

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC MỞ THÀNH PHỐ HỒ CHÍ MINH**



2151050455 - Trần Đăng Mỹ Tiên

2151050518 - Nguyễn Tuyễn

BÀI TẬP MÔN KHAI PHÁ DỮ LIỆU

**Chủ đề: Phân tích các vấn đề giáo dục của các quốc gia
trên thế giới**

Lớp: DH21IT03

TP. HỒ CHÍ MINH, 2023

MỤC LỤC

Chương 1. MÔ TẢ DỮ LIỆU BÀI TOÁN	7
Chương 2. TIỀN XỬ LÝ DỮ LIỆU.....	19
2.1. Thu thập dữ liệu	19
2.2. Sửa chữa dữ liệu.....	19
2.2.1. Dataset Global Education	19
2.3. Làm sạch dữ liệu	21
2.4. Chuẩn hóa dữ liệu	22
Chương 3. PHÂN TÍCH DỮ LIỆU.....	23
3.1. Sự tương quan giữa các cột trong dữ liệu	23
3.2. Sự tương quan giữa tỷ lệ sinh với các cột.....	24
3.3. Sự tương quan giữa tỷ thất nghiệp với các cột	26
3.4. Tỷ lệ bỏ học ở độ tuổi mầm non	27
3.5. Tỷ lệ thất nghiệp của các quốc gia.....	28
3.5.1. Top 10 quốc gia/ khu vực có tỷ lệ thất nghiệp cao nhất.....	28
3.6. Tỷ lệ tuyển sinh.....	30
3.6.1. Tỷ lệ tuyển sinh tiểu học	31
3.6.2. Tỷ lệ tuyển sinh đại học	32
3.6.3. Tỷ lệ nhập học trung bình.....	32
3.6.4. Tỷ lệ tuyển sinh chung	33
3.7. Khả năng đọc thành thạo, trình độ Toán học theo quốc gia	33
3.7.1. Khả năng đọc thành thạo	33
3.7.2. Thành thạo toán học	36
3.7.3. Mức độ thông thạo (đọc, toán học) trung bình ở các cấp độ giáo dục	39
3.8. Tỷ lệ biết chữ ở thanh niên	40
3.9. Tỷ lệ hoàn thành các cấp bậc giáo dục	42
3.9.1. Tỷ lệ hoàn thành ở bậc tiểu học.....	43
3.9.2. Tỷ lệ hoàn thành bậc trung học cơ sở.....	43
3.9.3. Tỷ lệ hoàn thành bậc trung học phổ thông	44
3.9.4. Tỷ lệ hoàn thành trung bình ở cấp độ giáo dục	45
3.10. Tỷ lệ ngoài nhà trường.....	46
3.10.1. Tỷ lệ ngoài nhà trường trung bình.....	47
3.10.2. Tỷ lệ ngoài nhà trường ở các cấp độ giáo dục	48
Chương 4. PHÂN CỤM	50
4.1. Phân cụm theo tỷ lệ ngoài nhà trường	50
4.2. Phân cụm theo tỷ lệ hoàn thành các bậc học	53
4.3. Phân cụm theo tỷ lệ biết chữ của thanh thiếu niên độ tuổi 15-24.....	55
Chương 5. CÂY QUYẾT ĐỊNH.....	56

Chương 6. NHẬT XÉT TỔNG QUAN.....60

DANH MỤC TỪ VIẾT TẮT

1. **Countries and Areas:** Name of the countries and areas.(Tên quốc gia và khu vực.)
2. **Latitude:** Latitude coordinates of the geographical location.(Tọa độ vĩ độ của vị trí địa lý.)
3. **Longitude:** Longitude coordinates of the geographical location.(Tọa độ kinh độ của vị trí địa lý.)
4. **OOSR_Pre0Primary_Age_Male:** Out-of-school rate for pre-primary age males.(Tỷ lệ bỏ học của nam giới độ tuổi mầm non.)
5. **OOSR_Pre0Primary_Age_Female:** Out-of-school rate for pre-primary age females.(Tỷ lệ bỏ học của nữ độ tuổi mầm non.)
6. **OOSR_Primary_Age_Male:** Out-of-school rate for primary age males.(Tỷ lệ bỏ học của nam độ tuổi tiểu học.)
7. **OOSR_Primary_Age_Female:** Out-of-school rate for primary age females.(Tỷ lệ bỏ học của nữ độ tuổi tiểu học.)
8. **OOSR_Lower_Secondary_Age_Male:** Out-of-school rate for lower secondary age males.(Tỷ lệ bỏ học của nam độ tuổi THCS.)
9. **OOSR_Lower_Secondary_Age_Female:** Out-of-school rate for lower secondary age females.(Tỷ lệ bỏ học của nữ độ tuổi THCS.)
10. **OOSR_Upper_Secondary_Age_Male:** Out-of-school rate for upper secondary age males.(Tỷ lệ bỏ học của nam độ tuổi THPT.)
11. **OOSR_Upper_Secondary_Age_Female:** Out-of-school rate for upper secondary age females.(Tỷ lệ bỏ học của nữ độ tuổi THPT.)
12. **Completion_Rate_Primary_Male:** Completion rate for primary education among males.(Tỷ lệ hoàn thành bậc tiểu học của nam giới.)
13. **Completion_Rate_Primary_Female:** Completion rate for primary education among females.(Tỷ lệ hoàn thành bậc tiểu học của nữ.)
14. **Completion_Rate_Lower_Secondary_Male:** Completion rate for lower secondary education among males.(Tỷ lệ hoàn thành bậc trung học cơ sở của nam giới.)

15. **Completion_Rate_Lower_Secondary_Female:** Completion rate for lower secondary education among females.(Tỷ lệ hoàn thành bậc trung học cơ sở của nữ.)
16. **Completion_Rate_Upper_Secondary_Male:** Completion rate for upper secondary education among males.(Tỷ lệ hoàn thành bậc trung học phổ thông của nam giới.)
17. **Completion_Rate_Upper_Secondary_Female:** Completion rate for upper secondary education among females.(Tỷ lệ hoàn thành bậc trung học phổ thông của nữ.)
18. **Grade_2_3_Proficiency_Reading:** Proficiency in reading for grade 2-3 students.(Thành thạo kỹ năng đọc dành cho học sinh lớp 2-3.)
19. **Grade_2_3_Proficiency_Math:** Proficiency in math for grade 2-3 students.(Thành thạo môn toán dành cho học sinh lớp 2-3.)
20. **Primary_End_Proficiency_Reading:** Proficiency in reading at the end of primary education.(Thành thạo đọc khi kết thúc bậc tiểu học.)
21. **Primary_End_Proficiency_Math:** Proficiency in math at the end of primary education.(Thành thạo toán khi kết thúc bậc tiểu học.)
22. **Lower_Secondary_End_Proficiency_Reading:** Proficiency in reading at the end of lower secondary education.(Thành thạo kỹ năng đọc khi kết thúc chương trình giáo dục trung học cơ sở.)
23. **Lower_Secondary_End_Proficiency_Math:** Proficiency in math at the end of lower secondary education.(Thành thạo môn toán khi kết thúc bậc trung học cơ sở.)
24. **Youth_15_24_Literacy_Rate_Male:** Literacy rate among male youths aged 15-24.(Tỷ lệ biết chữ của thanh niên nam độ tuổi 15-24.)
25. **Youth_15_24_Literacy_Rate_Female:** Literacy rate among female youths aged 15-24.(Tỷ lệ biết chữ của thanh niên nữ độ tuổi 15-24.)
26. **Birth_Rate:** Birth rate in the respective countries/areas.(Tỷ lệ sinh ở các quốc gia/khu vực tương ứng.)
27. **Gross_Primary_Education_Enrollment:** Gross enrollment in primary education.(Tổng số tuyển sinh tiểu học.)

28. **Gross_Tertiary_Education_Enrollment:** Gross enrollment in tertiary education.(Tổng số tuyển sinh vào giáo dục đại học.)
29. **Unemployment_Rate:** Unemployment rate in the respective countries/areas.(Tỷ lệ thất nghiệp ở các quốc gia/khu vực tương ứng.)

Chương 1. MÔ TẢ DỮ LIỆU BÀI TOÁN

1. Dataset Global Education

Tập dataset này cung cấp cái nhìn toàn cảnh về quy mô giáo dục trên toàn cầu, mang lại những hiểu biết sâu sắc về bối cảnh giáo dục năng động ở nhiều quốc gia và khu vực khác nhau. Bao trùm một tầm thẩm phong phú về các khía cạnh giáo dục, nó gói gọn các số liệu quan trọng bao gồm tỷ lệ ngoài nhà trường, tỷ lệ hoàn thành, trình độ thông thạo, tỷ lệ biết chữ, tỷ lệ sinh và thống kê tuyển sinh giáo dục tiểu học và đại học. Là một kho tàng kiến thức, bộ dữ liệu này là tài sản không thể thiếu đối với các nhà nghiên cứu sáng suốt, các nhà giáo dục tận tâm và các nhà hoạch định chính sách có tư duy tiên bộ, giúp họ bắt tay vào hành trình chuyển đổi nhằm đánh giá, nâng cao và định hình lại các hệ thống giáo dục trên toàn thế giới.

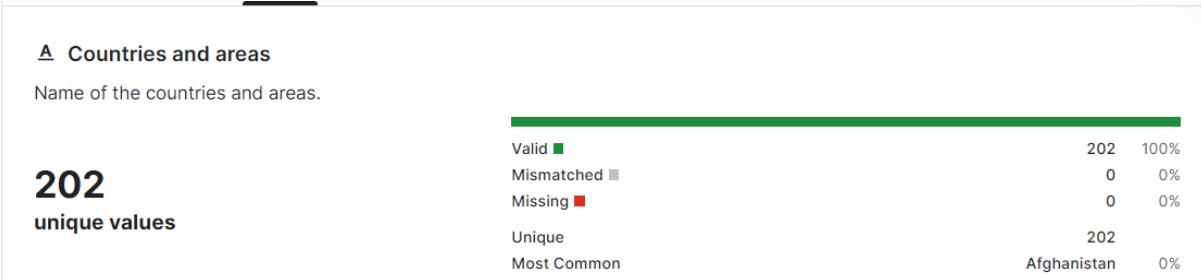
Tập dataset gồm 29 cột và 202 hàng dữ liệu. Mỗi hàng trong dữ liệu đại diện cho một quốc gia cụ thể. Các cột đó là:

- **Countries and Areas:** Name of the countries and areas.(Tên quốc gia và khu vực.)
- **Latitude:** Latitude coordinates of the geographical location.(Tọa độ vĩ độ của vị trí địa lý.)
- **Longitude:** Longitude coordinates of the geographical location.(Tọa độ kinh độ của vị trí địa lý.)
- **OOSR_Pre0Primary_Age_Male:** Out-of-school rate for pre-primary age males.(Tỷ lệ bỏ học của nam giới độ tuổi mầm non.)
- **OOSR_Pre0Primary_Age_Female:** Out-of-school rate for pre-primary age females.(Tỷ lệ bỏ học của nữ độ tuổi mầm non.)
- **OOSR_Primary_Age_Male:** Out-of-school rate for primary age males.(Tỷ lệ bỏ học của nam độ tuổi tiểu học.)
- **OOSR_Primary_Age_Female:** Out-of-school rate for primary age females.(Tỷ lệ bỏ học của nữ độ tuổi tiểu học.)
- **OOSR_Lower_Secondary_Age_Male:** Out-of-school rate for lower secondary age males.(Tỷ lệ bỏ học của nam độ tuổi THCS.)
- **OOSR_Lower_Secondary_Age_Female:** Out-of-school rate for lower secondary age females.(Tỷ lệ bỏ học của nữ độ tuổi THCS.)

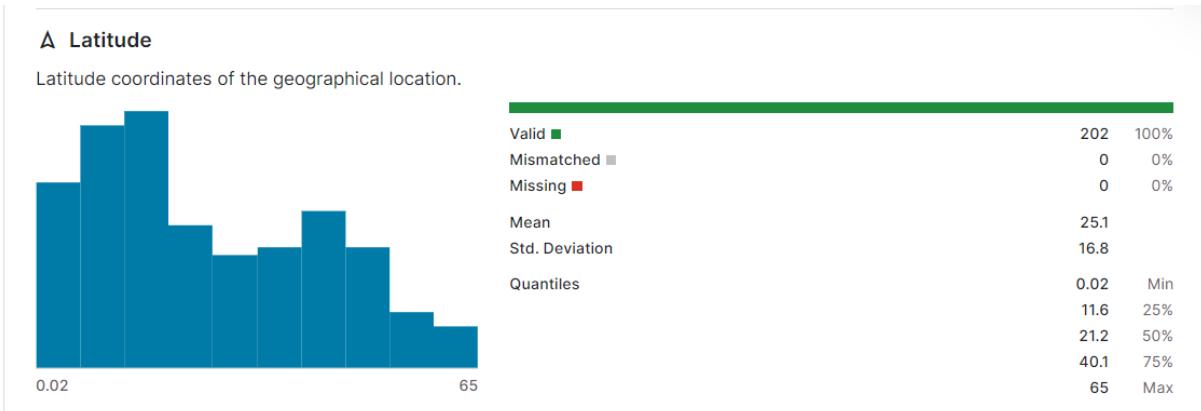
- **OOSR_Upper_Secondary_Age_Male:** Out-of-school rate for upper secondary age males.(Tỷ lệ bỏ học của nam độ tuổi THPT.)
- **OOSR_Upper_Secondary_Age_Female:** Out-of-school rate for upper secondary age females.(Tỷ lệ bỏ học của nữ độ tuổi THPT.)
- **Completion_Rate_Primary_Male:** Completion rate for primary education among males.(Tỷ lệ hoàn thành bậc tiểu học của nam giới.)
- **Completion_Rate_Primary_Female:** Completion rate for primary education among females.(Tỷ lệ hoàn thành bậc tiểu học của nữ.)
- **Completion_Rate_Lower_Secondary_Male:** Completion rate for lower secondary education among males.(Tỷ lệ hoàn thành bậc trung học cơ sở của nam giới.)
- **Completion_Rate_Lower_Secondary_Female:** Completion rate for lower secondary education among females.(Tỷ lệ hoàn thành bậc trung học cơ sở của nữ.)
- **Completion_Rate_Upper_Secondary_Male:** Completion rate for upper secondary education among males.(Tỷ lệ hoàn thành bậc trung học phổ thông của nam giới.)
- **Completion_Rate_Upper_Secondary_Female:** Completion rate for upper secondary education among females.(Tỷ lệ hoàn thành bậc trung học phổ thông của nữ.)
- **Grade_2_3_Proficiency_Reading:** Proficiency in reading for grade 2-3 students.(Thành thạo kỹ năng đọc dành cho học sinh lớp 2-3.)
- **Grade_2_3_Proficiency_Math:** Proficiency in math for grade 2-3 students.(Thành thạo môn toán dành cho học sinh lớp 2-3.)
- **Primary_End_Proficiency_Reading:** Proficiency in reading at the end of primary education.(Thành thạo đọc khi kết thúc bậc tiểu học.)
- **Primary_End_Proficiency_Math:** Proficiency in math at the end of primary education.(Thành thạo toán khi kết thúc bậc tiểu học.)
- **Lower_Secondary_End_Proficiency_Reading:** Proficiency in reading at the end of lower secondary education.(Thành thạo kỹ năng đọc khi kết thúc chương trình giáo dục trung học cơ sở.)

- **Lower_Secondary_End_Proficiency_Math:** Proficiency in math at the end of lower secondary education.(Thành thạo môn toán khi kết thúc bậc trung học cơ sở.)
- **Youth_15_24_Literacy_Rate_Male:** Literacy rate among male youths aged 15-24.(Tỷ lệ biết chữ của thanh niên nam độ tuổi 15-24.)
- **Youth_15_24_Literacy_Rate_Female:** Literacy rate among female youths aged 15-24.(Tỷ lệ biết chữ của thanh niên nữ độ tuổi 15-24.)
- **Birth_Rate:** Birth rate in the respective countries/areas.(Tỷ lệ sinh ở các quốc gia/khu vực tương ứng.)
- **Gross_Primary_Education_Enrollment:** Gross enrollment in primary education.(Tổng số tuyển sinh tiểu học.)
- **Gross_Tertiary_Education_Enrollment:** Gross enrollment in tertiary education.(Tổng số tuyển sinh vào giáo dục đại học.)
- **Unemployment_Rate:** Unemployment rate in the respective countries/areas.(Tỷ lệ thất nghiệp ở các quốc gia/khu vực tương ứng.)

1. Cột Countries and areas



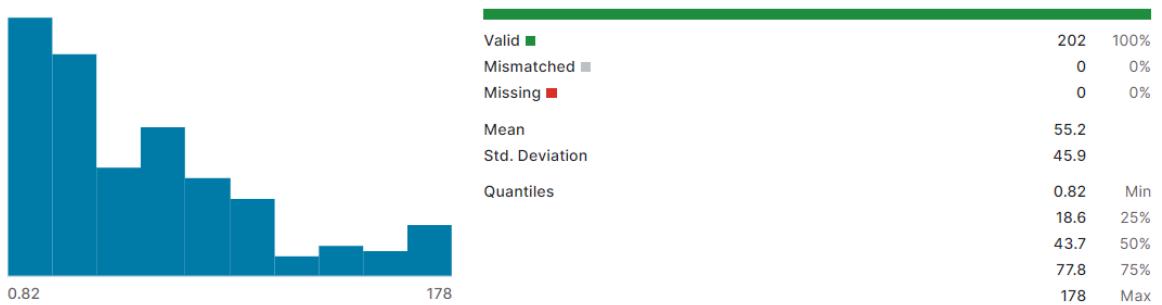
2. Cột Latitude



3. Cột Longitude

A Longitude

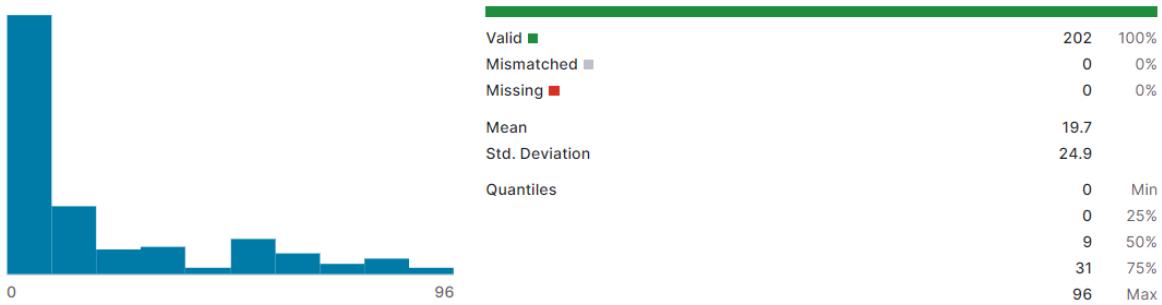
Longitude coordinates of the geographical location.



4. Cột OOSR_Pre0Primary_Age_Male

OOSR_Pre0Primary_Age_Male

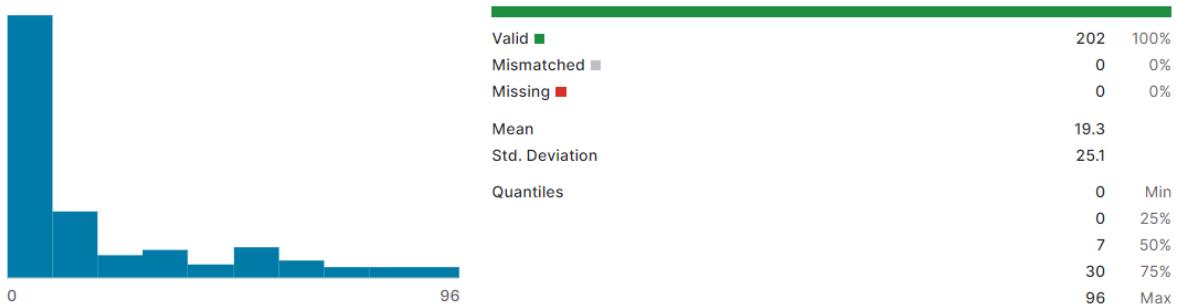
Out-of-school rate for pre-primary age males.



5. Cột OOSR_Pre0Primary_Age_Female

OOSR_Pre0Primary_Age_Female

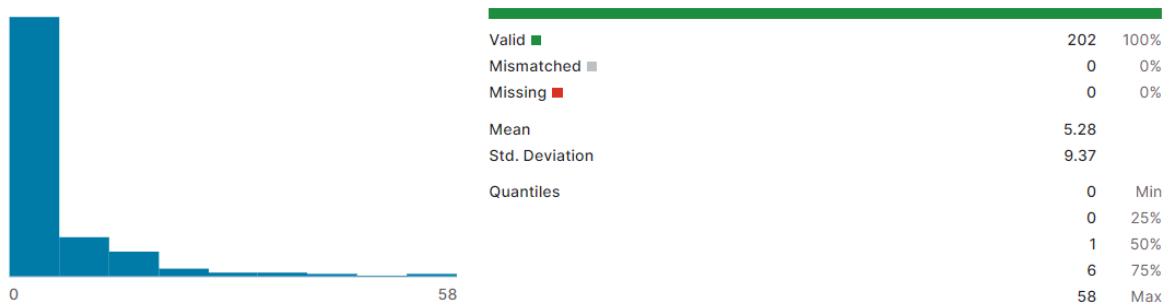
Out-of-school rate for pre-primary age females.



6. Cột OOSR_Primary_Age_Male

OOSR_Primary_Age_Male

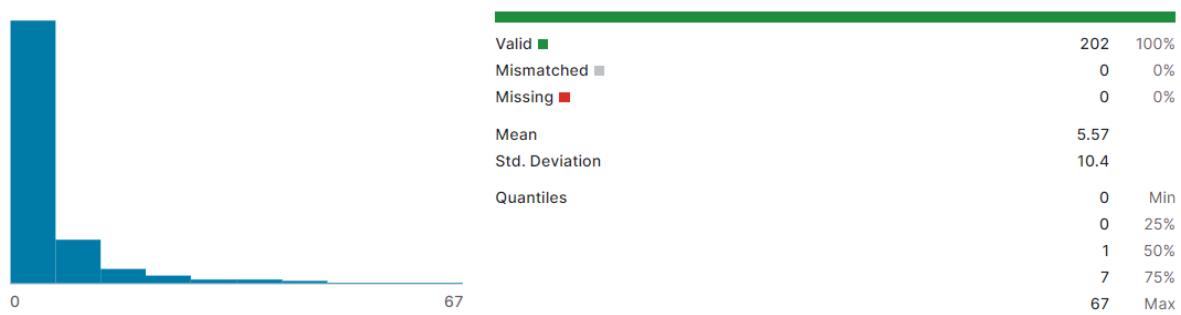
Out-of-school rate for primary age males.



7. Cột OOSR_Primary_Age_Female

OOSR_Primary_Age_Female

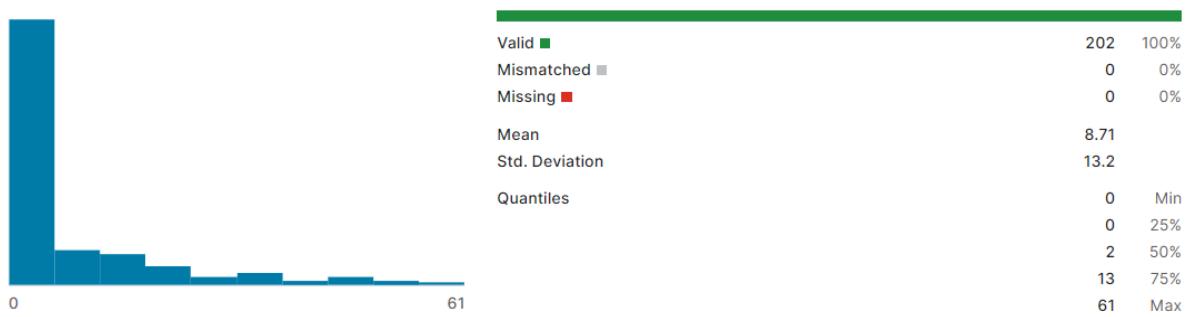
Out-of-school rate for primary age females.



8. Cột OOSR_Lower_Secondary_Age_Male

OOSR_Lower_Secondary_Age_Male

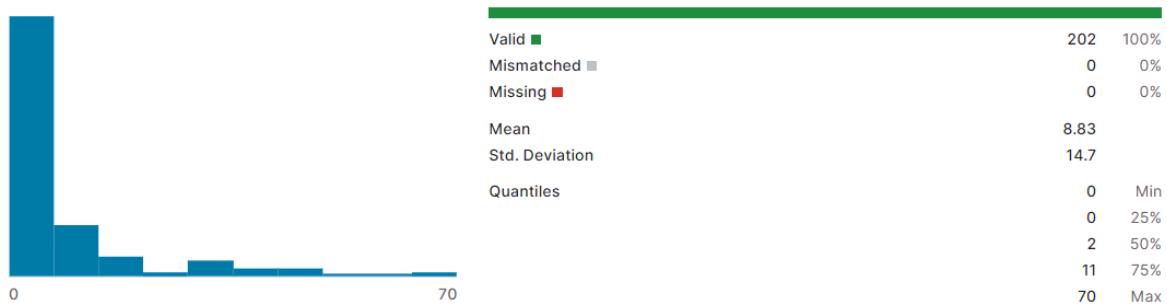
Out-of-school rate for lower secondary age males.



9. Cột OOSR_Lower_Secondary_Age_Female

OOSR_Lower_Secondary_Age_Female

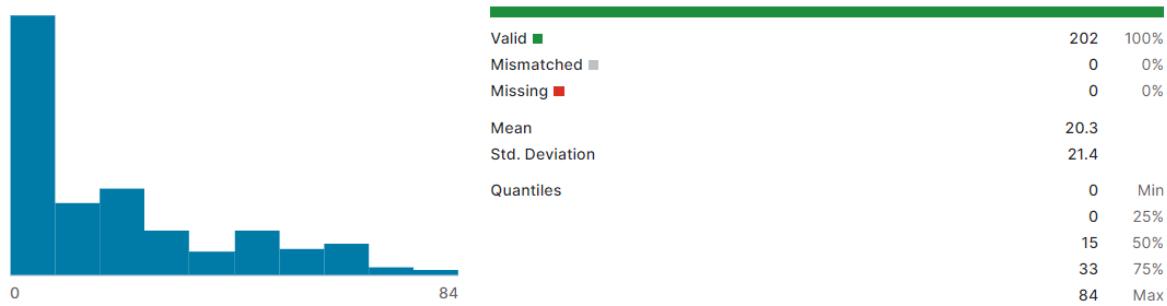
Out-of-school rate for lower secondary age females.



10. Cột OOSR_Upper_Secondary_Age_Male

OOSR_Upper_Secondary_Age_Male

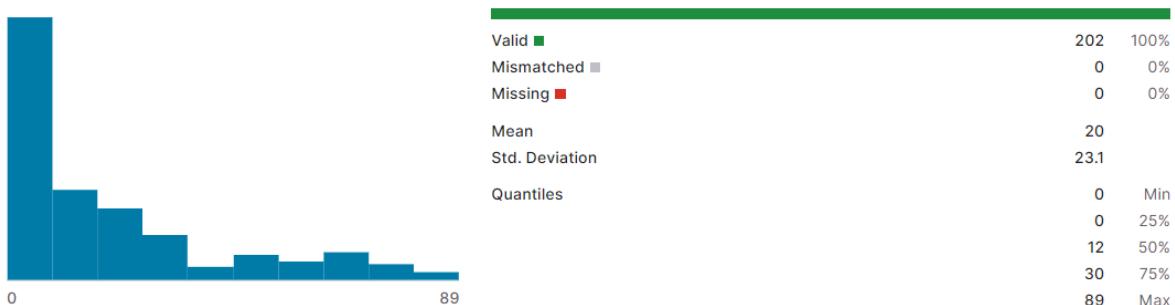
Out-of-school rate for upper secondary age males.



11. Cột OOSR_Upper_Secondary_Age_Female

OOSR_Upper_Secondary_Age_Female

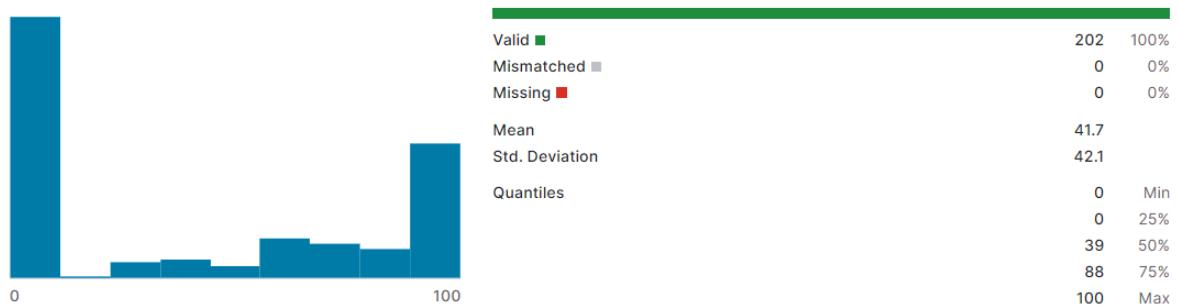
Out-of-school rate for upper secondary age females.



