

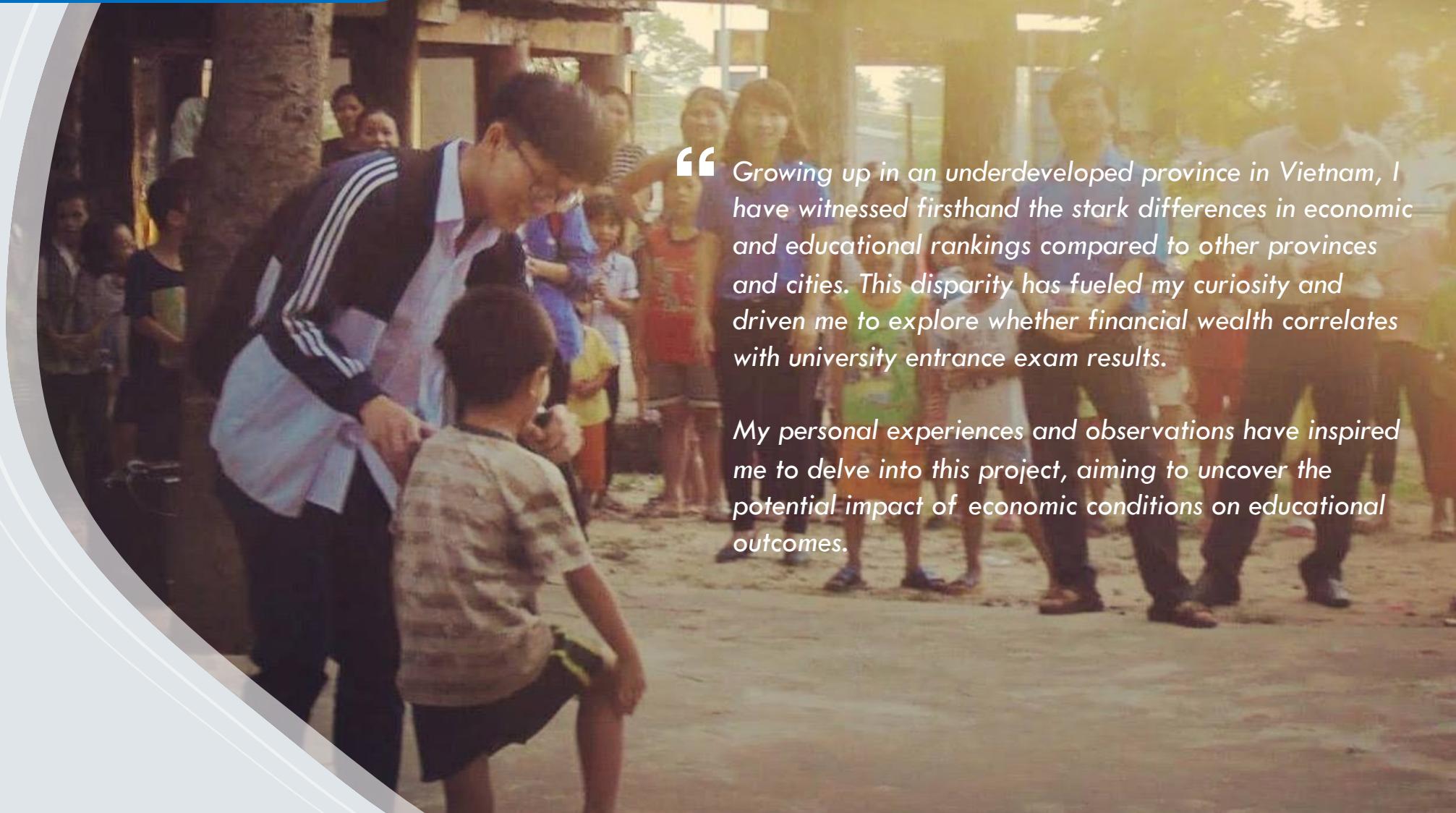


# Provincial Wealth and Exam Grades in Vietnam from 2020 to 2022

FINAL PORTFOLIO PROJECT

by Duc Tran Anh

# Project topic's motivation



**“** Growing up in an underdeveloped province in Vietnam, I have witnessed firsthand the stark differences in economic and educational rankings compared to other provinces and cities. This disparity has fueled my curiosity and driven me to explore whether financial wealth correlates with university entrance exam results.

My personal experiences and observations have inspired me to delve into this project, aiming to uncover the potential impact of economic conditions on educational outcomes.

# AGENDA

Project topic

## *“Provincial wealth and Exam Grades in Vietnam from 2020 to 2022”*

### Project Overview

- Objective
- Data collection
- Central questions

### Project Scoping

- Task scoping
- Deliverables

### Key findings

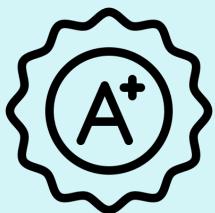
- Central questions
- Key findings

### Summary

- Key learnings
- Next-steps

## Project Objective

The project objective is to analyze university entrance exam (UEE) grade, monthly income per capita and the GINI index across 63 provinces and cities in Vietnam from 2020 to 2022.



### **University entrance exam score**

is the set of scores that a student receives on standardized exams by Vietnam Ministry of Education, used for admission to universities.



### **The Gini Index**

a measure of inequality in income distribution, is calculated based on the Lorenz curve, which plots the cumulative share of income received by the cumulative share of the population. (Source: World Bank Databank)

