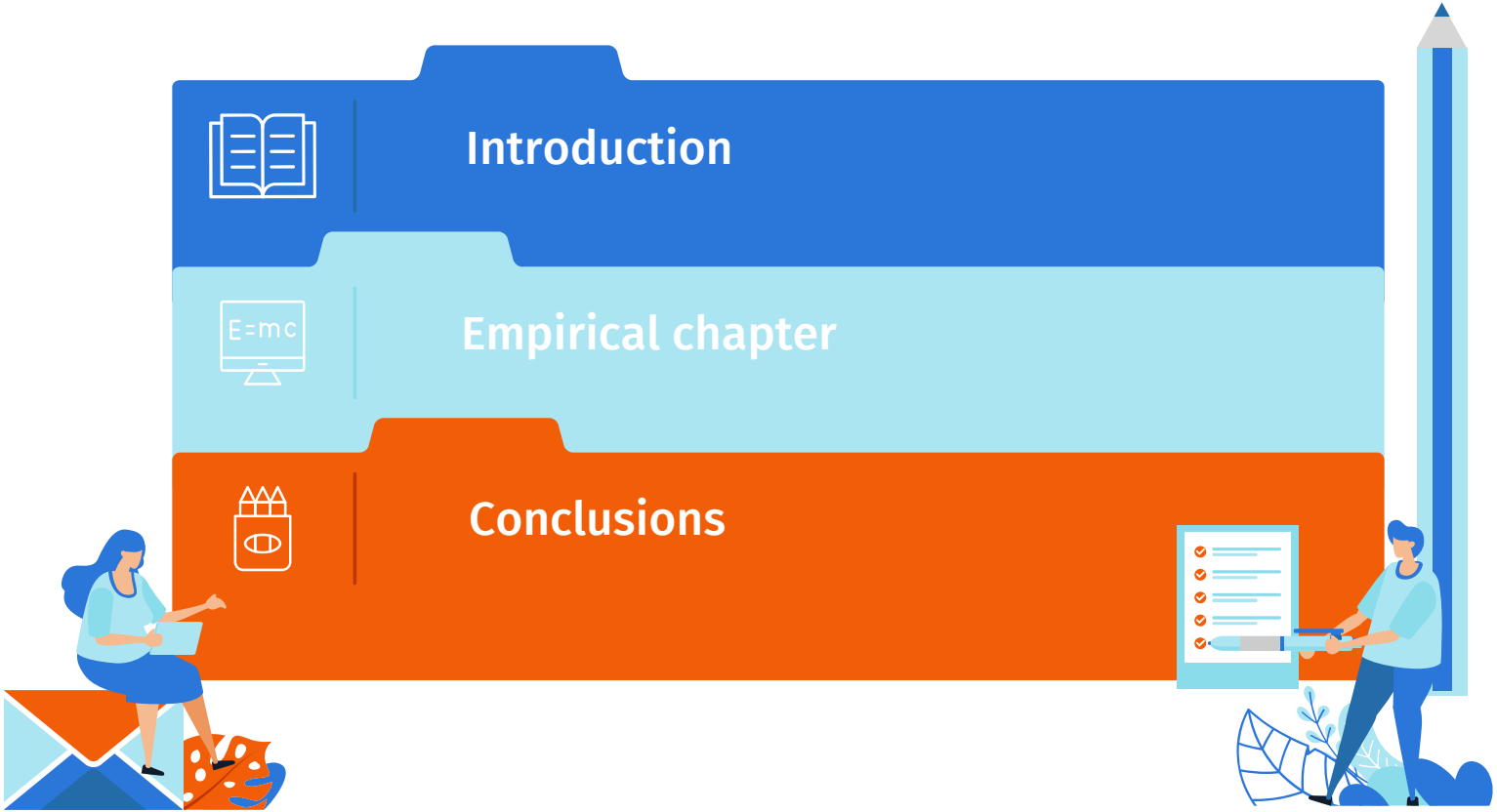
Early school leaving in Europe

Authors: Vîlceanu Diana-Maria,
Zaremba Daniel, Sava Nicolae

CONTENTS

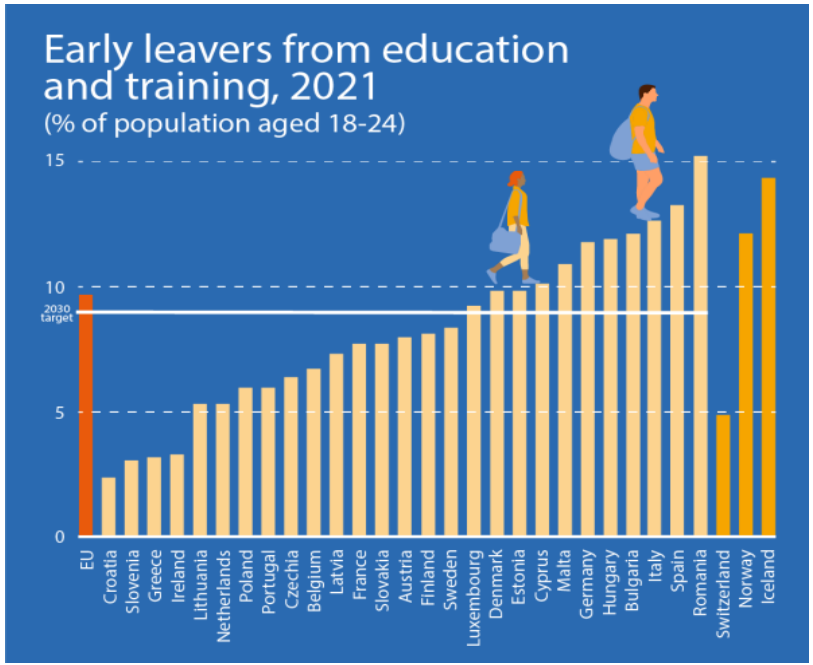


INTRODUCTION

Education is one of the most important factor in a person's career and life, but we all know that it is not mandatory and that for many countries abandoning education for gaining money faster is an easy way to survive due to their economic level. Do people abandon education for money, for an easy lifestyle with no pain in the brain? Many people get tricked by the idea that in a corrupt society you cannot have a good income just by having a high level of education. But is this true?

Throughout this detailed research that we made on our dear continent we wanted to see what is the cause of this prolonged problem. Is it an economic, geographic or cultural one? What actually makes young people dropout education from an early age in Europe?

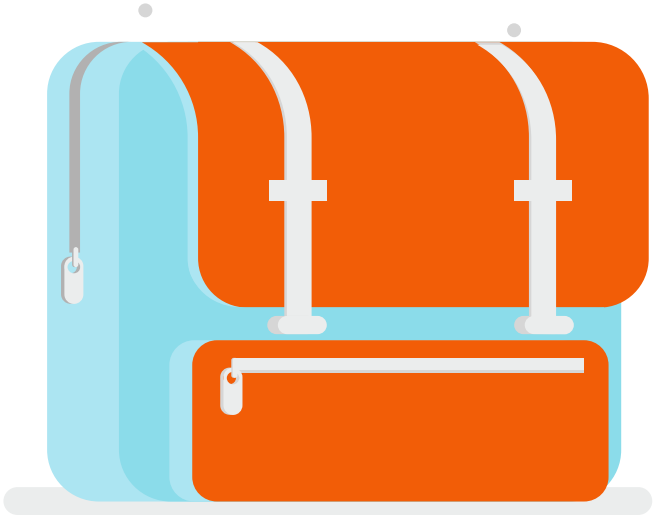
We made an analysis to see how the European countries are facing this problem, as we see in the chart down below:



This analysis is based only on the countries belonging to European Union.

The situation is slightly different when we analyze most of the European countries that are not currently part of EU.

We can observe that, in the European Union, on an analysis made last year, Romania occuppies the first place in the early school dropout rate.



Research methodology

Firstly, we consider the geographic area by European regions in our analysis of early school dropout for the stratified random sampling. We are going to take 5 strata and analyze each of them.

Then, we provide a complete analysis for the hypothesis testing. Here, we study the Europe as a whole entity and check if all europeans are on the same page in terms of school leaving rate by assuming we identify their mean.

Keeping this in mind, using ANOVA analysis we want to see if people from a specific region of Europe are liable to leave school earlier and compare this to another ANOVA analysis made on each country's predominant religion. In the last case, we take into consideration the mail religions that europeans have and the unaffiliated ones.

Moving on, we want to study the relationship between the labor force participation rate and the early school leaving rate using simple linear regression analysis because they seem to be different between them but correlated somehow to our dependent variable, the inflation rate.

Last but not least, for multiple regression, we combined some of the variables we mentioned above, the best ones for this particular study, with one dummy: the inflation rate. We considered it a high inflation rate if it is bigger than 3 and a low one if it is smaller than 3 and allocated the countries 1 for a high inflation rate and 0 for a small one.