12. Cột Completion_Rate_Primary_Male

Completion_Rate_Primary_Male

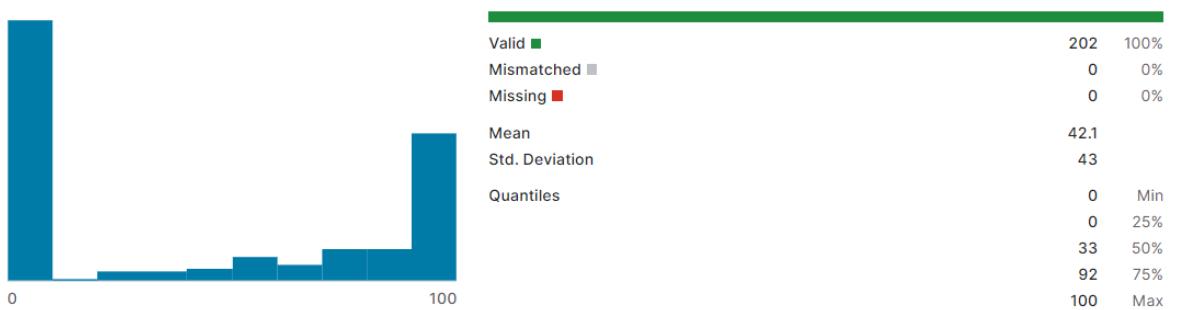
Completion rate for primary education among males.



13. Cột Completion_Rate_Primary_Female

Completion_Rate_Primary_Female

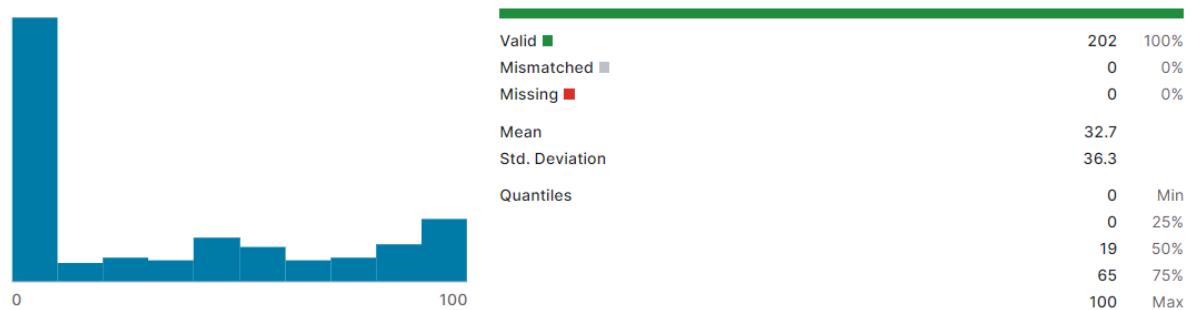
Completion rate for primary education among females.



14. Cột Completion_Rate_Lower_Secondary_Male

Completion_Rate_Lower_Secondary_Male

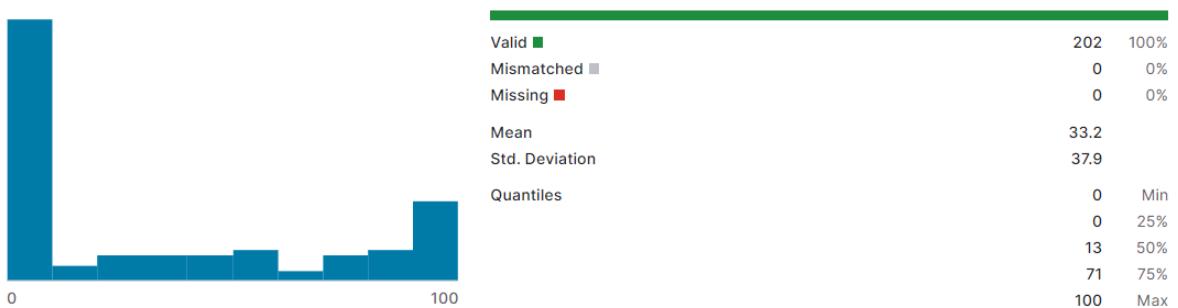
Completion rate for lower secondary education among males.



15. Cột Completion_Rate_Lower_Secondary_Female

Completion_Rate_Lower_Secondary_Female

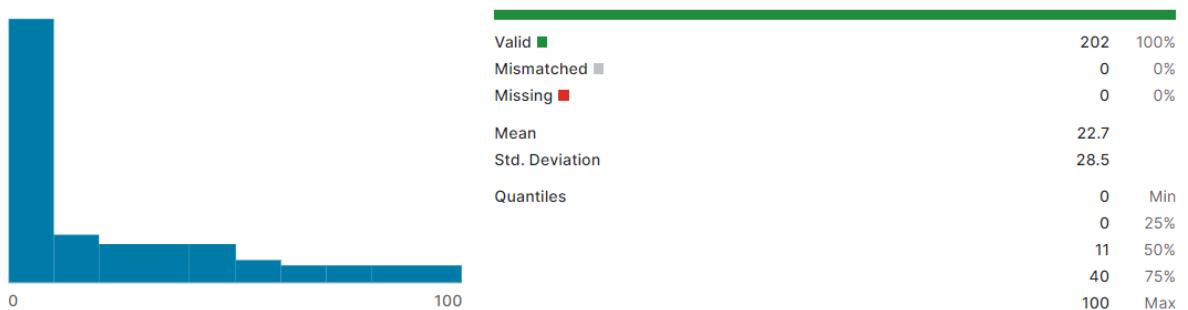
Completion rate for lower secondary education among females.



16. Cột Completion_Rate_Upper_Secondary_Male

Completion_Rate_Upper_Secondary_Male

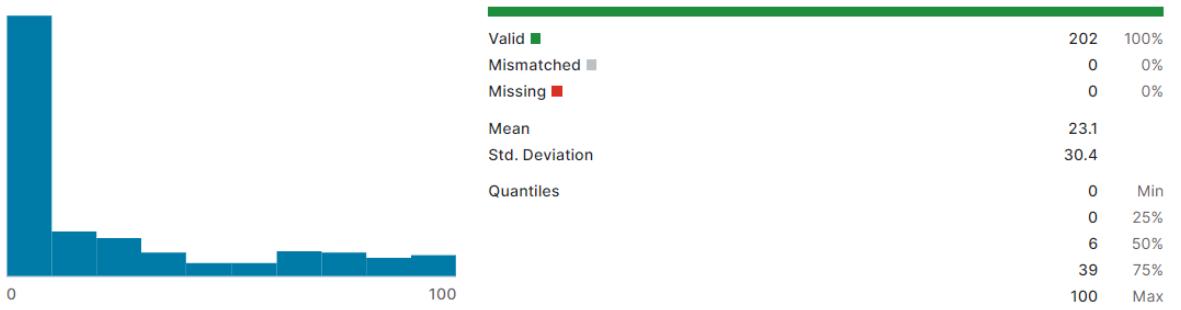
Completion rate for upper secondary education among males.



17. Cột Completion_Rate_Upper_Secondary_Female

Completion_Rate_Upper_Secondary_Female

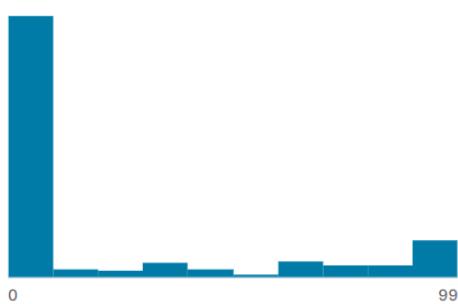
Completion rate for upper secondary education among females.



18. Cột Grade_2_3_Proficiency_Reading

Grade_2_3_Proficiency_Reading

Proficiency in reading for grade 2-3 students.

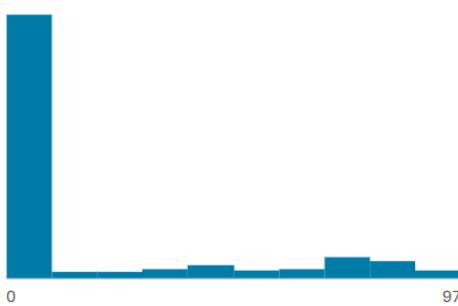


Valid	202	100%
Mismatched	0	0%
Missing	0	0%
Mean	22	
Std. Deviation	34.8	
Quantiles		
	0	Min
	0	25%
	0	50%
	39	75%
	99	Max

19. Cột Grade_2_3_Proficiency_Math

Grade_2_3_Proficiency_Math

Proficiency in math for grade 2-3 students.

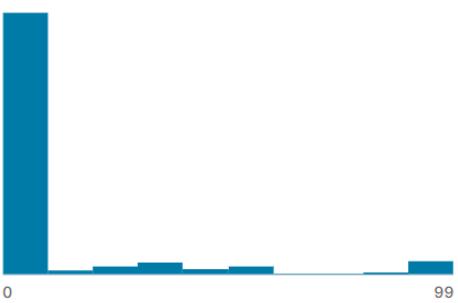


Valid	202	100%
Mismatched	0	0%
Missing	0	0%
Mean	17.4	
Std. Deviation	29.7	
Quantiles		
	0	Min
	0	25%
	0	50%
	33	75%
	97	Max

20. Cột Primary_End_Proficiency_Reading

Primary_End_Proficiency_Reading

Proficiency in reading at the end of primary education.

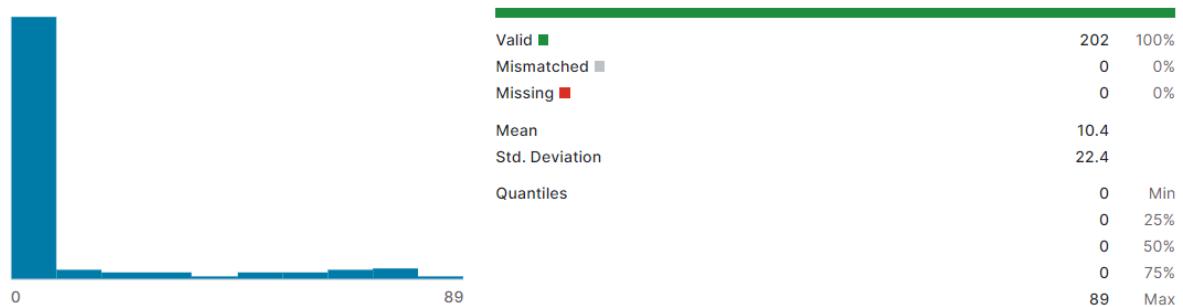


Valid	202	100%
Mismatched	0	0%
Missing	0	0%
Mean	10.7	
Std. Deviation	24.8	
Quantiles		
	0	Min
	0	25%
	0	50%
	0	75%
	99	Max

21. Cột Primary_End_Proficiency_Math

Primary_End_Proficiency_Math

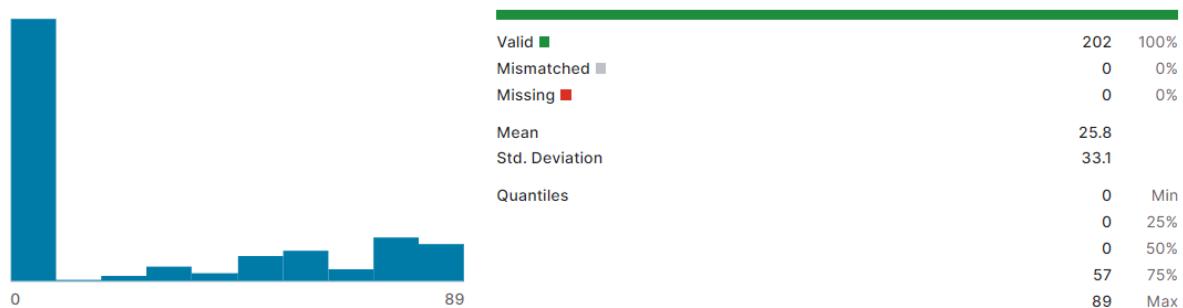
Proficiency in math at the end of primary education.



22. Cột Lower_Secondary_End_Proficiency_Reading

Lower_Secondary_End_Proficiency_Reading

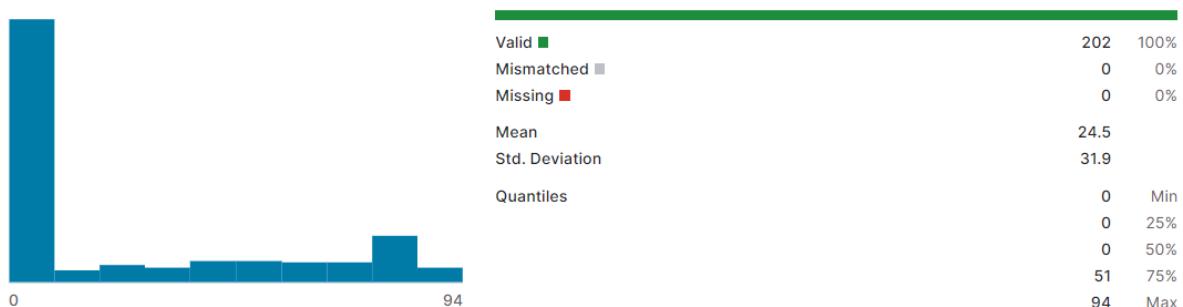
Proficiency in reading at the end of lower secondary education.



23. Cột Lower_Secondary_End_Proficiency_Math

Lower_Secondary_End_Proficiency_Math

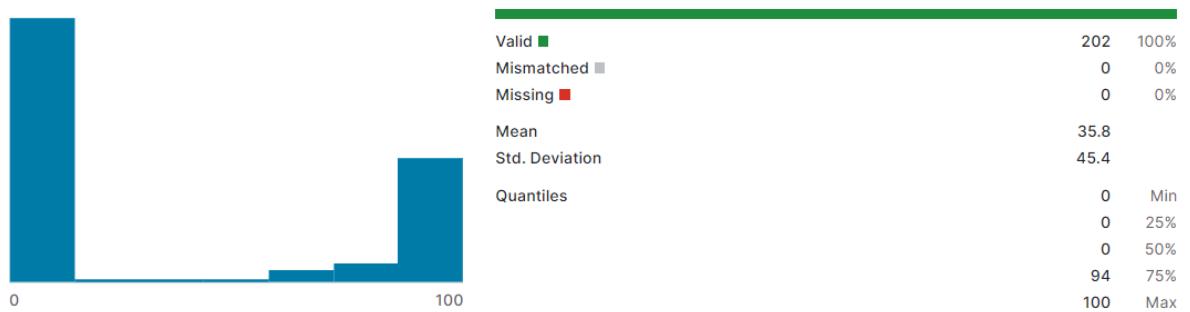
Proficiency in math at the end of lower secondary education.



24. Cột Youth_15_24_Literacy_Rate_Male

Youth_15_24_Literacy_Rate_Male

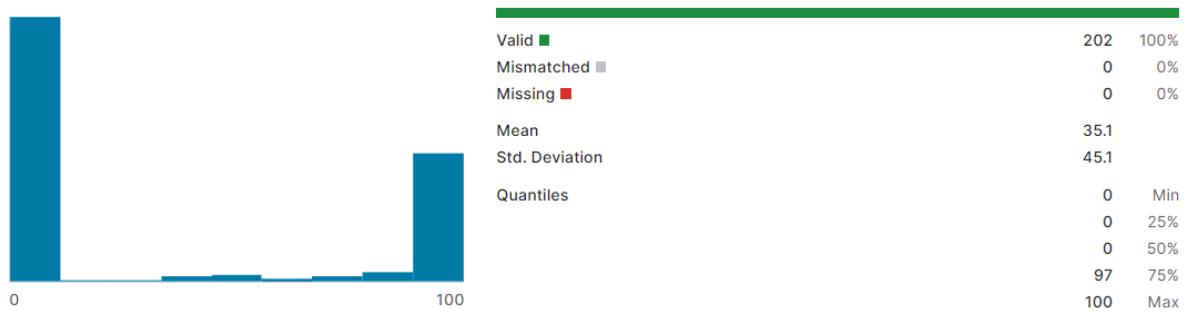
Literacy rate among male youths aged 15-24.



25. Cột Youth_15_24_Literacy_Rate_Female

Youth_15_24_Literacy_Rate_Female

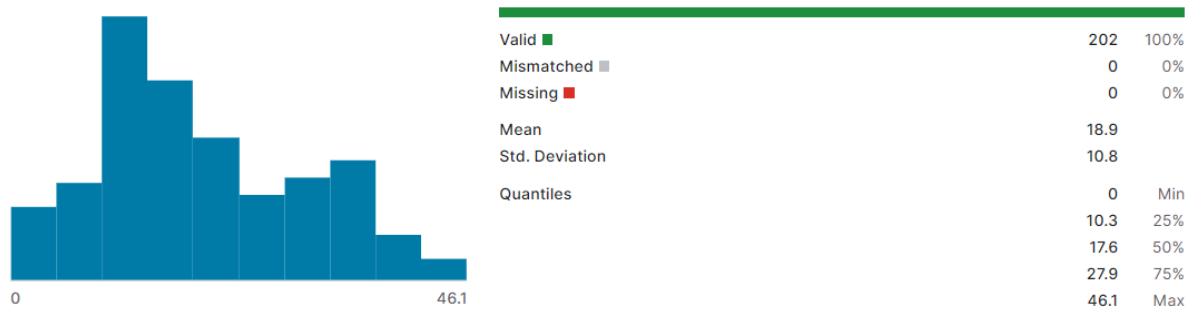
Literacy rate among female youths aged 15-24.



26. Cột Birth_Rate

Birth_Rate

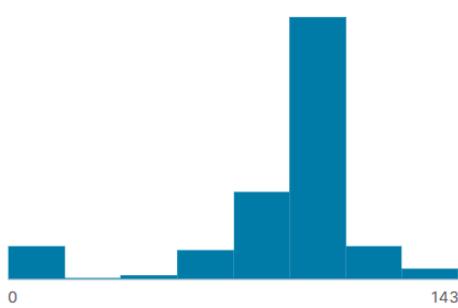
Birth rate in the respective countries/areas.



27. Cột Gross_Primary_Education_Enrollment

Gross_Primary_Education_Enrollment

Gross enrollment in primary education.

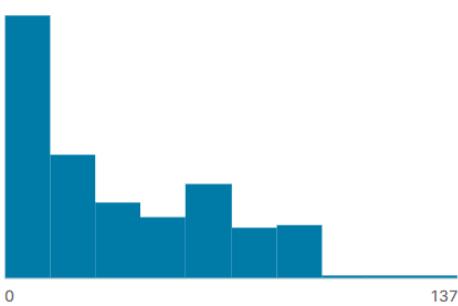


Valid	202	100%
Mismatched	0	0%
Missing	0	0%
Mean	94.9	
Std. Deviation	29.7	
Quantiles	0	Min
	97.2	25%
	102	50%
	107	75%
	143	Max

28. Cột Gross_Tertiary_Education_Enrollment

Gross_Tertiary_Education_Enrollment

Gross enrollment in tertiary education.

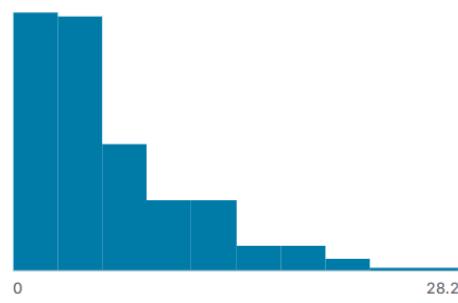


Valid	202	100%
Mismatched	0	0%
Missing	0	0%
Mean	34.4	
Std. Deviation	29.9	
Quantiles	0	Min
	9	25%
	24.9	50%
	60	75%
	137	Max

29. Cột Unemployment_Rate

Unemployment_Rate

Unemployment rate in the respective countries/areas.

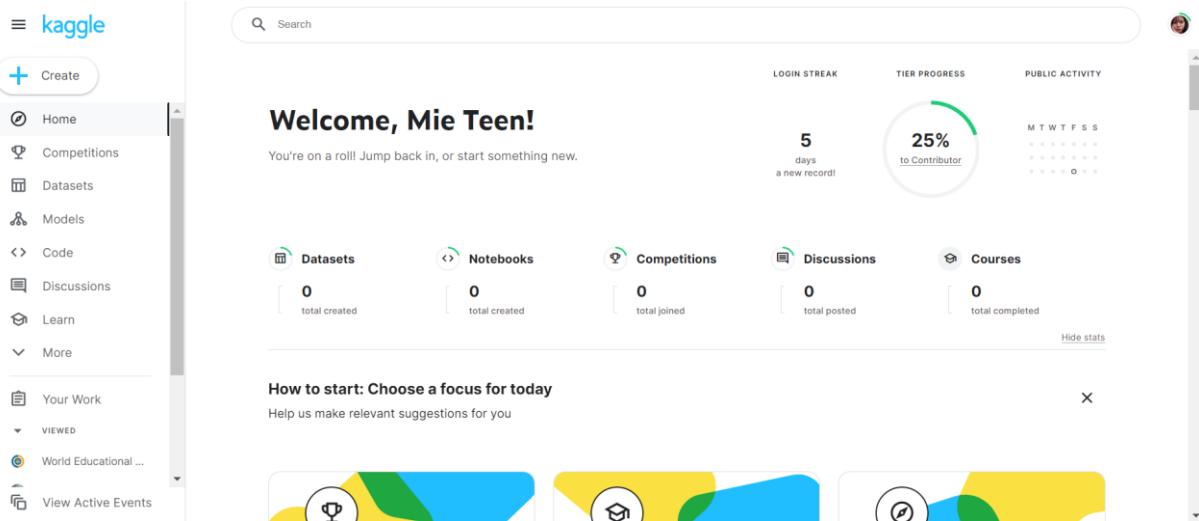


Valid	202	100%
Mismatched	0	0%
Missing	0	0%
Mean	6	
Std. Deviation	5.26	
Quantiles	0	Min
	2.29	25%
	4.59	50%
	8.73	75%
	28.2	Max

Chương 2. TIỀN XỬ LÝ DỮ LIỆU

2.1. Thu thập dữ liệu

- Dữ liệu được thu thập từ trang : <https://www.kaggle.com/>



2.2. Sửa chữa dữ liệu

2.2.1. Dataset Global Education

Kiểm tra dữ liệu thì các cột đều đầy đủ dữ liệu, cột **Countries and areas** mang kiểu dữ liệu object, các kiểu dữ liệu còn lại mang kiểu dữ liệu số .

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 202 entries, 0 to 201
Data columns (total 29 columns):
 #   Column           Non-Null Count Dtype  
 --- 
 0   Countries and areas    202 non-null   object  
 1   Latitude              202 non-null   float64 
 2   Longitude             202 non-null   float64 
 3   OOSR_Pre0Primary_Age_Male 202 non-null   int64  
 4   OOSR_Pre0Primary_Age_Female 202 non-null   int64  
 5   OOSR_Primary_Age_Male    202 non-null   int64  
 6   OOSR_Primary_Age_Female  202 non-null   int64  
 7   OOSR_Lower_Secondary_Age_Male 202 non-null   int64  
 8   OOSR_Lower_Secondary_Age_Female 202 non-null   int64  
 9   OOSR_Upper_Secondary_Age_Male 202 non-null   int64  
 10  OOSR_Upper_Secondary_Age_Female 202 non-null   int64  
 11  Completion_Rate_Primary_Male 202 non-null   int64  
 12  Completion_Rate_Primary_Female 202 non-null   int64  
 13  Completion_Rate_Lower_Secondary_Male 202 non-null   int64  
 14  Completion_Rate_Lower_Secondary_Female 202 non-null   int64  
 15  Completion_Rate_Upper_Secondary_Male 202 non-null   int64  
 16  Completion_Rate_Upper_Secondary_Female 202 non-null   int64  
 17  Grade_2_3_Proficiency_Reading 202 non-null   int64  
 18  Grade_2_3_Proficiency_Math    202 non-null   int64  
 19  Primary_End_Proficiency_Reading 202 non-null   int64  
 20  Primary_End_Proficiency_Math   202 non-null   int64  
 21  Lower_Secondary_End_Proficiency_Reading 202 non-null   int64  
 22  Lower_Secondary_End_Proficiency_Math   202 non-null   int64  
 23  Youth_15_24_Literacy_Rate_Male    202 non-null   int64  
 24  Youth_15_24_Literacy_Rate_Female  202 non-null   int64  
 25  Birth_Rate              202 non-null   float64 
 26  Gross_Primary_Education_Enrollment 202 non-null   float64 
 27  Gross_Tertiary_Education_Enrollment 202 non-null   float64 
 28  Unemployment_Rate        202 non-null   float64 