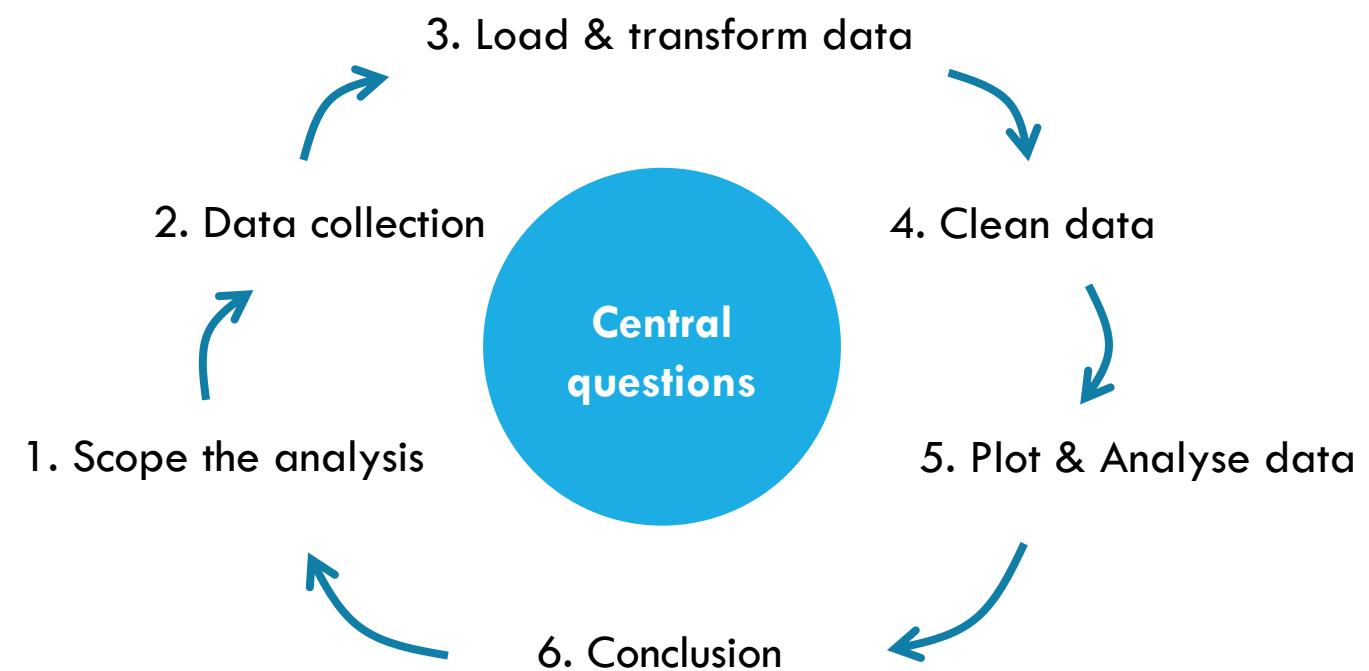


### **Monthly income per capita**

refers to the average income earned by each individual in a given area, such as a province or a city, on a monthly basis.

## Project scoping

The project is conducted in a circular process with 6 steps, centred around exploring insights towards central questions.

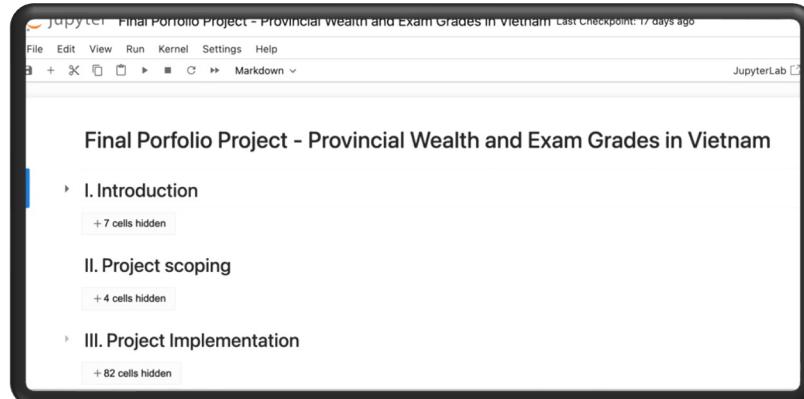


## Project scoping

The deliverables of the project include **3 key items**:

### Python code

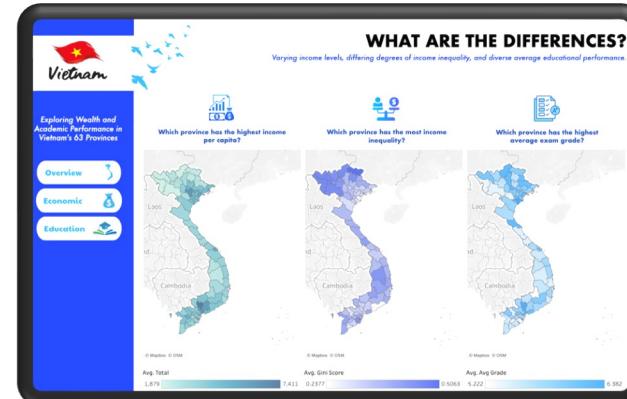
Use basic python prompts to load, clean, plot and analyze data to answer around central questions.



Link access: [here](#)

### Tableau Public

Discover more insights beyond central questions such as distribution, correlation etc via data visualization.



Link access: [here](#)

### PPT Deck

Project summary with purpose of sharing to recruiters and peers for feedback.



<You are reading this file>

## Data collection

### Source of data collection:

Data	Sub-data	Year	Source	Link
Economic	Average monthly income per capita at current prices by income source and by locality	2020	General Statistics Office of Vietnam	<a href="#">here</a>
Economic	Average monthly income per capita at current prices by income source and by locality	2021	General Statistics Office of Vietnam	
Economic	Average monthly income per capita at current prices by income source and by locality	2022	General Statistics Office of Vietnam	
Grade	National high school graduation exam scores	2020	Github - ngc-minh	<a href="#">here</a>
Grade	National high school graduation exam scores	2021	Github - chuongmep	<a href="#">here</a>
Grade	National high school graduation exam scores	2022	Github - anhdung98	<a href="#">here</a>
Social	Income inequality coefficient (GINI coefficient) by locality	2020	General Statistics Office of Vietnam	<a href="#">here</a>
Social	Income inequality coefficient (GINI coefficient) by locality	2021	General Statistics Office of Vietnam	
Social	Income inequality coefficient (GINI coefficient) by locality	2022	General Statistics Office of Vietnam	

## Python code

## Central questions

### Income Data

1. Which provinces have the highest and lowest income per capita in 2022?
2. What type of income is the highest, and how is each type distributed?
3. How has the income level changed over the years?