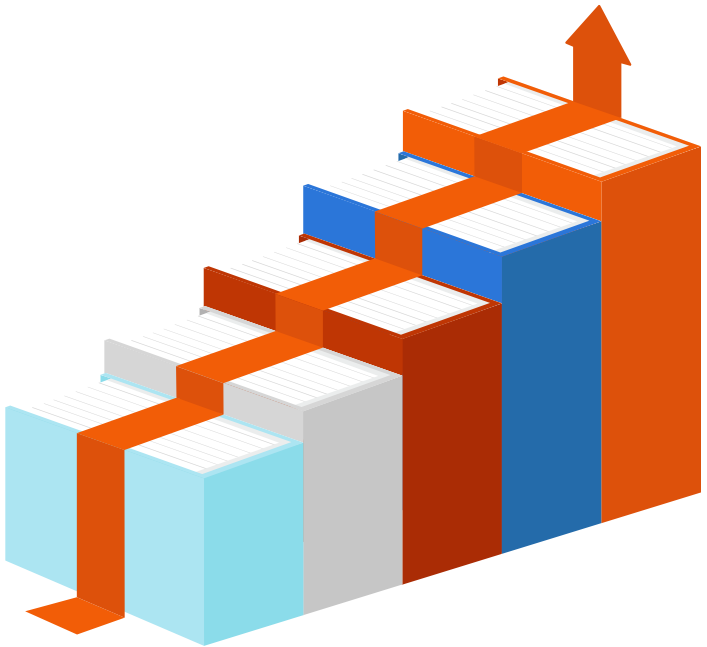
Data used

At first, let's take a look at our data. We collected data about the Early school leaving rate of young people (between 18-24 years) in Europe in the past 5 years. As you may notice, our periodicity is in an yearly form. We also took into account some countries that are not in the European Union such as Moldova, Bosnia and Herzegovina or Albania.

Early school leavers rate in Europe in the past 5 years

Nr crt	Country	Region	Years					Average
			2017	2018	2019	2020	2021	
1	Albania	Southern Europe	7.0	6.4	7.7	6.5	6.7	6.9
2	Austria	Western Europe	7.4	7.3	7.8	8.1	8.0	7.7
3	Belarus	Eastern Europe	13.2	13.3	12.0	12.5	9.9	12.2
4	Belgium	Western Europe	8.9	8.6	8.4	8.1	6.7	8.1
5	Bosnia and Herzegovina	Southern Europe	8.0	7.4	7.3	7.1	7.2	7.4
6	Bulgaria	Eastern Europe	12.7	12.7	13.9	12.8	12.2	12.9
7	Croatia	Southern Europe	3.1	3.3	3.0	2.2	2.4	2.8
8	Cyprus	Eastern Europe	8.5	7.8	9.2	11.5	10.2	9.4
9	Czech Republic (Czechia)	Eastern Europe	6.7	6.2	6.7	7.6	6.4	6.7
10	Denmark	Northern Europe	8.8	10.4	9.9	9.3	9.8	9.6
11	Estonia	Northern Europe	11.8	12.0	11.2	8.5	9.8	10.7
12	Finland	Northern Europe	8.2	8.3	7.3	8.2	8.2	8.0
13	France	Western Europe	8.8	8.7	8.2	8.0	7.8	8.3
14	Germany	Western Europe	10.1	10.3	10.3	10.1	11.8	10.5
15	Greece	Southern Europe	6.0	4.7	4.1	3.8	3.2	4.4
16	Hungary	Eastern Europe	12.5	12.5	11.8	12.1	12.0	12.2
17	Iceland	Northern Europe	17.8	21.5	17.9	14.8	14.4	17.3
18	Ireland	Northern Europe	5.0	5.0	5.1	5.0	3.3	4.7
19	Italy	Southern Europe	14.0	14.5	13.5	13.1	12.7	13.6
20	Latvia	Northern Europe	8.6	8.3	8.7	7.2	7.3	8.0
21	Lithuania	Northern Europe	5.4	4.6	4.0	5.6	5.3	5.0
22	Luxembourg	Western Europe	7.3	6.3	7.2	8.2	9.3	7.7
23	Malta	Southern Europe	14.0	14.0	13.9	12.6	11.0	13.1
24	Moldova	Eastern Europe	11.0	8.5	9.1	9.9	10.3	9.8
25	Montenegro	Southern Europe	5.4	4.6	5.0	3.6	3.6	4.4
26	Netherlands	Western Europe	7.1	7.3	7.5	7.0	5.3	6.8
27	North Macedonia	Southern Europe	8.5	7.1	7.1	5.7	5.7	6.8
28	Norway	Northern Europe	10.4	9.9	9.9	9.9	12.3	10.5
29	Poland	Eastern Europe	5.0	4.8	5.2	5.4	5.9	5.3
30	Portugal	Southern Europe	12.6	11.8	10.6	8.9	5.9	10.0
31	Romania	Eastern Europe	18.1	16.4	15.3	15.6	15.3	16.1
32	Serbia	Southern Europe	6.2	6.8	6.6	5.6	6.3	6.3
33	Slovakia	Eastern Europe	9.3	8.6	8.3	7.6	7.8	8.3
34	Slovenia	Southern Europe	4.3	4.2	4.6	4.1	3.1	4.1
35	Spain	Southern Europe	18.3	17.9	17.3	16.0	13.3	16.6
36	Sweden	Northern Europe	7.7	7.5	6.5	7.7	8.4	7.6
37	Switzerland	Western Europe	4.5	4.4	4.4	4.0	4.9	4.4
38	Turkey	Eastern Europe	32.5	31.0	28.7	26.7	26.7	29.1
39	United Kingdom	Northern Europe	10.6	10.7	10.9	10.9	10.9	10.8

However, we also computed the average for those 5 years, but for our statistics we used the index from 2020, according to our other data collected. Our project has around 4 other variables in use, including GDP per capita, the inflation rate, the labor force participation rate and religion. The first three of them are based on the economic environment from 2020 of each country and the last one is purely a qualitative variable which we used to see if the school abandonment is influence by people's beliefs or if it is only a economic matter. The result is seen in the numbers, we will not reveal that yet.



