DETERMINATION OF PERFORMANCE OF MATH STUDENTS

1. Problem Definition:

   What are the strength and weakness of the Math students under study?

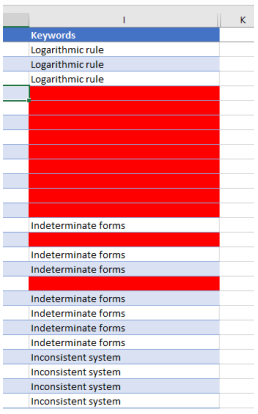
2. Data Collection:

   Dataset for Assessing Mathematics Learning in Higher Education [1]

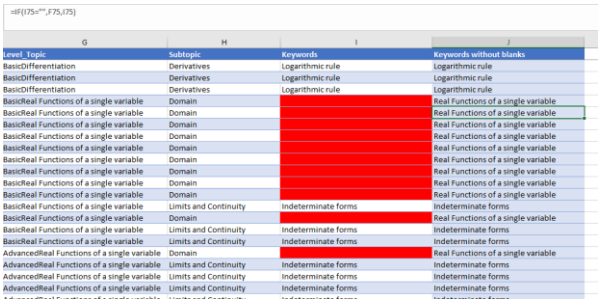
   MathE is a mathematical platform developed under the MathE project (mathe.pixel-online.org). The dataset has 9546 answers to questions in the Mathematical topics taught in higher education. The file has eight features, named: Student ID, Student Country, Question ID, Type of answer (correct or incorrect), Question level (basic or advanced), Math Topic, Math Subtopic, and Question Keywords. The question level was associated with the professor who submitted the question. The data was obtained from February 2019 until December 2023.

3. Data Cleaning:

   Removal of Missing Values: The *keywords* column contains some missing values that can potentially affect the analysis of the data. These missing values were highlighted with the Conditional Formatting feature of Excel as shown below.

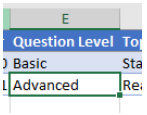


   The missing values were further replaced with their corresponding topics with the use of an IF function on a new column, *keywords without blanks,* as shown below. The formula used is also shown in the formula box.



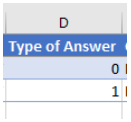
G	H	I	J
Level_Topic	Subtopic	Keywords	Keywords without blanks
BasicDifferentiation	Derivatives	Logarithmic rule	Logarithmic rule
BasicDifferentiation	Derivatives	Logarithmic rule	Logarithmic rule
BasicDifferentiation	Derivatives	Logarithmic rule	Logarithmic rule
BasicReal Functions of a single variable	Domain		Real Functions of a single variable
BasicReal Functions of a single variable	Domain		Real Functions of a single variable
BasicReal Functions of a single variable	Domain		Real Functions of a single variable
BasicReal Functions of a single variable	Domain		Real Functions of a single variable
BasicReal Functions of a single variable	Domain		Real Functions of a single variable
BasicReal Functions of a single variable	Domain		Real Functions of a single variable
BasicReal Functions of a single variable	Domain		Real Functions of a single variable
BasicReal Functions of a single variable	Domain		Real Functions of a single variable
BasicReal Functions of a single variable	Limits and Continuity	Indeterminate forms	Indeterminate forms
BasicReal Functions of a single variable	Domain		Real Functions of a single variable
BasicReal Functions of a single variable	Limits and Continuity	Indeterminate forms	Indeterminate forms
BasicReal Functions of a single variable	Limits and Continuity	Indeterminate forms	Indeterminate forms
AdvancedReal Functions of a single variable	Domain		Real Functions of a single variable
AdvancedReal Functions of a single variable	Limits and Continuity	Indeterminate forms	Indeterminate forms
AdvancedReal Functions of a single variable	Limits and Continuity	Indeterminate forms	Indeterminate forms
AdvancedReal Functions of a single variable	Limits and Continuity	Indeterminate forms	Indeterminate forms

4. Data Exploration:



E
Question Level
Basic
Advanced

Using advanced filters, the unique values in the *Question Level* column are Basic and Advanced. This means these are the two question levels that we have in the dataset.



