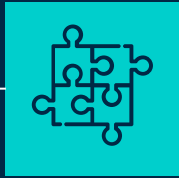




GROUP 12 FINAL PRESENTATION

By Grace Kane, Adrian
Jeziorski, Dom Garofola and
Caroline Favelson

TABLE OF CONTENTS



01

OUR STORY



02

QUESTIONS
AND METHODS



03

OUTCOMES AND
OVERVIEW

DATASET

LINK

<https://www.kaggle.com/datasets/lainguyn123/student-performance-factors>

OUR STORY: Education 575

What do we do:

- We are an education marketing firm designed to help schools optimize their recruitment, retention, and overall student success strategies

Services:

- Student Experience Analytics, Targeted Enrollment Campaigns, Curriculum and Program Improvement, Alumni Engagement

Empowering Schools
Enhancing Futures



We use advanced data analytics, student insights, and personalized feedback loops to create tailored solutions that improve the educational experience and outcomes for every learner

We help institutions understand their unique needs, preferences, and behaviors, enabling them to craft more effective academic offerings and improve overall results

OUR GOAL

By developing regressions and creating experiments, we could find the impact of one behavior on the others and evaluate any correlations between them. This information would be able to help the schools focus on certain areas to improve students scores and overall results maximizing the schools potential



DATASET INFO/DESCRIPTION

Number of Records: 6,607

Number of Features: 20

File Format: CSV

Hours_Studied

Attendance

Parental_Involvement

Access_to_Resources

Extracurricular_Activities

Sleep_Hours

School_Type

Peer_Influence

Learning_Disabilities

Parental_Education_Level

Motivation_Level

Internet_Access

Tutoring_Sessions

Previous_Scores

Family_Income

Teacher_Quality

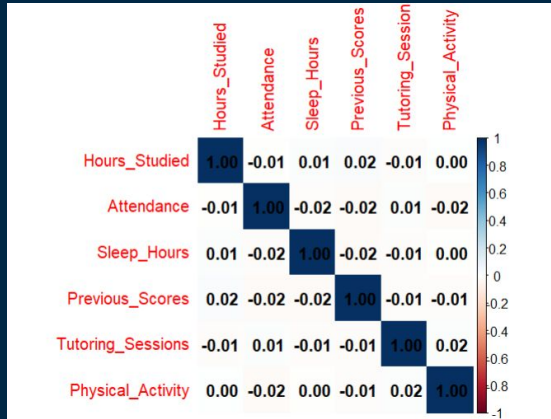
Gender

Physical_Activity

Distance_from_Home

Exam_Score

Data Preprocessing



Grid of 0s and 1s representing dummy variables for categorical data.

NO Missing Values

NO Multicollinearity

Categorical Variables Transformed to **n-1** dummy variables

Regression - What has the largest impact on scores?

- Method: regression analysis ($R^2 = 72.75\%$)
- Factors school can affect: access to resources, motivation level, teacher quality, extracurricular activities, tutoring sessions, hours studied, attendance, physical activity, and learning disabilities
- Invest in better access to resources, teacher quality, and extracurricular activities
- Prioritize mental health and academic interest of the students to increase motivation
- Increase resources for students with disabilities
- Incentivize students to go to tutoring sessions and increase attendance and study hours
- Incorporate more physical activity in curriculum

Coefficients:	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	35.005289	0.329783	106.147	< 2e-16
Hours_Studied	0.294834	0.004188	70.400	< 2e-16
Attendance	0.198753	0.002176	91.357	< 2e-16
Extracurricular_Activities	0.558975	0.051151	10.928	< 2e-16
Sleep_Hours	-0.001390	0.017090	-0.081	0.9352
Previous_Scores	0.048853	0.001745	27.998	< 2e-16
Internet_Access	0.929630	0.094940	9.792	< 2e-16
Tutoring_Sessions	0.496896	0.020376	24.386	< 2e-16
School_Type	0.029134	0.054559	0.534	0.5934
Physical_Activity	0.187809	0.024384	7.702	1.53e-14
Learning_Disabilities	-0.855836	0.081761	-10.467	< 2e-16
Gender	-0.040520	0.050762	-0.798	0.4248
Parental_InvolvementHigh	1.984975	0.072807	27.264	< 2e-16
Parental_InvolvementMedium	0.926382	0.065948	14.047	< 2e-16
Access_to_ResourcesHigh	2.055914	0.072644	28.301	< 2e-16
Access_to_ResourcesMedium	1.051786	0.066573	15.799	< 2e-16
Motivation_LevelHigh	1.060777	0.072807	14.570	< 2e-16
Motivation_LevelMedium	0.517726	0.058264	8.886	< 2e-16
Family_IncomeHigh	1.080151	0.069530	15.535	< 2e-16
Family_IncomeMedium	0.494506	0.055846	8.855	< 2e-16
Teacher_Quality	0.505949	0.244453	2.070	0.0385
Teacher_QualityHigh	1.051641	0.092035	11.427	< 2e-16
Teacher_QualityMedium	0.504725	0.086059	5.865	4.71e-09
Peer_InfluenceNeutral	0.519686	0.068043	7.638	2.53e-14
Peer_InfluencePositive	1.027375	0.067784	15.157	< 2e-16
Parental_Education_Level	0.154660	0.218087	0.709	0.4782
Parental_Education_LevelCollege	0.487117	0.058180	8.373	< 2e-16
Parental_Education_LevelPostgraduate	0.981954	0.066893	14.680	< 2e-16
Distance_from_Home	-0.467976	0.251413	-1.861	0.0627
Distance_from_HomeFar	-0.909371	0.085978	-10.577	< 2e-16
Distance_from_HomeModerate	-0.520908	0.056133	-9.280	< 2e-16

