Project overview

For this project, we are divided into groups. These groups consist of students from different disciplines. For instance, we have two computer scientists, one sociology student, and one media, arts, design and architecture student. One of our goals in this project is to learn how to collaborate with other disciplines and try to minimize misunderstandings. Additionally, our aim was to learn how to conduct data-driven research in a structured manner.

To achieve this goal, we discussed our strengths and weaknesses and looked at our differences. Then we studied our dataset and consulted each other about what we found interesting to investigate for our research question. Initially, we chose the research question: *is there a correlation between the unemployment rate of East-Africa and the number of pirate attacks in nearby waters from 2015-2020?*

However, while working with the data, we found that we were missing information about the unemployment rate in 2020. As a result, it was not possible to investigate the unemployment rate in 2020, and we had to adjust our research question to: *is there a correlation between the unemployment rate of East Africa and the number of pirate attacks in nearby waters from 2010-2019.*

After formulating our research question, we discussed what our thesis statement should be. Our thesis statement for our research is as follows: *this study will explore the correlation between unemployment rates in East African countries and the number of maritime piracy attacks in nearby waters between 2010 and 2019. A comparison of the data might begin to reveal the relationship between unemployment and the resurgence of piracy. As a means of survival. The findings are meant to demonstrate how economic instability and limited employment opportunities fuel piracy, highlighting the critical role of socio-economic conditions in maritime security dynamics in East Africa.*

We found it interesting to investigate this because it can be very relevant to study in relation to the occurrence of pirate attacks. This can tell us something about why pirate attacks occur and, if so, how we can lower the attack rate. Since we were interested in finding a connection between pirate attacks and the unemployment rate, we started with a literature review.

For this, we read various studies that could provide insights into unemployment in East African countries and how the unemployment rate in a country can affect crime.

Research shows that unemployment can lead to higher crime rates. This can have several reasons. Individuals may face financial hardships, which may lead them to commit crimes. Unemployment can also exacerbate social inequalities, which can lead to criminal behavior. Reduced opportunities or psychological factors can also push people toward illegal activities (Jawadi et al., 2021).

This is also confirmed in the work of Mora et al. In their article it was found that a 1% increase in unemployment corresponds to a 0.92% increase in the crime index (Mora et al., 2014).

In addition to Jawadi et al. and Mora et al., Okeahalam and Otowombe also discuss this. In their article, they state that areas with low economic development and high unemployment are often more vulnerable to piracy (Okeahalam & Otwombe, 2016). Due to this low economic development and the poverty that prevails in East African countries, an environment is created in which people are more likely to engage in piracy as a means of making a living.

Research into the correlation between unemployment and the number of pirate attacks is currently relevant because unemployment is rising in east Africa, particularly in countries like Kenya. It is also expected that the unemployment rate will continue to rie if no effective measures are taken to address the underlying causes. The expectation is that the number of unemployed will increase by approximately 1 million per year (Mathenge, 2021). Consequently, economic development is unlikely to improve.

Given this expected increase, it is important to investigate whether there is a casual relationship between unemployment and pirate attacks, so that the rise in the number of pirate attacks can be taken into account and a way can be found to mitigate this problem.

Data acquisition

There are three datasets that this team used. One is the country codes, since that is the way the author of the datasets tried to shorten the country names by having them take on a 3 letter name. The other is pirate attacks, which showcase the date of a pirate attack, and surrounding info, as in what vessel was attacked, what the type of attack was, the exclusive economic zone in which it took place, the nearest coastline and the coordinates of the attack. The 3rd dataset consisted of country indicators, which talk about the unemployment rate, general population, amount of fisheries, corruption index, homicide rate and other things. The author retrieved this data about the pirate attacks from the International maritime bureau and the country indicator dataset was retrieved from the World Bank.

The datasets are quite large but for the research question at hand the preprocessing has been made easy by using the filter function in excel. The first step was identifying the countries that can be considered East-African, several sources can be used for this. Next up is checking which countries are actually named in the country code table, if they were not then we chose not to fill

the data out since there might be a bias of like separatist states. Then we filtered only for these countries in the pirate attack dataset and the year that we wished to research 2010-2019. The date is made 2010-2019 since the country indicators dataset went on till 2019 while pirate attacks showcased the pirate attacks that happened in 2020. That leaves us with the East-African countries that we know have had pirate attacks in this period. The only downside is that the Seychelles had NA in its unemployment rate, therefore not having any data. Which means that we have to enrich it, we did it by using the IMF, the most reputable source we could find. The question of unemployment in the Seychelles is very niche and only a few sources are available and they all show different results from one another, therefore the choice was made to take the values given by the most reputable source. This completes the data set for our further visualization and computation needs.


