# Education Data and Trends

Ironhack Data Analytics Bootcamp - Ceci

# The Objective

Present a statistical analysis overview of Education and factors that influence it in diverse areas of the world is one part of this project.

Present a small machine learning model study using models learnt in class.

# The Questions

- What are the observable enrollment trends?
- How does the government expenditure in education look across the world?
- Are there enough teachers? And do all of them have the minimum necessary qualifications?
- Does the Capital Expenditure in Education influence the Overall enrollment ratios?
- What policy recommendations can be made?

01

# The Data

Where it came from and the processes it went through

# Data Handling Process

## Gathering and Cleaning

Data gathered from the public UNData site, the main 4 datasets did not need much cleaning.

## Handling

Data handled through normalizing columns, filling missing values and dropping non-needed columns

## Filtering

Filtered data into respective needed countries to proceed with analysis, EDA and ML testing and implementation

# Caveats

As not all the years in the Datasets were reflected for the whole list of countries being used. They were filtered in a way that kept the years that had the most amount of data for the largest number of countries.

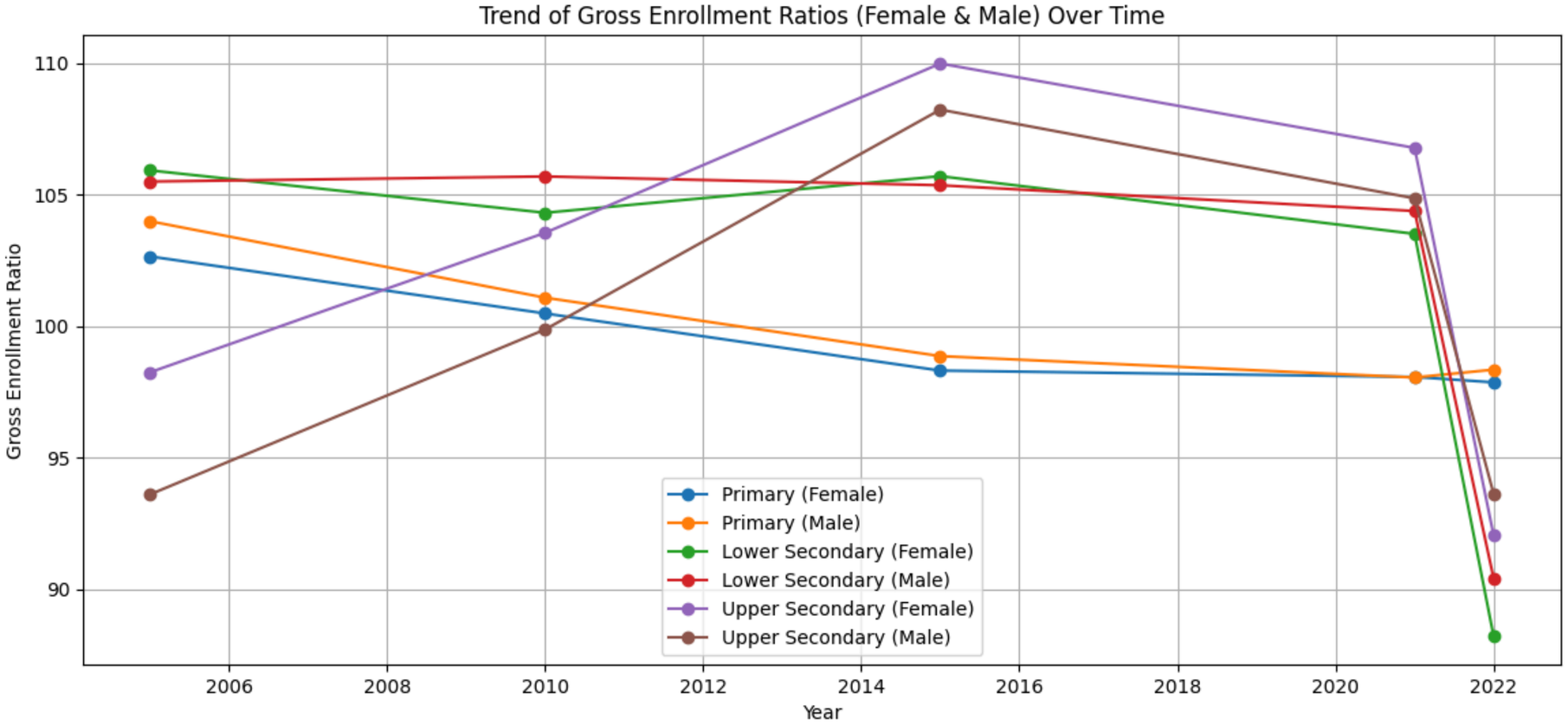
# Final Country Based Dataset

- Spans from 2005 to 2022, with most data points concentrated around the early 2010s.
- Primarily reflects stable educational systems in high-income countries, capturing trends during a dynamic period, including the effects of events like the global financial crisis.
- Mainly includes high-income countries from Europe, advanced Asian economies and the US.
- Countries like Albania, Gibraltar, Montenegro, and Moldova have fewer data points.
- Countries such as China, Russia, India, and the US appear more often.

# 02

# EDA

# Comparing Enrollment Ratios



# Enrollment Trend Discussion

## Primary

Stable access to primary education with minor declines likely due to demographic or economic shifts.

