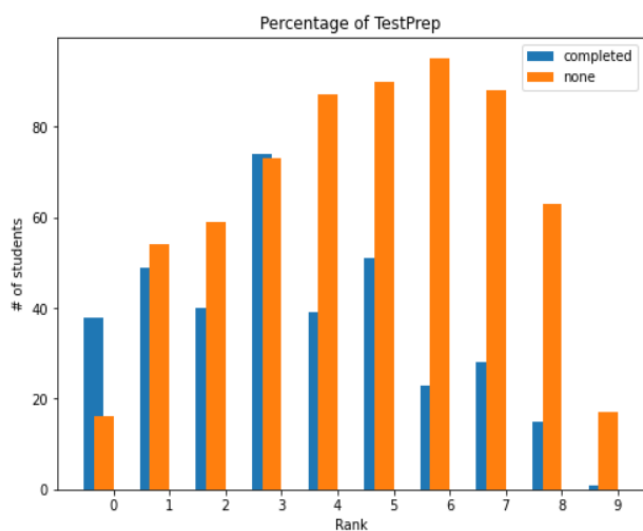
### 3.7 Role of test preparation, nutrition and gender among clusters.

In the end, to see the role of test preparation, lunch and gender in each cluster, we plotted the number of students vs rank of each cluster.

Keeping in mind that cluster 2 (rank 0) is the best cluster and cluster 9 (rank 9) is the worst we individually examine the graphs. Firstly, we can see that those clusters which have a greater number of students who prepared for the exams show better results, rank 0 has more than 50% students who have prepared for exams whereas rank 9 cluster has less than 10% students who have prepared. From this observation we can infer that preparation for the test is critical for having a better performance.

We can also see that almost 90% of rank 0 students had standard lunch whereas less than 20% of rank 9 students had lunch, which also proves the point that nutrition is also critical for good performance in exams.

We can now take a look of gender distribution among the clusters, both rank 0 and rank 9 clusters have twice the number of girls as they have boys and there is a lot of difference between their performances. From the graphs we can see no relation between performance of the cluster and the gender ratio, so its safe to assume that gender plays no role in performance in exams.



## **4 Conclusion**

In this report we tried to find correlation between different features and the performance in exams, and we found out some interesting insights. We saw that the two crucial features for performance in exams is test preparation and nutrition, which is intuitive to understand as concentration and retaining ability depends upon nutrition which are important factors for better performance. Nutrition and test preparation have strong correlation with the performance and are a deciding factor in the overall performance, the ethnicity and education level of parents do have some impact on the performance, but they aren't a crucial factor in determining the results, whereas gender seems to play no role in determining the performance of students.

In summary we can say that for better results, students should focus on test preparation and have proper nutrition, other things such as education of parents and ethnicity is not a deciding factor and its is not under our control, and gender plays no role whatsoever in performance in exams.

## **5 Link for the dataset**

<https://github.com/rashida048/Datasets/blob/master/StudentsPerformance.csv>