### Gini Data

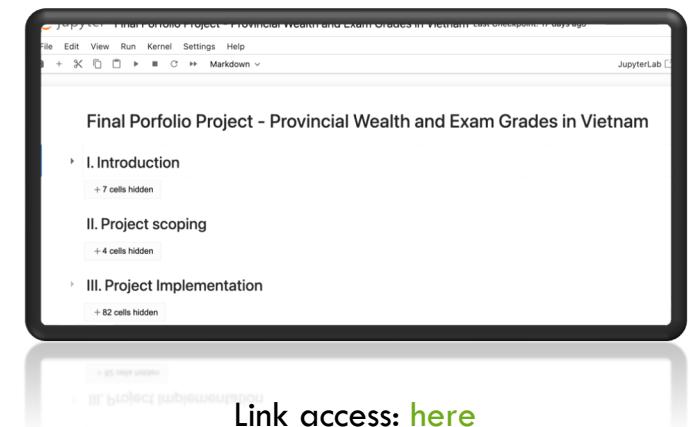
4. Which provinces have the highest and lowest Gini index?
5. What is the distribution of the Gini index?
6. How have the top 5 provinces with the highest Gini index changed over the years?

### University Entrance Exam Score Data

7. How has the UEE score distribution changed from 2020 to 2022?
8. Which 10 provinces have the most students with excellent grades (score > 9)?

### Relationships

9. Is there a relationship between changes in income per capita and changes in the Gini index?
10. Is there a relationship between changes in income per capita and changes in UEE grades?



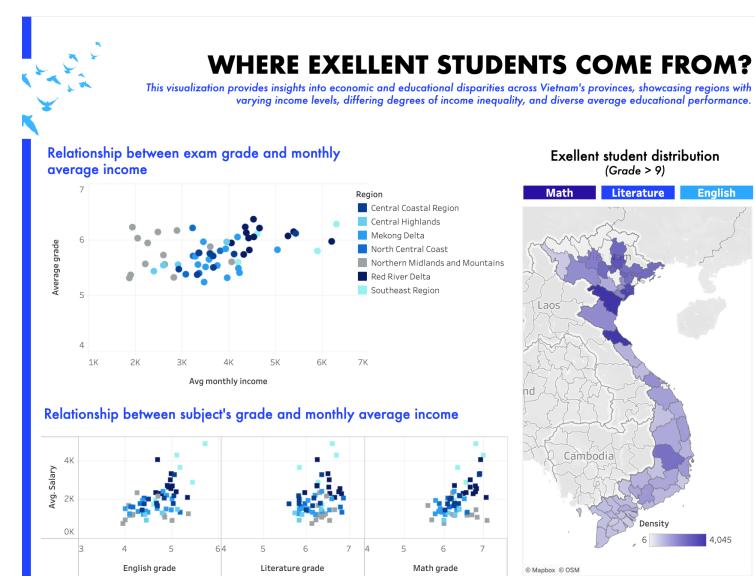
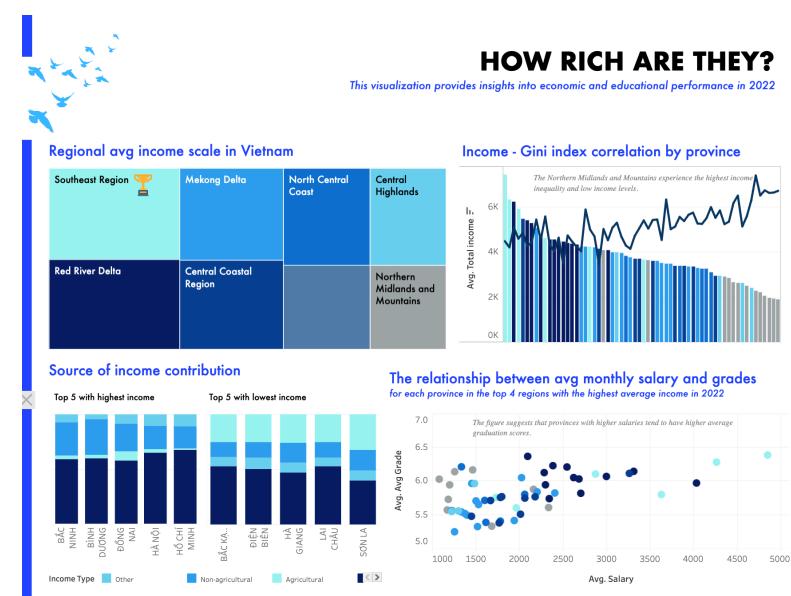
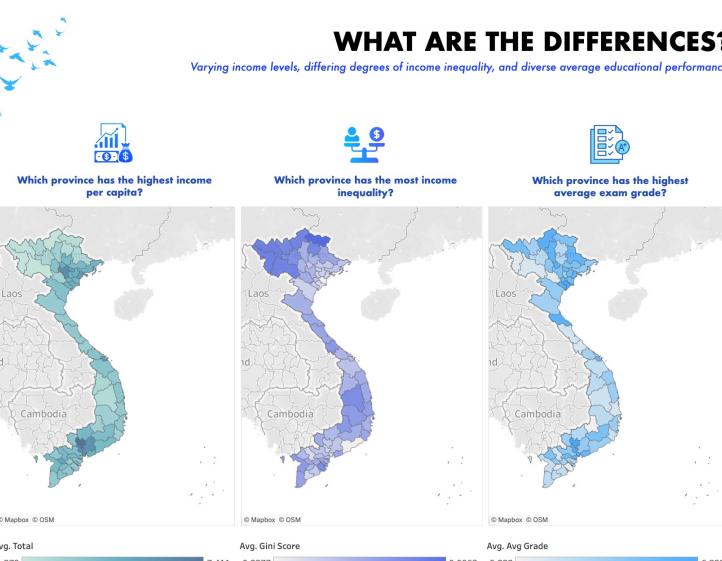
The screenshot shows a Jupyter Notebook interface with a title bar "Final Portfolio Project - Provincial Wealth and Exam Grades in Vietnam" and a menu bar "File Edit View Run Kernel Settings Help". Below the menu is a toolbar with icons for file operations like New, Open, Save, and Run. The main area displays a table of contents for the project:

Final Portfolio Project - Provincial Wealth and Exam Grades in Vietnam	
>	I. Introduction + 7 cells hidden
>	II. Project scoping + 4 cells hidden
>	III. Project Implementation + 82 cells hidden

Link access: [here](#)

# Tableau report

## Example



Link access: [here](#)