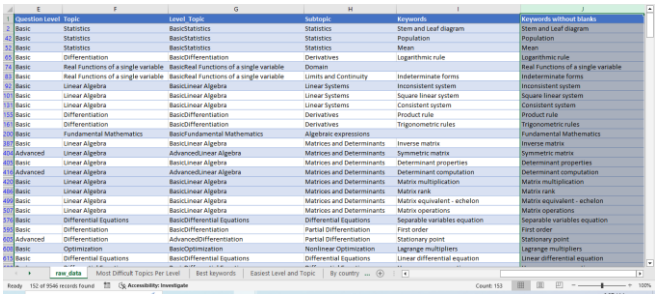
D
Type of Answer
0
1

Likewise, the *Type of Answer* column has only two unique values, 0 and 1 as shown above.

The students were tested on 14 different topics of Math as shown below.

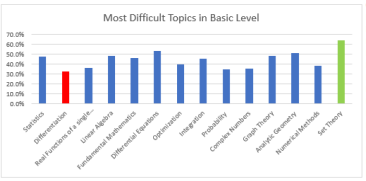
Topic
Statistics
Differentiation
Real Functions of a single variable
Linear Algebra
Fundamental Mathematics
Differential Equations
Optimization
Integration
Probability
Complex Numbers
Graph Theory
Analytic Geometry
Numerical Methods
Set Theory

After removing the missing values in the *keywords* column, the new column, *keywords without blanks,* contains 152 unique values as shown at the bottom left of the Excel worksheet shown below

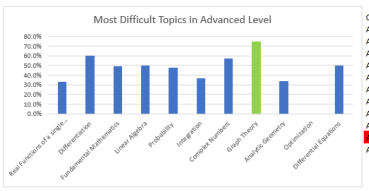


Question Level	Topic	Level_Topic	Subtopic	Keywords	Keywords without blanks
Basic	Statistics	BasicStatistics	Statistics	Stem and Leaf diagram	Stem and Leaf diagram
Basic	Statistics	BasicStatistics	Statistics	Population	Population
Basic	Statistics	BasicStatistics	Statistics	Mean	Mean
Basic	Differentiation	BasicDifferentiation	Derivatives	Logarithmic rule	Logarithmic rule
Basic	Real Functions of a single variable	BasicReal Functions of a single variable	Domain		Real Functions of a single variable
Basic	Real Functions of a single variable	BasicReal Functions of a single variable	Limits and Continuity	Indeterminate forms	Indeterminate forms
Basic	Linear Algebra	BasicLinear Algebra	Linear Systems	Inconsistent system	Inconsistent system
Basic	Linear Algebra	BasicLinear Algebra	Linear Systems	Square linear system	Square linear system
Basic	Linear Algebra	BasicLinear Algebra	Linear Systems	Consistent system	Consistent system
Basic	Differentiation	BasicDifferentiation	Derivatives	Product rule	Product rule
Basic	Differentiation	BasicDifferentiation	Derivatives	Trigonometric rules	Trigonometric rules
Basic	Fundamental Mathematics	BasicFundamental Mathematics	Algebraic expressions		Fundamental Mathematics
Basic	Linear Algebra	BasicLinear Algebra	Matrices and Determinants	Inverse matrix	Inverse matrix
Advanced	Linear Algebra	AdvancedLinear Algebra	Matrices and Determinants	Symmetric matrix	Symmetric matrix
Basic	Linear Algebra	BasicLinear Algebra	Matrices and Determinants	Determinant properties	Determinant properties
Advanced	Linear Algebra	AdvancedLinear Algebra	Matrices and Determinants	Determinant computation	Determinant computation
Basic	Linear Algebra	BasicLinear Algebra	Matrices and Determinants	Matrix multiplication	Matrix multiplication
Basic	Linear Algebra	BasicLinear Algebra	Matrices and Determinants	Matrix rank	Matrix rank
Basic	Linear Algebra	BasicLinear Algebra	Matrices and Determinants	Matrix equivalent - echelon	Matrix equivalent - echelon
Basic	Linear Algebra	BasicLinear Algebra	Matrices and Determinants	Matrix operations	Matrix operations
Basic	Differential Equations	BasicDifferential Equations	Differential Equations	Separable variables equation	Separable variables equation
Basic	Differentiation	BasicDifferentiation	Partial Differentiation	First order	First order
Advanced	Differentiation	AdvancedDifferentiation	Partial Differentiation	Stationary point	Stationary point
Basic	Optimization	BasicOptimization	Nonlinear Optimization	Lagrange multipliers	Lagrange multipliers
Basic	Differential Equations	BasicDifferential Equations	Differential Equations	Linear differential equation	Linear differential equation