The empirical results

Stratified random sampling. Confidence intervals

At first, we grouped the countries from Europe in european regions to have a clear image about what we have to analyze. There is a total of 39 countries, from which 31% are from Southern Europe, 26% are from Northern Europe, 26% are from Eastern Europe and 18% are from Western Europe.

Southern Europe		RAND	Northern Europe		RAND	Eastern Europe		RAND	Western Europe		RAND
Albania	6.5	0.9833	Denmark	9.3	0.6892509	Belarus	12.5	0.124615	Austria	8.1	0.69034293
Bosnia and Herzegovina	7.1	0.8555	Estonia	8.5	0.6552087	Bulgaria	12.8	0.46505	Belgium	8.1	0.62064512
Croatia	2.2	0.4052	Finland	8.2	0.3006174	Cyprus	11.5	0.844204	France	8.0	0.53838494
Greece	3.8	0.7788	Iceland	14.8	0.9193633	Czech Republic (Czechia)	7.6	0.118104	Germany	10.1	0.4573114
Italy	13.1	0.3471	Ireland	5.0	0.9410296	Hungary	12.1	0.770035	Luxembourg	8.2	0.41995396
Malta	12.6	0.3723	Latvia	7.2	0.7095036	Moldova	9.9	0.073954	Netherlands	7.0	0.37579565
Montenegro	3.6	0.2273	Lithuania	5.6	0.5764063	Poland	5.4	0.477805	Switzerland	4.0	0.53845662
North Macedonia	5.7	0.422	Norway	9.9	0.9723317	Romania	15.6	0.838065			
Portugal	8.9	0.7736	Sweden	7.7	0.2578239	Slovakia	7.6	0.367647			
Serbia	5.6	0.2339	United Kingdom	10.9	0.8740237	Turkey	26.7	0.353796			
Slovenia	4.1	0.5872									
Spain	16.0	0.0257									
12			10			10			7		
31%			26%			26%			18%		

Then, we decided to pick 5 strata samples of 20 countries by using the =RAND() function in Excel, followed by calculating the mean and standard deviation of each strat.

Our strata means and standard deviations are in connection to one another taking into account that we took samples of a little more than a half of our 39 countries.

For the confidence intervals we took a sample of our five created. Then we took a look at our descriptive statistics based on that sample:

Early school leaving	
Mean	9.53872098
Median	8.15
Mode	8.1
Standard Deviation	5.40904304
Sample Variance	29.25774661
Kurtosis	4.472866295
Skewness	1.751451601
Range	24.5
Minimum	2.2
Maximum	26.7
Sum	190.7744196
Count	20
Largest(1)	26.7
Smallest(1)	2.2

Confidence Level(95.0%)	2.53151
Confidence Level(90.0%)	2.09138
Confidence Level(80.0%)	1.60589

Now, we calculate the confidence level for three cases: 80%, 90% and 95%.

1) 80% confidence level	
$\alpha = 1 - 0.8$	0.2
Margin of error E:	1.550035072
Lower limit:	7.932835313
Upper limit:	11.14460665

we are 80% confident that the population mean is scored between 7.932835313 and 11.144460665

3) 95% confidence level	
$\alpha = 1 - 0.95$	0.05
Margin of error E:	2.370574074
Lower limit:	7.007210912
Upper limit:	12.07023105

we are 95% confident that the population mean is scored between 7.007210912 and 12.07023105.

2) 90% confidence level	
$\alpha = 1 - 0.9$	0.1
Margin of error E:	1.989448477
Lower limit:	7.447336931
Upper limit:	11.63010503

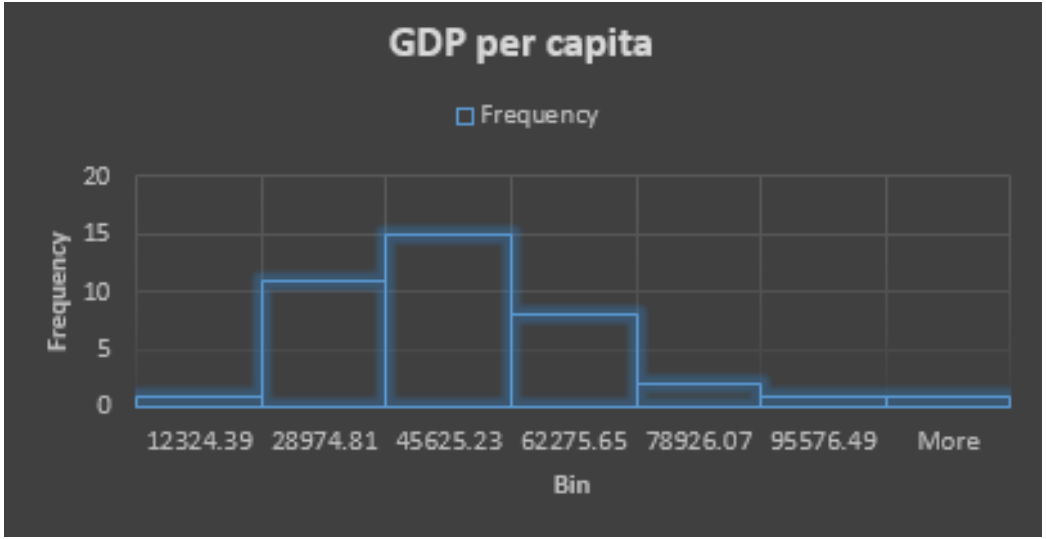
we are 90% confident that the population mean is scored between 7.447336931 and 11.63010503.