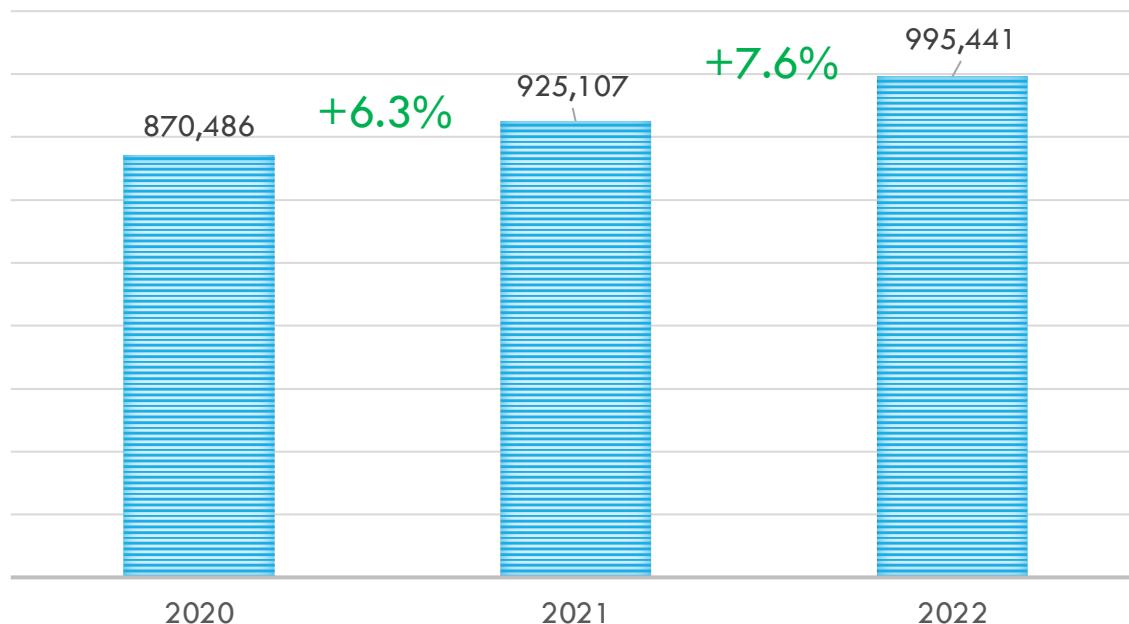
**Do you believe a student from a wealthy province would achieve higher scores than one from a modest area?**

**Is it true that the wealthier a province becomes over time,  
the higher a student's score will be?**

**Let's explore...**

# Every year, about 1 million students in Vietnam take the National Exam, but only half pursue higher education.

THE NUMBER OF ANNUAL NATIONAL EXAM CANDIDATES FROM 2020 TO 2022

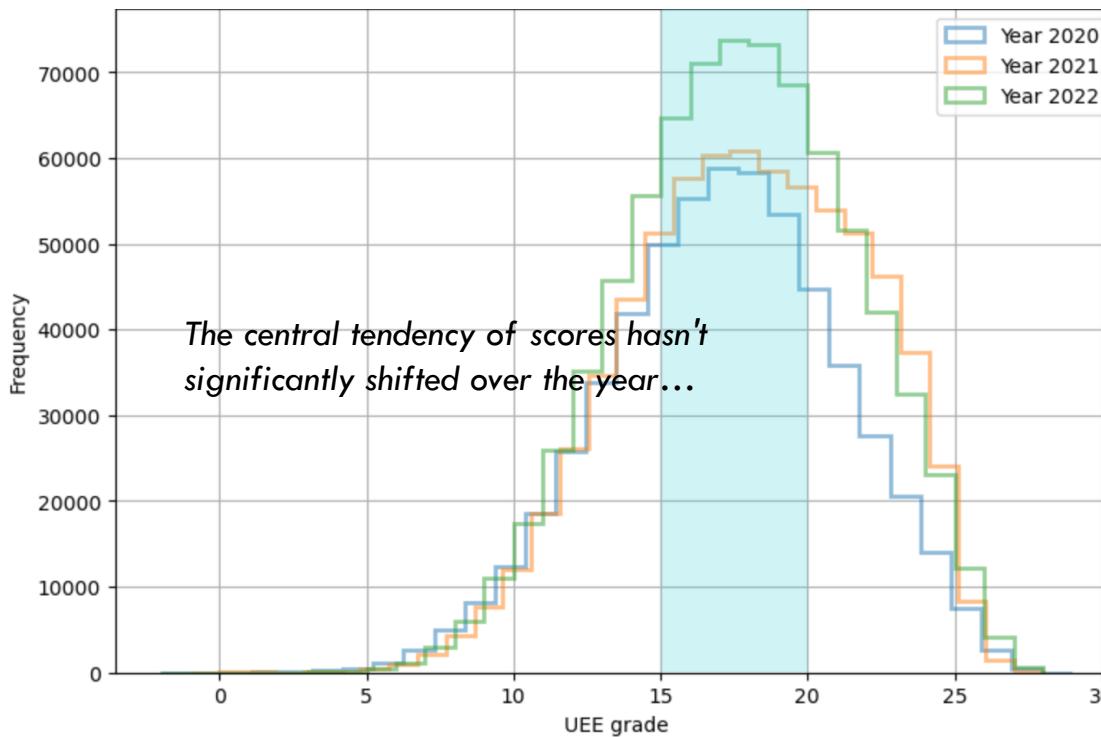


Among those candidates...