dtypes: float64(6), int64(22), object(1)
memory usage: 45.9+ KB

```

Tất cả các dòng của dữ liệu đều đầy đủ

Countries_and_areas	0
Latitude	0
Longitude	0
OOSR_Pre0Primary_Age_Male	0
OOSR_Pre0Primary_Age_Female	0
OOSR_Primary_Age_Male	0
OOSR_Primary_Age_Female	0
OOSR_Lower_Secondary_Age_Male	0
OOSR_Lower_Secondary_Age_Female	0
OOSR_Upper_Secondary_Age_Male	0
OOSR_Upper_Secondary_Age_Female	0
Completion_Rate_Primary_Male	0
Completion_Rate_Primary_Female	0
Completion_Rate_Lower_Secondary_Male	0
Completion_Rate_Lower_Secondary_Female	0
Completion_Rate_Upper_Secondary_Male	0
Completion_Rate_Upper_Secondary_Female	0
Grade_2_3_Proficiency_Reading	0
Grade_2_3_Proficiency_Math	0
Primary_End_Proficiency_Reading	0
Primary_End_Proficiency_Math	0
Lower_Secondary_End_Proficiency_Reading	0
Lower_Secondary_End_Proficiency_Math	0
Youth_15_24_Literacy_Rate_Male	0
Youth_15_24_Literacy_Rate_Female	0
Birth_Rate	0
Gross_Primary_Education_Enrollment	0
Gross_Tertiary_Education_Enrollment	0
Unemployment_Rate	0
dtype:	int64

Không có dữ liệu trùng lặp (theo tên quốc gia/ khu vực)

➡ Số dòng dữ liệu trùng lặp: 0

2.3. Làm sạch dữ liệu

Dataset đã sạch (không thiếu dữ liệu, format đúng, không chứa ký tự đặc biệt).

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD		
1	Countries	Latitude	Longitude	OOSR_Pre0Primary_Age_Male	OOSR_Pre0Primary_Age_Female	OOSR_Primary_Age_Male	OOSR_Primary_Age_Female	OOSR_Lower_Secondary_Age_Male	OOSR_Lower_Secondary_Age_Female	OOSR_Upper_Secondary_Age_Male	OOSR_Upper_Secondary_Age_Female	Completion_Rate_Primary_Male	Completion_Rate_Primary_Female	Completion_Rate_Lower_Secondary_Male	Completion_Rate_Lower_Secondary_Female	Completion_Rate_Upper_Secondary_Male	Completion_Rate_Upper_Secondary_Female	Grade_2_3_Proficiency_Reading	Grade_2_3_Proficiency_Math	Primary_End_Proficiency_Reading	Primary_End_Proficiency_Math	Lower_Secondary_End_Proficiency_Reading	Lower_Secondary_End_Proficiency_Math	Youth_15_24_Literacy_Rate_Male	Youth_15_24_Literacy_Rate_Female	Birth_Rate	Gross_Primary_Education_Enrollment	Gross_Tertiary_Education_Enrollment	Unemployment_Rate		
2	Afghanistan	33.93911	67.70995	0	0	0	0	0	44	69	67	40	49	26	32	14	22	25	13	11	0	0	74	56	32.49	104	9.7	11.7	Medium		
3	Albania	41.15333	20.16833	4	2	6	3	6	1	21	15	94	96	98	97	76	80	0	0	0	0	48	58	99	100	11.78	107	55	12.33	Medium	
4	Algeria	28.03898	1.659626	0	0	0	0	0	0	0	0	93	93	49	65	22	37	0	0	0	0	21	19	98	97	24.28	109.9	51.4	11.7	Medium	
5	Andorra	42.50629	1.521801	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Low		
6	Angola	11.20269	17.87389	31	39	0	0	0	0	0	0	0	63	57	42	32	24	15	0	0	0	0	0	0	0	0	40.73	113.5	9.3	6.89	Low
7	Anguilla	18.22058	63.06862	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Low	
8	Antigua and Barbuda	17.06082	61.79643	14	4	4	1	1	2	14	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15.33	105	24.8	0	Low
9	Argentina	38.4161	63.61667	2	2	0	0	0	0	15	7	91	94	70	79	46	53	76	71	46	56	48	31	99	100	17.02	109.7	90	9.79	Medium	
10	Armenia	40.0691	45.03819	52	50	9	9	11	9	16	4	99	99	95	99	69	79	0	0	55	0	50	0	0	13.99	92.7	54.6	16.99	High		
11	Australia	25.2744	133.7751	13	14	0	0	2	3	10	6	0	0	0	0	0	0	94	70	0	64	80	78	0	0	12.6	100.3	113.1	5.27	Low	
12	Austria	47.51623	14.55007	0	0	0	0	1	0	10	9	0	0	0	0	0	0	0	0	98	0	76	79	0	0	9.7	103.1	85.1	4.67	Low	
13	Azerbaijan	40.14311	47.57693	32	19	10	7	0	0	0	0	0	0	0	0	0	0	0	0	81	0	0	0	0	0	0	14	99.7	27.7	5.51	Low
14	The Bahamas	25.03428	77.39628	0	0	0	0	23	21	29	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13.97	81.4	15.1	10.36	Medium
15	Bahrain	26.0667	55.5577	31	28	2	3	7	0	18	6	0	0	0	0	0	0	69	40	0	0	0	39	100	99	13.99	99.4	50.5	0.71	Low	
16	Bangladesh	23.68499	90.35633	0	0	0	0	0	0	41	36	76	89	59	71	32	27	47	34	44	32	54	57	94	96	18.18	116.5	20.6	4.19	Low	
17	Barbados	13.19389	59.5432	6	10	1	2	7	3	7	3	99	99	98	98	91	97	0	0	0	0	0	0	0	0	10.65	99.4	65.4	10.33	Medium	
18	Belarus	53.70981	27.95339	0	4	1	2	1	1	6	4	100	100	100	100	91	94	0	0	0	0	77	71	100	100	9.9	100.5	87.4	4.59	Low	
19	Belgium	50.50389	4.6469936	3	2	1	0	1	1	1	2	0	0	0	0	0	0	0	0	0	79	80	0	0	10.3	103.9	79.7	5.59	Low		
20	Belize	17.18988	88.49765	18	16	1	1	9	11	38	35	95	96	55	66	48	51	0	0	0	0	0	0	0	0	20.79	111.7	24.7	6.41	Low	
21	Benin	9.30769	2.315834	15	16	3	10	27	43	46	66	51	44	25	13	12	5	9	34	23	11	0	0	70	52	36.22	122	12.3	2.23	Low	
22	Bhutan	27.51416	90.4336	58	59	5	2	18	6	33	22	67	71	41	38	25	18	0	0	0	0	56	0	0	0	17.26	100.1	15.6	2.34	Low	
23	Bolivia	16.29015	63.58865	9	8	5	5	13	14	22	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21.75	98.2	0	3.5	Low	
24	Bosnia and Herzegovina	43.91589	17.67908	72	73	0	0	0	0	22	19	99	100	97	97	92	92	0	0	0	0	46	42	0	0	8.11	0	23.3	18.42	High	
25	Botswana	22.32847	24.68487	79	78	12	10	0	0	0	0	95	98	92	92	55	66	0	0	0	16	0	0	0	0	24.82	103.2	24.9	18.19	High	
26	Brazil	14.235	51.9528	0	1	1	1	2	3	16	14	95	97	81	89	65	75	80	71	53	52	50	32	99	99	13.92	115.4	51.3	12.08	Medium	
27	British Virgin Islands	18.4207	64.63997	9	2	4	3	8	7	24	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Low	

2.4. Chuẩn hóa dữ liệu

Các cột đã được chuẩn hóa về kiểu dữ liệu số và kiểu dữ liệu object. Cột Countries and Areas là kiểu dữ liệu object, các cột còn lại là kiểu dữ liệu số.

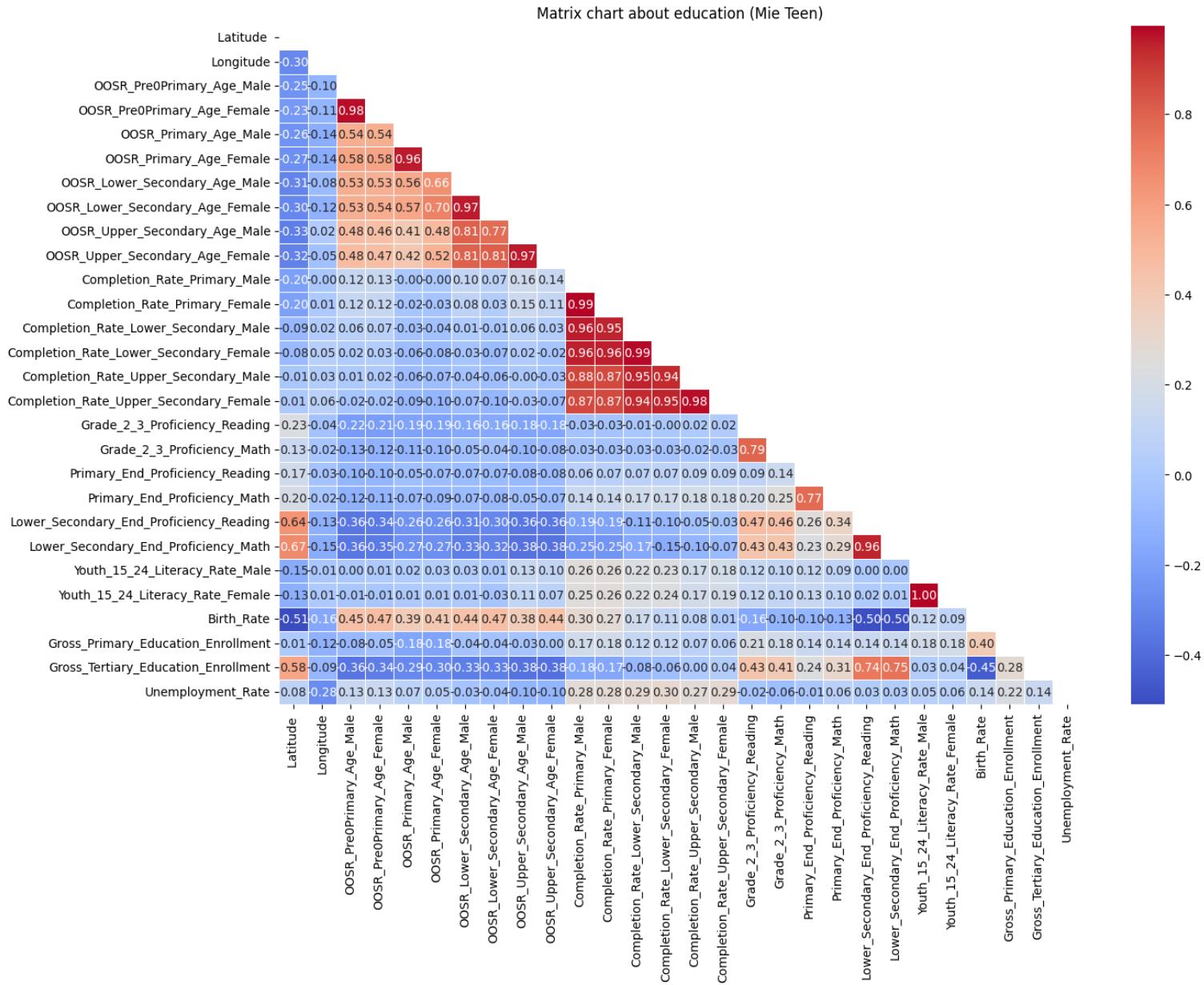
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	
1	Countries	Latitude	Longitude	OOSR_Pre	OOSR_Pre	OOSR_Pri	OOSR_Pri	OOSR_Low	OOSR_Low	OOSR_Upp	OOSR_Upp	Completo1	Completo1	Completo1	Completo1	Completo1	Completo1	Grade_2_;	Grade_2_;	Primary_E	Lower_Sec	Lower_Sec	Youth_15	Youth_15	Birth_Rate	Gross_Prin	Gross_Ter	Unemploy	Class	
2	Afghanista	33.93911	67.70959	0	0	0	0	0	44	69	67	40	49	26	32	14	22	25	13	11	0	0	74	56	32.49	104	9.7	11.12	Medium	
3	Albania	41.15333	20.16833	4	2	6	3	6	1	21	15	94	96	98	97	76	80	0	0	0	48	58	99	100	11.78	107	55	12.33	Medium	
4	Algeria	28.03389	1.659626	0	0	0	0	0	0	0	93	93	49	65	22	37	0	0	0	21	19	98	97	24.28	109.9	51.4	11.7	Medium		
5	Andorra	42.50629	1.521801	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7.2	106.4	0	0	Low	
6	Angola	11.20268	17.87389	31	39	0	0	0	0	0	0	63	57	42	32	24	15	0	0	0	0	0	0	0	40.73	113.5	9.3	6.89	Low	
7	Anguilla	18.22058	63.06862	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Low	
8	Antigua and	17.06082	61.79643	14	4	4	1	1	2	14	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15.33	105	24.8	0	Low
9	Argentina	38.4161	63.61667	2	2	0	0	0	0	15	7	91	94	70	79	46	53	76	71	46	56	48	31	99	100	17.02	109.7	90	9.79	Medium
10	Armenia	40.0691	45.03819	52	50	9	9	11	9	16	4	99	99	95	99	69	79	0	0	0	55	0	50	0	0	13.99	92.7	54.6	16.99	High
11	Australia	25.2744	133.7751	13	14	0	0	2	3	10	6	0	0	0	0	0	0	94	70	0	64	80	78	0	0	12.6	100.3	113.1	5.27	Low
12	Austria	47.51623	14.55007	0	0	0	0	1	0	10	9	0	0	0	0	0	0	0	0	98	0	76	79	0	0	9.7	103.1	85.1	4.67	Low
13	Azerbaijan	40.14311	47.57693	32	19	10	7	0	0	0	0	0	0	0	0	0	0	0	0	81	0	0	0	0	0	14	99.7	27.7	5.51	Low
14	The Baham	25.03428	77.39628	0	0	0	0	23	21	29	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13.97	81.4	15.1	10.36	Medium
15	Bahrain	26.0667	50.5577	31	28	2	3	7	0	18	6	0	0	0	0	0	69	40	0	0	0	39	100	99	13.99	99.4	50.5	0.71	Low	
16	Bangladesh	23.68499	90.35633	0	0	0	0	0	0	41	36	76	89	59	71	32	27	47	34	44	32	54	57	94	96	18.18	116.5	20.6	4.19	Low
17	Barbados	13.19389	59.5432	6	10	1	2	7	3	7	3	99	99	98	98	91	97	0	0	0	0	0	0	0	0	10.65	99.4	65.4	10.33	Medium
18	Belarus	53.70981	27.95339	0	4	1	2	1	1	6	4	100	100	100	100	91	94	0	0	0	0	77	71	100	100	9.9	100.5	87.4	4.59	Low
19	Belgium	50.50389	4.469936	3	2	1	0	1	1	1	2	0	0	0	0	0	0	0	0	0	0	79	80	0	0	10.3	103.9	79.7	5.59	Low
20	Belize	17.18986	88.49765	18	16	1	1	9	11	38	35	95	96	55	66	48	51	0	0	0	0	0	0	0	0	20.79	111.7	24.7	6.41	Low
21	Benin	9.30769	2.315834	15	16	3	10	27	43	46	66	51	44	25	13	12	5	9	34	23	11	0	0	70	52	36.22	122	12.3	2.23	Low
22	Bhutan	27.51416	90.4336	58	59	5	2	18	6	33	22	67	71	41	38	25	18	0	0	0	56	0	0	0	0	17.26	100.1	15.6	2.34	Low
23	Bolivia	16.29015	63.58865	9	8	5	5	13	14	22	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21.75	98.2	0	3.5	Low
24	Bosnia and	43.91589	17.67908	72	73	0	0	0	0	22	19	99	100	97	97	92	92	0	0	0	0	46	42	0	0	8.11	0	23.3	18.42	High
25	Botswana	22.32847	24.68487	79	78	12	10	0	0	0	0	95	98	92	92	55	66	0	0	0	16	0	0	0	0	24.82	103.2	24.9	18.19	High
26	Brazil	14.235	51.92528	0	1	1	1	2	3	16	14	95	97	81	89	65	75	80	71	53	52	50	32	99	99	13.92	115.4	51.3	12.08	Medium
27	British Virgin Islands	18.4207	64.63997	9	2	4	3	8	7	24	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Low	

Global Education



Chương 3. PHÂN TÍCH DỮ LIỆU

3.1. Sự tương quan giữa các cột trong dữ liệu



Sử dụng biểu đồ heatmap hiển thị mối quan hệ giữa các biến trong data trở nên rõ ràng hơn. Heatmap sử dụng màu sắc để biểu thị giá trị, phân loại dữ liệu thành các nhóm và hiển thị sự phân phối của các giá trị trong từng nhóm. Các giá trị cao hoặc thấp có thể được nhận diện dễ dàng thông qua màu sắc. Dựa vào biểu đồ ta thấy:

- ❖ Tình trạng thiếu thốn về giáo dục của **nam** và **nữ** luôn có mối liên quan chặt chẽ với nhau.

```
# OOSR_PrePrimary_Age_Male and OOSR_Primary_Age_Female correlation 0.98
# ==> Tình trạng thiếu thốn về giáo dục của nam và nữ luôn có mối liên quan chặt chẽ với nhau.
correlation1 = round(df['OOSR_PrePrimary_Age_Male'].corr(df['OOSR_Primary_Age_Female']), 2)
print(f"Correlation Ratio: {correlation1}")

Correlation Ratio: 0.98
```

- ❖ Có mối tương quan dương và cao giữa trình độ đọc và trình độ tính toán

```
# Correlation of Grade_2_3_Proficiency_Reading and Grade_2_3_Proficiency_Math is 0.79
# Đối với học sinh lớp 2 và lớp 3, có mối tương quan dương và cao giữa trình độ đọc và trình độ tính toán
correlation2 = round(df['Grade_2_3_Proficiency_Reading'].corr(df['Grade_2_3_Proficiency_Math']),2)
print(f"Correlation Ratio: {correlation2}")

