Bridging Gaps for Inclusive Development in Peruvian Districts

Digital Portfolio Report

Critical Data Representation and Analysis
MSc Data Science & Artificial Intelligence
University of the Arts London

Santiago Won Siu
March 2024

Table of Contents

I. INTRODUCTION 3

II. PROJECT RESEARCH AND STATISTICAL ANALYSIS 4

1. Research sources 4
2. Scope of the project 5
3. Analysis variables scheme – Study Design 6
4. Statistical Calculations and Analysis – Study Design 7
5. Evaluation: Results and Findings 8
6. Conclusions 15

III. DATA SCIENCE METHODOLOGY, QUALITY AND ETHICS 16

9. Data Exploration, Evaluating and Cleansing 16
10. Data Quality and Ethics (bias, ethics, limitations, considerations) 17
11. Data Visualization / UX Data Visualization and Website 18
12. Future additional ideas to complement anlaysis 18

IV. TOWARDS A DIGITAL EDUCATION PROPOSAL 19

13. Context 19
14. Challenge and Idea Proposal: 19
15. Benchmark Analysis – Opportunity for Value Proposal 20
16. Key Model Features Ideas: 22
17. Data Ethics on the Project to consider 22
18. KPIs on Ethics of the project 22

V. SKETCHBOOKS' REFERENCE GLOSSARY 23

VI. BIBLIOGRAPHY / DATASET REFERENCES 23

I. INTRODUCTION

As a Peruvian citizen, I am deeply concerned about the pervasive issue of inequality within our country. Peru comprises 24 departments and approximately 1,800 districts withing those, many of which face significant challenges in areas such as poverty, education, discrimination, and money laundering.

The root of these problems lies in the centralized nature of our economy, which disproportionately favors the capital city, leaving other regions overlooked and under-resourced. Consequently, there is a lack of comprehensive understanding and analysis of the data scattered across various sources, hindering efforts to address these pressing social issues.

This project aims to address this gap through a two-part approach:

1. Data Analysis: We will undertake a thorough examination of social variables across all 1,800 districts to gain insights into their unique challenges and vulnerabilities. By clustering districts based on these factors, we can identify those most in need of targeted interventions, regardless of their geographical location.

2. Project Exploration: Recognizing the pivotal role of education in driving social development, we will explore potential solutions to address the identified issues. This includes investigating the feasibility and impact of digital education initiatives and developing a robust business model to support these efforts.

Throughout this project, we will leverage a variety of datasets, including demographic, financial, connectivity, and educational information. By employing statistical analysis techniques and data visualization tools, we aim to uncover meaningful correlations and insights that can inform evidence-based policymaking and/or the ideal focus on projects to entrepreneur.

Our findings underscore the importance of prioritizing teacher training, addressing educational disparities, and targeting interventions based on vulnerability rather than geographical considerations. Additionally, we highlight the potential of technology, particularly cellphone ownership, in reaching marginalized communities and reducing barriers to education.

By adopting innovative strategies and reevaluating traditional approaches, Peru can make significant strides towards lasting poverty alleviation and inclusive development.

Potential audience for this project includes government institutions, investors interested in the education industry, entrepreneurs in education and EdTech startups in developing countries.

II. Project research and Statistical Analysis

1. Research sources

For this project, several sources served as input regarding data, theory for analysis and inspiration.

Datasets

For main data on demographics, financials, connectivity, and education (18 datasets) the System of District Information for Public Management of the National Institute of Statistics and Information of Peru was key (INEI, Sistema de Informacion Distrital para la Gestion Publica, 2024).

For money laundry and discrimination, the Integrated System of Statistics for Criminality and Citizen Safety of INEI was referred (6 datasets) (INEI, 2022).

For geographical data, towards understand the surface area of each district and its location we leverage a GeoJSON file, and a GitHub repository elaborated by Juan Eladio Sanchez Rosas (Sanchez, 2016) which at the same time leveraged information from INEI.

For agricultural and topology data, as a reference to analyze potential topological complexity by district, we leveraged the dataset by Diletta Topazio at the Hand in Hand Initiative in the Food and Agriculture Organization of the United Nations (Topazio, 2022).

2. Scope of the project

The first part of the project tries to assess all Peruvian districts (~1,800) to gather a more holistically and with deeper understanding under the variables of interest.