-  **50-55%** pursue higher education
-  **20-35%** attend vocational training
-  **10-20%** enter labor market
-  **5-10%** unemployment

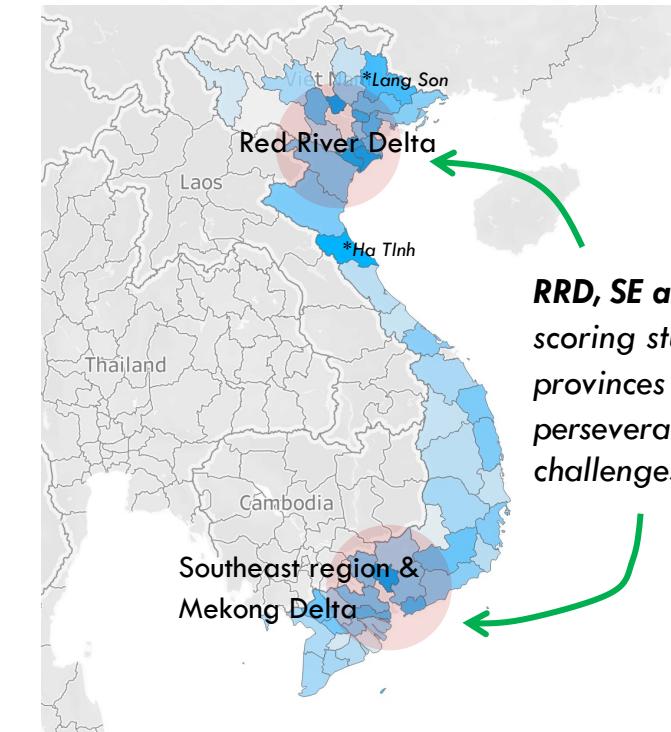
**Distribution peaks appear to be around 15-20 range each year, with high scores mainly come from metropolitan areas, except for Lang Son & Ha Tinh.**

Distribution of UEE grade from 2020 to 2022



The histograms for all three years are roughly **bell-shaped**, indicating a normal distribution of grades with some skewness. This suggests that most students score around the middle range, with fewer students achieving very high or very low scores.

Average score by province in 2022



**RRD, SE and Mekong Delta** have high-scoring students, Lang Son and Ha Tinh provinces are notable for their students' perseverance despite economic challenges.

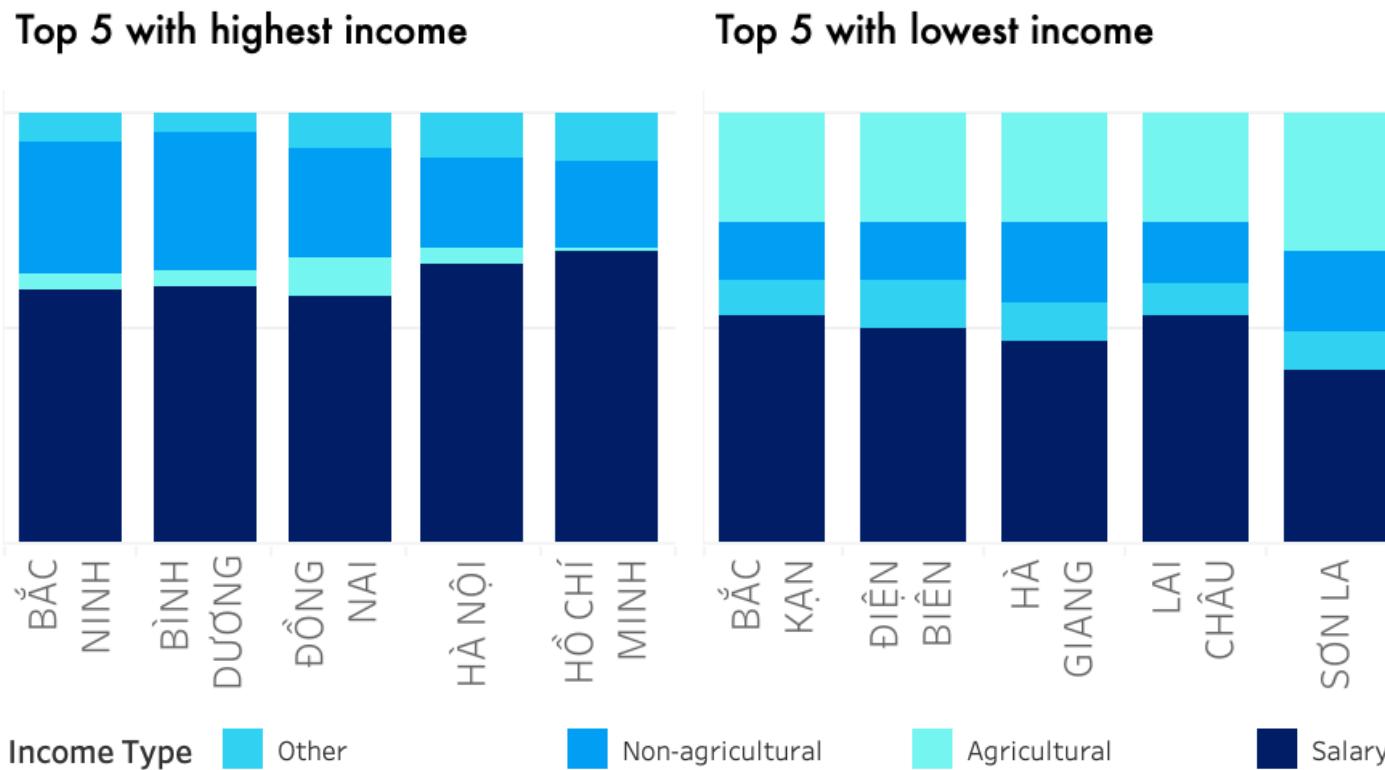
**When looking deeper by subject and excellence, northern provinces have most of brilliant students across key subjects, especially Lang Son.**

Density of excellent students (Grade > 9) in 2022



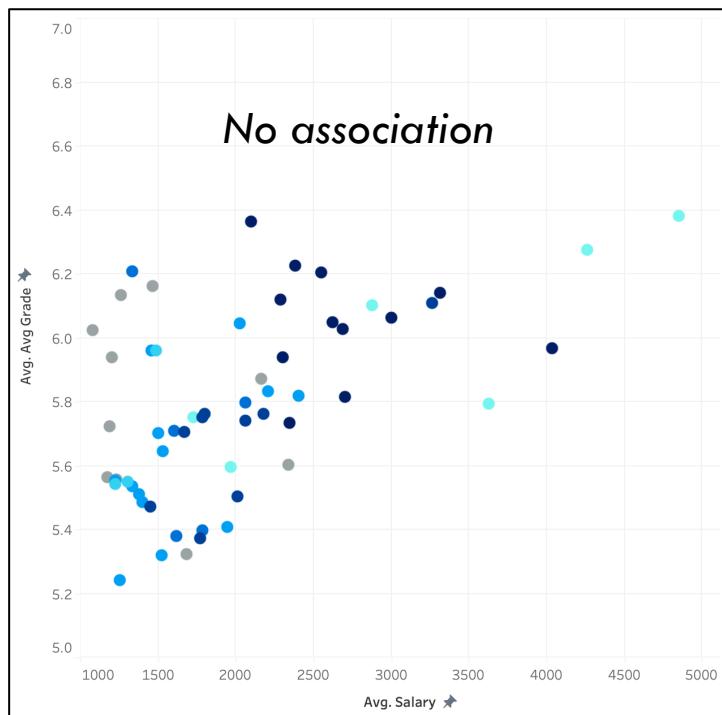
**It is clear that wealthier provinces predominantly earn from salaries and non-agricultural income, while more modest provinces, in addition to salaries, rely more on agricultural income.**