Correlation Ratio: 0.79
```

3.2. Sự tương quan giữa tỷ lệ sinh với các cột

- ❖ Top 5 mối tương quan cao nhất:

OOSR_PrePrimary_Age_Female	0.469859
OOSR_Lower_Secondary_Age_Female	0.465821
OOSR_PrePrimary_Age_Male	0.451898
OOSR_Upper_Secondary_Age_Female	0.443845
OOSR_Lower_Secondary_Age_Male	0.438541
Name: Birth_Rate, dtype: float64	

- Cho thấy tỷ lệ ngoài nhà trường có tác động đến tỷ lệ sinh, nhưng những tác động này rất yếu.
- Các yếu tố khác cũng có thể ảnh hưởng đến những mối quan hệ này.

- ❖ Top 5 mối tương quan thấp nhất:

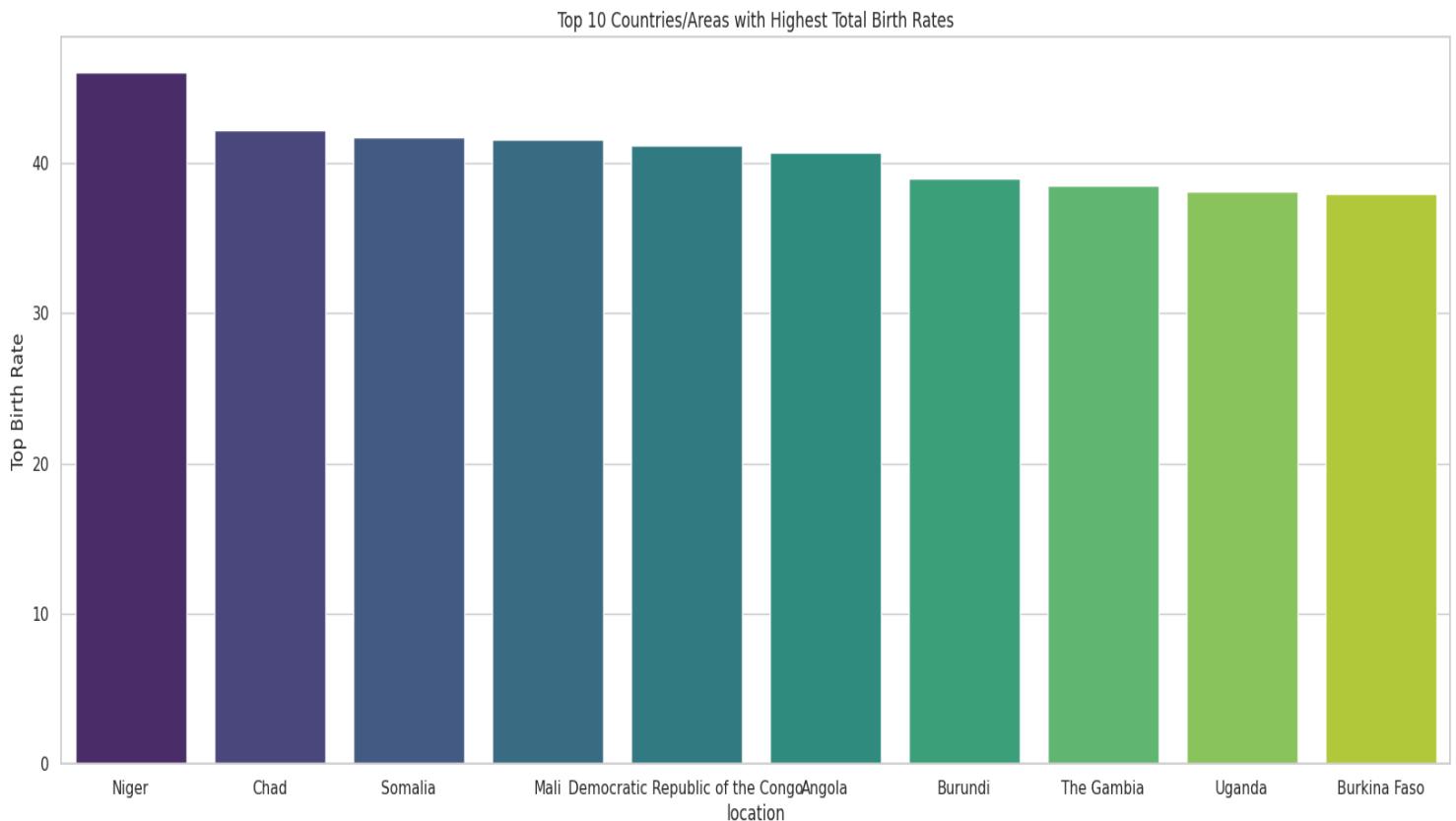
Latitude	-0.509112
Lower_Secondary_End_Proficiency_Math	-0.501194
Lower_Secondary_End_Proficiency_Reading	-0.495208
Gross_Tertiary_Education_Enrollment	-0.454775
Grade_2_3_Proficiency_Reading	-0.164135
Name: Birth_Rate, dtype: float64	

- Khi trình độ học vấn của học sinh tăng lên thì tỷ lệ sinh giảm.
- Nhìn chung, ta thấy rằng tỷ lệ sinh giảm ở các quốc gia gần Xích đạo.

- ❖ Top 10 quốc gia có tỷ lệ sinh cao nhất

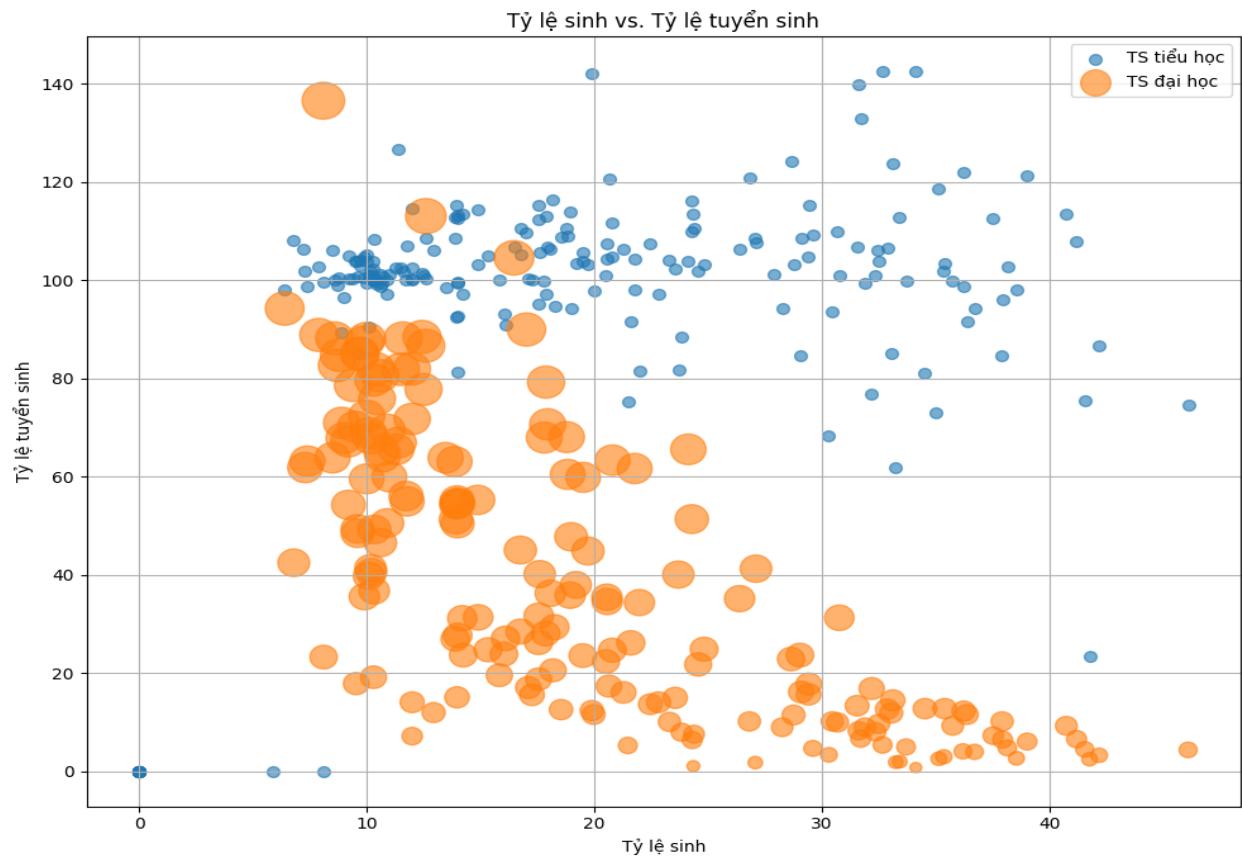
1. Niger
2. Chad
3. Somalia

4. Mali
5. Democratic Republic of the Congo
6. Angola
7. Burundi
8. The Gambia
9. Uganda
10. Burkina Faso



Các nước trên nằm chủ yếu ở Châu Phi. Tỉ lệ sinh cao ở các quốc gia châu Phi có thể được giải thích bởi sự kết hợp của các yếu tố như điều kiện kinh tế khó khăn, giáo dục thấp, giá trị văn hóa tích cực đối với gia đình lớn, và các thách thức y tế công cộng. Những yếu tố này tạo nên môi trường mà trong đó việc có nhiều con được xem là cần thiết cho sự bền vững gia đình và vượt qua các khó khăn kinh tế và xã hội.

- ❖ So sánh tỷ lệ sinh và tỷ lệ tuyển sinh vào tiểu học, đại học



Khi tỷ lệ sinh càng cao thì tỷ lệ trẻ em được tuyển sinh vào tiểu học và đại học khá thấp.

Tỷ lệ sinh ở mức từ khoảng 5-15% thì tỷ lệ tuyển sinh vào tiểu học, đại học đạt mức cao nhất.

Tỷ lệ sinh khoảng dưới 5% thì hầu như tỷ lệ tuyển sinh vào tiểu học và đại học gần như bằng 0%

3.3. Sự tương quan giữa tỷ thất nghiệp với các cột

- ❖ Top 5 mối tương quan cao nhất:

Completion_Rate_Lower_Secondary_Female	0.297181
Completion_Rate_Lower_Secondary_Male	0.285564
Completion_Rate_Upper_Secondary_Female	0.285201
Completion_Rate_Primary_Female	0.281272
Completion_Rate_Primary_Male	0.277663
Name: Unemployment_Rate, dtype: float64	

➤ Tỷ lệ thất nghiệp giữa nam và nữ không quá khác biệt

- Học vấn cao giúp giảm tỷ lệ thất nghiệp
 - ❖ Top 5 mối tương quan thấp nhất:
- | | |
|---|-----------|
| Longitude | -0.278258 |
| OOSR_Upper_Secondary_Age_Female | -0.104612 |
| OOSR_Upper_Secondary_Age_Male | -0.103774 |
| Grade_2_3_Proficiency_Math | -0.057321 |
| OOSR_Lower_Secondary_Age_Female | -0.042566 |
| Name: Unemployment_Rate, dtype: float64 | |
- Nó cho thấy có mối quan hệ tiêu cực yếu giữa tọa độ địa lý của các quốc gia và tỷ lệ thất nghiệp.
 - Vì vậy, chúng ta có thể nói rằng những thay đổi về vị trí đông-tây có liên quan đến việc giảm tỷ lệ thất nghiệp.
 - Cần phải phân tích sâu hơn cũng như thông tin theo ngữ cảnh để xác định xem các mối quan hệ có phải là nguyên nhân hay không.

3.4. Tỷ lệ bỏ học ở độ tuổi mầm non

- ❖ Giới tính nam

Countries and areas	OOSR_PrePrimary_Age_Male
0	Afghanistan
92	Kenya
161	Singapore
185	Turkmenistan
33	Canada
...	...
35	Chad
51	Djibouti
176	Tajikistan
122	Myanmar
199	Yemen
202 rows × 2 columns	

- Quốc gia có số lượng nam sinh mầm non ít nhất là Afghanistan và cao nhất là Yemen

❖ Giới tính nữ

Countries and areas	OOSR_PrePrimary_Age_Female	
0	Afghanistan	0
86	Israel	0
186	Turks and Caicos Islands	0
33	Canada	0
34	Central African Republic	0
...
35	Chad	87
176	Tajikistan	88
122	Myanmar	88
51	Djibouti	88
199	Yemen	96

202 rows × 2 columns

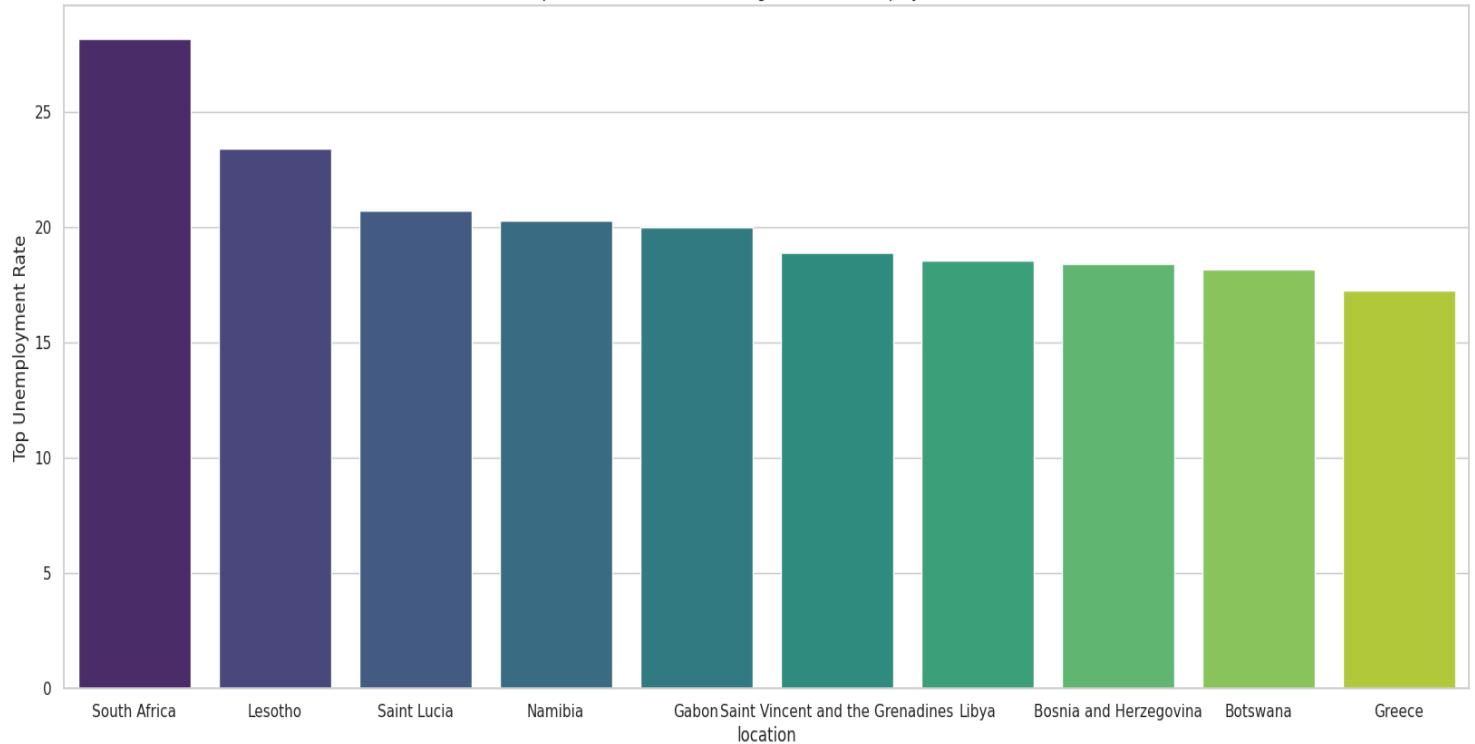
➤ Quốc gia có số lượng nữ sinh mầm non ít nhất là Afghanistan và cao nhất là Yemen

3.5. Tỷ lệ thất nghiệp của các quốc gia

3.5.1. Top 10 quốc gia/ khu vực có tỷ lệ thất nghiệp cao nhất

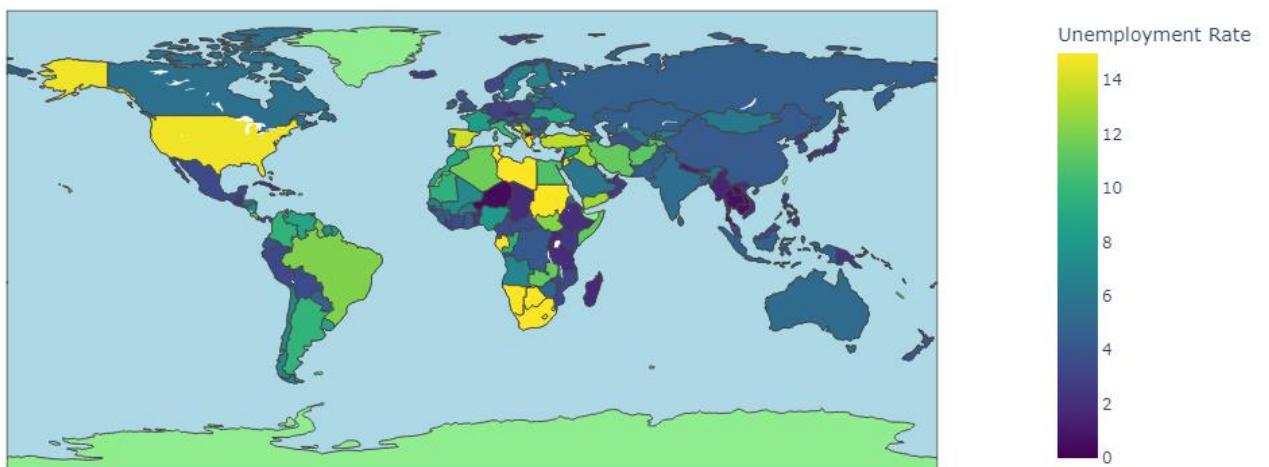
1. South Africa
2. Lesotho