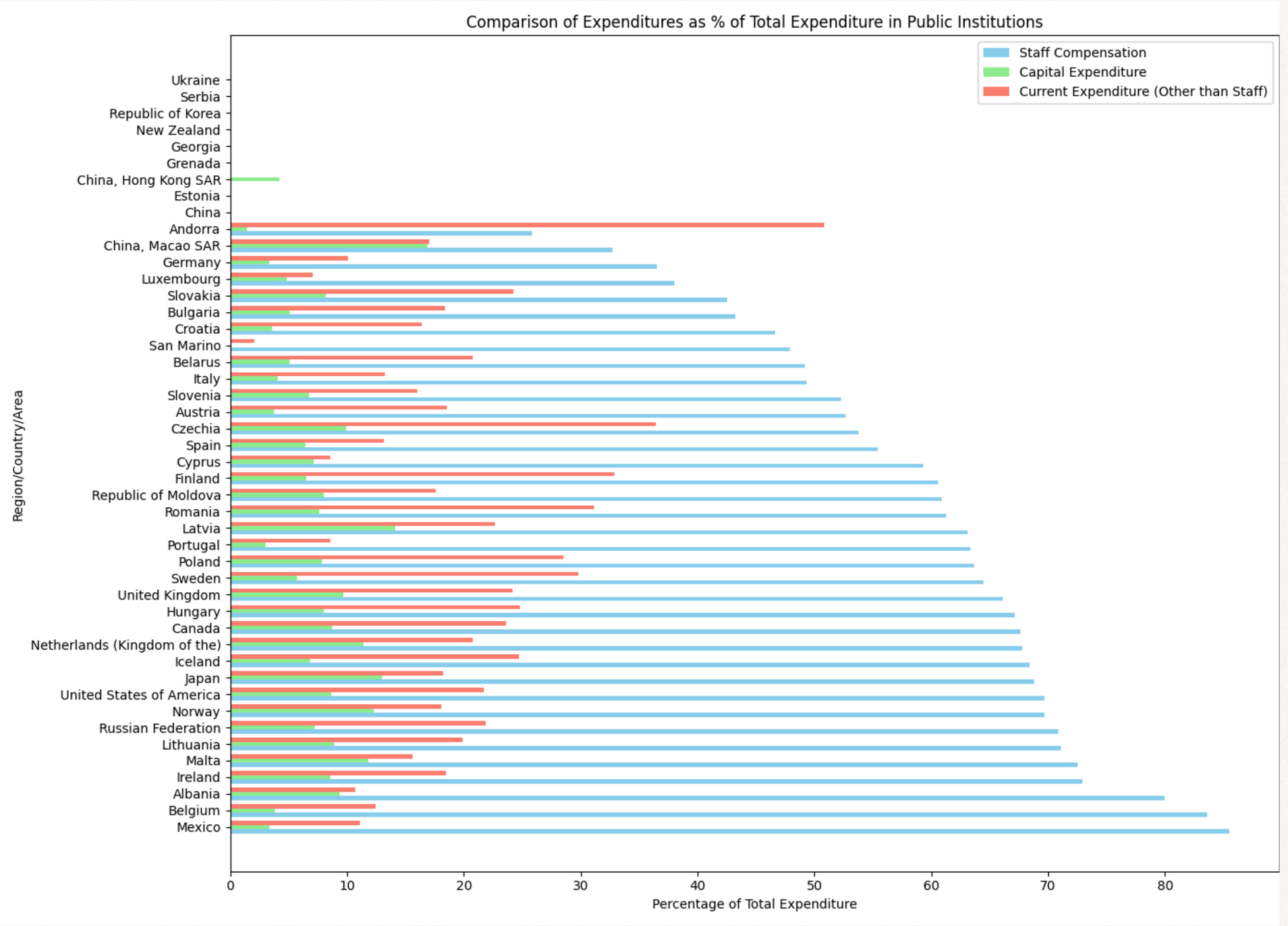
## Lower Secondary

Declining enrollment suggests challenges like economic instability, policy changes, or infrastructure issues

## Upper Secondary

Variability suggests changing access or economic factors influencing enrollment

Possible Hypothesis: Factors like the Unschooling and Homeschooling movements are affecting the enrollment in secondary education


Comparison of Expenditures as % of Total Expenditure in Public Institutions
Staff Compensation
Capital Expenditure
Current Expenditure (Other than Staff)
Region/Country/Area
Ukraine
Serbia
Republic of Korea
New Zealand
Georgia
Grenada
China, Hong Kong SAR
Estonia
China
Andorra
China, Macao SAR
Germany
Luxembourg
Slovakia
Bulgaria
Croatia
San Marino
Belarus
Italy
Slovenia
Austria
Czechia
Spain
Cyprus
Finland
Republic of Moldova
Romania
Latvia
Portugal
Poland
Sweden
United Kingdom
Hungary
Canada
Netherlands (Kingdom of the)
Iceland
Japan
United States of America
Norway
Russian Federation
Lithuania
Malta
Ireland
Albania
Belgium
Mexico
0
10
20
30
40
50
60
70
80
Percentage of Total Expenditure

# Expenditure Comparison Explanation

## Staff Compensation

Mexico (85.5%) and Belgium (83.6%) allocate the highest percentages to staff compensation, indicating that a significant portion of the public sector budget goes towards salaries and benefits