Source of income by selected provinces in 2022

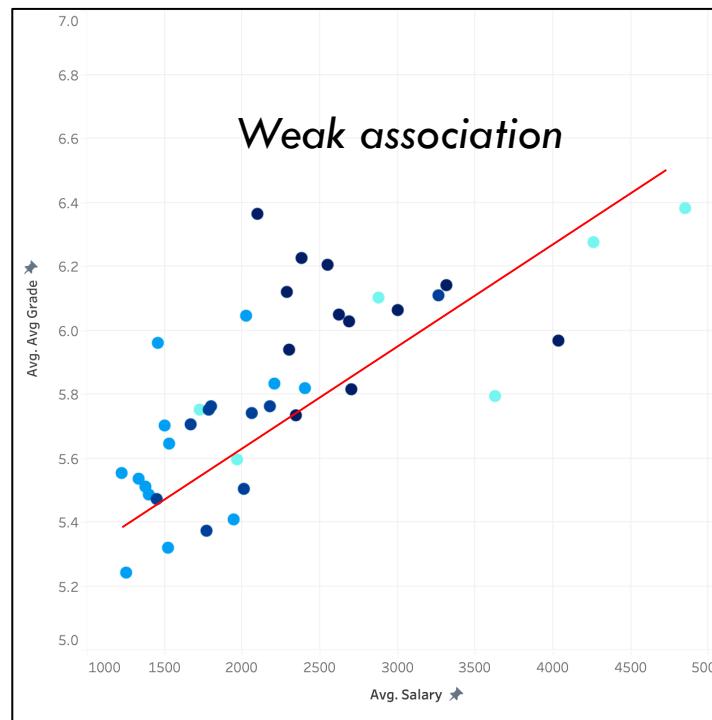


# Wealthier regions show a weak positive relationship among income level and grade, indicating that economic resources may have some influence on educational outcomes, though not strongly.

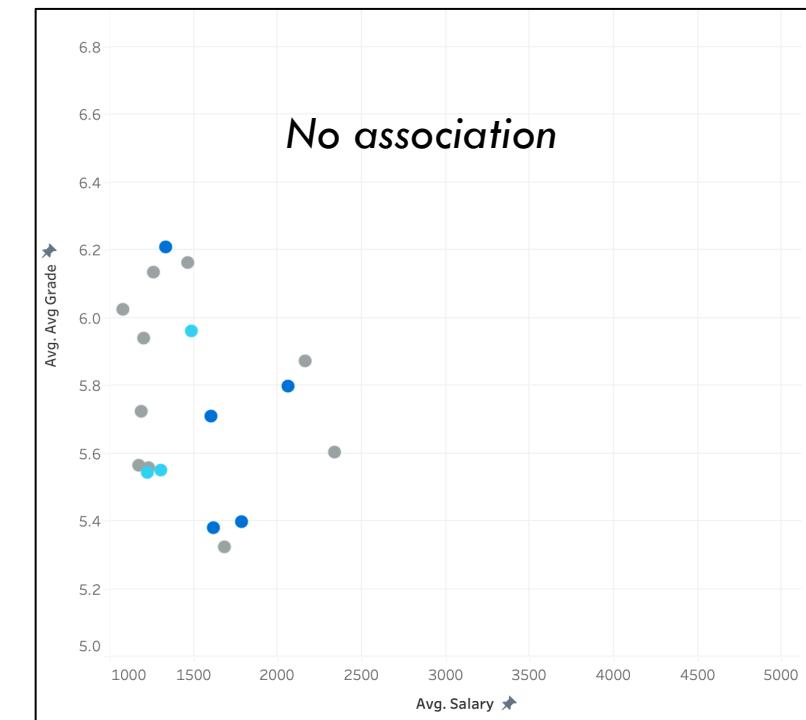
Relationship between average salary and average grade across different regions and provinces.



All regions and provinces



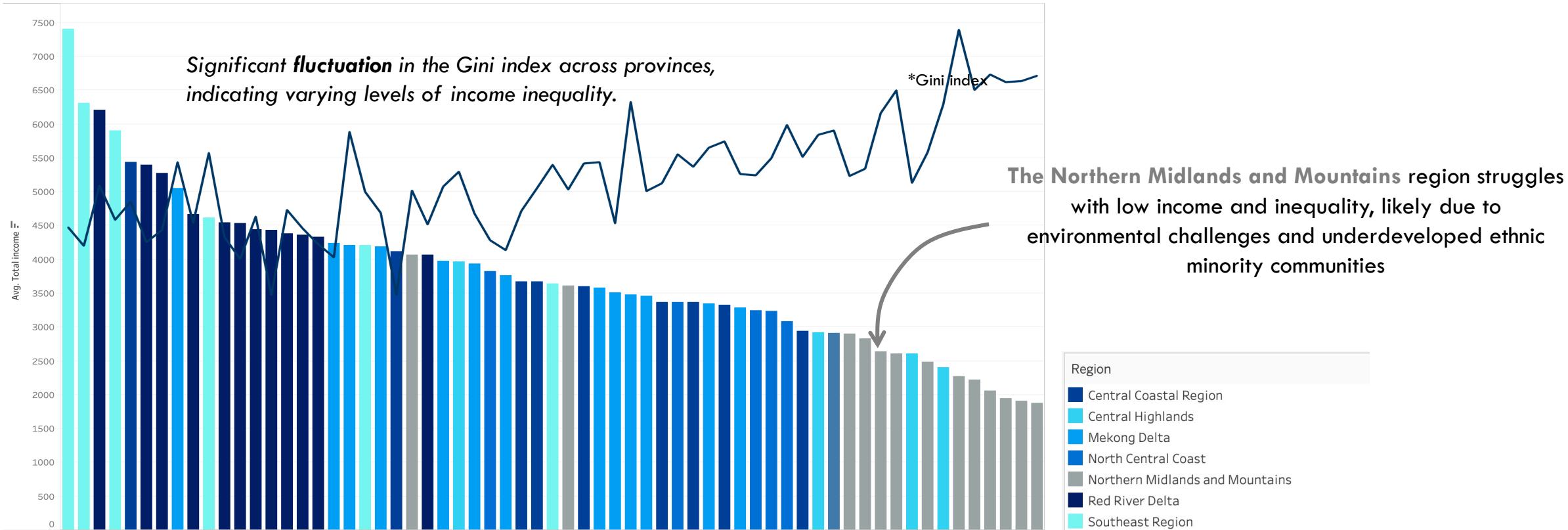
Provinces in Top 4 regions having highest avg monthly income per capita: Southeast, Red River Delta, Mekong Delta and Central Coastal regions