K-Means Cluster Analysis - Hours Studied and Hours Slept

Descriptives

[DataSet2]

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Hours_Studied	6607	1	44	19.98	5.991
Sleep_Hours	6607	4	10	7.03	1.468
Valid N (listwise)	6607				

Quick Cluster

Initial Cluster Centers

	Cluster		
	1	2	3
Zscore(Hours_Studied)	-3.16752	-.16281	3.00883
Zscore(Sleep_Hours)	-2.06322	2.02364	-2.06322

Iteration History^a

Iteration	Change in Cluster Centers		
	1	2	3
1	2.221	1.593	2.169
2	.313	.236	.243
3	.107	.123	.127
4	.070	.063	.044
5	.068	.077	.056
6	.000	.049	.054
7	.051	.047	.000
8	.023	.069	.053
9	.014	.015	.001
10	.000	.000	.000

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is .000. The current iteration is 10. The minimum distance between initial centers is 5.073.

Final Cluster Centers

	Cluster		
	1	2	3
Zscore(Hours_Studied)	-.93968	.08357	.78396
Zscore(Sleep_Hours)	-.50380	1.08975	-.56890

ANOVA

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Zscore(Hours_Studied)	1651.349	2	.500	6604	3301.397	<.001
Zscore(Sleep_Hours)	1938.184	2	.413	6604	4689.192	<.001

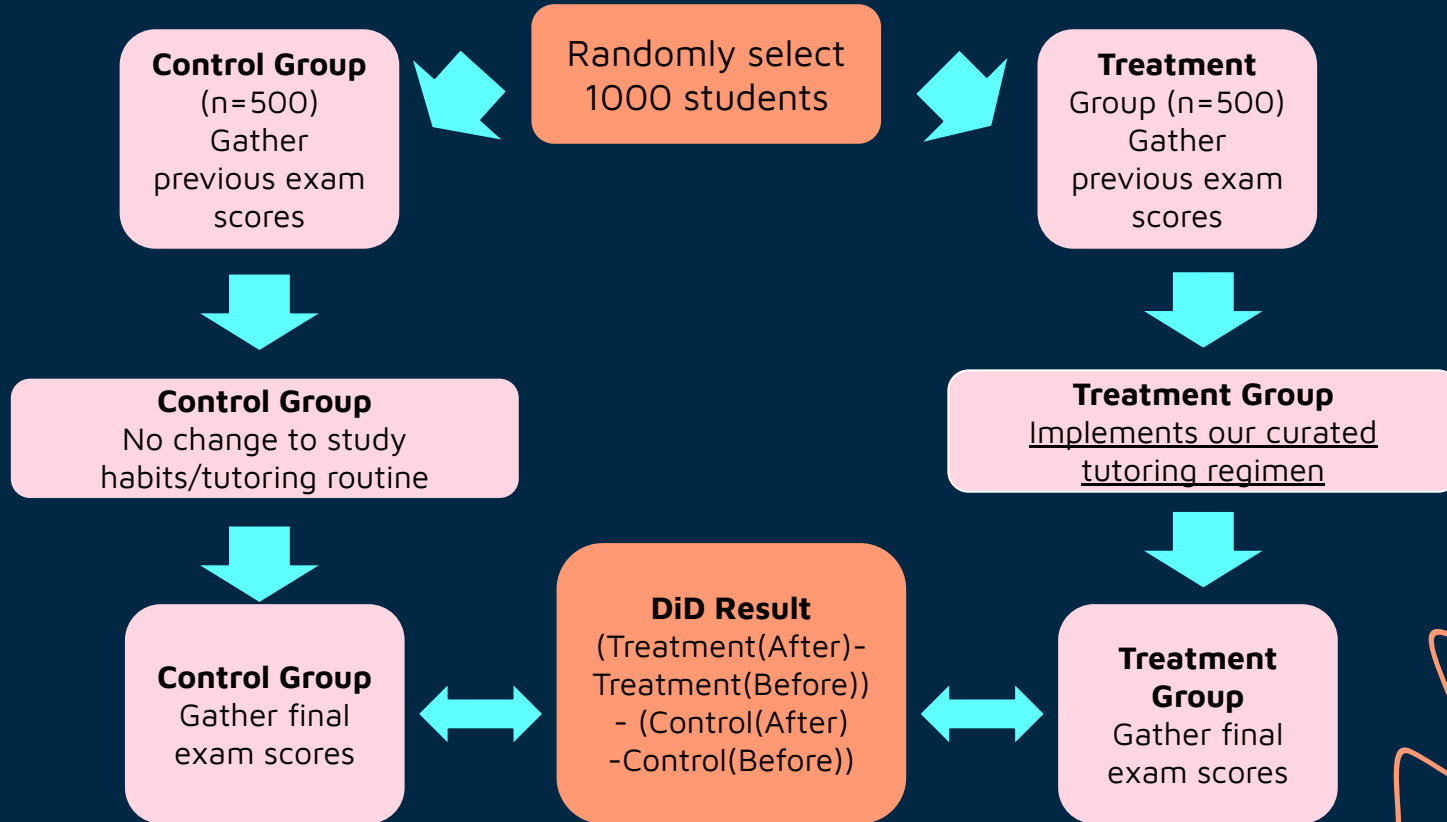
The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Number of Cases in each Cluster

Cluster	1	2118.000
	2	2183.000
	3	2306.000
Valid		6607.000
Missing		.000

Potential Intervention: A DiD Design

How can we determine if our intervention improves exam scores?



Managerial Implications

Regression Analysis

- Help the school maximize their resources given the budget
- Lead to improved exam scores among students

K Means Clustering

- Figure out why certain students have lower levels of study hours and sleep
- Sleep less + study more cluster vs. well balanced cluster



QUESTIONS?

THANK YOU