Methodology

The main and only tool used for the data visualization was google spreadsheets, it has identical capabilities of excel. The choice to just stick to this was convenience and efficiency. The choices were either remain with the excel like program or code in r studio or python. However we thought that it would be convenient to have the dataset and the work together in one place, in case we need to change stuff in the dataset further along the line, as we actually have done. The reasoning for efficiency is that with code it will take a longer period for data visualization and will be looking off or less standardized than expected. This of course is subjective but we thought in order to achieve the same result our team would lose more time coding than using pre existing methods from google spreadsheets to have the visualization. So now to handle the graphs chronologically, the first graph will now be explained and justified.
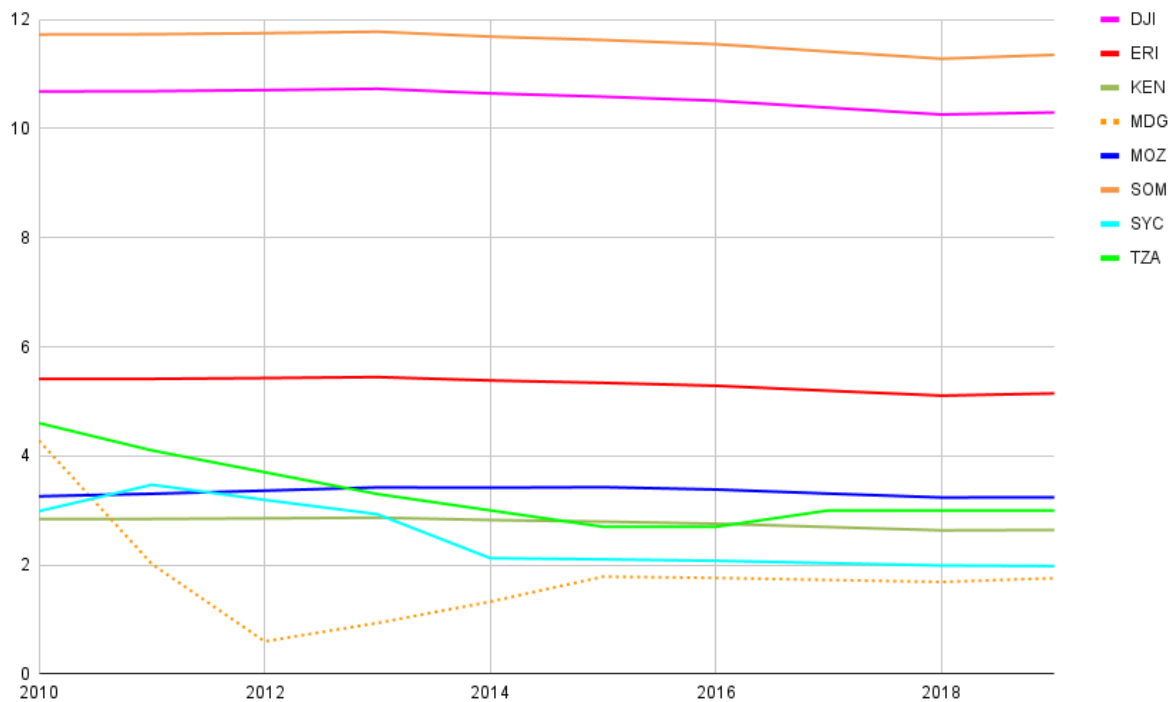
The x-axis is the years we wish to research and this graph serves more of an introductory figure. The y-axis shows the percentage of the unemployment rate. This graph depicts how some countries will affect the average more than the others like Djibouti and Somalia for example. For other digital humanists it would be interesting to find out why this would be the cause, whether it might be due to political reasons or other conditions. The graph has been made using lines, in order to show the chronological story of unemployment of the countries we wished to research.
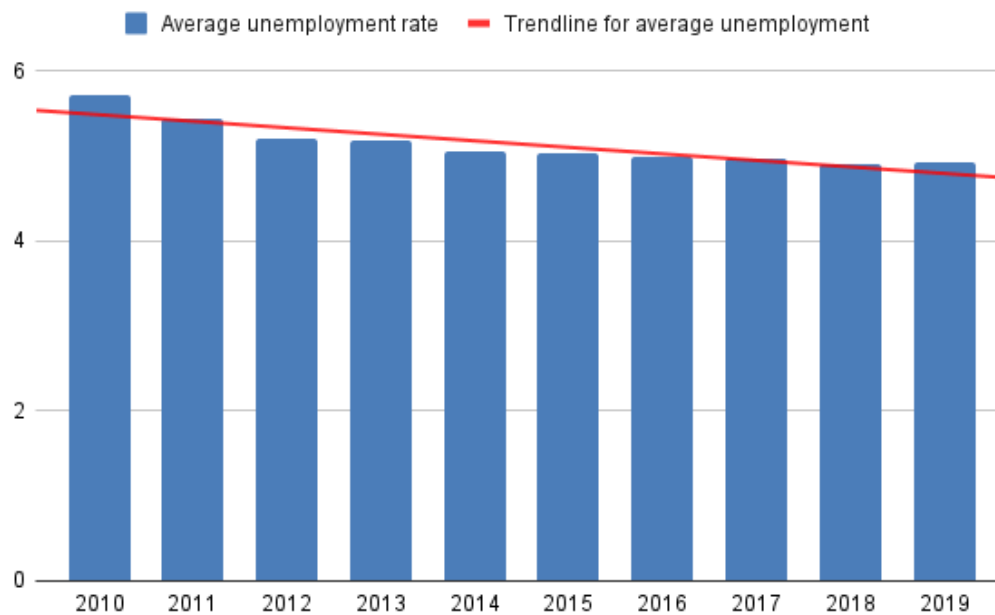