3. Saint Lucia
4. Namibia
5. Gabon
6. Saint Vincent and Grenadines
7. Libya
8. Bosnia and Herzegovina
9. Botswana
10. Greece

Top 10 Countries/Areas with Highest Total Unemployment Rates



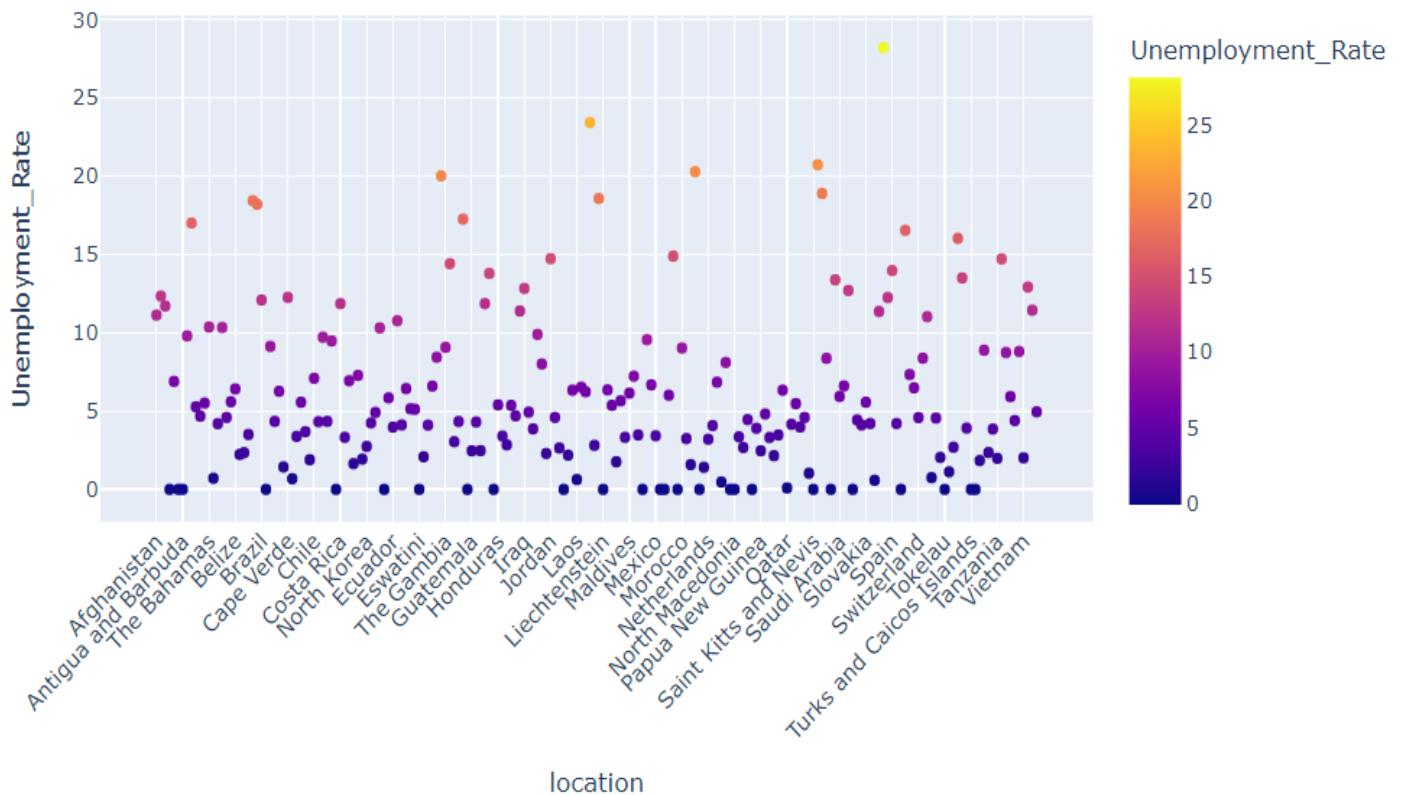
Bộ dữ liệu đã xác định Nam Phi là quốc gia có tỷ lệ thất nghiệp cao nhất. Trong số 10 quốc gia có tỷ lệ thất nghiệp cao nhất, 6 quốc gia nằm ở Châu Phi và 2 trong số đó nằm ở Caribe. Nhìn chung, tỷ lệ thất nghiệp trên toàn cầu là tương đương nhau.

Unemployment Rate by Country



Sơ đồ tỷ lệ thất nghiệp chung (Biểu đồ choropleth)

Unemployment Rates Across Countries



Tỷ lệ thất nghiệp trên khắp các quốc gia (biểu đồ scatter-biểu đồ phân tán)

Tỷ lệ thất nghiệp trên toàn cầu nằm trong khoảng 0-10% là chủ yếu. Đây cũng chưa phải là tình hình thực tế sau đại dịch Covid. Chỉ khoảng dự đoán tích cực

3.6. Tỷ lệ tuyển sinh

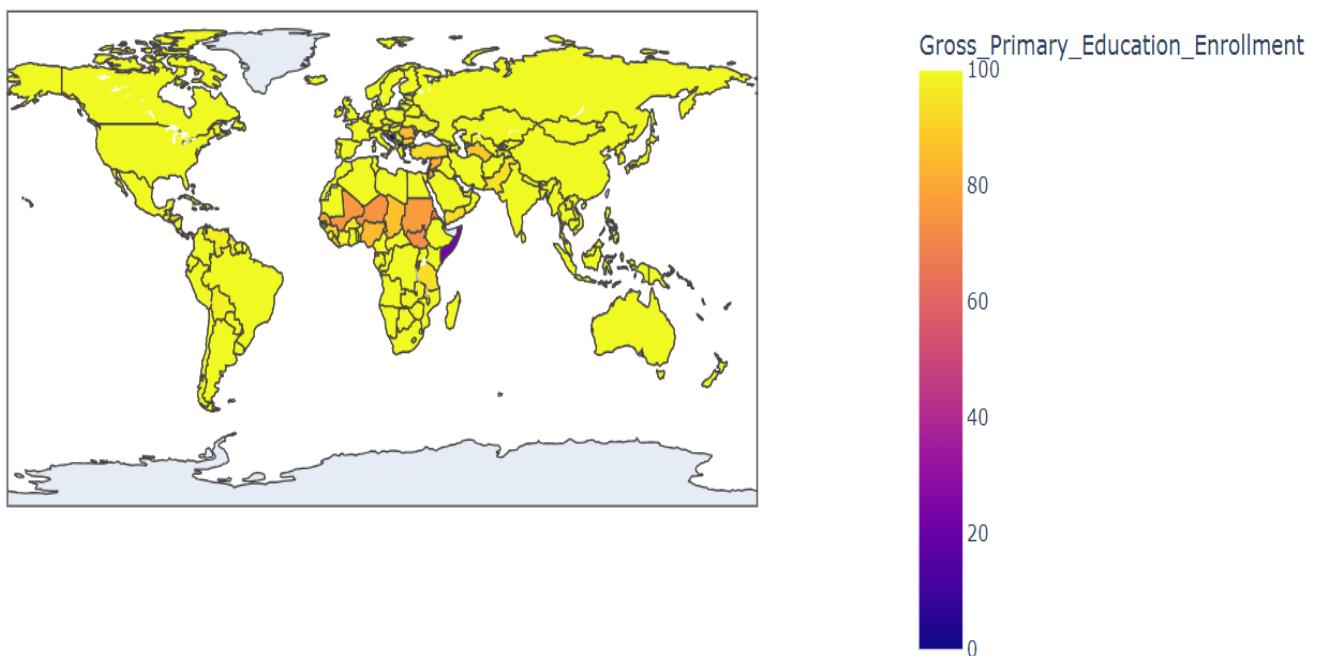
Bộ dữ liệu chỉ ra rằng, trái ngược với một số quốc gia châu Phi, tỷ lệ nhập học chung cho giáo dục tiểu học trong khu vực nhìn chung ở mức thuận lợi. Tuy nhiên, các quốc gia như Somalia lại có tỷ lệ nhập học chung thấp hơn.

Tổng số tuyển sinh toàn cầu vào giáo dục đại học thể hiện nhiều đặc điểm khác nhau. Úc có tổng số sinh viên ghi danh cao nhất toàn cầu, trong khi Hoa Kỳ, Canada và một số nước châu Âu có số lượng sinh viên theo học trung bình. Các quốc gia Nam Mỹ có tổng số tuyển sinh giảm, trong khi một số quốc gia châu Á giảm xuống dưới mức trung bình và các nước châu Phi có tổng số tuyển sinh thấp nhất.

Tỷ lệ nhập học chung ở cấp Tiểu học thấp hơn đáng kể so với tỷ lệ nhập học chung ở cấp Đại học, cho thấy sự chênh lệch đáng kể giữa cấp tiểu học và cấp đại học. Đây có thể là dấu hiệu của một vấn đề quan trọng trong bộ giáo dục.

3.6.1. Tỷ lệ tuyển sinh tiểu học

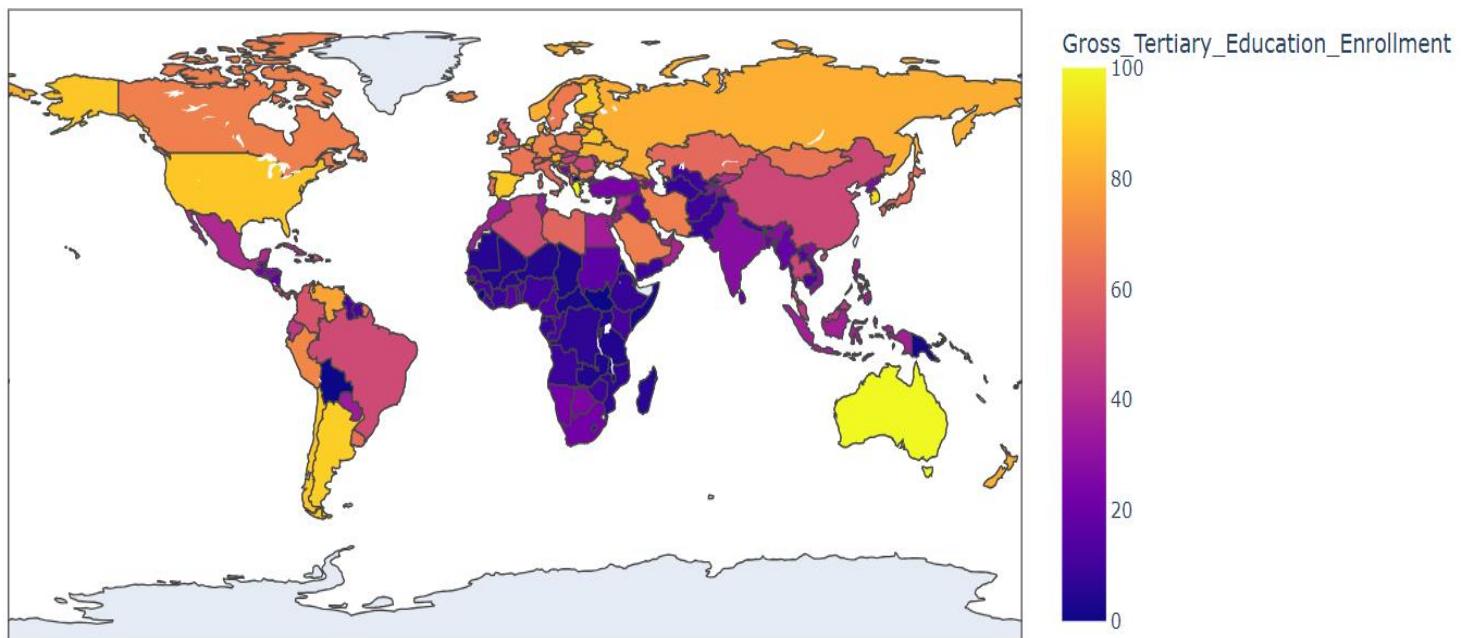
Gross Primary Education Enrollment



(Biểu đồ choropleth)

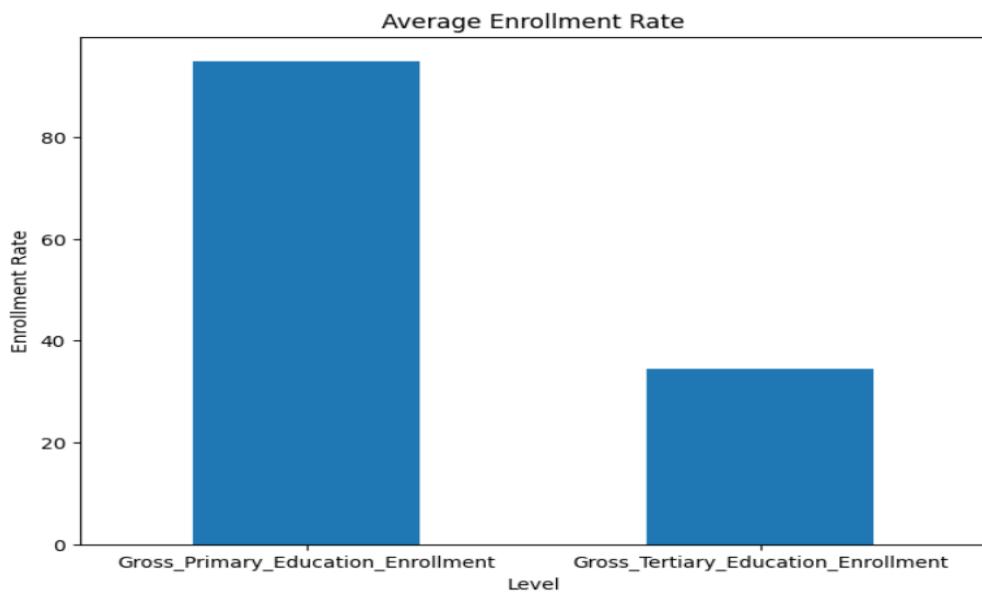
3.6.2. Tỷ lệ tuyển sinh đại học

Gross Tertiary Education Enrollment



(Biểu đồ choropleth)

3.6.3. Tỷ lệ nhập học trung bình



Tỷ lệ nhập học trung bình đầu vào tiểu học luôn cao hơn tỷ lệ nhập học đầu vào đại học, điều này có thể được giải thích bởi nhiều yếu tố như sự phổ cập giáo dục cơ bản, chính sách giáo dục, hay nguy cơ tiếp tục học vấn ở đại học...

3.6.4. Tỷ lệ tuyển sinh chung

Overall Enrollment Rate



Biểu đồ cột (bar plot)

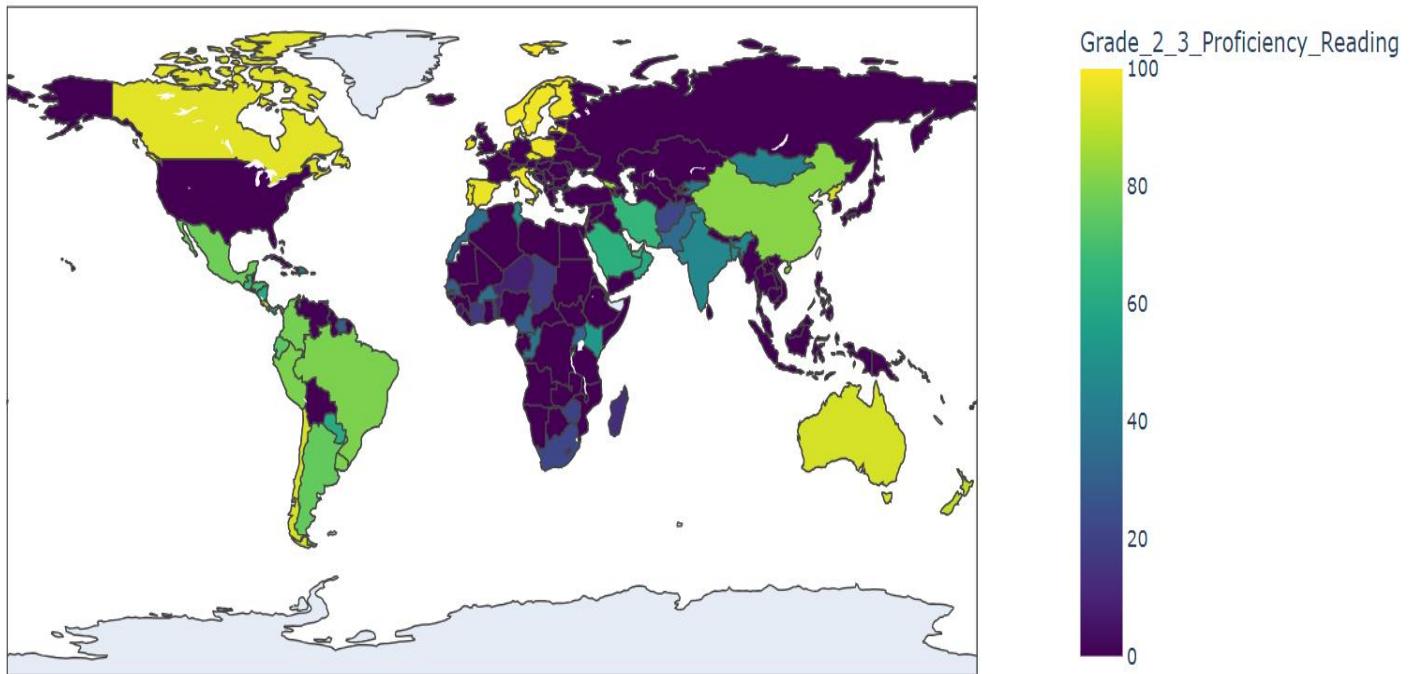
Tỷ lệ nhập học đại học ở các quốc gia luôn thấp hơn tỷ lệ nhập học tiểu học

3.7. Khả năng đọc thành thạo, trình độ Toán học theo quốc gia

3.7.1. Khả năng đọc thành thạo

- Mức độ thành thạo đọc dành cho học sinh lớp 2-3

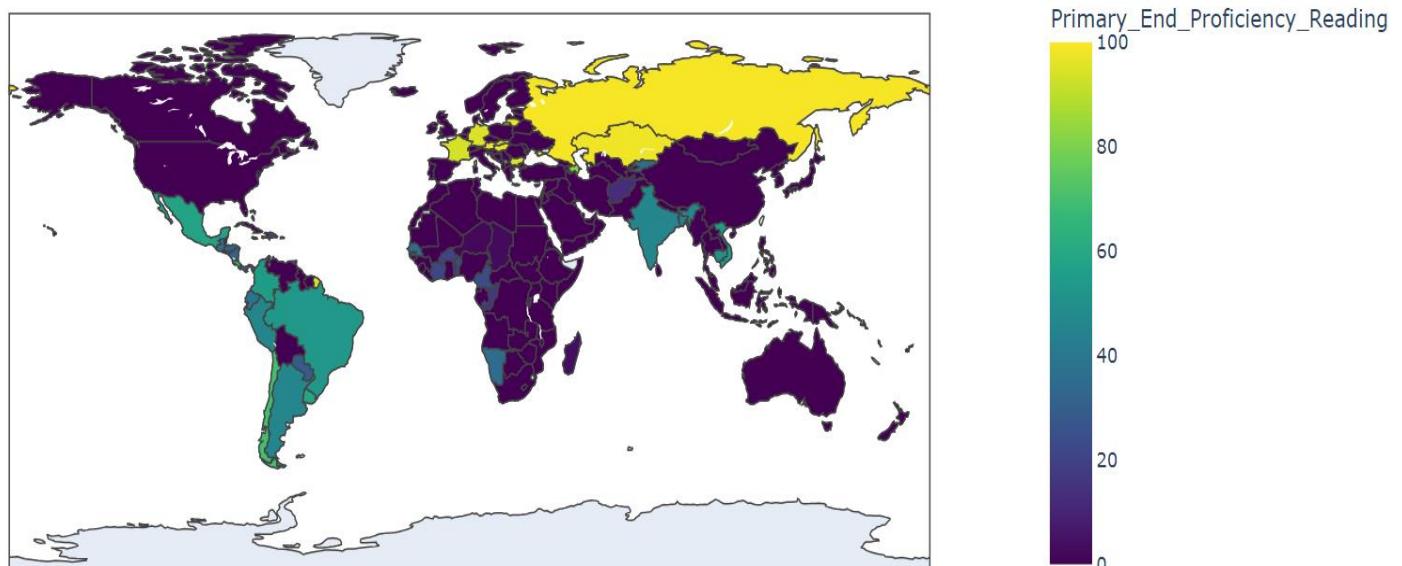
Grade 2-3 Reading Proficiency



Canada, Úc và một số quốc gia Châu Âu có khả năng đọc cao nhất ở lớp hai và lớp ba. Nam Mỹ, Trung Quốc và Châu Phi có trình độ đọc thấp nhất. Nga và Hoa Kỳ không có bất kỳ dữ liệu nào cho thấy rằng có thể đọc thành thạo của học sinh lớp hai và lớp ba.

- Mức độ thành thạo đọc khi kết thúc tiểu học

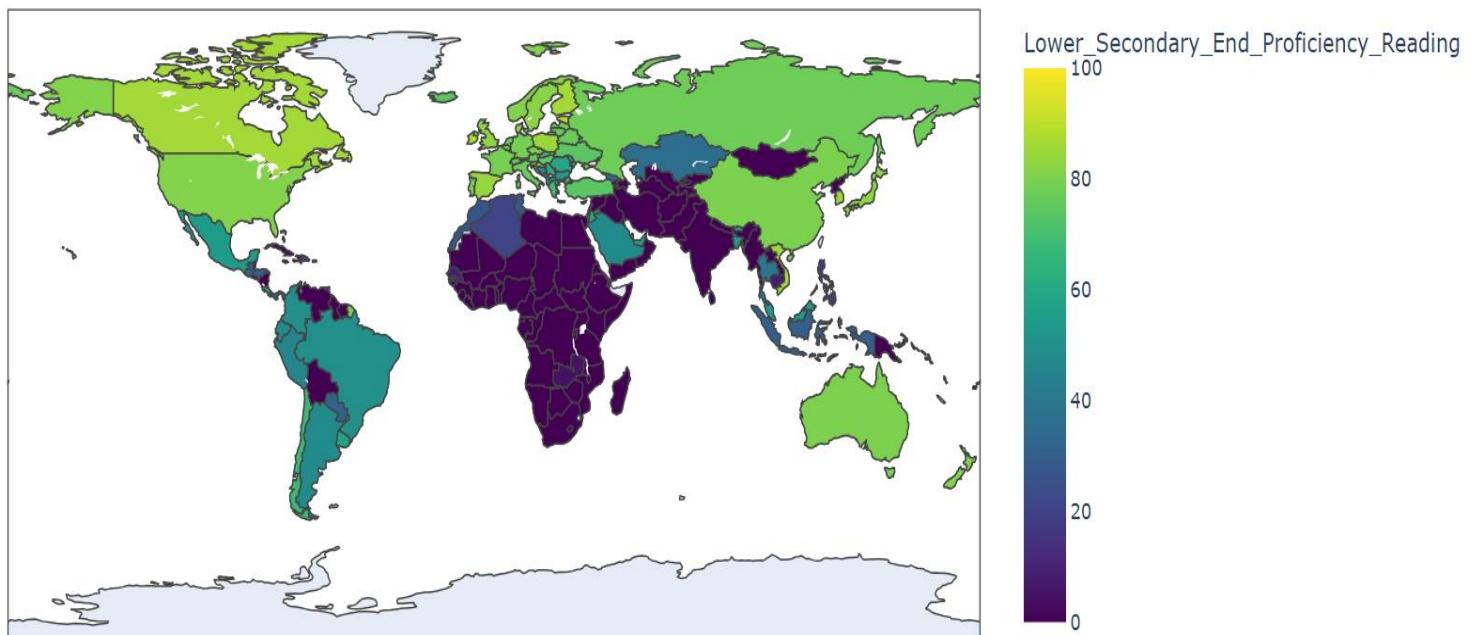
Primary End Reading Proficiency



Nga đứng đầu về trình độ đọc ở cấp tiểu học, tiếp theo là Mexico, các nước Nam Mỹ và Ấn Độ. Tuy nhiên, một số nước, chẳng hạn như Hoa Kỳ, Canada và một số quốc gia Châu Âu và Châu Á, không có dữ liệu cho thấy rằng học sinh có thể đọc thành thạo ở độ tuổi tiểu học.

- Mức độ thành thạo đọc khi kết thúc THCS

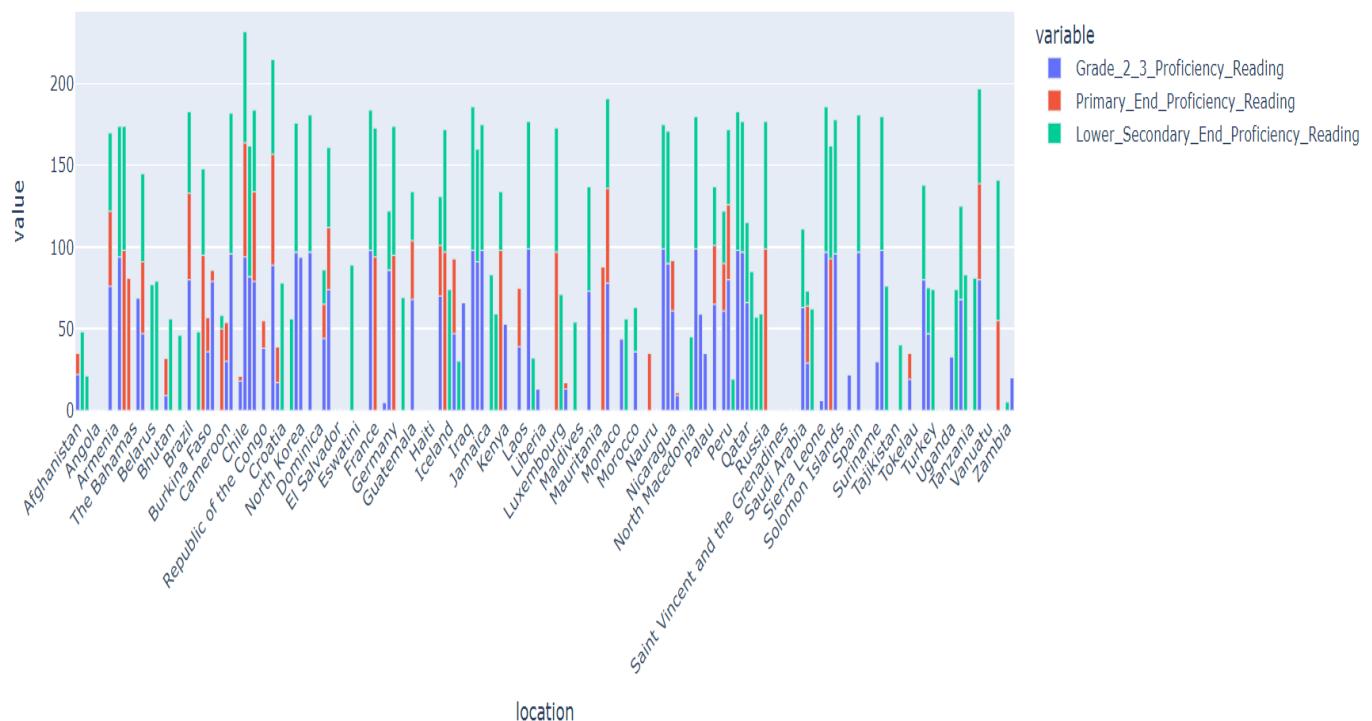
Lower Secondary Reading Proficiency



Bộ dữ liệu chỉ ra rằng Canada và Hoa Kỳ, cũng như các nước ở Châu Âu, Úc và một số quốc gia ở Châu Á, có mức độ đọc thông thạo cao nhất. Các nước Nam Mỹ kém thành thạo hơn một chút. Hơn nữa, một số quốc gia Châu Phi và Trung Đông không có bất kỳ dữ liệu nào được ghi lại cho thấy rằng học sinh có thể đọc thành thạo ở cấp trung học cơ sở.

- Tỷ lệ thành thạo đọc ở các cấp độ giáo dục

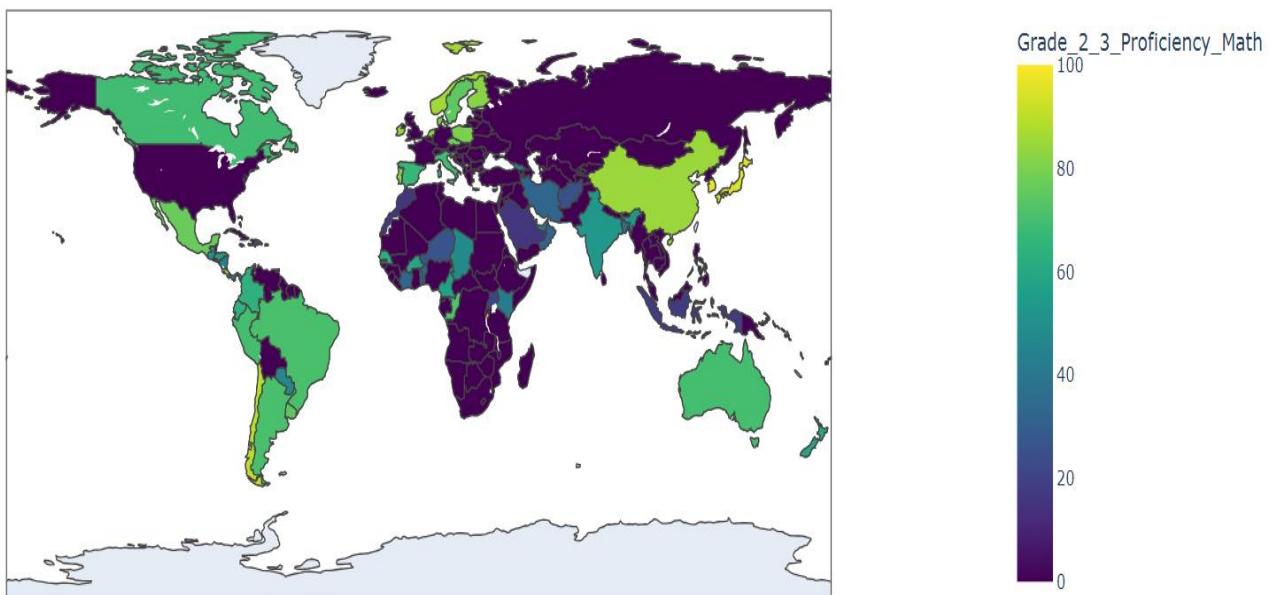
Reading Proficiency Rates Over Different Education Levels