5. Data Analysis:



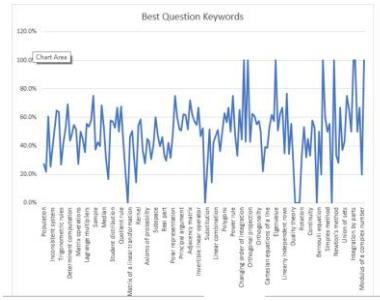
Question Level	Topic	Correct Answers	Total Answers	Pass Rate
Basic	Statistics	163	340	47.9%
Basic	Differentiation	183	554	33.0%
Basic	Real Functions of a single variable	44	122	36.1%
Basic	Linear Algebra	2216	4538	48.8%
Basic	Fundamental Mathematics	350	755	46.4%
Basic	Differential Equations	57	106	53.8%
Basic	Optimization	70	154	45.5%
Basic	Integration	57	125	45.6%
Basic	Probability	37	105	35.2%
Basic	Complex Numbers	118	328	36.0%
Basic	Graph Theory	17	35	48.6%
Basic	Analytic Geometry	158	308	51.3%
Basic	Numerical Methods	120	310	38.7%
Basic	Set Theory	27	42	64.3%



Question Level	Topic	Correct Answers	Total Answers	Pass Rate
Advanced	Real Functions of a single variable	14	42	33.3%
Advanced	Differentiation	15	25	60.0%
Advanced	Fundamental Mathematics	31	63	49.2%
Advanced	Linear Algebra	591	1188	49.7%
Advanced	Probability	11	23	47.8%
Advanced	Integration	7	19	36.8%
Advanced	Complex Numbers	151	264	57.2%
Advanced	Graph Theory	15	20	75.0%
Advanced	Analytic Geometry	17	50	34.0%
Advanced	Optimization	0	6	0.0%
Advanced	Differential Equations	1	2	50.0%

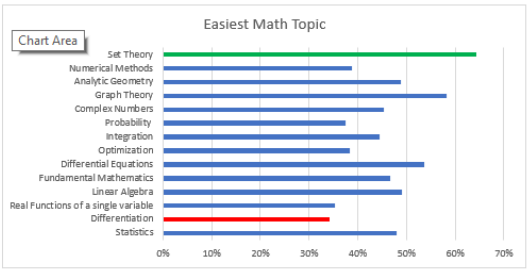
The chart above suggests that the Set Theory is the easiest topic for students who did the test for Basic Level. But the topic that students found most difficult in the Basic level is Differentiation. Topics like Statistics, Differential Equations, and Analytic Geometry have quite a high pass rate, which means students also find them easy in the basic Level.

For the advanced level questions, Graph theory is a doddle for the students. But it is pathetic that no student scored a point in Optimization questions.

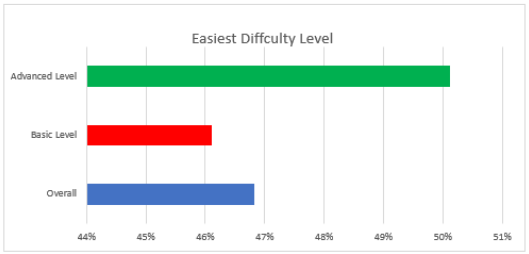


Keywords	Correct Answers	Total Answers	Pass Rate
Stem and Leaf diagram	11	40	27.5%
Population	5	23	21.7%
Mean	26	43	60.5%
Logarithmic rule	16	63	25.4%
Indeterminate forms	10	25	40.0%
Inconsistent system	118	233	50.6%
Square linear system	48	74	64.9%
Consistent system	37	58	63.8%
Product rule	22	82	26.8%
Trigonometric rules	48	115	41.7%
Inverse matrix	13	25	52.0%
Symmetric matrix	9	13	69.2%
Determinant properties	7	16	43.8%
Determinant computation	36	76	47.4%
Matrix multiplication	23	42	54.8%
Matrix rank	15	29	51.7%
Matrix equivalent - echelon	3	11	27.3%
Matrix operations	21	42	50.0%
Separable variables equation	9	20	45.0%
First order	69	194	35.6%
Stationary point	5	9	55.6%
Lagrange multipliers	8	15	53.3%
Linear differential equation	41	71	57.7%
Homogeneous equation	6	8	75.0%
Second order	12	32	37.5%

The math students found (but not limited to) Iterate Integral, y simple region, eigenvalue, chromatic number, nonlinear equation, area of a planar region, easiest during the test. But they found some other keywords difficult. They include Euclidean space, population, linear application and others as illustrated in the chart.