Hypothesis testing

Our distribution is leptokurtic, being concentrated at the midpoint and it also have a positive skewness, the distribution having a longer tail towards -infinity. There is a wide amplitude because the largest GDP country is Luxembourg with a score of 112226.91 and at the opposite side we have Moldova as the lowest GDP country in Europe with a value of 12324.39.

GDP per capita	
Mean	39364.20026
Standard Error	3258.070494
Median	37091
Mode	#N/A
Standard Deviation	20346.64371
Sample Variance	413985910.3
Kurtosis	3.731496252
Skewness	1.549737747
Range	99902.52
Minimum	12324.39
Maximum	112226.91
Sum	1535203.81
Count	39
Largest(1)	112226.91
Smallest(1)	12324.39
Confidence Level(95.0%)	6595.618893

Distribution of GDP on European countries:

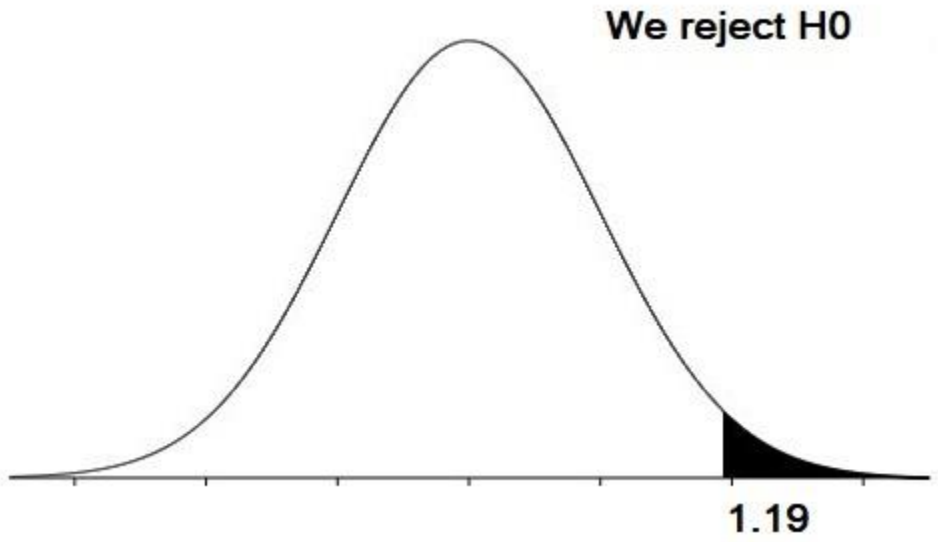


We proceeded with the hypothesis analysis in two ways:

1. One-tailed test

One-tailed test		One-tail
Sample size (n)	39	Critical value for t is 1.64485
Average	39364.20026	We have a right-tail distribution, which means that we should verify if t * > 1.19
H0:	μ <= 18918.72	It is, so we reject the null hypothesis
H1:	μ > 18918.72	
variance	413985910.3	Interpretation: At 95% confidence we can conclude that the average GDP per capita in Europe is bigger than the average world's GDP per capita
a =	0.05	

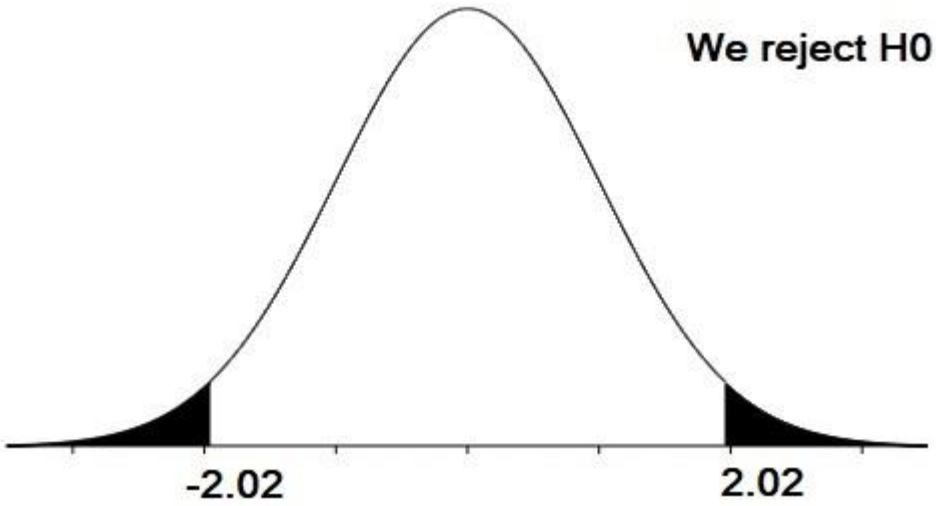
For a more visual idea, here is the chart output:



2. Two-tailed test

Two-tailed test		Two-tail
H0:	μ = 18918.72	Critical value for t is 2.02
H1:	μ != 18918.72	We should verify if t* < -2.02 or t* > 2.02
		It is, so we reject the null hypothesis
		Interpretation: At 95% confidence we cannot conclude that the average GDP per capita in Europe value is equal to the average world's GDP per capita

Moreover, for a visual purpose, the chart output can be seen here:



ANOVA analysis

For the ANOVA analysis, we tried to do two different things: we distributed our countries based on regions and also on religions to see if the early school leaving rate is influenced by those two different factors: the geopolitical factor and the cultural factor. In this regard, we computed ANOVA analysis two times and we had seen some interesting results.

a. *ANOVA analysis on regions from Europe*

Southern Europe	Northern Europe	Eastern Europe	Western Europe
6.5	9.3	12.5	8.1
7.1	8.5	12.8	8.1
2.2	8.2	11.5	8.0
3.8	14.8	7.6	10.1
13.1	5.0	12.1	8.2
12.6	7.2	9.9	7.0
3.6	5.6	5.4	4.0
5.7	9.9	15.6	
8.9	7.7	7.6	
5.6	10.9	26.7	
4.1			
16.0			

Is the early school leaving rate affected by the geographical aspect?

SUMMARY				
Groups	*Count*	*Sum*	*Average*	*Variance*
Southern Europe	12	89.2376	7.43647	18.9148
Northern Europe	10	87.1	8.71	7.85433
Eastern Europe	10	121.64	12.164	35.1934
Western Europe	7	53.5	7.64286	3.42952

As we see here, in the SUMMARY table, we can conclude that the geographic region where a person is born really affects his/her education. Looking at the averages, we see a similarity between Southern, Northern and Western Europe, but the difference is made by the Eastern Europe, which has an average rate significantly bigger than the others with approximately 50%.

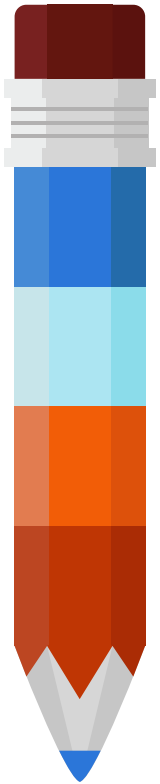
We made then a more ”in dept” analysis which we can see here:

Is the early school leaving rate affected by religion?

Islam	Catholic	Orthodox	Lutheran	Uaffiliated	Protestant
6.5	8.1	12.5	9.3	7.6	10.1
7.1	8.1	12.8	8.2	8.5	
26.7	2.2	11.5	14.8		
	8.0	3.8	7.2		
	12.1	9.9	9.9		
	5.0	3.6	7.7		
	13.1	5.7			
	5.6	15.6			
	8.2	5.6			
	12.6				
	7.0				
	5.4				
	8.9				
	7.6				
	4.1				
	16.0				
	4.0				
	10.9				

SUMMARY				
Groups	*Count*	*Sum*	*Average*	*Variance*
Islam	3	40.3856899	13.4618966	131.52241
Catholic	18	146.9	8.16111111	13.19545752
Orthodox	9	80.9402596	8.99336217	19.46510923
Lutheran	6	57.1	9.51666667	7.701666667
Uaffiliated	2	16.1	8.05	0.405
Protestant	1	10.1	10.1	#DIV/0!

As we see in this SUMMARY table, the averages are slightly different and the difference is made by Islam even though we see a similarity between the others, which are also different but not as much as the average of Islam is. The Islam average early leaving rate is approximately 30% bigger than the other religions rate.



We now came up to the conclusion that religion definitely matters when analysing the Early school leaving rate in Europe taking into account that Islam has the highest average mean and those unaffiliated states occupy the lowest Early school leaving rate in terms of average.

Simple linear regression

This time, for this analysis, we checked the relationship between exactly 2 variables, a dependent one, namely the inflation rate, and an independent one. For the independent variable, we chose 2 different situations, explained below. Finally, we checked the multi-collinearity between the independent variables to choose the best ones for the next analysis. Throughout this analysis we wanted to see if the inflation rate is influenced differently firstly by the labor force participation and secondly by the early school leaving rate in Europe.