The second graph depicts the countries as a unit, as a whole "East-Africa" which gets affected by the pirate attacks. The graph was achieved by taking the unemployment rates of all countries and the years and selecting "average" function in the plotting menu for google spreadsheet. Again it is visible how Djibouti and Somalia have in general caused a higher than expected unemployment rate. The y-axis shows the average unemployment rate in percentages and the x-axis shows the years. The choice to have this be bar graphs is because it now is nicer to visualize the unemployment as a whole, and showing a trendline to see what is going on and how it might shift in the future. Which is important for the relevancy of this research question that we chose, since if there is a correlation between unemployment and piracy, whether or not both will go down with one another. The choice for the colors was to go with the default color scheme given. Trying to stay uniform with all the bars, so no unnecessary implications will arise with the readers. The trendline being red, is just for the transparency and visibility of what way we are headed.
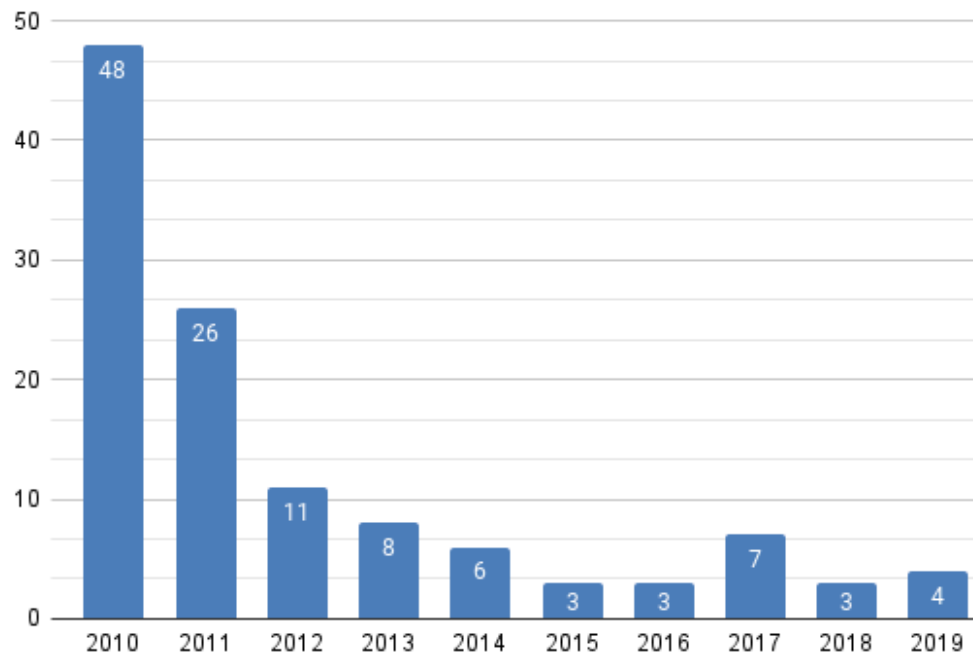
The third graph shows the total number of pirate attacks on all countries in "East-Africa". This was achieved by collecting the data of the pirate attacks and the years and selecting the "total" option in the plotting menu of google spreadsheets. The y-axis shows the actual number of attacks and the x-axis is just the years again. The choice to include the numbers in the bar is for better visibility, since the data values are very different from one another like in 2010 vs 2018. The numbers help the unnecessary need for the reader to scale the bars to assume how big the bar actually is. The numbers were chosen as a total since it's discrete data, it is binary so it happens or it doesn't happen.