The current **ratios**, **correlations,** and **estimated effect / impact** between society / district's performance (e.g., district's income, poverty, illiterate population) by context (e.g., demographic variables) and resources (e.g., educational and telecommunications / connectivity).

The **evident relevance on education** over society performance is also deeply analyzed and compared towards all the other variables, specifically trying to validate the strongest correlation towards social variables.

Through the previous analysis, I was able to classify the **most vulnerable and disadvantaged districts** and **run a clustering k-means from the SKLearn library** to analyze new clusters based on the evaluated variables.

Also, other additional analysis was run while data was processed and analyzed as several intuitions on the performance after analysis were raised to potential branches of analysis towards understanding the districts (e.g., towards eventually being able to analyze corruption patterns since it is one of the main variables affecting development) (Fhima, 2023)

Research question:

> *how can we best analyze and re-classify the 1,800 districts in Peru to better prioritize and tackle current social gaps and issues?*

The second part of the project tries to bring inspiration and feasibility potential on a specific area of solution for the societal issues encountered previously. First it tries to explain the **potential of digital education** across Peruvian districts and implications / assumptions backed by research papers. Then, a **benchmark analysis** on remote / digital education and success cases is made. Finally, **some exploration of potential features and ethical considerations with KPIs.**

3. Analysis variables scheme – Study Design

After reviewing around different datasets, the most interesting datasets were included and classified by how the relation should be analyzed according to dependency.

Analyzed Variables

- **Context per district**
  - Demographics - Population
  - Geographics – surface in KM2
  - Agricultural Taxonomy – (based on agricultural potential*)
  - Type of district – rural vs urban

- **Vulnerability** / Inequality Analysis (context ~ (resources ~ resources))
  - Educational resources (educational centres, teachers, years of education)
  - Telecommunications resources (internet, cellphone)

- **Society / District Performance**
  - Poverty
  - Illiterate population (%, gender %)
  - Money Laundry
  - Discrimination
  - District's municipality Income

4. Statistical Calculations and Analysis – Study Design

i. **Calculations to identify ratios, correlations, and potential clusters**

All analysis were ran on Python by leveraging different libraries such as pandas, numpy, matplotlib, scipy, seaborn, sklearn and geopandas. To refer the workbook in Jupyter Notebook format see Workbook_1.ipynb. Analysis can be found in comment by the alphabetical index expressed below.

- Relations Analysis (Ratio)
    a. Educational Centres / inhabitants
    b. Educational Centres / km2
    c. Population density
    d. Teachers / 100 inhabitants
    e. Teachers / Educational Centres
    f. District's municipality monthly Income / Population (i.e., per capita)
    g. Cellphone and internet user population (%)

- Negative Effect Analysis by context (Correlation)
    h. Illiterate population AND (Agricultural Taxonomy)
    i. Poverty (%) AND Agricultural Taxonomy

- Negative Effect Analysis by resources (Correlation)
    j. Poverty (%) AND District's Municipality Income per Capita
    k. Poverty (%) AND (#Education Centres / …)
    l. Illiterate population AND (# Education Centres / …)
    m. Illiterate population AND Teachers
    n. Money Laundry AND (# Education Centres / …)
    o. Discrimination (%) AND (# Education Centres / …)
    p. Discrimination (%) AND (Telecommunications Resources/population)

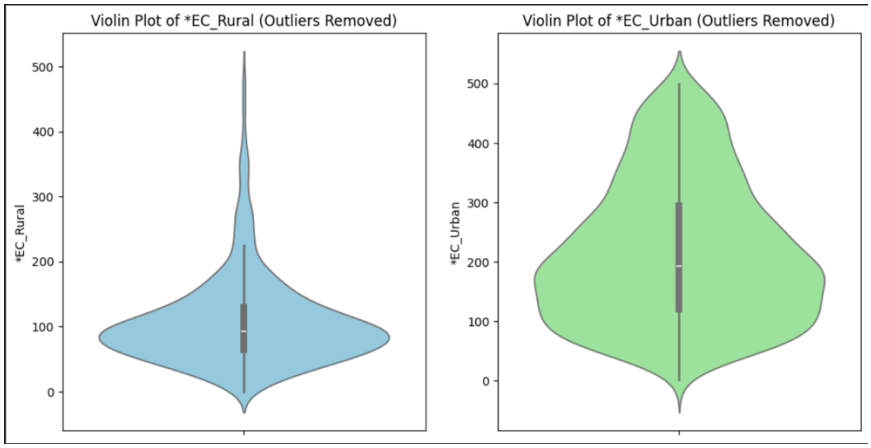
- Positive Effect Analysis
    q. District's Municipality Income AND (#Education Centres / …)
    r. District's Municipality Income AND Telecommunications Resources (*potential bias)

5. Evaluation: Results and Findings

Following up, we will show some of the main points from the previously mentioned analysis.