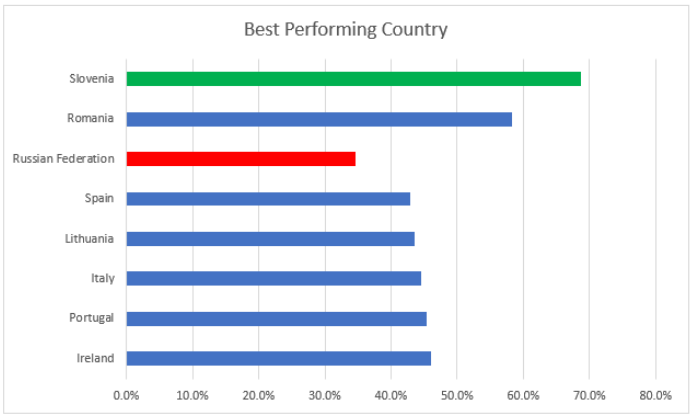


	Correct Answers	Incorrect Answers	Total Answers	Pass Rate
Statistics	163	177	340	48%
Differentiation	198	381	579	34%
Real Functions of a single variab	58	106	164	35%
Linear Algebra	2807	2919	5726	49%
Fundamental Mathematics	381	437	818	47%
Differential Equations	58	50	108	54%
Optimization	70	112	182	38%
Integration	64	80	144	44%
Probability	48	80	128	38%
Complex Numbers	269	323	592	45%
Graph Theory	32	23	55	58%
Analytic Geometry	175	183	358	49%
Numerical Methods	120	190	310	39%
Set Theory	27	15	42	64%



	Correct Answers	Incorrect Answers	Total Answers	Pass Rate
Overall	4470	5076	9546	47%
Basic Level	3617	4227	7844	46%
Advanced Level	853	849	1702	50%

Without considering the question difficulty level, the math students passed well in questions under the Set theory topic and did directly otherwise (that is, worst performance) in the questions under Differentiation topic. Also the overall pass rate of the math students across all topics and difficulty level is 47%, close to average. Even though questions in the advanced level are advanced questions, students still found their ways to pass at advanced level questions that basic level questions percent margin of 4%, with the pass rate of Advanced level questions on the average benchmark of 50%.



Country	Correct Answe	Incorrect A	Total Ans	Pass Rate
Ireland	138	162	300	46.0%
Portugal	2494	3001	5495	45.4%
Italy	606	752	1358	44.6%
Lithuania	629	814	1443	43.6%
Spain	12	16	28	42.9%
Russian Federation	37	70	107	34.6%
Romania	35	25	60	58.3%
Slovenia	519	236	755	68.7%

Evaluating the performance of students across the countries reveals that Math students who came from Slovenia are the Math geniuses in this case study. Followed tightly by Romania. Students from the Russian Federation however need help on their Math because they have a low pass rate of 34.6%.

6. Interpretation of Results:

"I love Math" is the slogan of some students. While some of them will not even dare befriending you if you have a math note. In essence, students can have different perspectives about math and this can affect their performance in a math test. This

analysis project doesn't check the validity of this perspective-result relationship. It rather aims to evaluate the performance of a given set of students in a math test to identify their weaknesses and strengths.

This analysis shows that the difficulty level of the question – basic or advanced - is not the most important factor that determines the strength or weakness of a student in a math test. Students in this study shockingly passed better at advanced questions than basic questions.

Set theory and Graph theory are a cinch for students in the basic and advanced levels respectively. In the basic level questions, students likewise found like Statistics, Differential Equations, and Analytic Geometry easy for them. Complex numbers, differentiation and differential equations were also easy for students in the advanced level during the test.

Differentiation is the nightmare of students who did the basic level questions. Others topics like the Complex numbers, probability, and real functions of single variable are all difficult for students in this difficulty level too. For the advanced level, Optimization is a myth! Her sisters are the real function of a single variable and integration; they were all difficult for the students who did advanced level questions during the test.

7. References

[1] Flamia Azevedo, B., Pacheco, M., P. Fernandes, F., & Pereira, A. (2024). Dataset for Assessing Mathematics Learning in Higher Education [Dataset]. UCI Machine Learning Repository. https://doi.org/10.34620/dadosipb/PW3OWY.