1. *Labor force participation rate*

SUMMARY OUTPUT

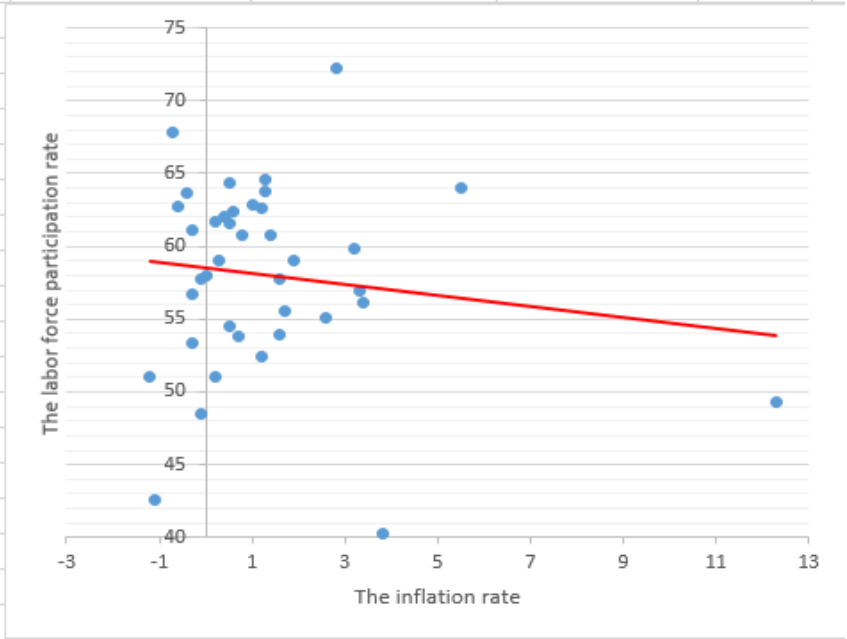
Regression Statistics	
Multiple R	0.134006304
R Square	0.01795769
Adjusted R Square	-0.008583994
Standard Error	2.331011274
Observations	39

ANOVA

	df	SS	MS	F	Significance F
Regression	1	3.676298217	3.676298	0.676584	0.416035102
Residual	37	201.0437018	5.433614		
Total	38	204.72			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	4.097749389	3.421742017	1.197562	0.238703	-2.835358497	11.0308573	-2.835358497	11.03085727
X Variable 1	-0.048246017	0.058654383	-0.82255	0.416035	-0.167091085	0.07059905	-0.167091085	0.070599052

The model is valid since the value of **significance F** is very low and also we can identify the same property for the **p-value** regarding the labor force participation rate. More than that, if we take a look at **R Square**, we can see that it translates to 1%, which means that there is a small correlation between the 2 variables and a negative one for sure.



This being said, <u>as the labor force rate increases, the inflation decreases</u>. On the other hand, taking a look at the adjusted ratio of determination is pretty low when we have a confidence interval of 95%.

*The **scatter plot** for this independent variable looks like this:*

2. Early school leaving rate

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.663850804
R Square	0.44069789
Adjusted R Square	0.425581617
Standard Error	1.759148503
Observations	39

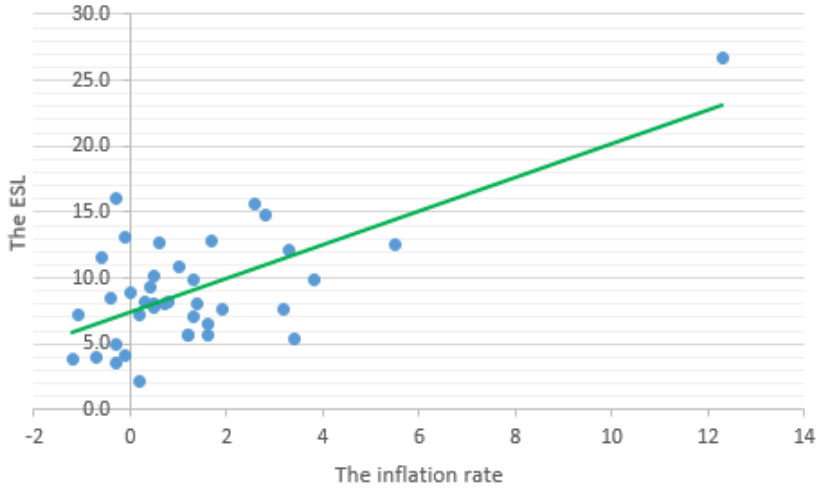
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	90.21967209	90.21967	29.15387	4.08426E-06
Residual	37	114.5003279	3.094603		
Total	38	204.72			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-1.807572087	0.640774069	-2.82092	0.007655	-3.105903676	-0.5092405	-3.105903676	-0.509240498
ESL 2020	0.344769174	0.063852853	5.399433	4.08E-06	0.215391004	0.47414734	0.215391004	0.474147345

Same as in the previous case, the **significance F** has a small value and **p-values** are also in our range. There is an increase in the value of **adjusted R squared**, ESL explaining the inflation in 42.55% of the cases in our sample and this assumption is valid in 95% of our observations in this specific sample.

The R Square coefficient is 0.44 which represents a strong linear and positive relationship between the ESL and the inflation rate.



This being said, as the early school leaving rate increases, the inflation increases as well.