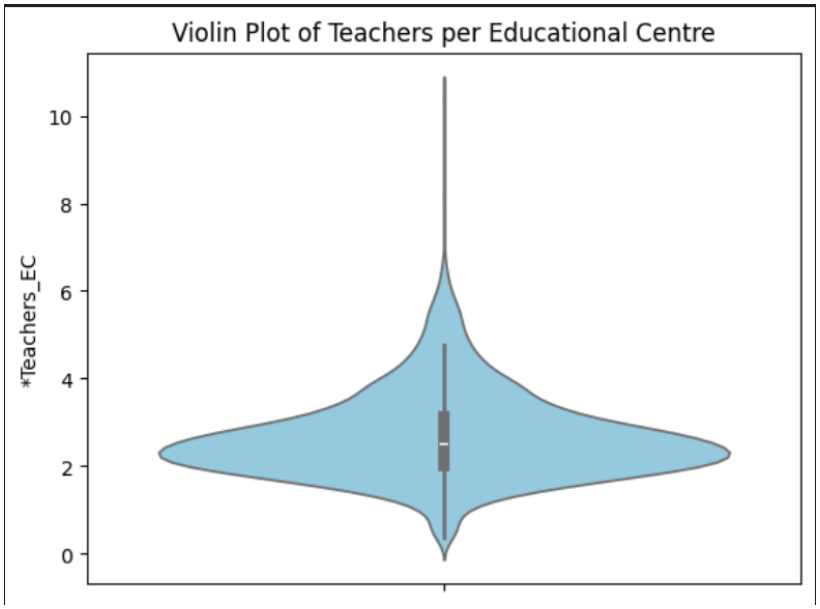
- **Regarding the number of inhabitants per Educational Centres**

  Through a violin plot, I was able to identify that not necessarily the number of educational centres per inhabitant is an indicator of quality, since clearly, urban areas have more inhabitants per Educational Centre, but at the same time they still have lower levels of poverty (as shown on later analysis results).



- **Regarding the number of teachers per Educational Centre**

  This specific ratio analysis is very interesting, because since the number of Educational Centres can be limited and the quality of education relies on the teachers, then it is important to understand if that reliance is somehow diversified or concentrated.



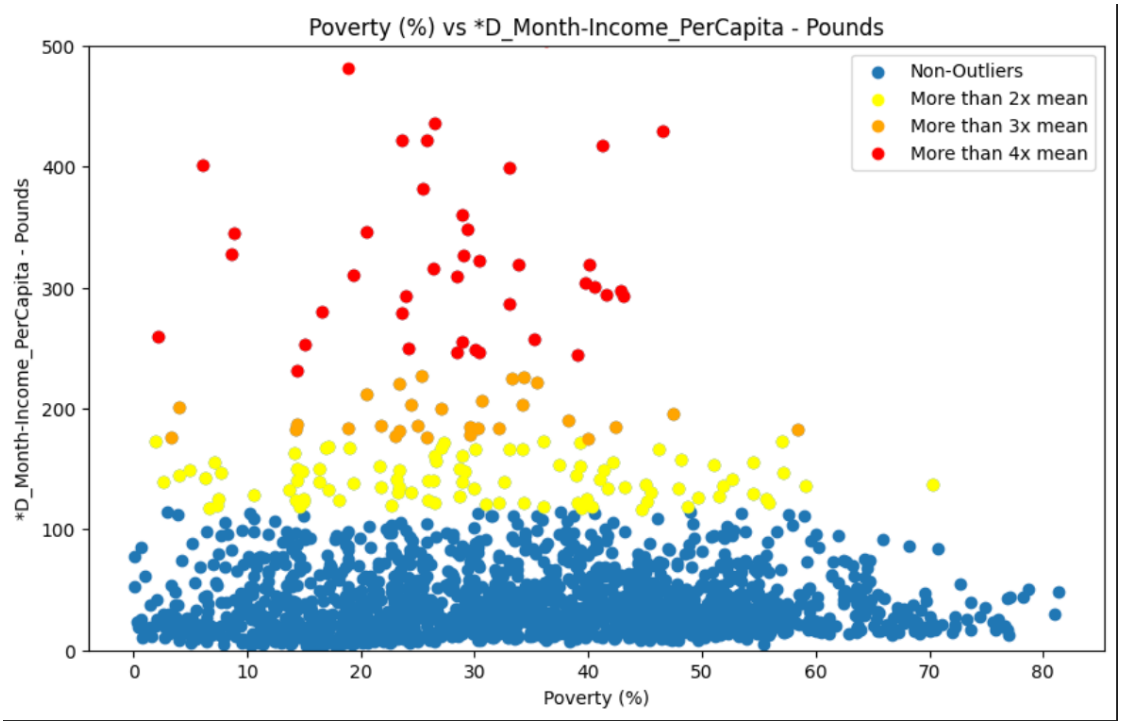
Considering that the mean of teachers per educational centre is 2.7 there is a high reliance and concentration of the education of a town. There are 385 districts with less that 2.7 teachers per education centre which involves around 11 Mn inhabitants (~30% of the total population of Peru).

