PDAS

Hello, my name is Hubert this is my data analysis, “World Education Analysis”.

//

The objective of this analysis is to find out how does the school life expectancy in countries affects the daily mean income/expenditure, the unemployment rates and the poverty rates in those countries.

//

There are a total of 4 factors in this analysis, they are school life expectancy, mean daily income, unemployment rate and poverty rate. I will be using a total of 5 datasets for this analysis.

//

Here are the links to those 5 datasets that I will be using. And now, I will be explaining these datasets in detail.

//

The first dataset is “school-life-expectancy-annual.csv”. It shows the expected number of years of education that a 6 year old child can expect to enrol in over his/her lifetime, given the current enrolment rates.

The second dataset is “expected-years-of-schooling.csv”. It is similar to the first dataset, except that it contains the expected years of schooling for children worldwide.

\*Repeat for the rest of the datasets\*

//

To display information about the datasets, I used some functions and methods like len(), type() and np.unique. This picture here shows my code.

//

Now, moving on the cleaning the datasets. In the expected-years-of-schooling.csv dataset, there are some entities that are regions instead of countries.

We have to clean this data out as it will affect our result findings later on in the data analysis. For example, it will cause the mean school life expectancy to change and become incorrect.

In order to prevent this, we need to remove the rows containing the regions.

First, we read the csv file into memory.

We then make a list named “lines” without the unwanted rows.

We then make another list named “x” to remove the indexes in the list “lines”

We then finally write the list “x” back to the csv file.

This data cleaning process only needs to be done once as it changes the original dataset “expected-years-of-schooling.csv” permanently. If we try to clean the cleaned dataset again, it will give us an error. Hence, we will need to comment out the code meant to clean the dataset, in this case, lines 11 to 30.

//

There was also the need to change the headers for the datasets expect-years-of-schooling.csv and unemployment-rate.csv as they have special characters like commas or spaces that messes with the delimiter. Since the delimiter is a comma, it would separate the last header of the dataset “unemployment-rate.csv” into two headers. We would not want that, as it will mess with our code.

Hence, to solve this, simply rewrite the new header without commas, in the “names” field when extracting the datasets with np.genfromtxt.

//

Now, moving onto the data analysis. I will firstly present a visual overview of the data.

The histogram on the left shows the distribution of the mean school life expectancy of each country. The distribution is approximately symmetrical at the centre, representing a normal distribution. The mean, median and mode are about the same in a normal distribution, that is, 12.5 years of school life expectancy.

The line graph on the right shows the mean worldwide school life expectancy over the years. It shows an increasing trend. Hence, we can conclude that as time pasts, the mean school life expectancy increases.

//

Before we draw up the graphs for each country, we need to extract the necessary rows from the expect-years-of-schooling.csv dataset, in this case, school life expectancy of each country from 2000 to 2019.

We will do the same for daily mean income, unemployment rate and the poverty rate.

We will only use the data for Singapore, USA and Indonesia as we will only be comparing these three countries.

//

Firstly, we have the visualisation of Singapore.

On the top left we have a line graph displaying the change of school life expectancy over time. In Singapore, as the years go on, the school life expectancy increases from around 12.75 to 16.5. There was a huge increase from around the year 2004 to 2005, an increase of school life expectancy from 12.6 to 14.0, an increase of 1.4 years.

On the top right we have a line graph displaying the rate of employment over time. Unfortunately, it does not seem to have a relation with school life expectancy. It seems to be affected by other external factors, like economic recessions etc.

At the bottom we have a pie chart displaying the mean proportion of the unemployed to the employed over the years in Singapore. The percentage of unemployed in Singapore is 4.3%.

Now, before I move onto the visualisation of the US and Indonesia, let me first explain why the rate of poverty and daily mean income is not present for Singapore.

Firstly, the rate of poverty and daily mean income for Singapore were not present in the datasets.

Secondly, Singapore does not have a fixed poverty line so it would be inaccurate and wrong to make up one myself.

To counter this, I decided to bring the US into the comparison as USA has many similar traits to Singapore, having the same mean unemployment rate and about the same mean school life expectancy. Both of them are first world countries too.

//

Now, onto the visualisation of the US.

On the top left we have a line graph displaying the change of school life expectancy over time. In USA, as the years go on, the school life expectancy increases from around 15.4 to 16.5. There’s no huge increases from 2000 to 2019, the increase remained steady.

On the top right we have another line graph displaying the change of daily mean income over time. The line graph resembles an exponential curve, with the years 2000 to 2012 seeing hardy any increase in the daily mean income at all. The years 2012 to 2019, however, saw a huge spike increase, from 70 dollars daily to around 86 dollars daily.