## Capital Expenditure as per GDP

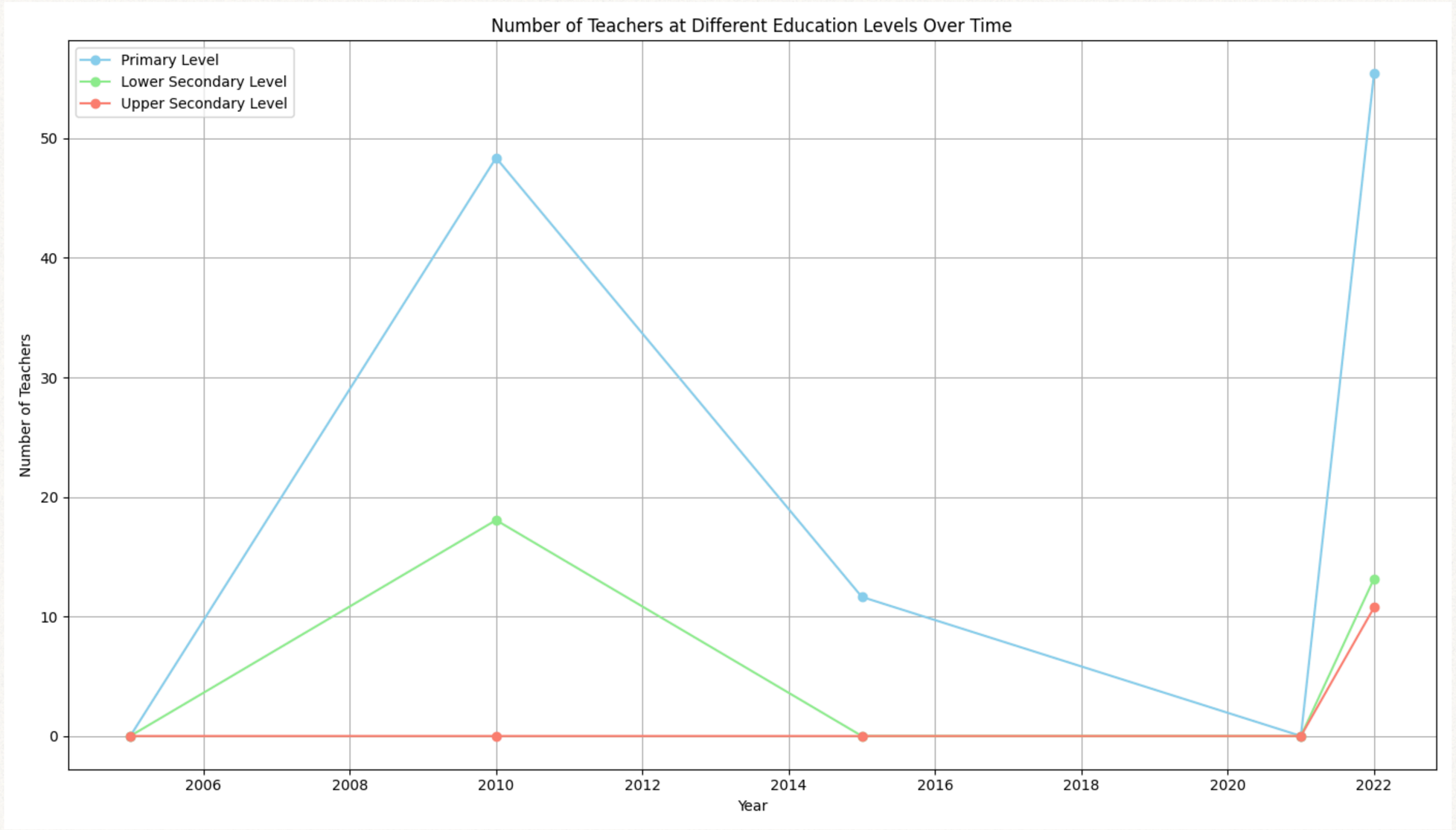
Japan (13%) and Norway (12.3%) have the highest capital expenditure percentages, suggesting substantial investments in infrastructure and long-term projects