average correlation of unemployment rate and pirate attacks

-0.0285*x + 0.337 R² = 0.19
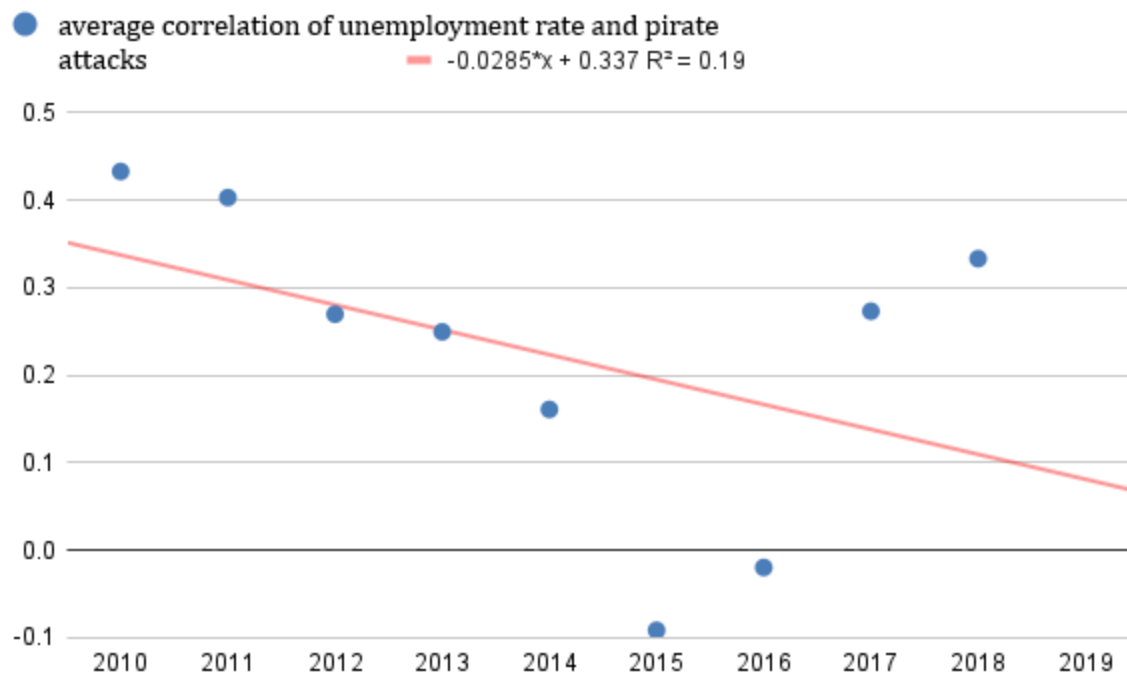
The third graph is almost finalizing the value we want to find. The y-axis plots the average correlation of unemployment and amount of attacks. This was achieved by making a column in the country indicator dataset and counting the pirate attacks and manually filling them out. The counting could be automated by using the pivot table, and creating a column in the pivot table for the counta, this shows how many countries got attacked in each year, making it easier to manually fill out. Our team could not figure out how to automate the process so that could have gone better. Then another column was made and the function =correl(), was used. The values to plug into here were the unemployment rate and the pirate attack amount, it will calculate the Pearson correlation for all the data points. Then the graph was made into scatter plots, to more accurately show how well the data aligns with the trend. So the points were achieved by using the average function in the plot menu, and then we decided to run a trendline through here. Since we wish to do further calculations on the R^2. to get the actual correlation. The one issue is though we compare continuous data points to discrete data points, so to resolve that issue for consistency we consider the pirate attacks from now on to be continuous data points.
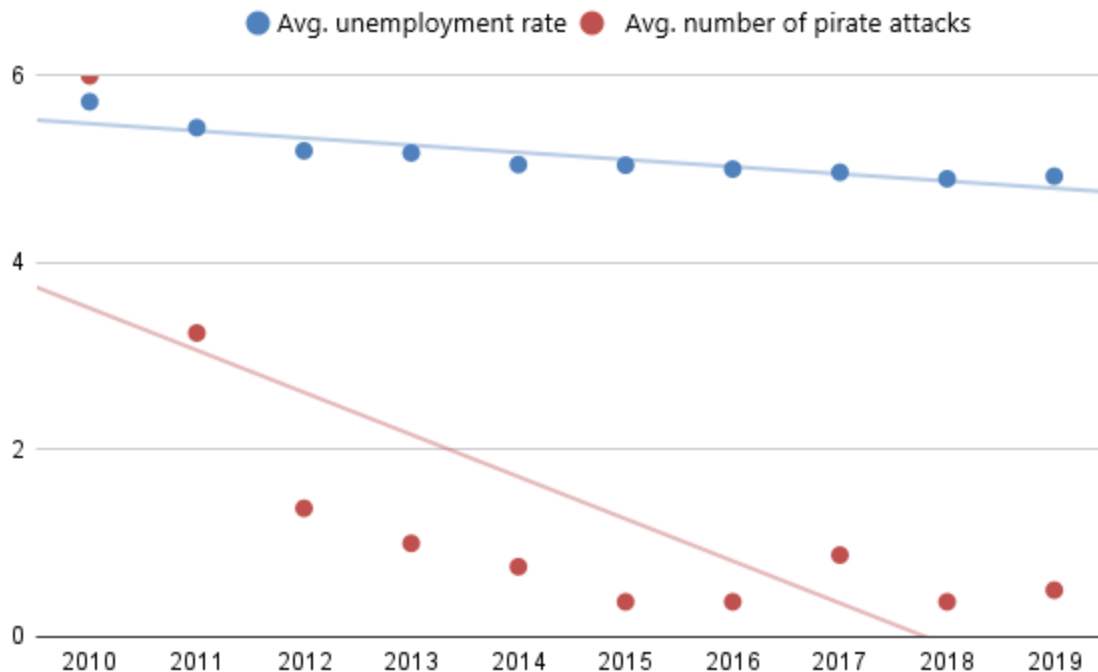
The fifth and final graph is more of a visual graph as a source just to check. Since our team has the R^2 value and it is non-zero. Meaning there is a correlation, but the question is what kind of correlation is it. Since we get the actual correlation by taking the square root of R^2, this could mean