3.7.2. Thành thạo toán học

- Mức độ thành thạo toán học dành cho học sinh lớp 2-3

Grade 2-3 Math Proficiency

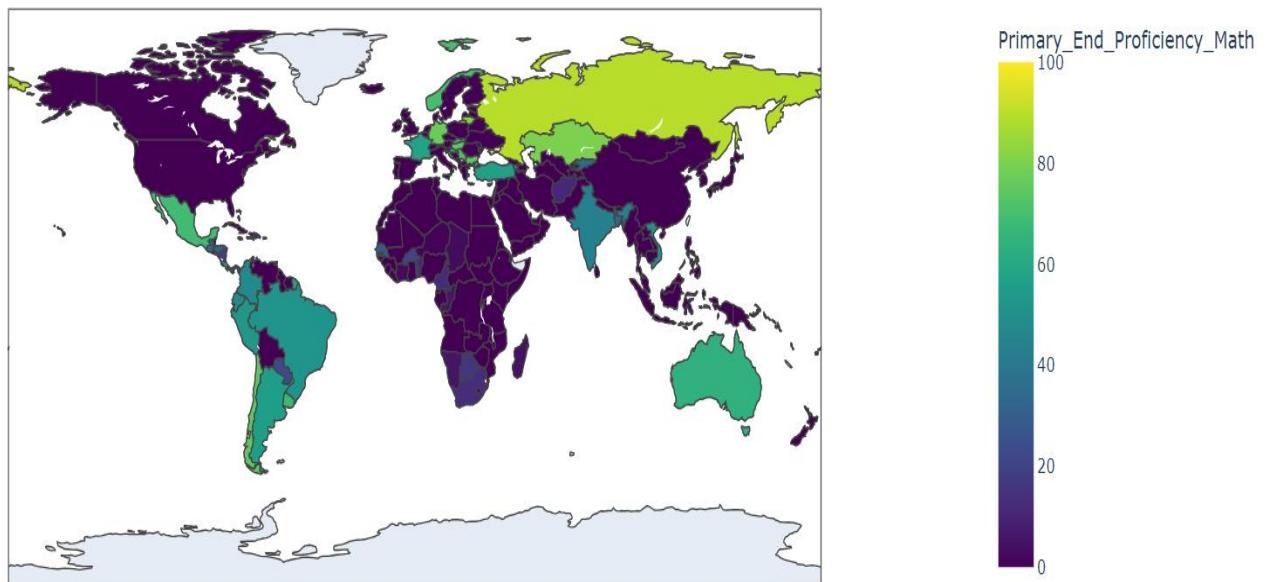


Ở Nam Mỹ, phần lớn các quốc gia đều thể hiện trình độ thông thạo môn Toán ở lớp 2-3. Nhật Bản đã thể hiện trình độ thông thạo cao nhất, sau là Trung Quốc và

Na Uy, Canada và Úc đã chứng tỏ được một số mức độ thành thạo, trong khi Hoa Kỳ và phần lớn các nước châu Âu không có dữ liệu để chứng minh bất kỳ mức độ thành thạo nào. Mặt khác, Châu Phi có tỷ lệ thông thạo thấp nhất từng được ghi nhận.

- Mức độ thành thạo toán học khi kết thúc tiểu học

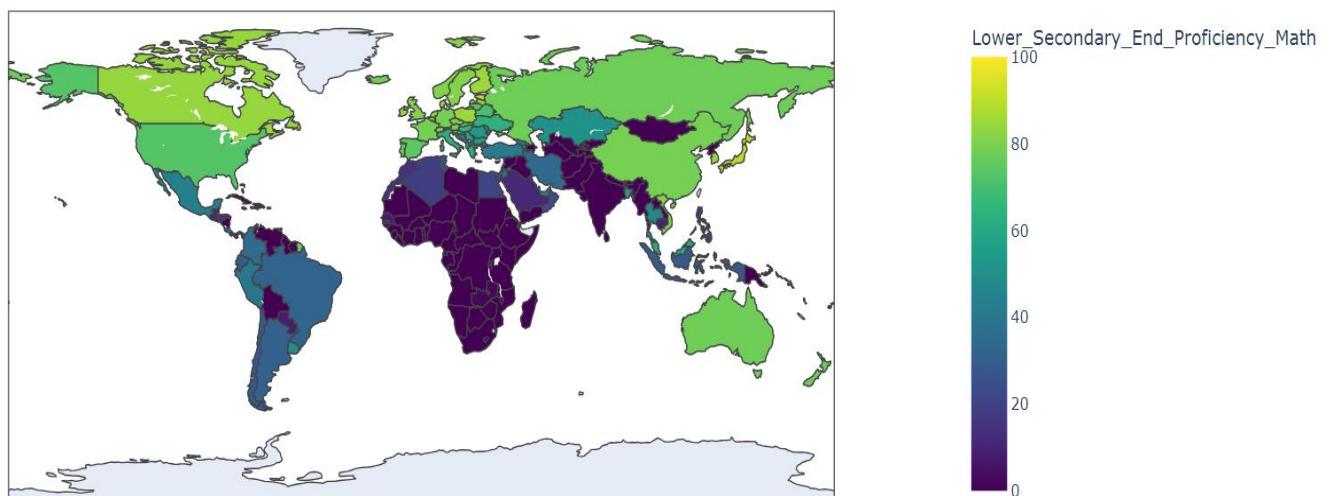
Primary End Math Proficiency



Bộ dữ liệu chỉ ra rằng Nga là quốc gia có trình độ Toán cấp tiểu học cao nhất, tiếp theo là Úc, Nam Mỹ như Mexico và Chile, Bắc Mỹ và Châu Âu, những nơi khác không có dữ liệu.

- Mức độ thành thạo toán học khi kết thúc THCS

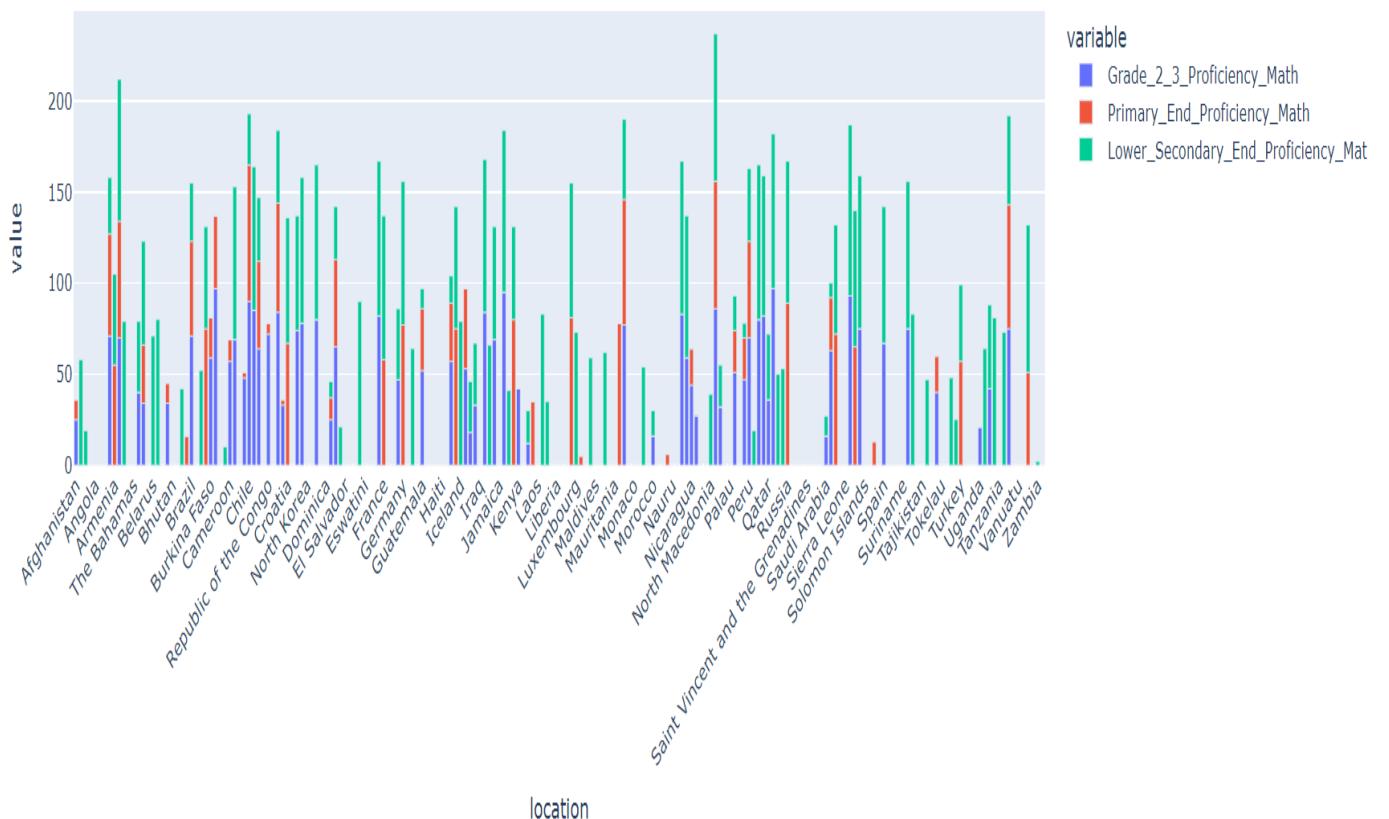
Lower Secondary Math Proficiency



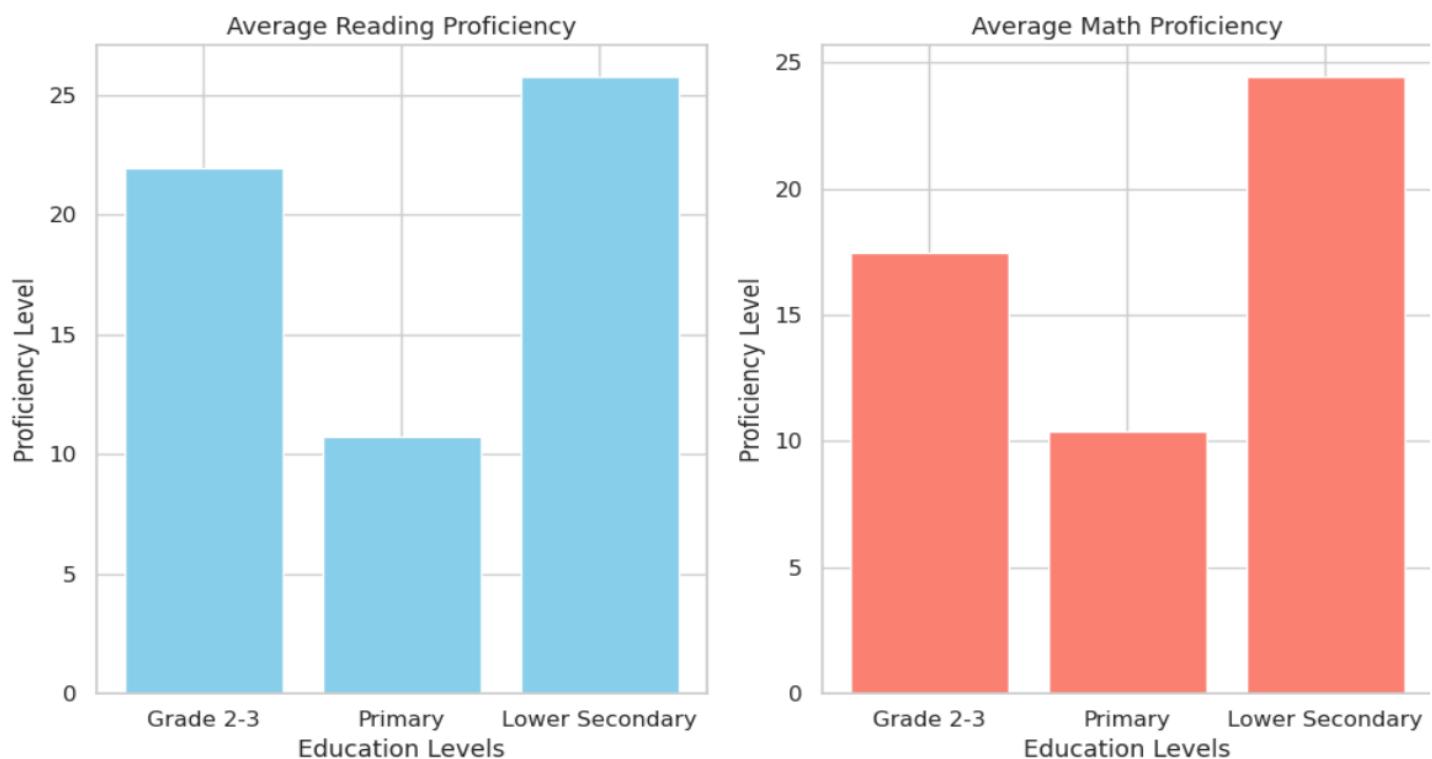
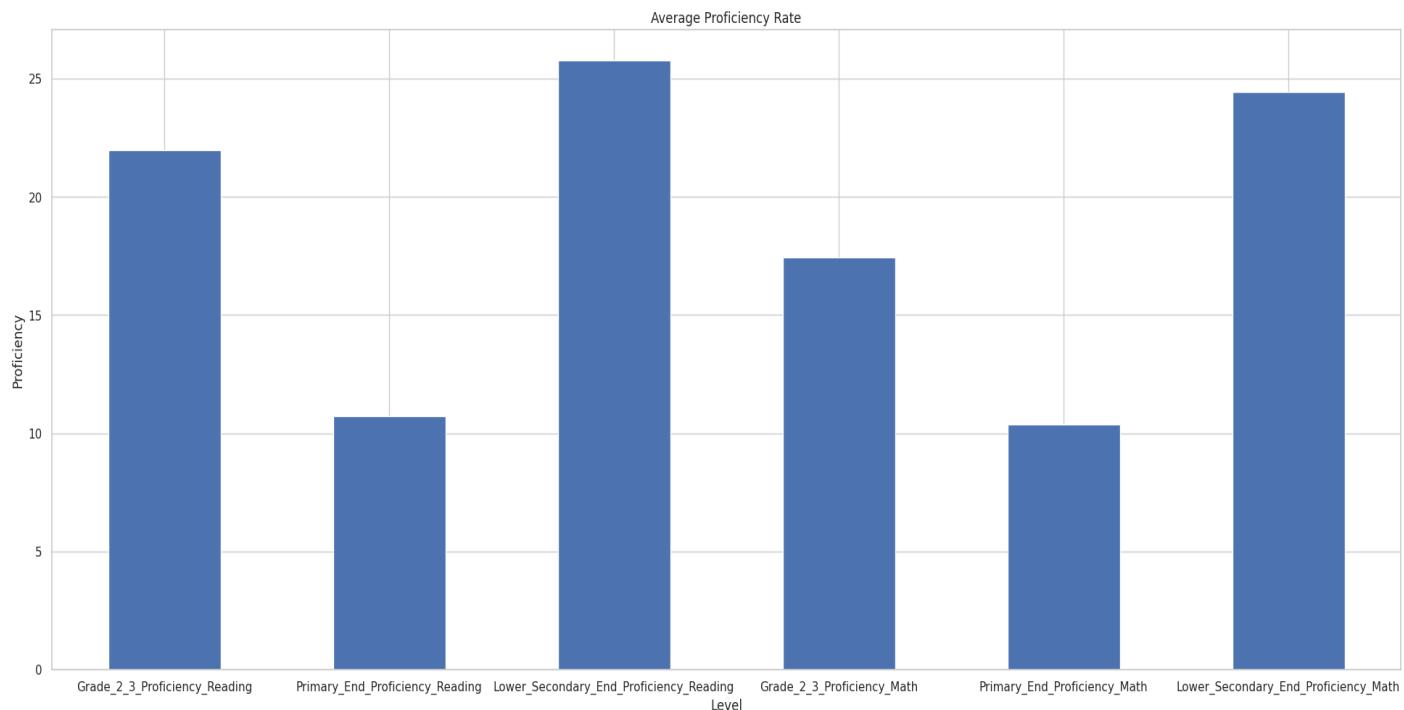
Canada và Nhật Bản có mức độ thông thạo cao nhất ở cấp trung học cơ sở, trong khi Phần Lan là quốc gia gần nhất với vị trí dẫn đầu. Phần lớn các nước châu Âu, cũng như Hoa Kỳ và Úc, có trình độ thông thạo thấp hơn Nga và Trung Quốc. Mexico, Brazil và một số quốc gia ở Nam Mỹ có trình độ thông thạo thấp hơn một chút so với Hoa Kỳ. Nam Phi cũng như châu Á không có sẵn bất kỳ dữ liệu nào về mức độ thành thạo.

- Tỷ lệ thành thạo toán học ở các cấp độ giáo dục

Math Proficiency Rates Over Different Education Levels



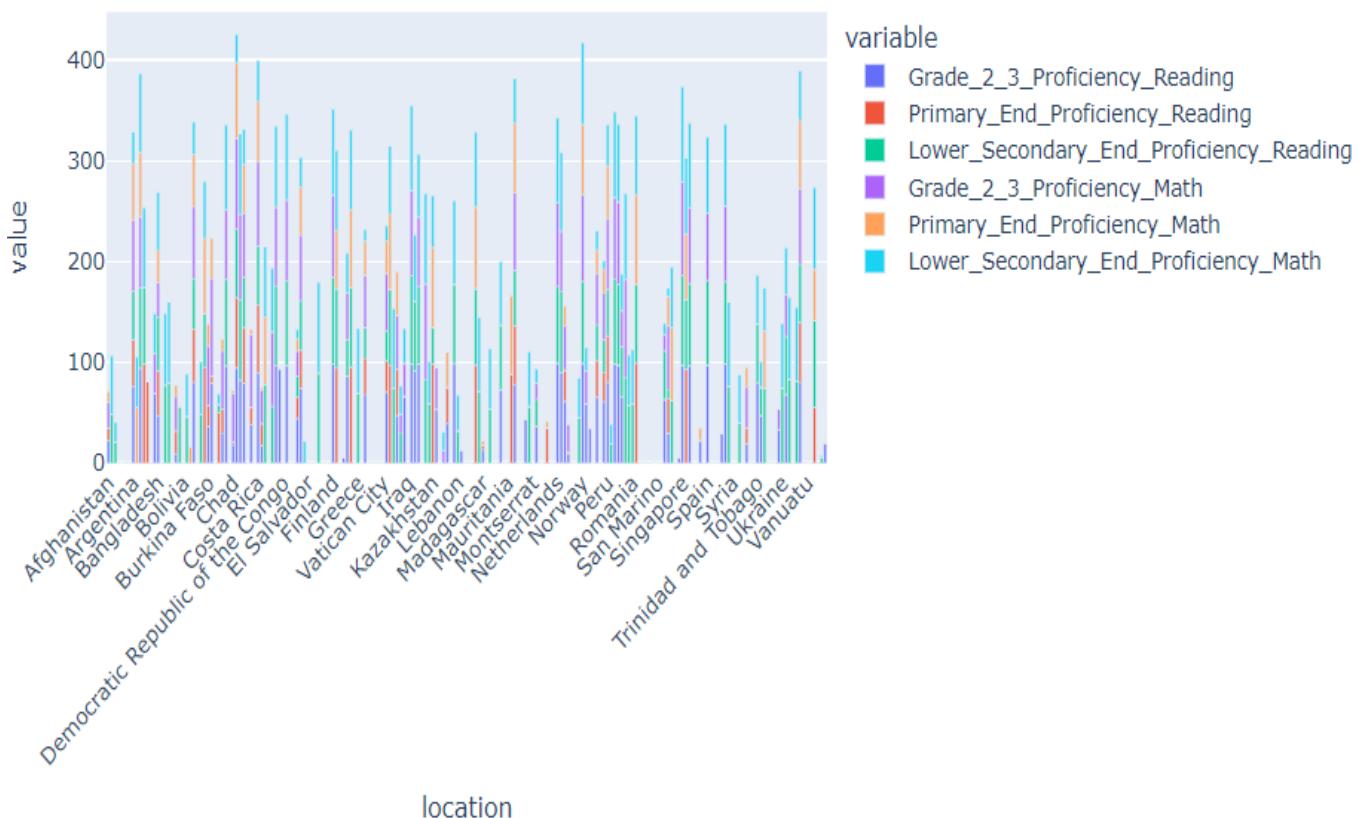
3.7.3. Mức độ thông thạo (đọc, toán học) trung bình ở các cấp độ giáo dục



Phân tích này kiểm tra mức độ thành thạo trung bình trong cả môn đọc và toán. Cấp tiểu học được đặc trưng bởi mức độ thông thạo thấp nhất, tiếp theo là cấp độ học từ lớp 2-3, và mức độ thông thạo cao nhất là ở cấp trung học cơ sở.

3.7.4. Tổng thể mức độ thành thạo ở các cấp độ giáo dục

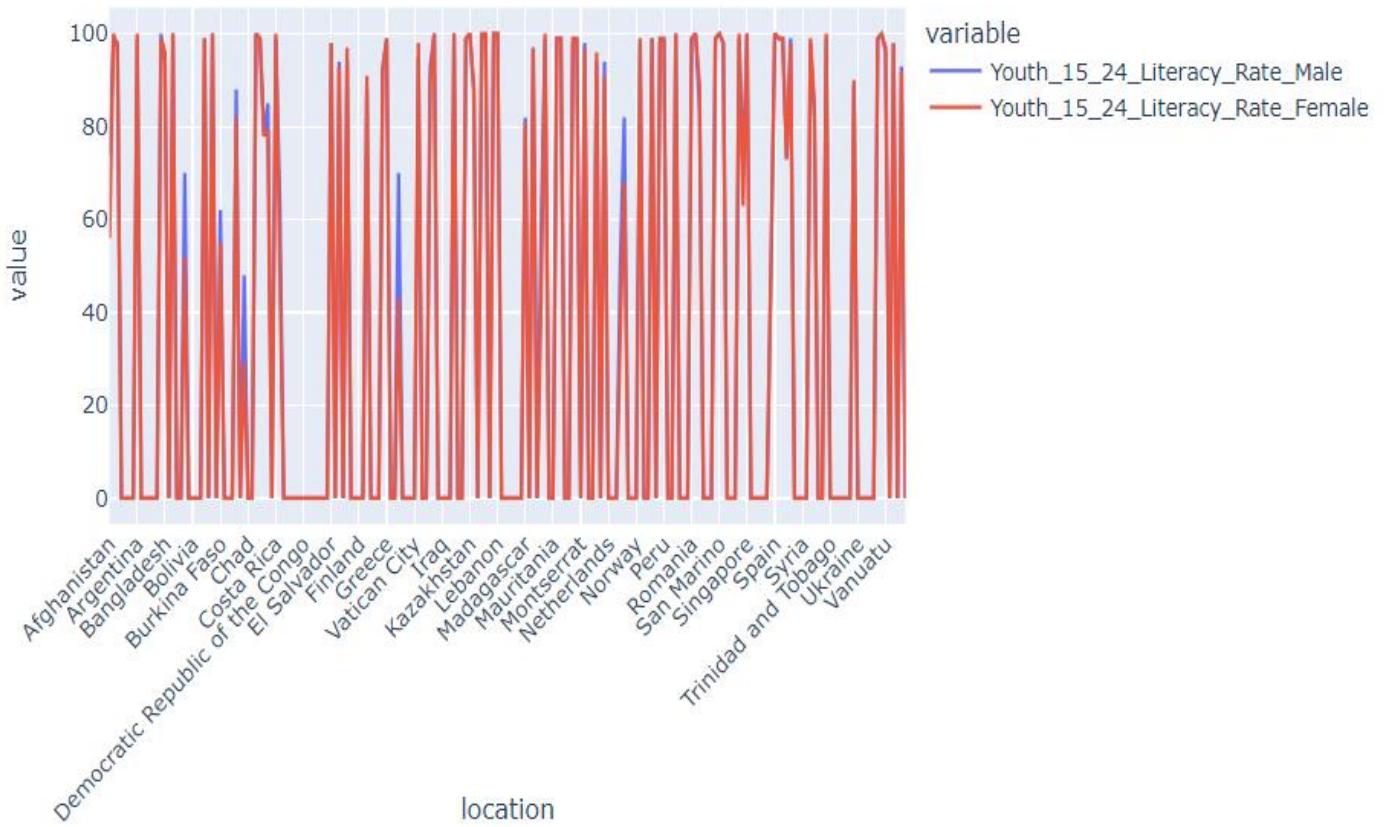
Proficiency Rates Over Different Education Levels



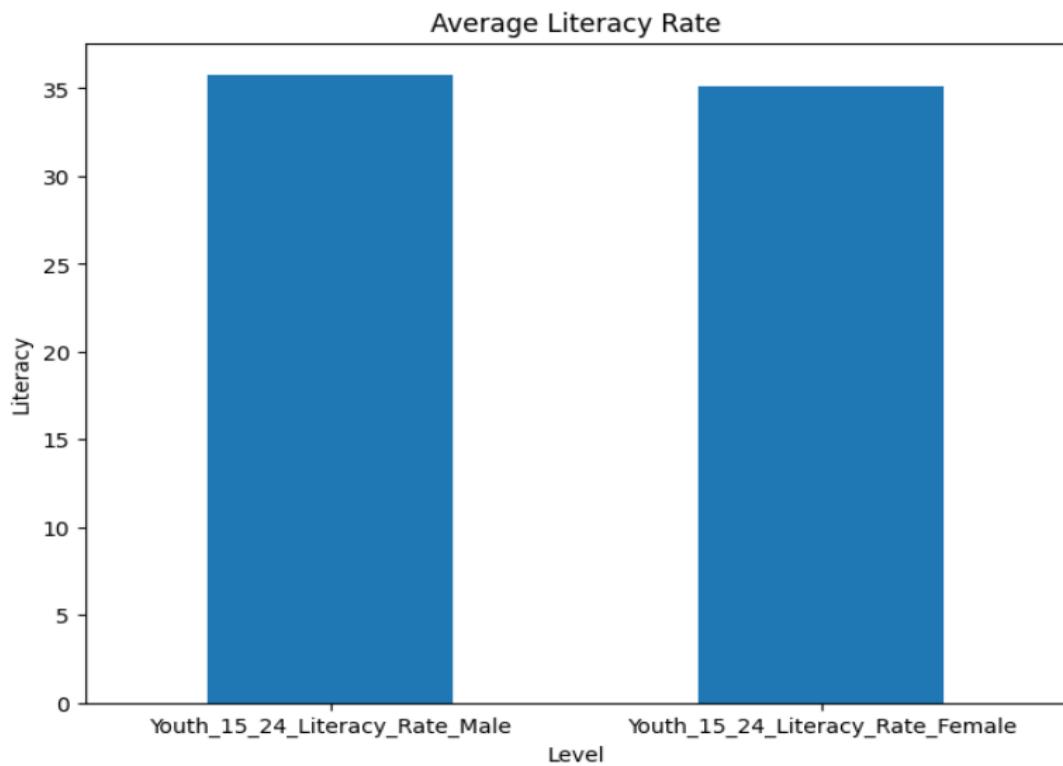
3.8. Tỷ lệ biết chữ ở thanh niên

Tỷ lệ biết chữ của học sinh trong độ tuổi 15-24 từ các quốc gia nơi dữ liệu được thu thập không cho thấy sự khác biệt đáng kể về giới tính.

Youth Literacy Rate



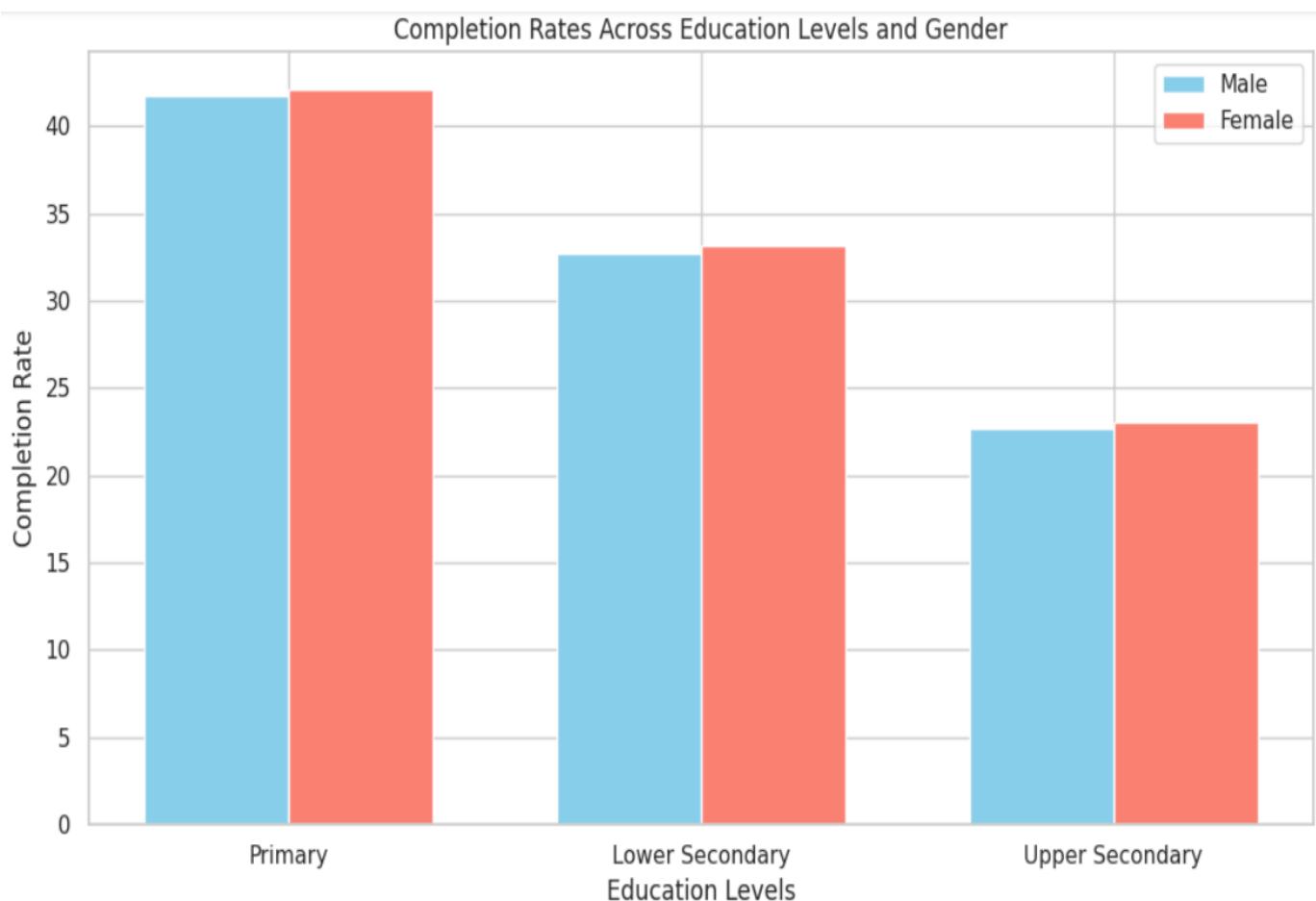
- Tỷ lệ biết chữ trung bình



Nhìn chung, tỷ lệ biết chữ của học sinh trong độ tuổi 15-24 ở nam cao hơn so với nữ 1 ít, nhưng không quá khác biệt.

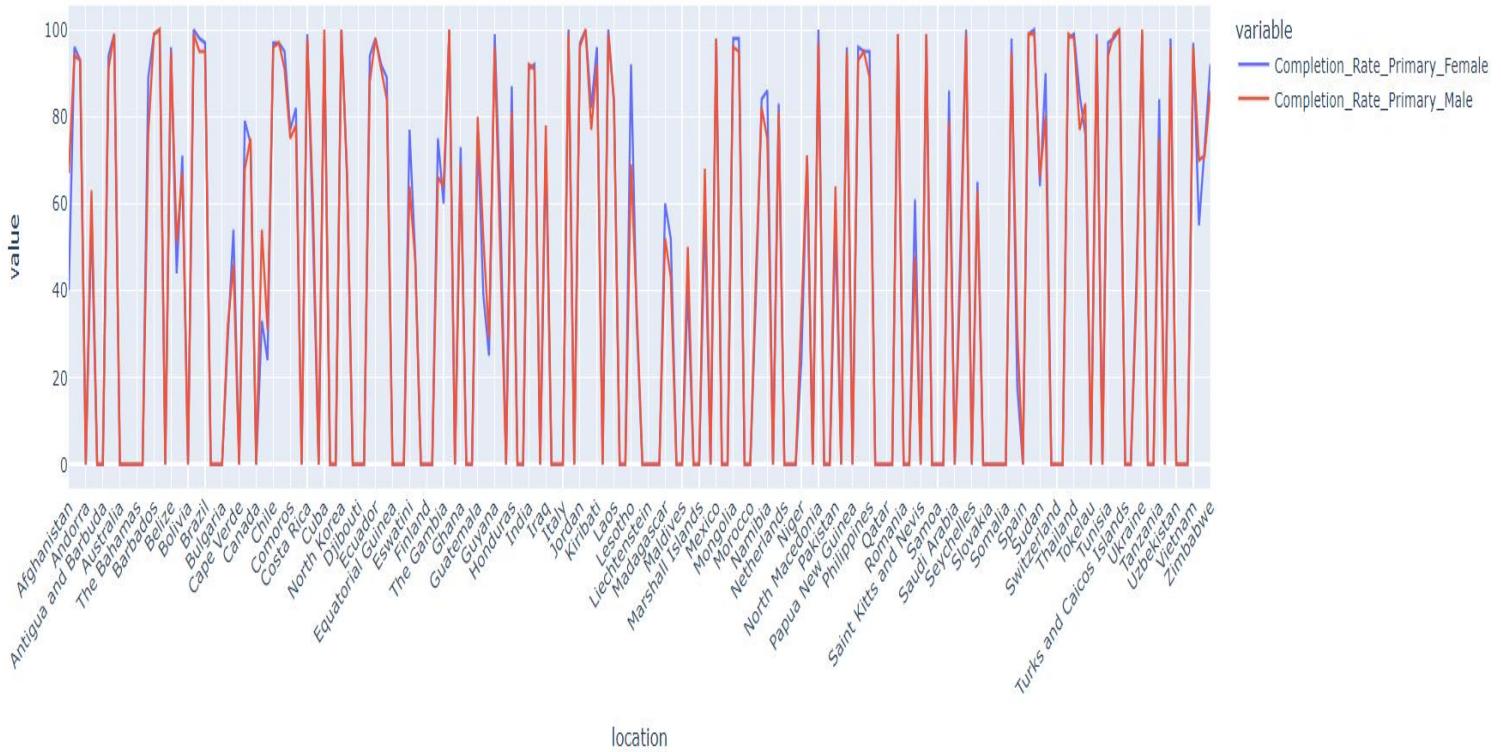
3.9. Tỷ lệ hoàn thành các cấp bậc giáo dục

Khi xem xét mức độ hoàn thành chương trình, nhìn chung không có sự khác biệt đáng kể giữa các giới tính. Cấp tiểu học có tỷ lệ hoàn thành cao nhất, tiếp theo là cấp trung học cơ sở và trung học phổ thông, trong đó cấp trung học phổ thông có tỷ lệ hoàn thành thấp nhất. Điều này cho thấy giáo dục tiểu học tập trung cao độ ở hầu hết các quốc gia, tuy nhiên, tỷ lệ hoàn thành các bậc giáo dục giảm dần khi trình độ học vẫn tăng lên.



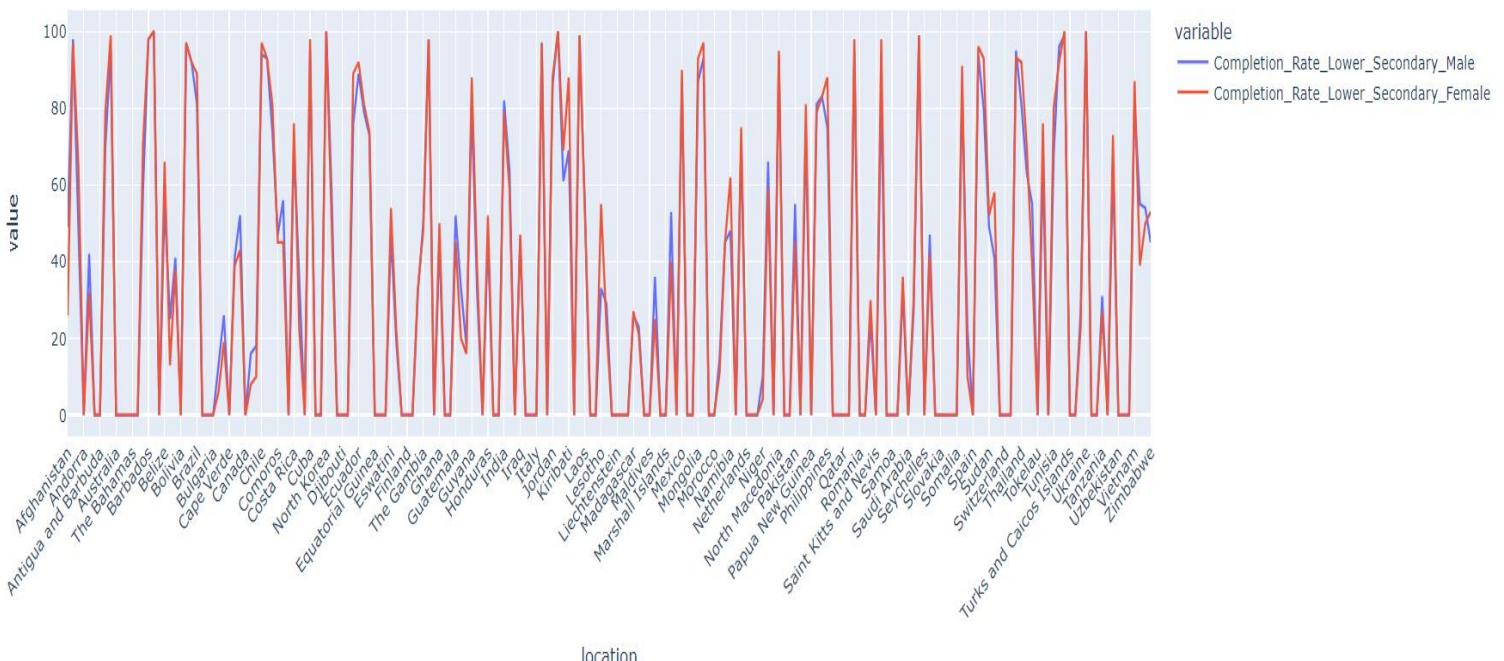
3.9.1. Tỷ lệ hoàn thành ở bậc tiểu học

Primary Completion Rate



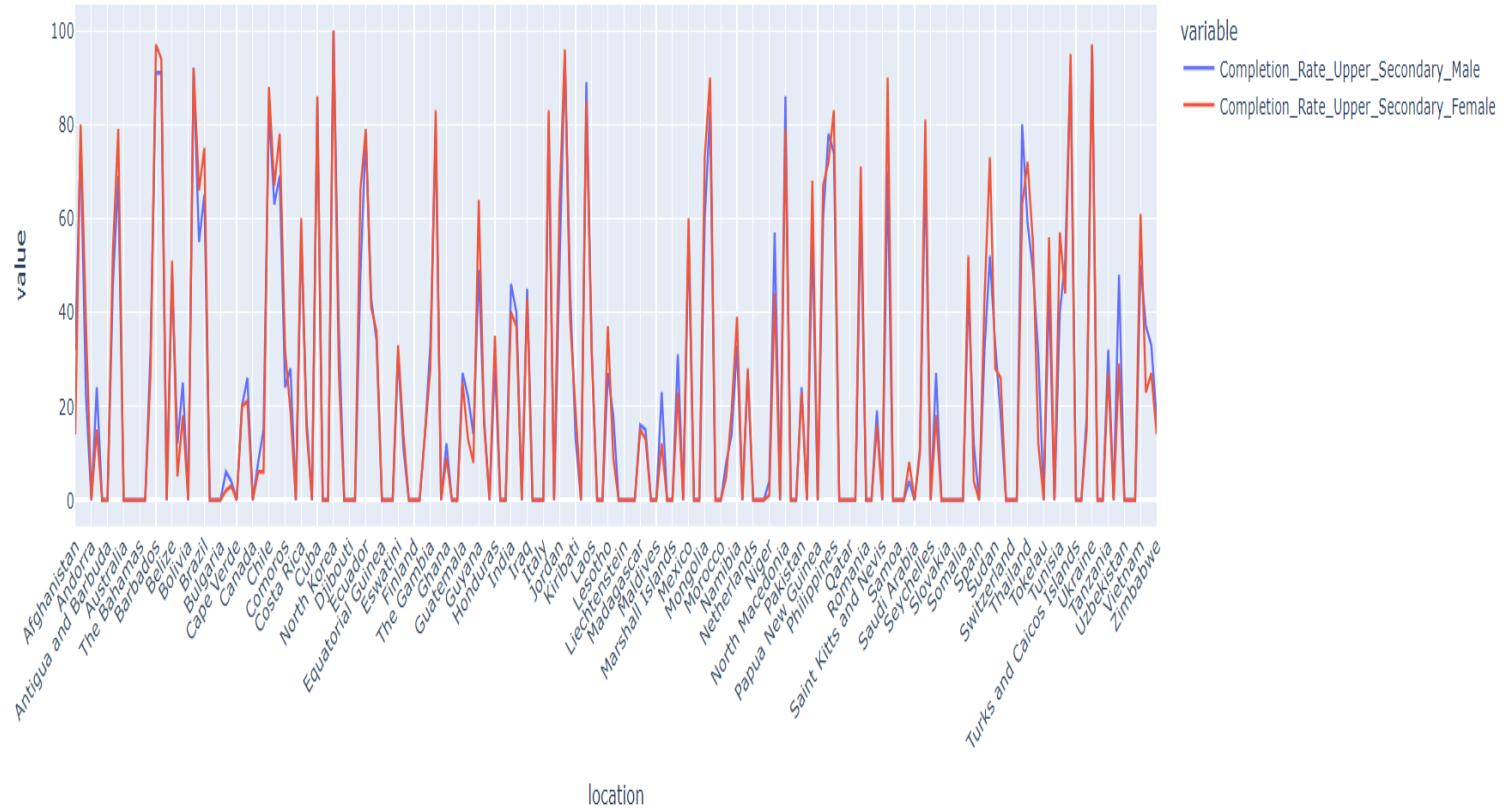
3.9.2. Tỷ lệ hoàn thành bậc trung học cơ sở

Lower Secondary Completion Rate

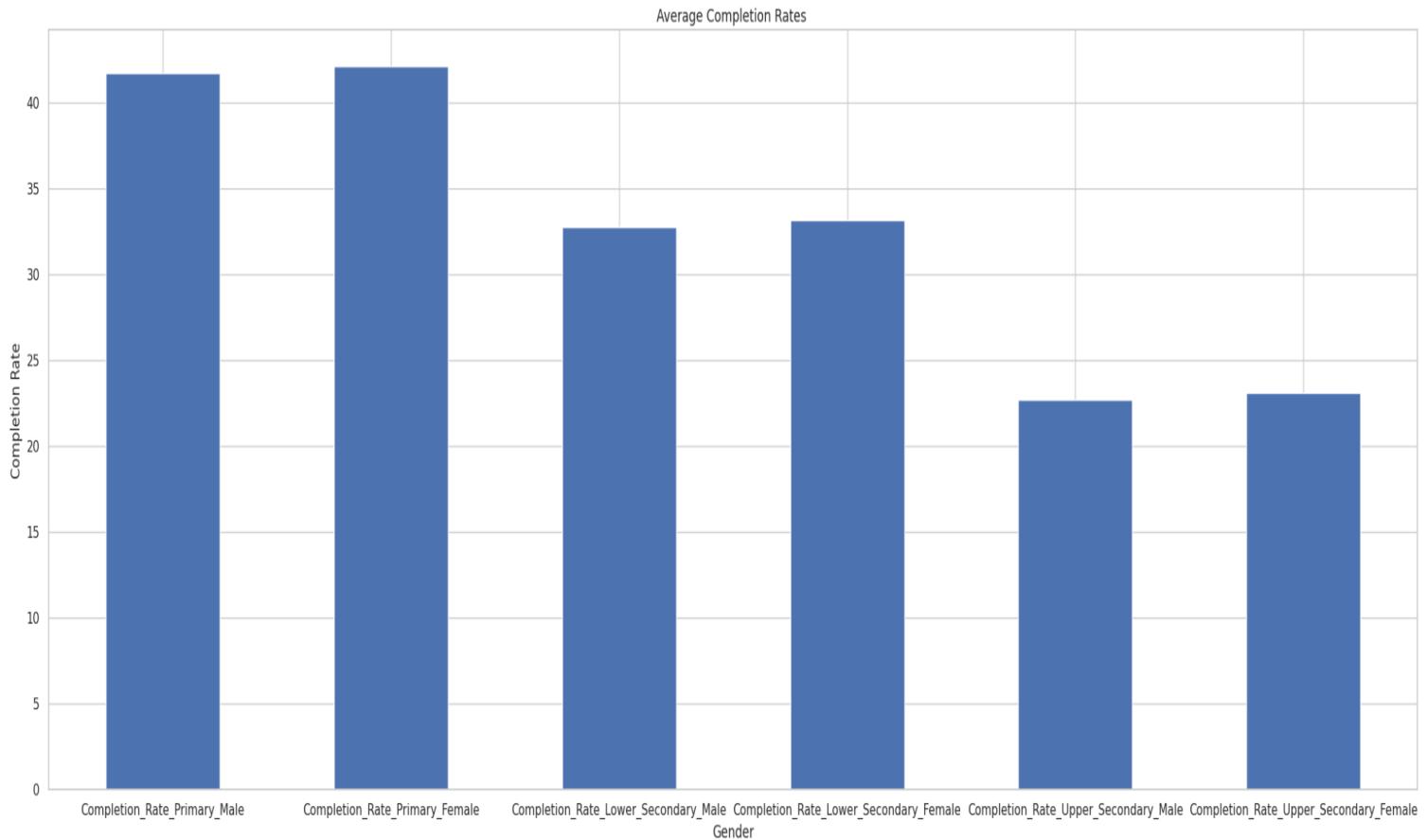


3.9.3. Tỷ lệ hoàn thành bậc trung học phổ thông

Upper Secondary Completion Rate



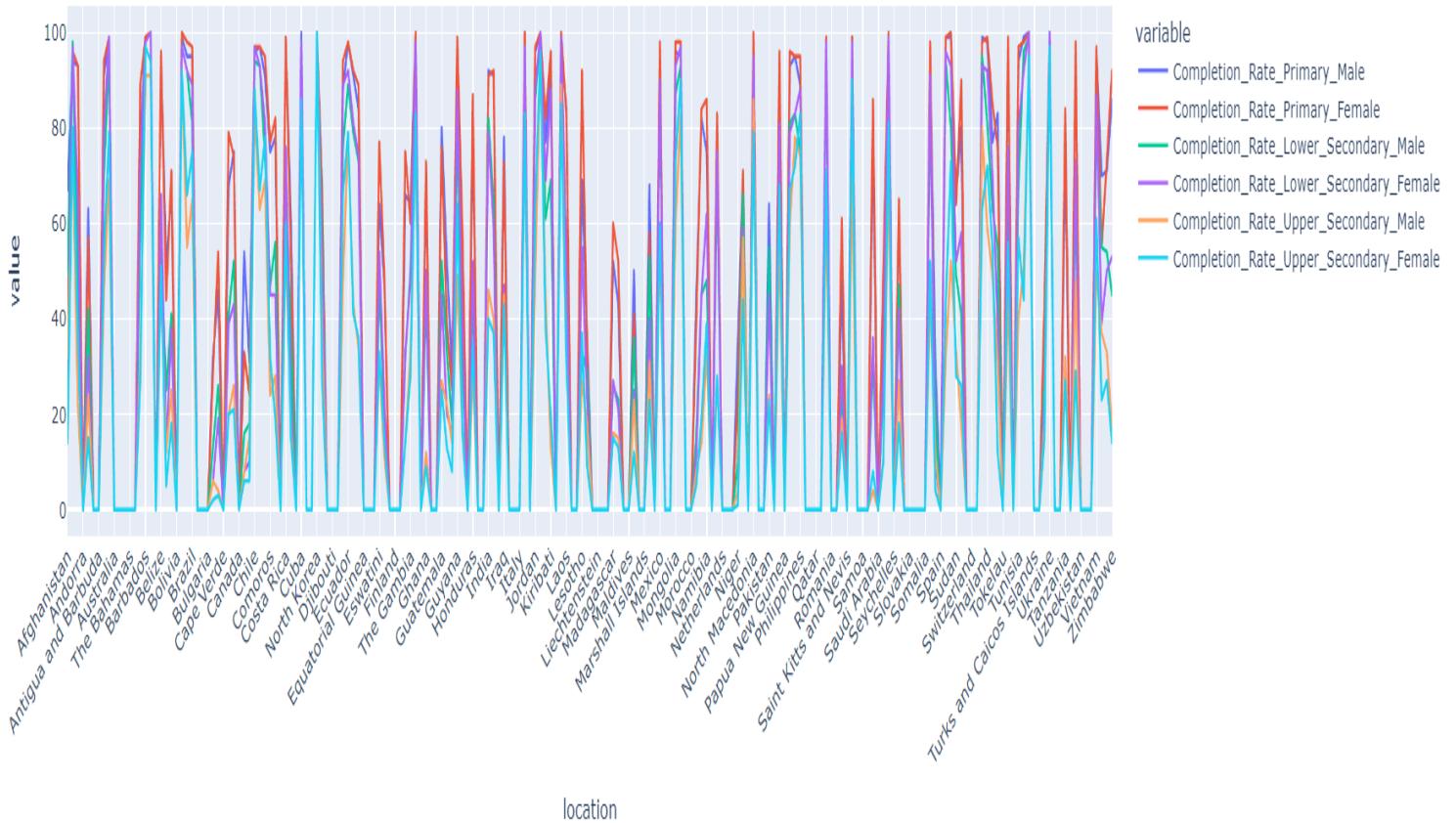
3.9.4. Tỷ lệ hoàn thành trung bình ở cấp độ giáo dục



Sự khác biệt giữa nam và nữ hầu như không lớn (nữ nhìn chung cao hơn 1 tí, nhưng ko nhiều)

3.9.5. Tỷ lệ hoàn thành các cấp bậc giáo dục tổng quát

Completion Rates Over Different Education Levels

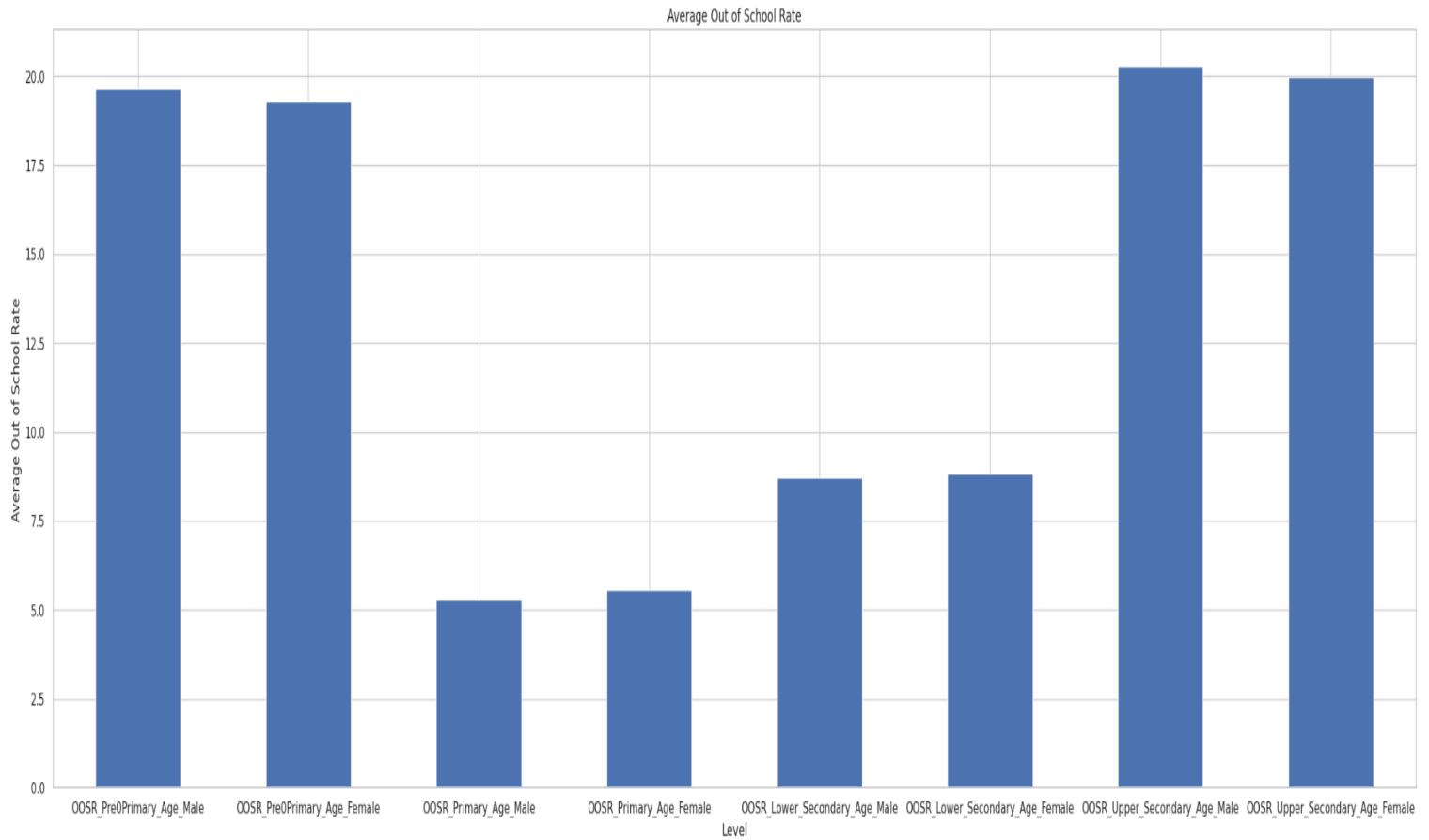


3.10. Tỷ lệ ngoài nhà trường