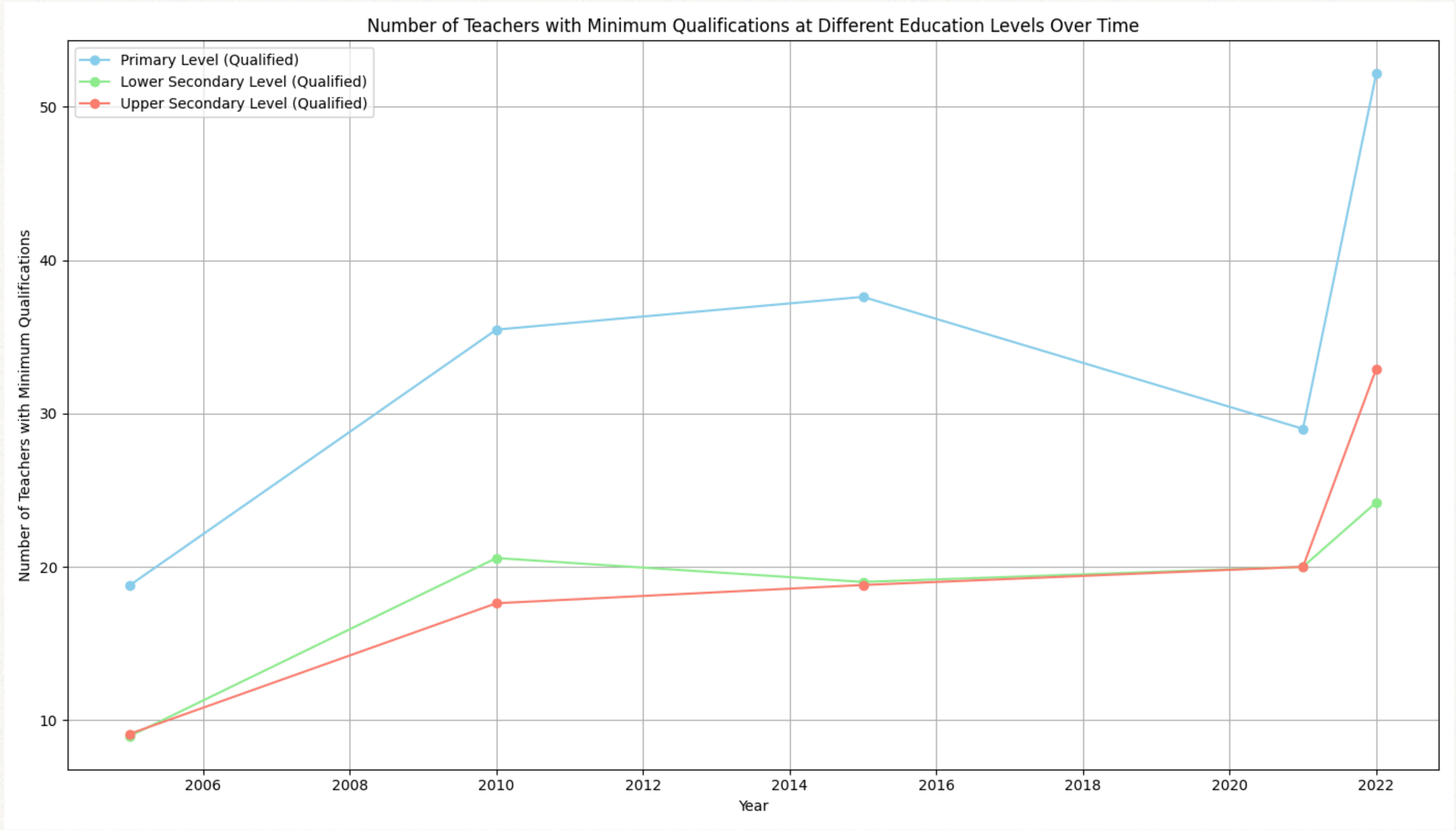
## Expenditure Other than Staff

Finland (32.85%) and Sweden (29.77%) dedicate the largest shares of their budgets to current expenditure other than staff compensation, highlighting robust spending on public services and operational costs.