that the square root is either positive or negative. So in order to see which it would be, would mean to see what the trends look like for both average unemployment rate and average pirate attack amount, as aforementioned for consistency we treat the pirate attacks as continuous therefore allowing us to take the average. Furthermore this graph should just get the representation of what direction both trends are going in order to make an assumption about the Pearson correlation coefficient. The x-axis shows the years and the y-axis this time is quite an interesting concept as it means different things for the different data points. For the unemployment rate it is percentages and for avg number of pirate attacks it's just the number of attacks. The choices for blue and red were made just because it gives visibility and won't confuse with the trendlines. The most important thing to note though is that unemployment went down and the number of pirate attacks went down. This means that there is a positive correlation.

Now for calculation it is simple, it is just taking the square root of 0.19 as was shown to be the R^2 value in the 4th graph. So this gives a value of 0.4358898944, which can be rounded off to 0.44 since for the categorization of correlation, it is usually done either with just one or two decimals usually. 0.44 indicates a moderate positive correlation between piracy and unemployment(Tutik Sri Wahyuni & Purwanto, 2020)(Statistics Solutions, 2024b). This means that there is some evidence that the one moderately affects the other. Meaning that according to the dataset, if piracy lowers or unemployment the other will moderately lower too. This would be a very positive outlook on East-Africa as the trends for both of these negative statistics are downwards.