At the bottom left we have a pie chart displaying the mean proportion of those in poverty to those not in poverty over the years in USA. The percentage of those in poverty is disturbingly high for a first world country, at 24.6%.

At the bottom right we have a pie chart displaying the mean proportion of the unemployed to the employed over the years in USA. The percentage of unemployed in USA is 5.9%.

//

Lastly, the visualisation of Indonesia.

In Indonesia, as the years go on, the school life expectancy increases from around 10.6 to 13.9. There’s no huge increases from 2000 to 2019, the increase remained steady.

Similar to the line graph displaying change of school life expectancy over time, there is a steady increase in daily mean income over the years, from 3 dollars daily to about 8 dollars daily.

At the bottom left we can see that the percentage of those in poverty is at 16.7%.

At the bottom right we can see that the percentage of unemployed in Indonesia is 5.7%.

//

Now, I will compare the school life expectancy by income and poverty.

Below, we have bubble plots for the US and Indonesia. The US is on the left and Indonesia is on the right. As Singapore does not have a rate of poverty and daily mean income, I will not be drawing a bubble plot for Singapore. I did not include unemployment rate as it is clear that from previous findings, school life expectancy hardly affects the unemployment rate in a country.

Generally, for both countries, as the school life expectancy increases, the mean daily income increases. This is seen from the increasing trend lines in red.

For both graphs, the higher poverty rates are generally concentrated at the bottom left and the lower poverty rates concentrated at the top right.

This means that a lower school life expectancy results in both lower mean daily income as well as a higher poverty rate.

When comparing the US and Indonesia, there is a steeper increase in mean daily income in Indonesia as compared to USA, though the increase in dollars is lesser in Indonesia, with an increase of 5 dollars in comparison to USA’s increase of 8 dollars. However, the proportion of increase in mean daily income is a lot greater in Indonesia, hence the steeper regression line.

In USA, the high and low poverty rates feel randomly displaced, with some high poverty rates having high school life expectancies. The relationship between the two factors is hard to see. There are also some outliers on the top right that have deviated far away from the trend line.

However, in Indonesia, the relationship between school life expectancy and poverty rates is easier to see. All the high poverty rate points are at the bottom left of the graph while the low poverty rate points are at the top right of the graph.

These could mean that the school life expectancy does not affect the poverty rate in the US as much as it is in Indonesia.

//

Now, I will compare the three countries together.

The mean school life expectancies differences between the three countries are rather small, with the difference between the highest and lowest being 4. USA has the highest school life expectancy, followed by Singapore, and lastly Indonesia.

Because of the mean school life expectancy, USA has a higher mean daily income as compared to Indonesia.

However, for some reason, the rate of poverty in the USA is a lot higher than in Indonesia, even though the unemployment rate is the same for the two of them. This is probably due to other factors, like the country’s economic or social situation.

The unemployment rate remains about the same for all three countries, despite the differences in school life expectancies.

//

To conclude, the school life expectancy does have some impact on the daily mean income and the poverty rate. It however does not seem to have much of an impact on unemployment rate

When the school life expectancy increases, daily mean income increases while the unemployment and poverty rate decreases.

The amount of impact on each factor, however, varies between countries. A lower school life expectancy could have a great effect in mean income on a country, yet maybe for another country, it may not be as big. For example, though the school life expectancy in Indonesia is not much lower as compared to the US, the daily mean income in Indonesia is drastically lower than that in the US. This could be due to some other external factors, like as previously mentioned earlier, the country’s economic situation.

Another example of the impact of a factor varying between countries is the poverty rate. Like I mentioned earlier, even though the school life expectancy in the US is higher than that in Indonesia, it has a higher poverty rate. It is like school life expectancy has an inverse impact on the poverty rate, contrary to the norm.

However, I feel that this could be due to external factors. Reading online, perhaps it could be due to the student loans in the US. Although the education in the US is good, it is very expensive, especially for college. This resulted in many students taking up huge loans that take many years to pay off. This caused them to be stuck in huge debt many years after they have graduated. Hence, even though they are well educated, their debts caused them to be stuck in poverty.

//Another reason could also be the house mortgages (maw-guh-juhz) in the US. Houses are ridiculously expensive in the US as compared to Indonesia. Even though the mean daily income in the US is a lot higher, it takes many years to pay off one’s house mortgages. Hence, even though they are well educated, their debts caused them to be stuck in poverty.

This data analysis hence strengthens the importance of education, it can serve as an opportunity for some to climb up the socio-economic ladder and seek a better life. I recommend governments around the world to place greater emphasis and funding in education to maintain a good school life expectancy. It is best to be above the mean world school life expectancy of 12.5 years.

And that brings me to the end of my data analysis. Thanks for listening.