Provinces in Top 3 regions having lowest avg monthly income per capita: Northern midlands and mountains, North Central Coast, Central Highlands.

**Wealthier provinces do not consistently exhibit lower income inequality, and poorer provinces do not consistently show higher inequality. This reflecting diverse economic conditions within each region.**

The correlation between avg monthly income per capita and the Gini index by province



**From 2020 to 2021 and from 2021 to 2022, there are changes in the income level, grade, and Gini index across provinces...**

*Are there any relationships in those changes?*

**Changes in income per capita do not appear to be associated with changes in grade based on the given data. However, the p-value is quite close to the significant level, which suggests further research on the effect.**

**Conduct Chi-square test to examine the relationship among 2 categorical variables**

### HYPOTHESIS

H0: There is no statistically significant relationship between the change in income per capita and the change in grade over year.

H1: There is statistically significant relationship between the change in income per capita and the change in grade over year.

### OBSERVED TABLE:

Grade trend Income trend	Decrease	Increase
Decrease	120,915	101,878
Increase	589,875	492,870

### OUTPUT:

Chi2 Statistic: 3.191

p-value: 0.074



Since the p-value is greater than 0.05, we reject the null hypothesis. This means that there is **no statistically significant relationship** between the change in income and the change in grade.

However, the p-value is relatively close to 0.05, suggesting a trend towards significance. This implies that there might be a weak association among variables. For **exploratory research purpose**, this potential effect should be delved further.

# No statistically significant relationship between the change in income per capita and change in the Gini index.

Conduct Chi-square test to examine the relationship among 2 categorical variables

## HYPOTHESIS

H0: There is no statistically significant relationship between the change in income per capita and the change in Gini index over year.  
H1: There is statistically significant relationship between the change in income per capita and the change in Gini index over year.

## OBSERVED TABLE:

Gini trend Income trend	Decrease	Increase
Decrease	13	10
Increase	47	51



The p-value is 0.6118, which is significantly higher than the typical significance level of 0.05. Hence, there is **no statistically significant relationship** between the change in income and the change in Gini index.

## OUTPUT:

Chi2 Statistic: 0.2575  
p-value: 0.6118

# Summary



Wealthier provinces in Vietnam primarily derive income from salaries and non-agricultural activities, while less affluent provinces rely more on agricultural income.

The distribution of income showed significant disparities across regions, with the Southeast and Red River Delta regions having the highest average income levels.



The analysis of UEE scores from 2020 to 2022 revealed that the majority of students scored in the middle range, with a normal distribution of scores across the three years.

Wealthier provinces generally had higher average UEE scores, but the correlation between income and educational outcomes was weak.

Lang Son and Ha Tinh provinces stood out for their high number of excellent students despite their economic challenges.



The Gini index, which measures income inequality, varied significantly across provinces. Wealthier provinces did not always have lower income inequality, and poorer provinces did not consistently exhibit higher inequality.

The Northern Midlands and Mountains region, characterized by low income and high inequality, faced challenges likely due to underdeveloped infrastructure and ethnic minority communities.



The Chi-square tests conducted to explore relationships between changes in income per capita, the Gini index, and UEE grades showed no statistically significant associations. However, the test for income and grade changes suggested a potential weak relationship that could warrant further exploration.

# Further suggestions and next-steps

## Expand Data Collection

Collect additional data points, especially for more recent years, to enhance the robustness of the analysis. A larger dataset might reveal more subtle relationships and trends.

## Investigate Additional Factors

Explore other socioeconomic factors that might influence the relationship between income, Gini index, and UEE grades. This could include factors like government spending on education, infrastructure development, and population density.