Tỷ lệ bỏ học ở bậc mầm non, tiểu học, THCS và THPT được biểu thị bằng tỷ lệ trung bình của mỗi quốc gia và giới tính. Một số quốc gia không bắt buộc giáo dục mầm non, đây có thể là nguyên nhân khiến tỷ lệ bỏ học ở bậc mầm non cao. Tuy nhiên, không rõ tại sao tỷ lệ này lại cao hơn ở cấp trung học phổ thông. Ngoài ra, giáo dục tiểu học dành cho nam và nữ có tỷ lệ bỏ học thấp nhất, tiếp theo là giáo dục trung học cơ sở.

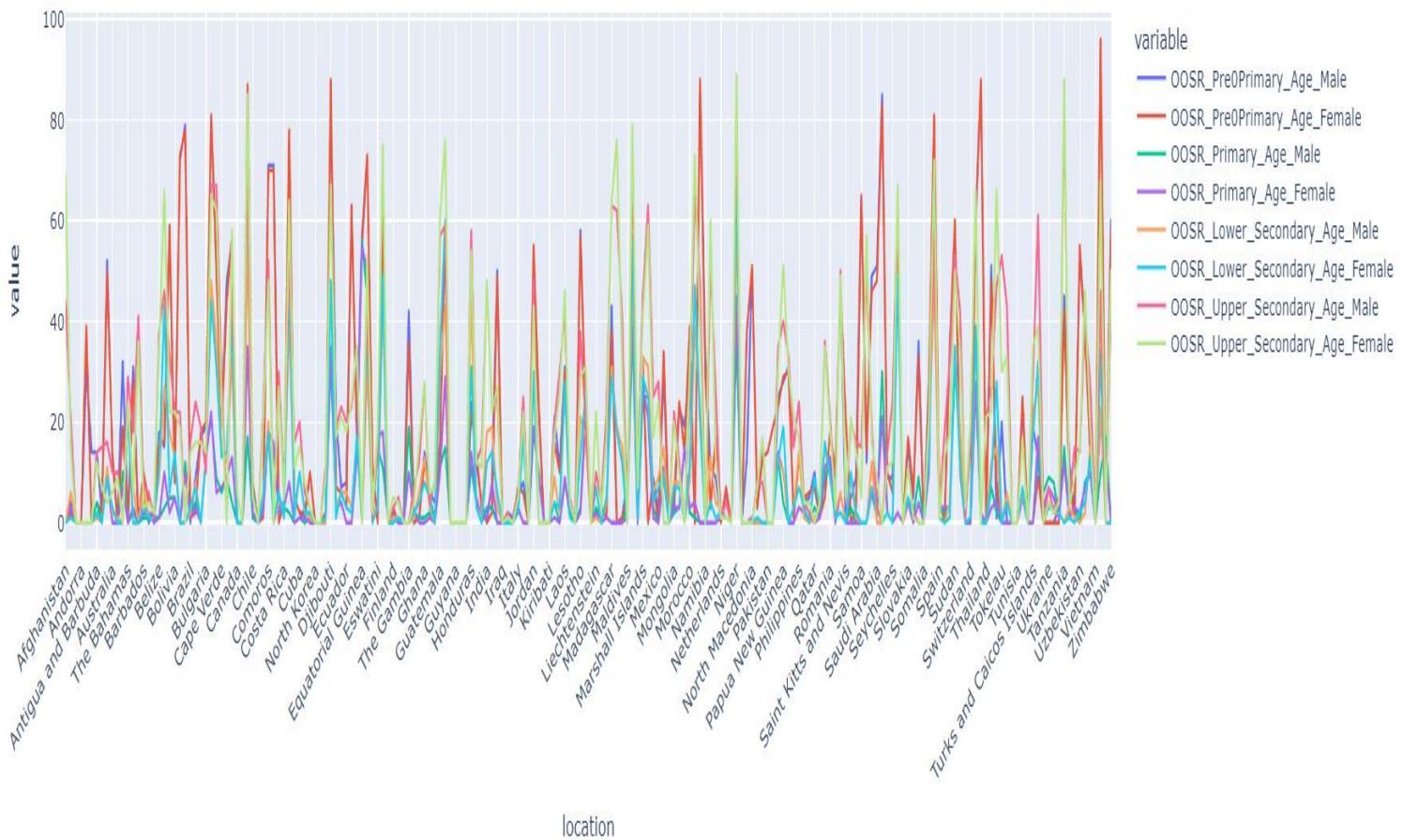
3.10.1.

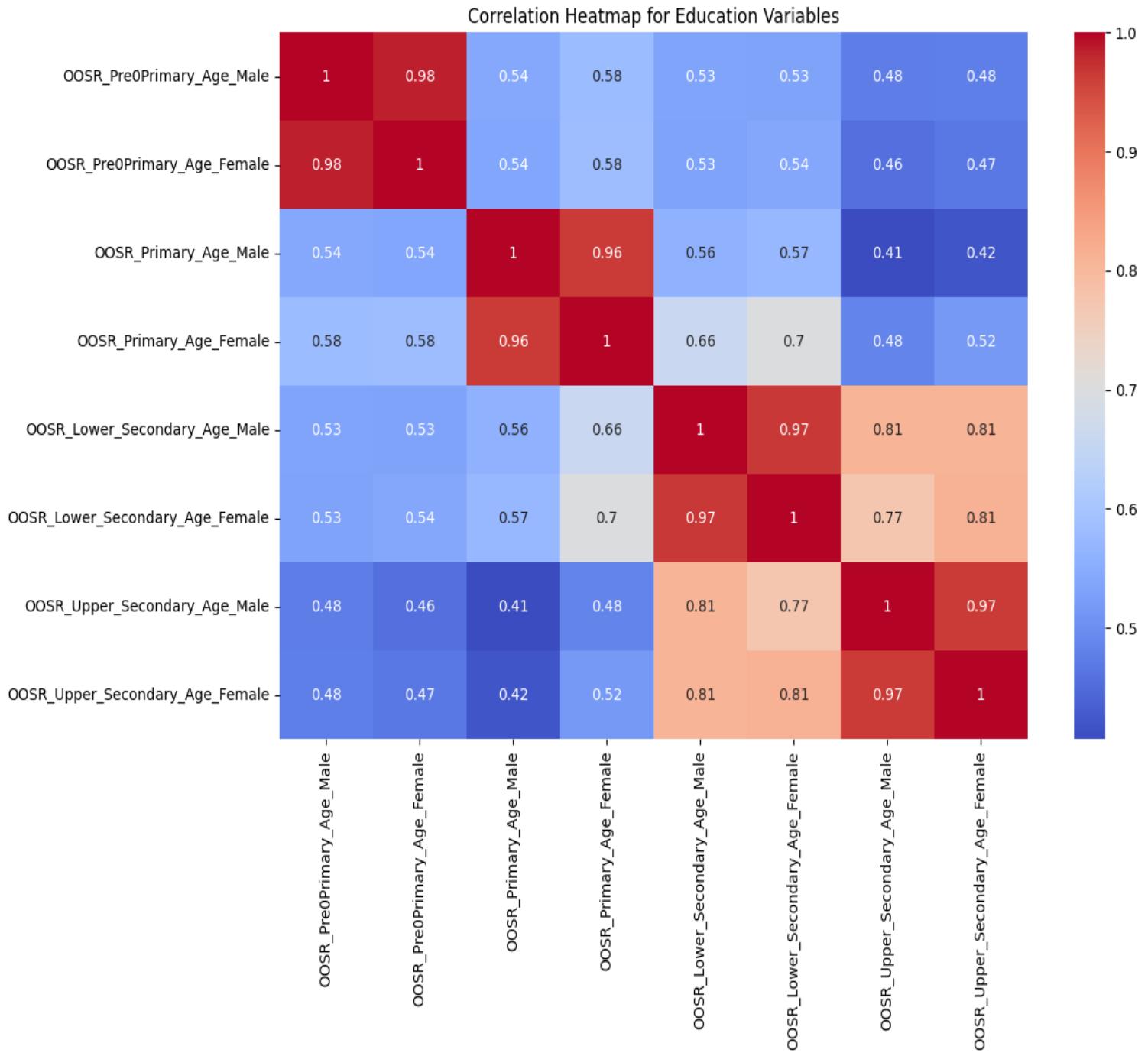
Tỷ lệ ngoài nhà trường trung bình



3.10.2. Tỷ lệ ngoài nhà trường ở các cấp độ giáo dục

Out of School Rates Over Different Education Levels





Sự tương quan

- OOSR_PrePrimary_Age_Female vs OOSR_PrePrimary_Age_Male = 0.98
- OOSR_PrePrimary_Age_Female vs OOSR_Primary_Age_Male = 0.96
- OOSR_Lower_Secondary_Age_Female vs
OOSR_Lower_Secondary_Age_Male = 0.97

→ OOSR_Upper_Secondary_Age_Female vs
OOSR_Upper_Secondary_Age_Male = 0.97

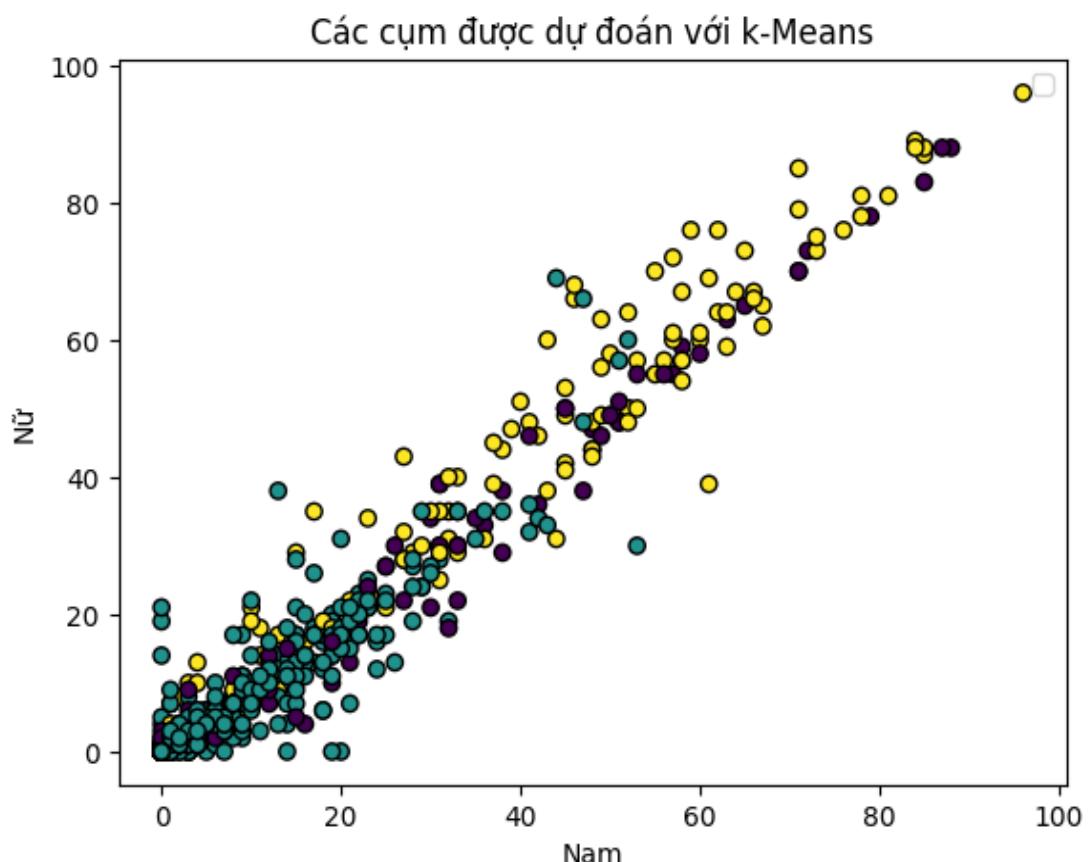
=> Không có nhiều sự khác biệt về giới tính

Chương 4. PHÂN CỤM

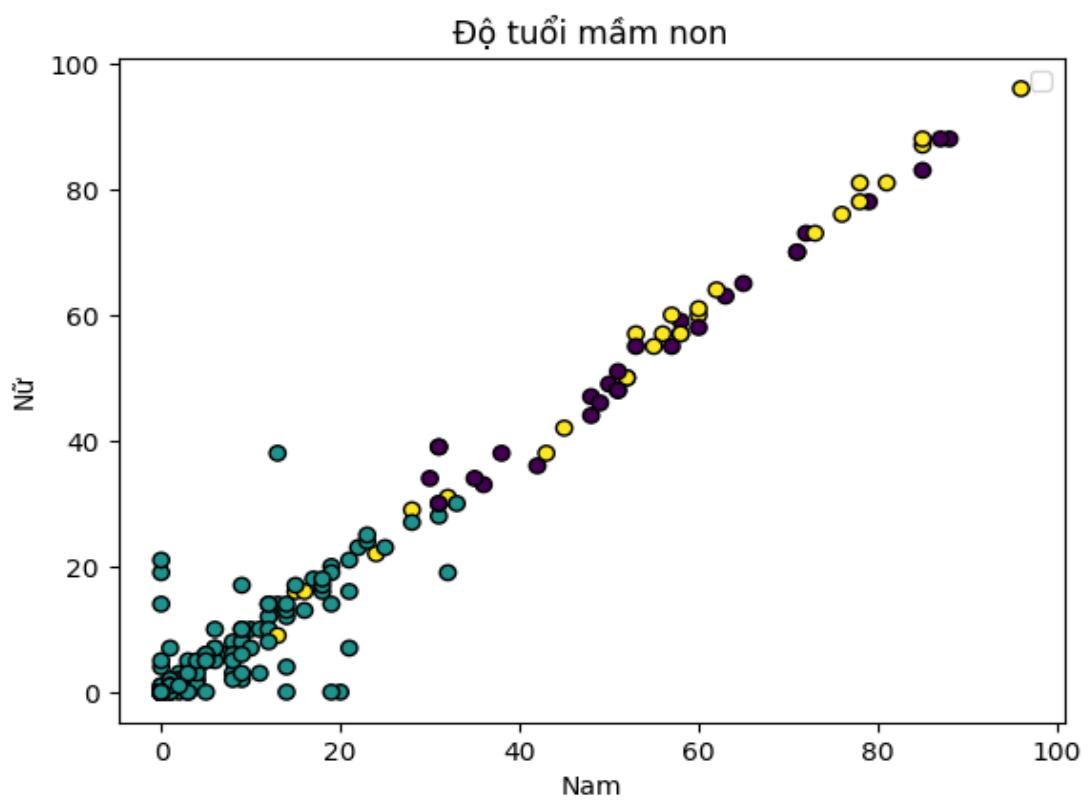
4.1. Phân cụm theo tỷ lệ ngoài nhà trường

Sử dụng K-Mean để vẽ biểu đồ (phân thành 3 cụm)

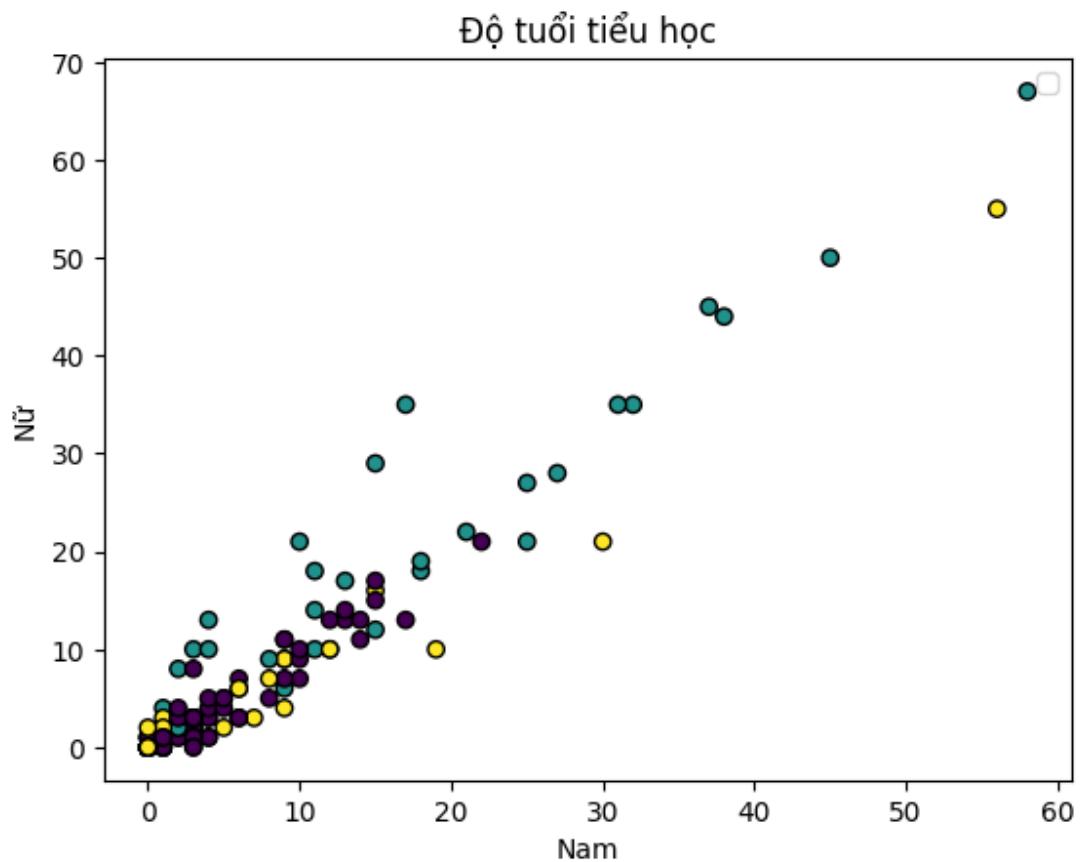
- Tổng quát tỷ lệ ngoài nhà trường



- Tỷ lệ ngoài nhà trường độ tuổi mầm non

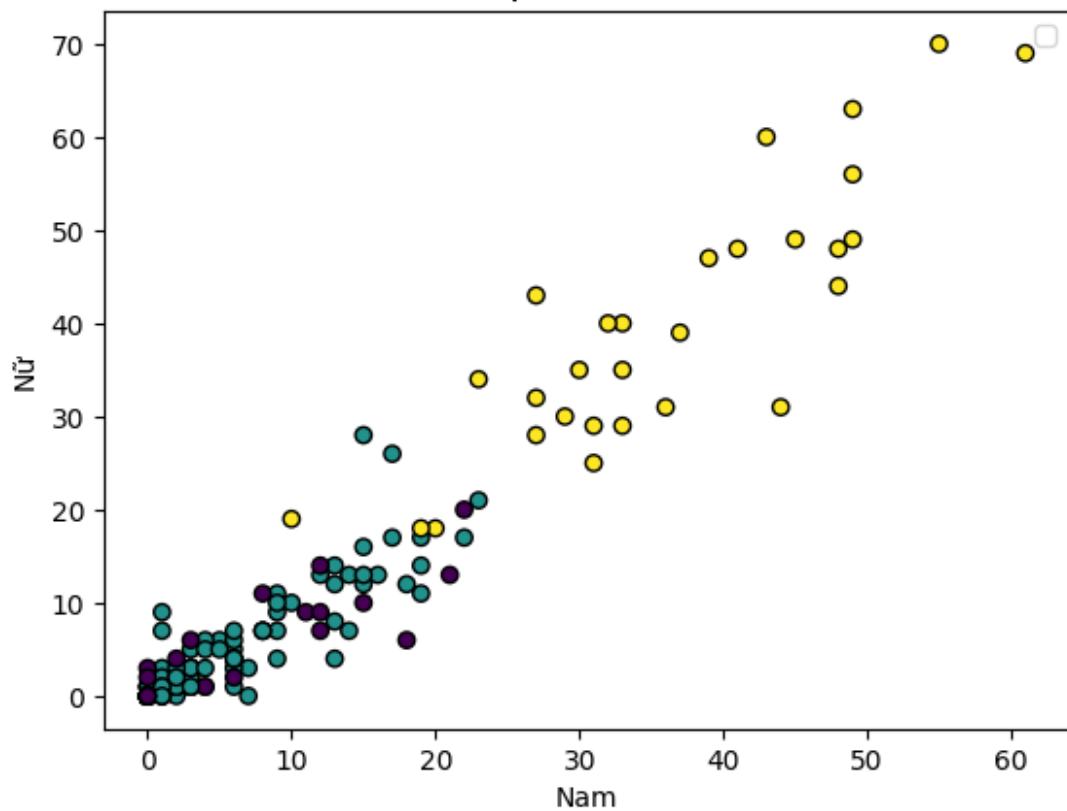


- Tỷ lệ ngoài nhà trường độ tuổi tiểu học



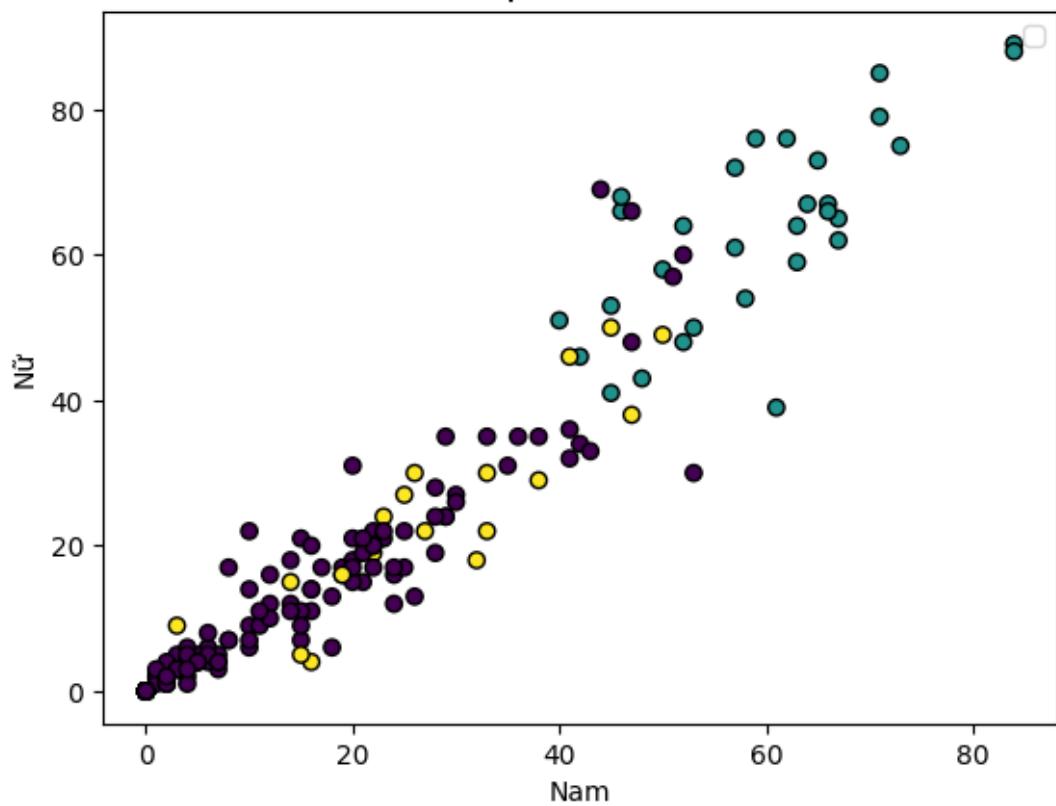
- Tỷ lệ ngoài nhà trường độ tuổi THCS

Độ tuổi THCS



- Tỷ lệ ngoài nhà trường độ tuổi THPT

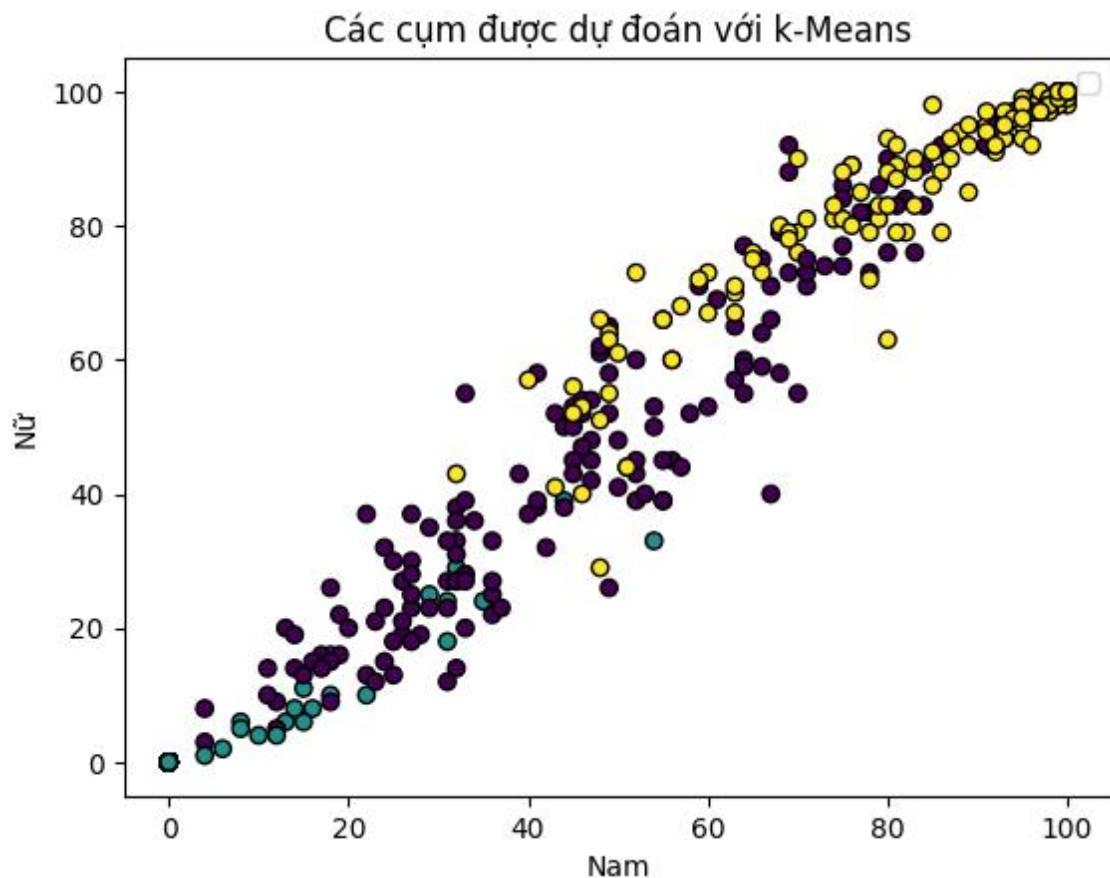
Độ tuổi THPT



4.2. Phân cụm theo tỷ lệ hoàn thành các bậc học

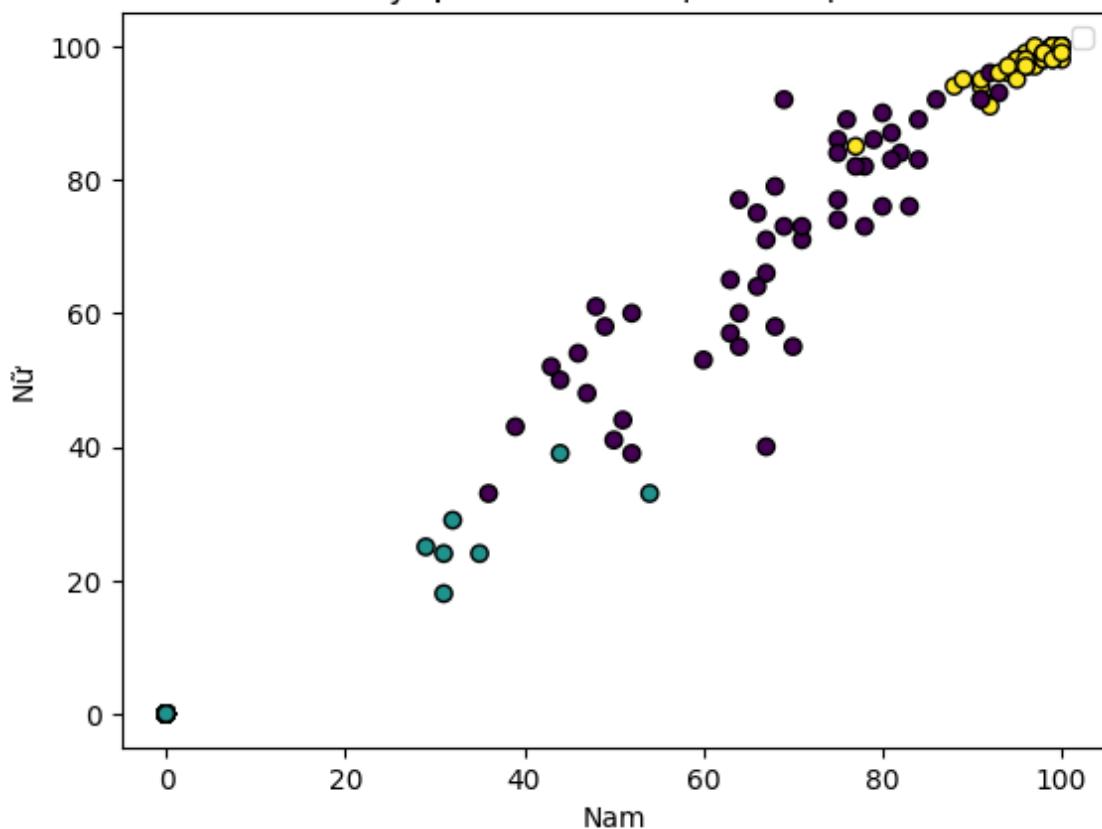
Sử dụng K-Mean để vẽ biểu đồ (phân thành 3 cụm)

- Tổng quát tỷ lệ hoàn thành các bậc học

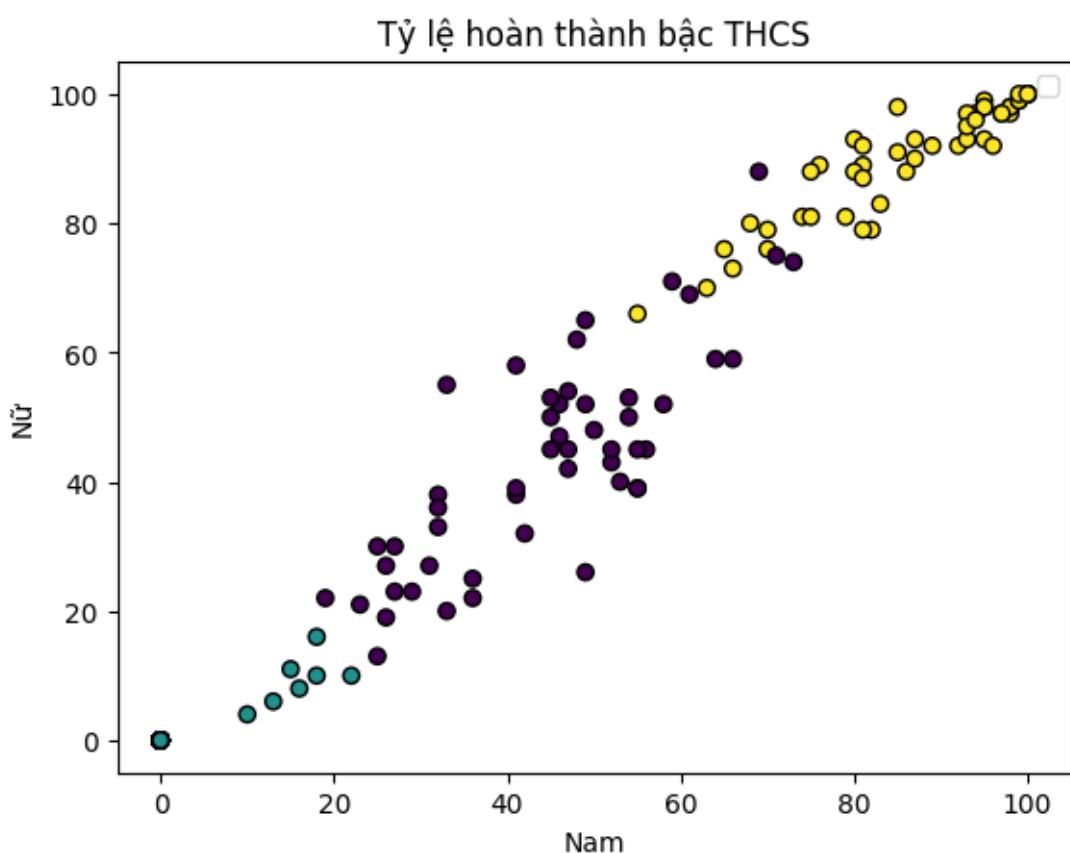


- Tỷ lệ hoàn thành bậc tiểu học

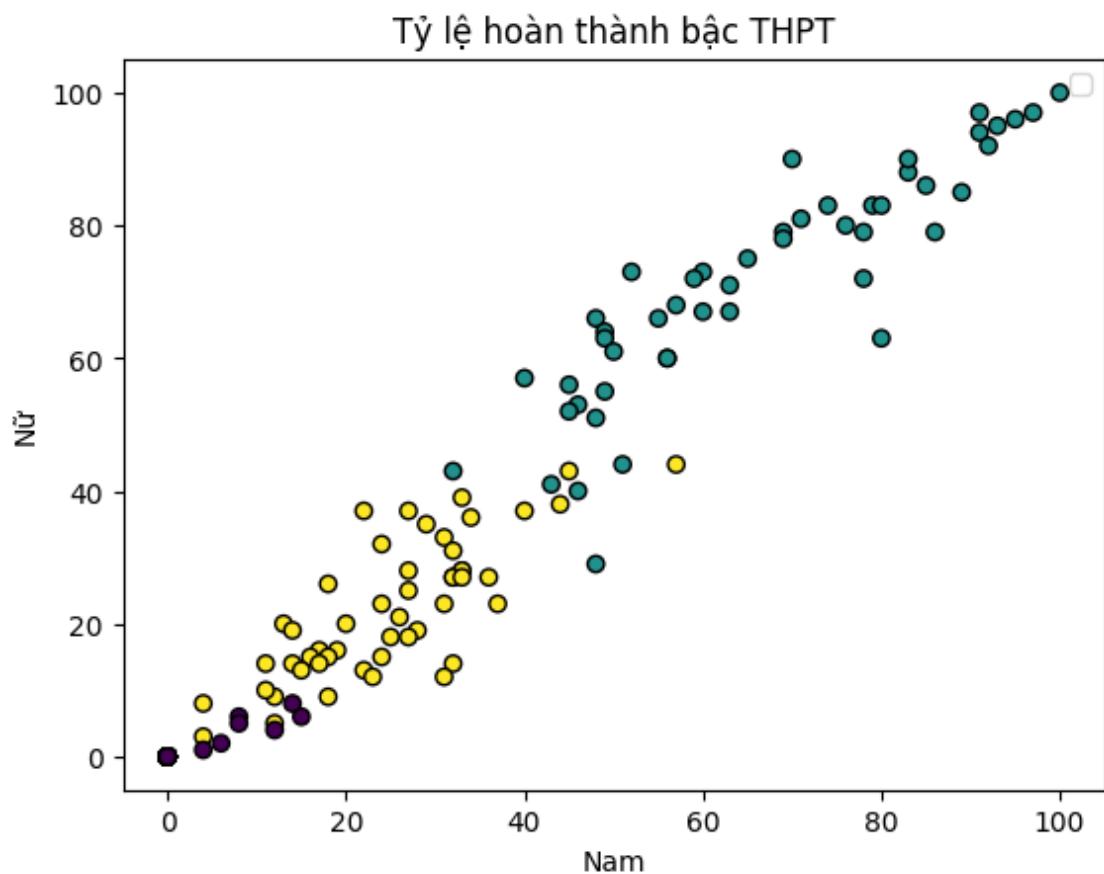
Tỷ lệ hoàn thành bậc tiểu học



- Tỷ lệ hoàn thành bậc THCS

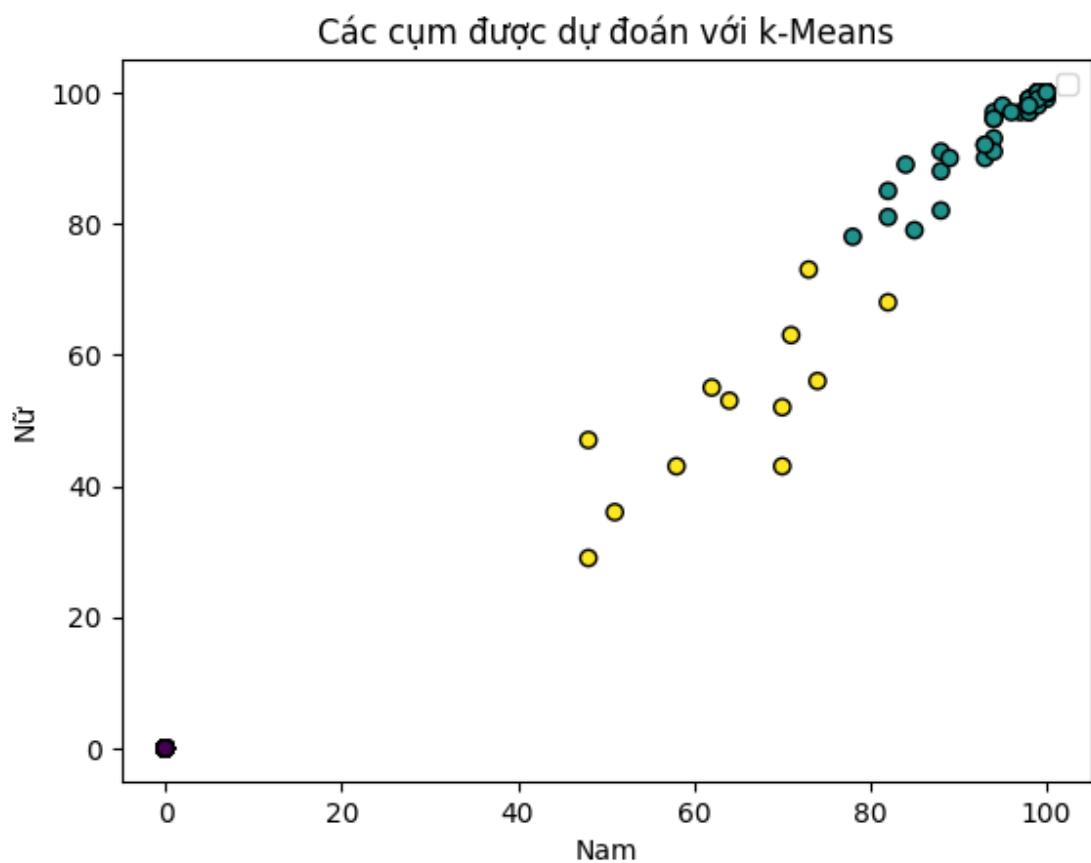


- Tỷ lệ hoàn thành bậc THPT



4.3. Phân cụm theo tỷ lệ biết chữ của thanh thiếu niên độ tuổi 15-24

Sử dụng K-Mean để vẽ biểu đồ (phân thành 3 cụm)



Chương 5. CÂY QUYẾT ĐỊNH

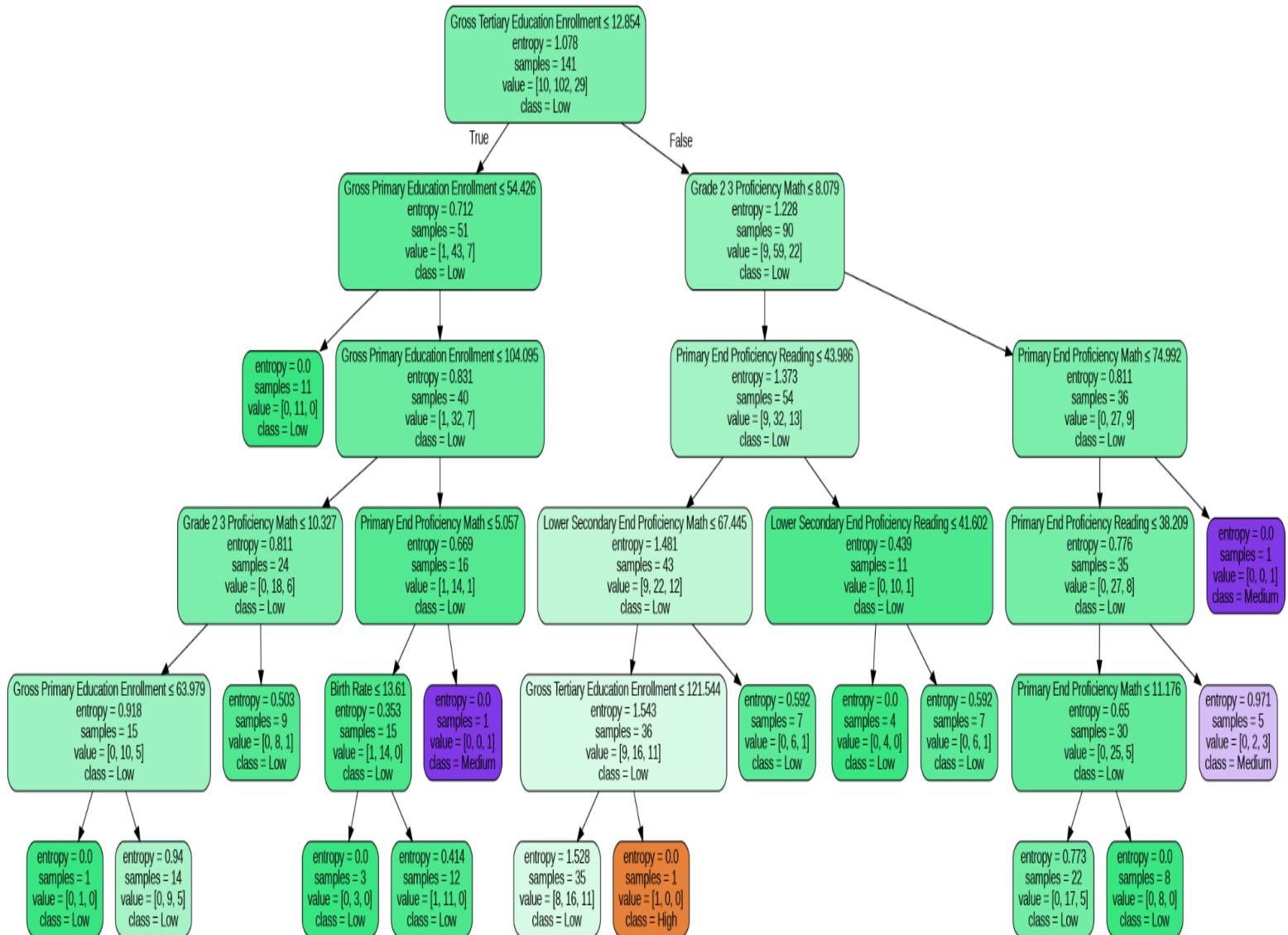
Ở đây ta sử dụng thư viện graphviz để in ra đồ thị cây quyết định với độ sâu là 5 cho dễ nhìn và tránh những trường hợp cây có quá nhiều nhánh con gây khó hiểu cho mô hình cây quyết định.

Kết quả độ chính xác sau khi chạy Python:

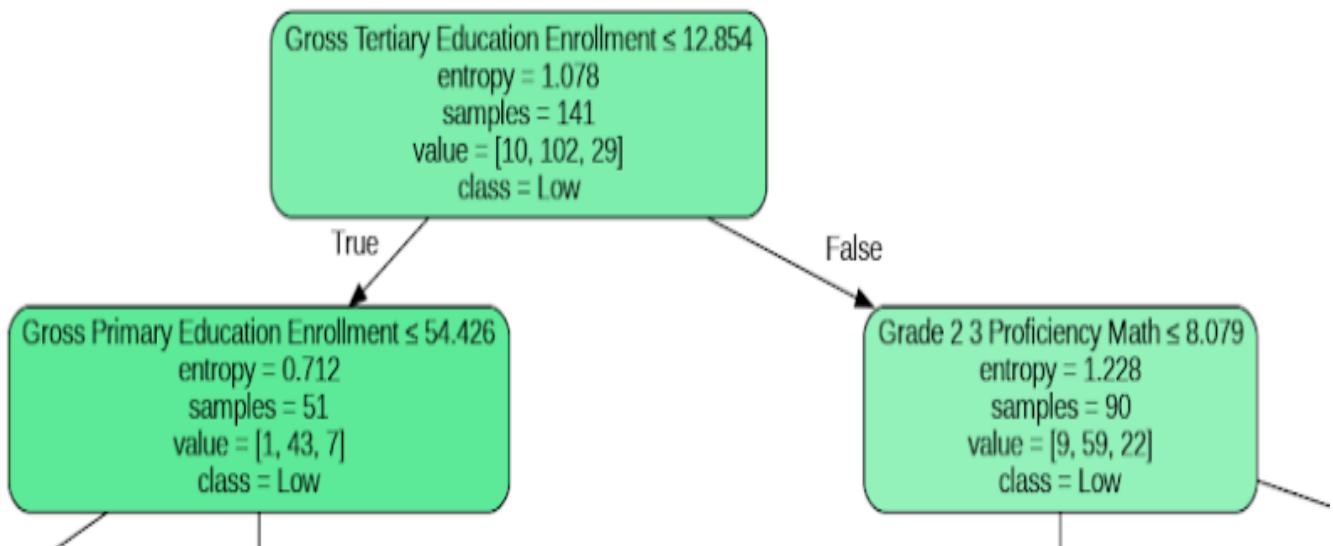
Best Accuracy: 0.7376847290640395

Accuracy on Test Set: 0.6229508196721312

CÂY QUYẾT ĐỊNH :

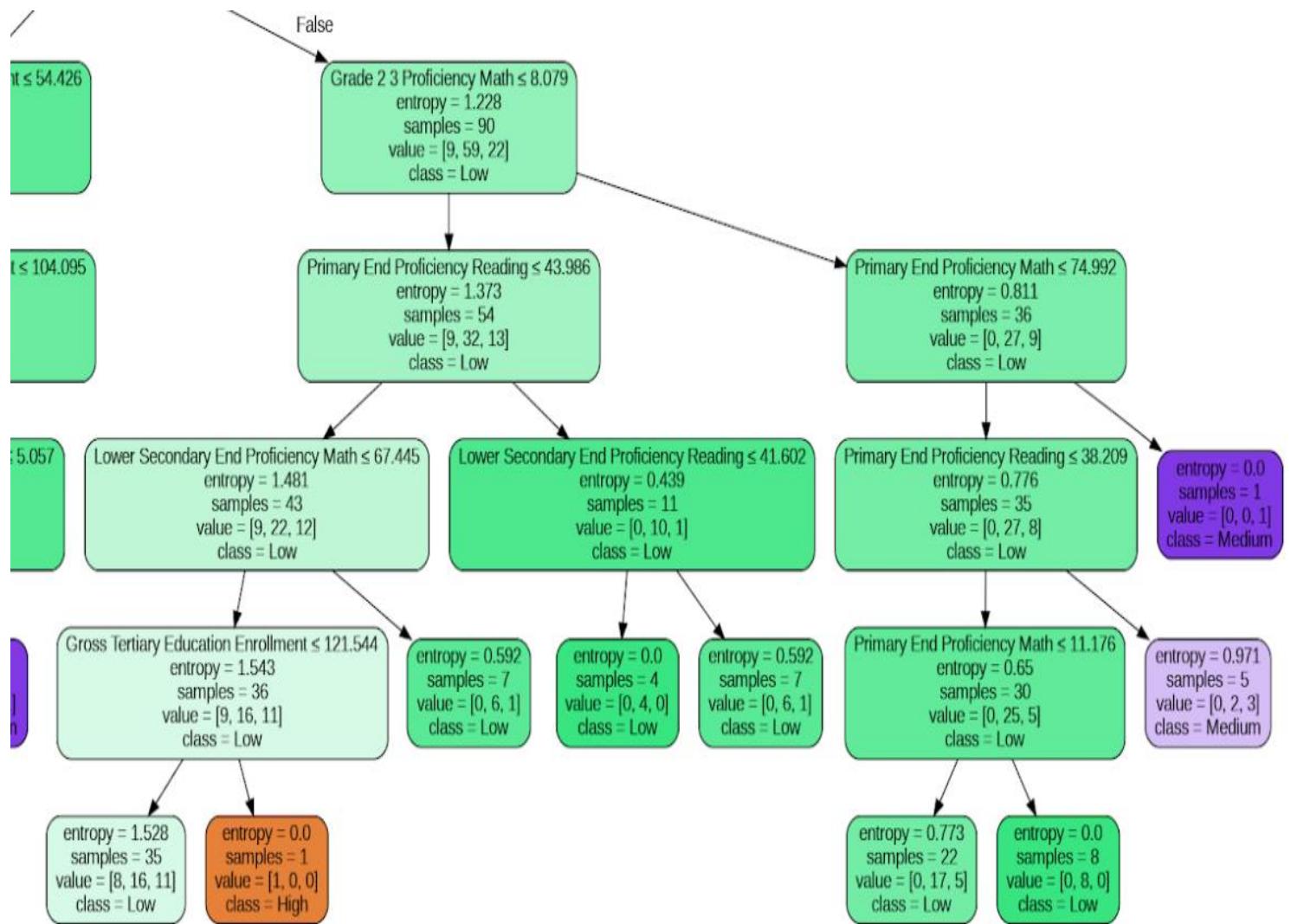


Model cây quyết định:



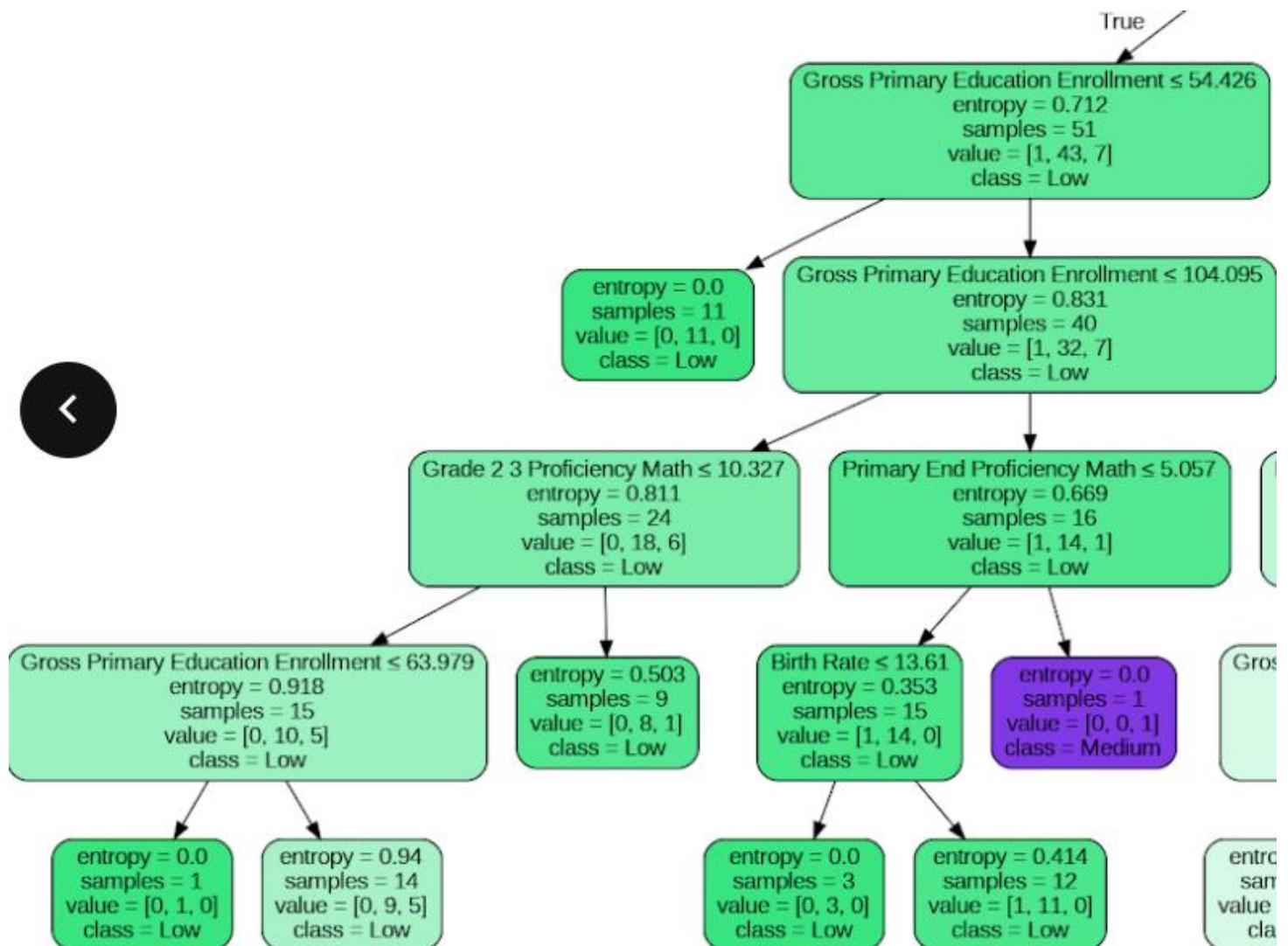
Model cây quyết định 2 nhánh đầu

Nhánh bên phải



Model cây quyết định nhánh bên phải

Nhánh bên trái



Model cây quyết định nhánh bên trái

Dựa theo cây quyết định này, chúng ta có thể khẳng định Gross_Tertiary_Education_Enrollment (Tổng số tuyển sinh vào giáo dục đại học) là nhân tố quyết định đánh giá tỷ lệ thất nghiệp (High, Medium, Low)

Kiểm tra kiểm thử để nhìn rõ bản chất đúng sai của mô hình khi cho vào tập dữ liệu test. Và đây là kết quả phân tích độ chính xác của cây:

[[1 1 1]				
[7 26 7]				
[2 13 3]]				
	precision	recall	f1-score	support
High	0.10	0.33	0.15	3
Low	0.65	0.65	0.65	40
Medium	0.27	0.17	0.21	18
accuracy			0.49	61
macro avg	0.34	0.38	0.34	61
weighted avg	0.51	0.49	0.49	61

Nhìn thấy, độ chính xác khi quyết định tỷ lệ thất nghiệp High là 10%, Low là 65%, Medium là 27%

Ta có thể kết luận rằng mô hình cây quyết định này có độ chính xác khá tốt khi ta huấn luyện 70% tập dữ liệu và dùng 30% tập dữ liệu ban đầu để kiểm thử tỷ lệ thất nghiệp ra 3 loại High, Medium, Low.

Chương 6. NHẬT XÉT TỔNG QUAN

Việc dựa vào phân cụm có thể giúp chúng ta nhìn thấy các nhóm tương đồng về mức độ phát triển giáo dục của các quốc gia và khu vực các nhóm quốc gia có mức độ phát triển giáo dục tương tự. Điều này có thể giúp chúng ta hiểu rõ hơn về những đặc điểm chung và khác biệt giữa các quốc gia trong cùng một nhóm phân loại. Ngoài ra, phân cụm cũng có thể là cơ sở để thực hiện các phân tích so sánh và đánh giá mức độ phát triển giáo dục giữa các nhóm khác nhau.