## Conclusions

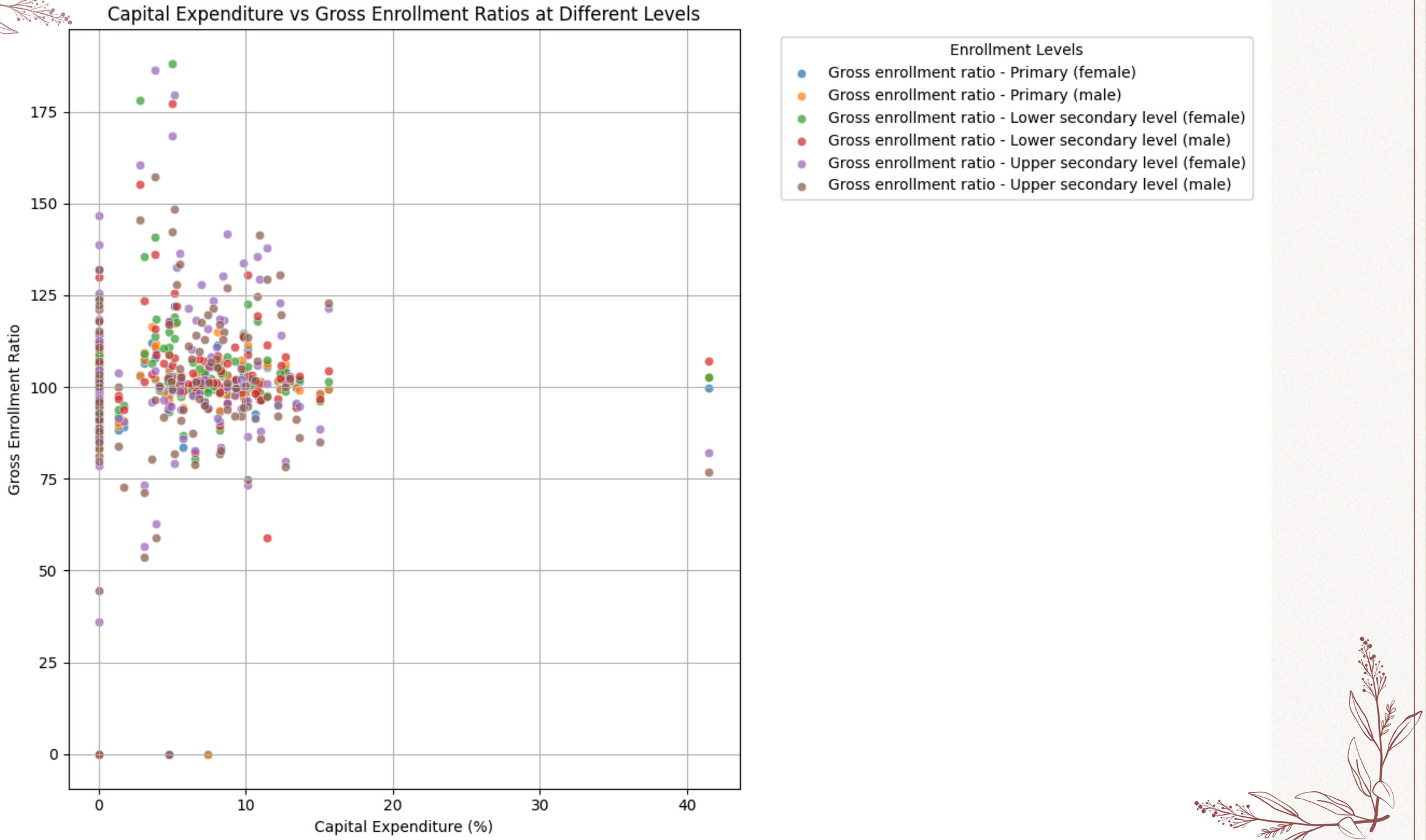
- Mexico, Belgium and Norway exhibit high allocations for staff compensation, suggesting a greater focus on personnel costs in public institutions.
- Japan, Norway and Latvia emphasize capital expenditure in education, indicating a long-term investment in infrastructure and public assets.
- Finland and Romania allocate substantial amounts to other expenditure, signaling broad operational funding beyond staffing needs.


Number of Teachers at Different Education Levels Over Time
Primary Level
Lower Secondary Level
Upper Secondary Level
Number of Teachers
0
10
20
30
40
50
2006
2008
2010
2012
2014
2016
2018
2020
2022
Year


Number of Teachers with Minimum Qualifications at Different Education Levels Over Time
Primary Level (Qualified)
Lower Secondary Level (Qualified)
Upper Secondary Level (Qualified)
Number of Teachers with Minimum Qualifications
Year
10
20
30
40
50
2006
2008
2010
2012
2014
2016
2018
2020
2022

# Teacher Evolution over time

- Teacher numbers see a sharp rise in 2010, especially at the primary and lower secondary levels, followed by a drop in 2015 and a recovery by 2022.
- The percentage of teachers meeting minimum qualification standards improved over time, particularly at the primary and upper secondary levels.
- The data reveals fluctuating trends, likely influenced by changes in policies, economic conditions, or disruptions in education between 2010 and 2021.
- The absence of teacher data for 2005 and 2010 suggests potential reporting issues or temporary setbacks in the education system during those years.