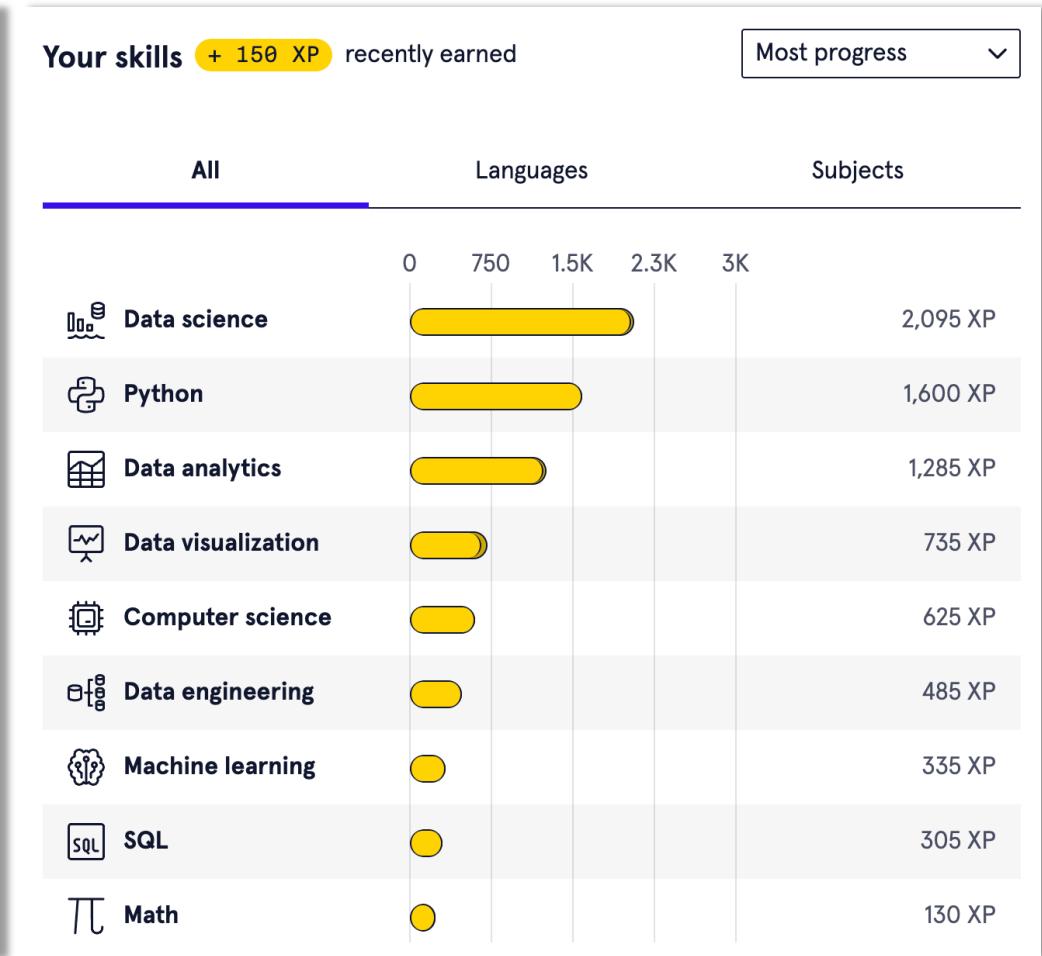
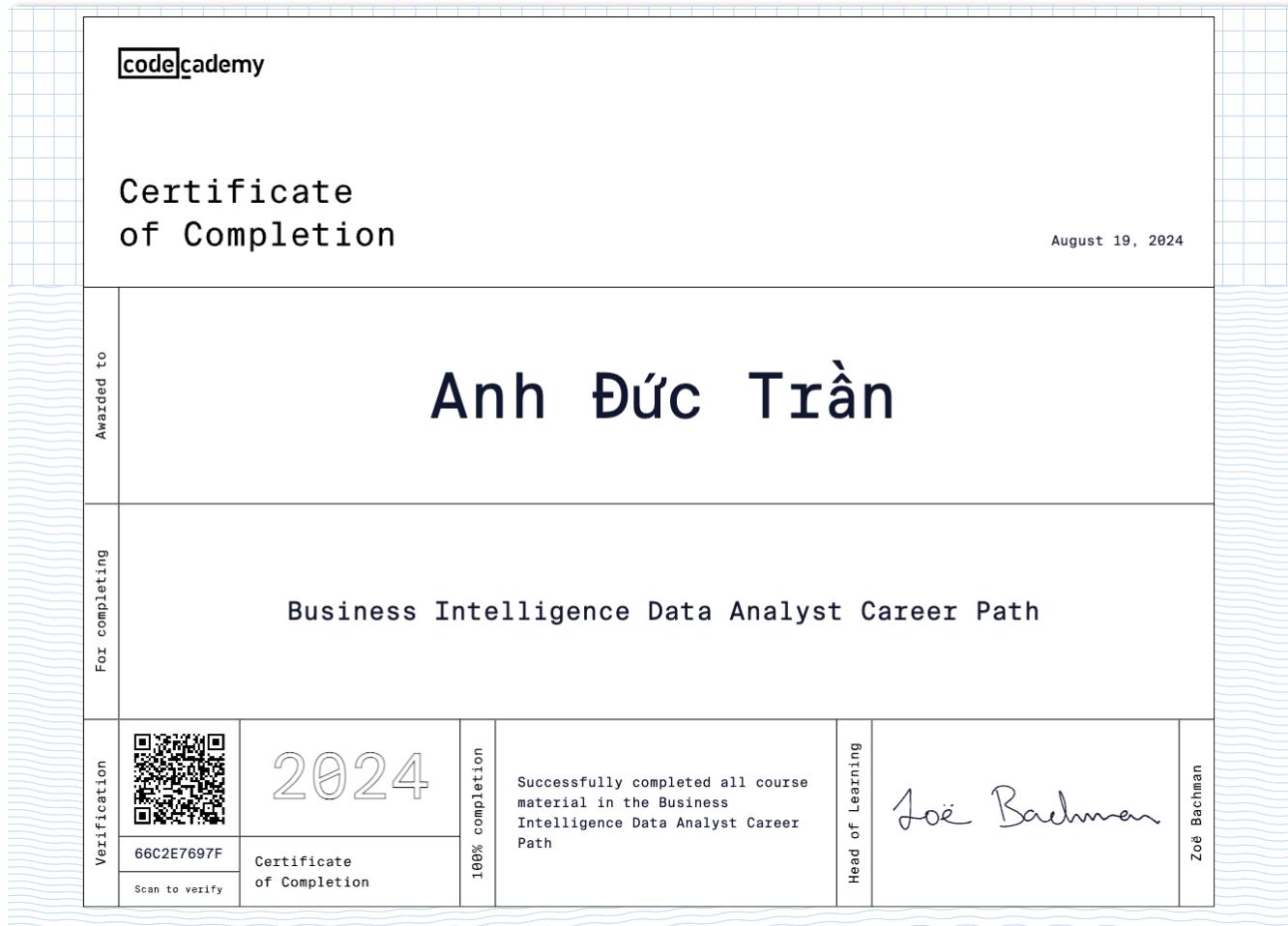
## Refine Analytical Methods

- Consider employing more advanced statistical techniques, such as multivariate regression analysis etc, to better understand the complex interactions between variables.
- Reassess the significance levels used in hypothesis testing, possibly adopting a more relaxed threshold (e.g., 10%) for exploratory analyses.

## Explore Regional Differences

Conduct a more detailed regional analysis to understand the specific challenges and opportunities within each province. This could involve case studies of provinces like Lang Son and Ha Tinh, which showed unexpected educational outcomes relative to their economic status.

# Certificate credentials



# THANK YOU FOR YOUR READING!

---

Any feedback and suggestions are welcomed to help me further improve my skills ^^!



+84 931 627 266



ducanhtrandt@gmail.com



/in/ducanhtranx/