Work Packages:

Daniela did the literature study for the project. Also she was responsible for close collaboration with the team to set the research question and hypothesis.

Yevgeniy did the preprocessing of the data and the visualization and computation of correlation coefficient.

Stefanos is responsible for the project workflow which in order help us gain structuralisation in our work.

Yuki did the deepdive into certain aspects of our research, like ethical considerations, biases and the reflection.

Workflow Steps

Description of Software and Tools Used at Each Stage:
For this project, we utilized Google Sheets to manage the dataset and create visualizations. This tool facilitated data organization and allowed for immediate visualization updates, crucial for ensuring the accuracy of our analyses.

Key Decisions and Rationale:

- **Filtering the Data:** We focused on East African countries by identifying relevant countries from the dataset provided to us.
- **Time Frame Selection:** Although our initial plan was to analyze data from 2010 to 2020, we limited our analysis to 2019 due to missing unemployment data for 2020 in the dataset, which represents a limitation.
- **Country Inclusion/Exclusion:** We excluded landlocked countries that could not logically experience pirate attacks, focusing on those with access to the sea and keeping the Exclusive Economic Zone (EEZ) as a reference.
- **Data Completion:** For Seychelles, where unemployment data was missing in the provided dataset, we manually searched for reliable unemployment rates from trusted sources to fill the gap.
- **Graph Creation:** We visualized the data in several ways:

1. The unemployment rate for each country, to better understand regional differences.
2. The average unemployment rate across all countries, as it directly addresses our research question.
3. The total number of pirate attacks, calculated to later find any correlations.
4. Graphing the correlation between average unemployment rates and the total number of pirate attacks. This produced an R^2 value, which was then used to calculate the correlation coefficient.
5. Finally, we visualized whether the correlation between unemployment and pirate attacks was positive or negative, confirming that the value was not zero, indicating some level of correlation.

Alternative Approaches Considered but Not Chosen:
We considered using Python or R for this analysis. However, we opted for Google Sheets to have both the data and calculations in one accessible location. This allowed us to quickly visualize and verify any changes or issues, which was more efficient than using separate software for analysis and data management.

Challenges and Solutions

Challenge 1: Missing unemployment data for Seychelles in the provided dataset.
Solution: We manually searched reliable sources to fill in the missing data.

Challenge 2: Difficulty exporting data on pirate attacks from one dataset to another.
Solution: The data was manually transferred.

Challenge 3: More work due to a small team and team members leaving.
Solution: The remaining team took on additional responsibilities and worked harder to complete the project.

Results

The unemployment data and the pirate attacks are represented in different ways. The unemployment data is continuous in nature, while pirate attacks are sporadic events.

Our data analysis includes eight East African countries. The analysis shows that the average unemployment rate of these countries has a declining trend over the years, highlighting economic improvement in the regions.

To check for a correlation between pirate attacks and unemployment, we conducted a correlation analysis using the Pearson correlation coefficient. The result yielded an R^2 value of 0.19. An R^2 value of 0.19 indicates a moderate positive correlation between the variables.

In addition to the R^2 value of 0.19, we also calculated the square root of R^2, which is 0.44. This further confirms that there is a moderate positive correlation.

This moderate correlation is also reflected in the trend lines. These trend lines show that as unemployment decreases, the number of pirate attacks also decreases. This suggests that unemployment may influence pirate attacks, but it is likely not the only factor affecting the number of attacks.

What is noteworthy about our results is that they do not fully align with our literature research. The literature indicates that there is an upward trend in the unemployment rate in East African countries (Mathenge, 2021), while our research shows a slight decrease.

However, the correlation we found in our research is also reflected in the literature. It was indicated in the literature that a 1% increase in unemployment is associated with a 0.92% increase in the crime index (Mora et al., 2014)4. This thus provides confirmation for the correlation we found.

This correlation also supports our thesis statement, as it suggests that there is indeed a relationship between the two factors: unemployment and pirate attacks. This supports the assertion that economic instability can force individuals to consider illegal activities, such as pirate attacks.