Capital Expenditure vs Gross Enrollment Ratios at Different Levels
Gross Enrollment Ratio
Capital Expenditure (%)
0
25
50
75
100
125
150
175
0
10
20
30
40
Enrollment Levels
Gross enrollment ratio - Primary (female)
Gross enrollment ratio - Primary (male)
Gross enrollment ratio - Lower secondary level (female)
Gross enrollment ratio - Lower secondary level (male)
Gross enrollment ratio - Upper secondary level (female)
Gross enrollment ratio - Upper secondary level (male)

# 03

# Machine Learning

# Conclusions

1. Underfit models
2. Will need to attempt different models
3. There is a very long list of models but not enough of them were attempted
4. Attempt further with different dataset filters as a lot of filters were used and using a less filtered dataset would absolutely be benefitial for the overall performance for the model usage

## Result analysis conclusions:

- For Classification: The KNN model struggles with both StandardScaler and MinMaxScaler. Data preprocessing, such as handling class imbalance, and hyperparameter tuning are recommended for improving performance.
- For Regression: The KNN Regressor with StandardScaler performs slightly better than with MinMaxScaler. Further model tuning and possibly exploring alternative regression models may help achieve better results.

# 04

# Conclusions

# Comparative Conclusions

Model Performance and Practical Implications

- KNN Models (ML Results):
  - KNN classifiers and regressors exhibited limited predictive power, with the best accuracy being 52.94% for classification and an $R^2$ of 0.34 for regression. These modest results highlight limitations in the current dataset's predictive capacity or model appropriateness.
  - Hyperparameter tuning (e.g., with SMOTE or RandomOverSampler) marginally improved accuracy but did not resolve issues like class imbalance or poor performance on certain subsets.
- EDA Trends:
  - The dataset shows significant variation across countries, particularly in educational access and resources. Disparities in data quality, class sizes, and feature distributions could explain the machine learning models' limited efficacy. For instance, stark differences in digital access or teacher qualifications likely create noisy patterns that the models struggle to generalize.

# Educational Resource Allocation

- Developed countries allocate significant resources to staff compensation and operational costs (e.g., United States: 69.7%, Belgium: 83.6%). Developing countries like Albania or Moldova face tighter constraints, reflected in lower compensation or fewer qualified teachers.
- Access to digital resources is highly unequal. Countries like Germany or Sweden achieve near-universal digital access in schools, contrasting sharply with countries reporting 0% access.

# Enrollment and Gender Parity Trends

- Enrollment:
  - Primary education shows stable near-universal enrollment globally. Secondary education, however, has seen fluctuations, with significant declines after 2015 in both lower and upper secondary levels, particularly for females.
- Gender Parity:
  - While gender parity is high in primary education, discrepancies appear in lower secondary levels and widen in some regions at upper secondary levels. These imbalances likely reflect cultural or economic barriers in certain regions. This imbalance affects mostly female students.

# 05

# Policy Recomendations

# Recommendations

1. Improve Educational Data Reporting
    a. Standardize and improve data collection and reporting systems.
2. Invest in Teacher Training
    a. Increase investment in teacher training programs, especially in low-income regions.
3. Bridge the Digital Divide
    a. Invest in technology and digital access for secondary education.
4. Promote Gender Equality in Education
    a. Implement targeted programs to reduce gender disparities in secondary and upper-secondary education.
5. Address Teacher Shortages
    a. Develop policies to recruit and retain teachers, especially in underserved and conflict-affected areas

# Recommendations 11

1. Balance Investment in Education
    a. Allocate a fair share of funding to both operational costs and infrastructure.
2. Support Long-Term Education Planning
    a. Encourage governments to create sustainable education policies aligned with global frameworks like SDGs.
3. Facilitate Global Knowledge Sharing
    a. Promote international collaboration for best practices in education.
4. Targeted Programs for Vulnerable Regions
    a. Design education programs tailored to the needs of vulnerable, conflict-affected, or economically unstable regions.
5. Increase International Education Investment
    a. Mobilize more financial resources for education, particularly in low-income countries

# Learnings and Challenges

Finding rewards in unexpected places
Being glad when things work
Realizing I'm more capable than I think
Taking off pressure because I CAN do it - I AM DOING IT

Choosing a theme
Setting an Objective
Deciding what to keep and what to drop
Making Choices
Time Management
Pressure
Method used was the good old *AAAAAAAAAAAAAAA*

# Thank You!

Any questions?
Comments?
Concerns?

EducationStudyApp
GitHub