Về việc chất lượng giáo dục dữ liệu cung cấp thông tin về trình độ đọc và tính toán của học sinh ở các cấp học khác nhau. Tỷ lệ trình độ đọc và tính toán có thể biểu thị mức độ chất lượng giáo dục trong mỗi quốc gia. Các quốc gia có tỷ lệ cao hơn có thể cho thấy học sinh có trình độ giáo dục tốt hơn.

Về tỷ lệ hoàn thành giáo dục ở các cấp học khác nhau. Tỷ lệ này cho biết tỷ lệ học sinh hoàn thành các cấp học cụ thể. Các quốc gia với tỷ lệ hoàn thành cao có thể có hệ thống giáo dục hiệu quả và đảm bảo cơ hội giáo dục cho tất cả học sinh.

Tỷ lệ biết chữ ở nhóm tuổi 15-24. Tỷ lệ biết chữ cao cho thấy mức độ biết chữ và giáo dục của người trẻ trong mỗi quốc gia. Điều này có thể tương quan với sự phát triển kinh tế và xã hội của quốc gia.

Tỷ lệ nhập học cao cho thấy cơ hội giáo dục đang mở rộng và đảm bảo sự tiếp cận giáo dục cho tất cả học sinh.

Tỷ lệ thất nghiệp thấp có thể chỉ ra rằng hệ thống giáo dục đang tạo ra nguồn nhân lực có khả năng làm việc và đáp ứng nhu cầu của thị trường lao động.

Tính ứng dụng của chủ đề:

1. Giúp đánh giá chất lượng giáo dục: So sánh các chỉ số và tiêu chí giáo dục giữa các quốc gia giúp xác định chất lượng giáo dục ở mỗi quốc gia. Bằng cách so sánh các chỉ số như tỷ lệ học sinh hoàn thành, chỉ số biết chữ, kỹ năng toán học và các chỉ số khác, ta có thể đánh giá mức độ thành công và hiệu quả của hệ thống giáo dục trong từng quốc gia.
2. Định hướng chính sách giáo dục: So sánh giữa các quốc gia có thể cung cấp thông tin quan trọng để hỗ trợ quyết định chính sách giáo dục. Bằng việc nghiên cứu các quốc gia có thành công trong một lĩnh vực cụ thể, các quốc gia khác có thể học hỏi và áp dụng các phương pháp, chương trình và chính sách tương tự để nâng cao chất lượng giáo dục của mình.
3. Phân tích sự khác biệt văn hóa và xã hội: So sánh giáo dục giữa các quốc gia cũng giúp hiểu rõ hơn về sự khác biệt văn hóa và xã hội. Các chỉ số giáo dục có thể phản ánh các yếu tố như độ tuổi học, tình trạng tài trợ, phân bổ giữa các cấp học, giới tính và tình trạng học vấn của thanh thiếu niên. Thông qua việc so sánh, ta có thể nhận thức được tác động của các yếu tố này đến chất lượng giáo dục và phát triển của quốc gia.
4. Nghiên cứu và phân tích: Dữ liệu so sánh giáo dục cũng cung cấp cơ sở cho các nghiên cứu và phân tích trong lĩnh vực giáo dục. Các nhà nghiên cứu và chuyên gia có thể sử dụng dữ liệu này để tìm hiểu các mô hình, xu hướng và tương quan giữa các yếu tố giáo dục và các chỉ số xã hội khác. Điều này có thể đóng góp vào việc phát triển các chiến lược và giải pháp cải thiện giáo dục trên toàn cầu.

5. Trao đổi kinh nghiệm và học hỏi: So sánh các vấn đề giáo dục giữa các quốc gia khác nhau cho phép trao đổi kinh nghiệm và học hỏi nhau. Các quốc gia có thể chia sẻ những phương pháp giảng dạy hiệu quả, các chương trình đào tạo giáo viên thành công, và các biện pháp hỗ trợ học sinh trong quá trình học tập.
6. Xác định ưu tiên và cơ hội phát triển: So sánh giữa các quốc gia giúp xác định các vấn đề ưu tiên và cơ hội phát triển trong lĩnh vực giáo dục. Các quốc gia có thể tìm ra những lĩnh vực cần đầu tư và cải thiện để nâng cao chất lượng giáo dục và tạo ra cơ hội tốt hơn cho học sinh.
7. Theo dõi tiến bộ và đánh giá hiệu quả: So sánh các chỉ số giáo dục theo thời gian giúp theo dõi tiến bộ và đánh giá hiệu quả của các biện pháp cải thiện giáo dục. Bằng cách so sánh dữ liệu theo năm, ta có thể xác định xem các biện pháp đã đạt được kết quả như mong đợi hay chưa và điều chỉnh chiến lược nếu cần thiết.

Tóm lại, việc so sánh các vấn đề giáo dục của các quốc gia trên toàn cầu giúp đánh giá chất lượng giáo dục, mang lại lợi ích lớn trong việc đánh giá, trao đổi kinh nghiệm, xác định ưu tiên và theo dõi tiến bộ, định hướng chính sách giáo dục, hiểu sự khác biệt về văn hóa và xã hội, cũng như nghiên cứu và phân tích trong lĩnh vực giáo dục. Qua đó, các quốc gia có thể nâng cao chất lượng giáo dục và đảm bảo cơ hội học tập tốt cho học sinh.