Ethical considerations

In our project, we encountered several factors that could lead to bias. Firstly, we had to deal with the limitations of the dataset itself. We found gaps in the data and we found a lack of transparency regarding the extent to which employment in informal economies, seasonal work, and underemployment had been accounted for. These factors could greatly impact the true nature of unemployment in these regions and thus distort the accuracy of the employment rate. These data gaps need to be addressed, or else the analysis may give a misleading representation of the unemployment trends.This incomplete representation can lead to inaccurate conclusions. As we worked with pre-selected datasets, we did occasionally decide to make alterations where possible, but we primarily chose to identify and articulate limitations for transparency purposes.

Secondly, Our thesis statement assumes that analyzing our data might reveal a relationship between piracy and unemployment. It is, however, important to acknowledge that our methods search for correlation, not causation. This means that when we find a relationship, we still can't claim that unemployment is a cause of piracy, as that would be a misleading interpretation of the data. Other factors likely play significant roles in influencing both unemployment and piracy rates. For this reason our findings should be interpreted as identifying patterns or associations that require further investigation before drawing any further conclusions. As we would like to avoid:

- Ineffective policy
- Linking unemployment directly to crime, which could lead to the stigmatization of individuals
- Overlooking systemic causes of unemployment

Reflections

Going through the process of completing this project has taught us a lot. Given the constraints of the assignment, we made conscious choices about how to approach the dataset and how to handle its limitations. Our research, however, could still greatly benefit from a few rounds of revisions. This became especially clear when we were confronted with the fact that our research question seems to pose a challenge that our current methodology cannot fully address yet. If we had had more time, we could have started exploring statistical methods that potentially could have made our research slightly more robust. For example, we could have looked into correlation methods that account for potential time lags, or we could have considered methods that calculate partial correlation. This might be desirable here, as including additional socio-economic variables such as the level of education, income inequality, or regional conflict could offer more insights into the drivers of piracy and help clarify the specific role unemployment plays within this complex network, without claiming causality. Along with this, our research could have been improved by embracing the interdisciplinary nature of the field by drawing more from the humanities to provide context for the gathered data. Adding qualitative methods, such as interviews or case studies, could provide context and deeper insights into the dynamics of unemployment and piracy. Additionally, if unemployment data for 2020 and years onward becomes available, it would be sensible to extend the analysis to include this period. To conclude, at the end of our project, the main takeaway is that the research process is never linear, and every choice made throughout requires some degree of compromise.

Documentation and Sustainability

The data can be accessed on GitHub for both transparency and future research. Version control within the repository can be used to track updates. Hopefully, the detailed documentation of the workflow allows for the repetition of the research, and notes will be made if the data is updated.

link:

BIBLIOGRAPHY:

- Anna Triantafillou, Ioanna Bardaka, Ioannis Vrettakos & George
Zombanakis (2023) Maritime piracy: Determining factors and the role of
deterrence, African
Security Review, 32:2, 166-183, DOI: 10.1080/10246029.2023.2179411

- *IMF Search Hub*. (2024, April 24). IMF.

  https://www.imf.org/en/Search#q=seychelles%20unemployment%20rate&sort=relevancy

- Jawadi, F., Mallick, S. K., Cheffou, A. I., & Augustine, A. (2021). Does higher

  unemployment lead to greater criminality? Revisiting the debate over the business cycle.

  Journal Of Economic Behavior & Organization, *182*, 448–471.

  https://doi.org/10.1016/j.jebo.2019.03.025

- Mathenge, G. S. (2021b). Determinants of Unemployment in East African Community

  Countries. *Account And Financial Management Journal*, *06*(10).

  https://doi.org/10.47191/afmj/v6i10.06

- Mora, J.J, Sandova, L.E., & Riano, L.C.(2024). Spatial relationship between

  UnemploymentImmigration and criminality in developing city. *Latin American Economic*

  Review,33(5).

- Okeahalam, C., & Otwombe, K. (2016). Socioeconomic development and the risk of

  maritime piracy. *Journal Of Transportation Security, 9*(3-4), 125-160.

  https://doi.org/10.1007/s12198-016-0171-4

- Statistics Solutions. (2024, 17 april). *Understanding Pearson's Correlation Coefficient: A*

  Guide to Statistical Association.

  https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/pearsons-correlation-coefficient/

- Tutik Sri Wahyuni, & Purwanto, K. K. (2020). *Students' conceptual understanding on acid-base titration and its relationship with drawing skills on a...* ResearchGate; IOP Publishing. https://www.researchgate.net/publication/338505421_Students