Multiple regression analysis

For this specific analysis, we have decided to use the data we already have. We are going to use the GDP per capita and the labor force participation rate as the main *independent variables* to see the overall impact on our main interest character: the early school leaving rate.

For our *dummy variable*, having a consistent sample, we decided to use the inflation rate. Our population is formed of 39 observations: countries from all the regions from Europe. Below we can see the results:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.677743447
R Square	0.45933618
Adjusted R Square	0.412993567
Standard Error	3.395588996
Observations	39

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	342.8480043	114.2826681	9.911745359	7.13883E-05
Residual	35	403.550862	11.53002463		
Total	38	746.3988662			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	5.8916	5.3474	1.1018	0.2781	-4.964297391	16.74746802	-4.964297391	16.74746802
GDP per capita 2020	0.00002	0.0000	0.7577	0.4537	-4.10526E-05	8.99437E-05	-4.10526E-05	8.99437E-05
Labor force participation rate	0.00307	0.1005	0.0305	0.9758	-0.200908967	0.207039334	-0.200908967	0.207039334
Inflation rate 2020	1.32229	0.2428	5.4460	0.0000	0.829383578	1.815196151	0.829383578	1.815196151

We can write the equation for multiple regression as following:

$$y = 5.8916 + 0.00002 * x_1 + 0.00307 * x_2 + 1.32229 * x_3$$

where:

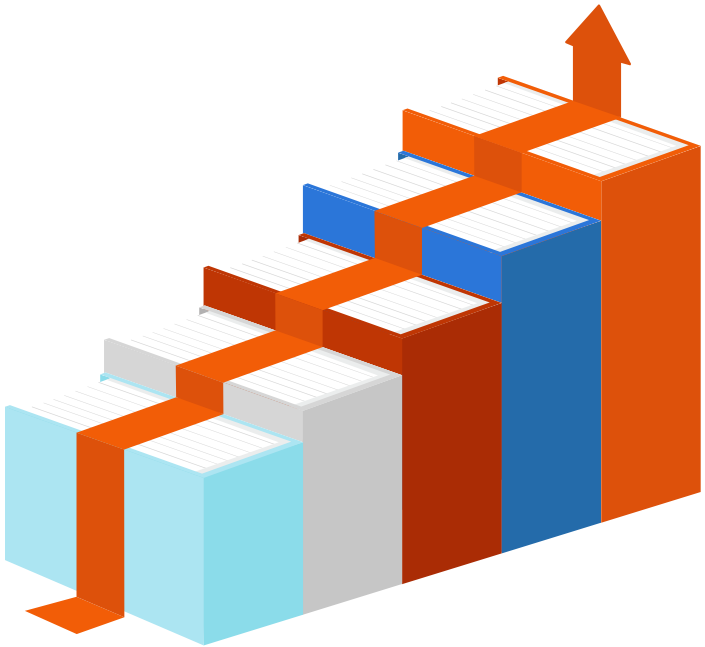
- y is the dependent variable, namely the early school leaving rate
- x1 is the first independent variable: GDP per capita
- x2 is the second independent variable: Labor force participation rate
- x3 is a dummy variable: Inflation rate

All the coefficients were taken from the table inserted above are marked with green.

CONCLUSIONS

This project was a journey for all of us for various reasons. A reason would be the fact that we learned a lot about our continent, about the countries cultural and economic backgrounds, which made us see our country in a slightly different position than before. Most of the study that we made is showing some deep rooted problems of some of the countries from Europe and a good economic structure for the other ones. The main thing is, all of the countries have this problem, even if it is a concerning one, or an insignificant one.

The most important achievement that we got from this project is knowledge and understanding some patterns, some statistical measures, understanding the reasons why people in Europe are drawn to leaving higher education. After this research we can say that education is extremely important and the fact that some countries are not investing in this sector is really painful. We hope, though to see some changes in this rates in the following years, because we will definitely look up for it.



REFERENCES

Bibliography in the field:

https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Early_leavers_from_education_and_training#Do_early_leavers_have_a_job.3F_Analysis_by_the_individual.E2.80.99s_labour_status
https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=edat_lfse_14&lang=en
https://ec.europa.eu/eurostat/databrowser/view/edat_lfso_00t3/default/table?lang=en
https://www.theglobaleconomy.com/rankings/Labor_force_participation/Europe/
https://www.theglobaleconomy.com/Turkey/Inflation/
https://www.theglobaleconomy.com/rankings/Inflation/
https://www.theglobaleconomy.com/rankings/GDP_per_capita_PPP/Europe/
https://www.theglobaleconomy.com/rankings/Economic_growth/
https://comtrade.un.org/

Recent bibliography:

https://www.frontiersin.org/articles/10.3389/fpsyg.2016.01993/full
https://www.tandfonline.com/doi/abs/10.1080/10564934.2015.1098265
https://www.ingentaconnect.com/content/intellect/pjss/2014/00000013/00000001/art00002
https://onlinelibrary.wiley.com/doi/abs/10.1111/ejed.12034