This analysis results raises even more concern as Peruvian news highlight that over 130 thousand teachers (58%) do not have a real “teacher degree” or any other “higher education” (Comercio, 2020).

- **Regarding the monthly income that each District’s Government institution earns and finding the per capita value in contrast with poverty**

  An analysis was made on how much “money” a public institution has, with the ideal that the government’s earnings should be in favor of the society, we try to analyze if there is a linear relation between less percentage of a population living in poverty and those earnings.

  After running a Pearson correlation analysis, the is a statistically significant negative correlation meaning that poverty percentage increases when there is less income per capita (-0.105). However, when looking at the data behavior, the effect on income per capita is not strictly visible, with significant inexplicable outliers. Therefore, it would be relevant to analyze them deeply since those could be classified even more as with inequality within the same district. So, causes could be explored, such as corruption.



- **Regarding the behavior and spread relation towards poverty (%) in districts between analysis on "Teachers per Education Centre" and "Education Centres per squared kilometers"**

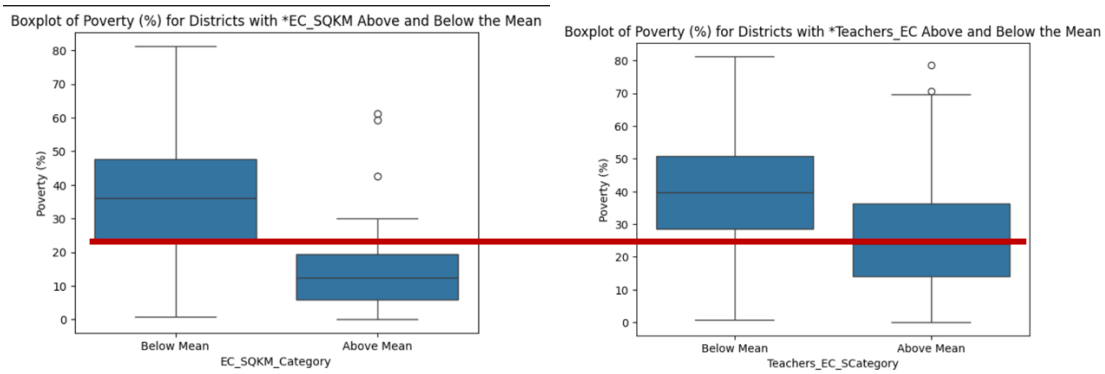
  When analyzing mean boxplot of teachers per EC or EC per km2, we see that the group of districts that perform with a below average numbers of Teachers per EC range has a smaller percentage of poverty and a higher minimum in the range.

  Therefore, we could establish a hypothesis of study that increasing the number of teachers per EC could be even more beneficial towards poverty that building more ECs per KM2 (which has a higher cost) … although we know it is not causality, but a correlation, it is worth exploring.

  Potential research questions:

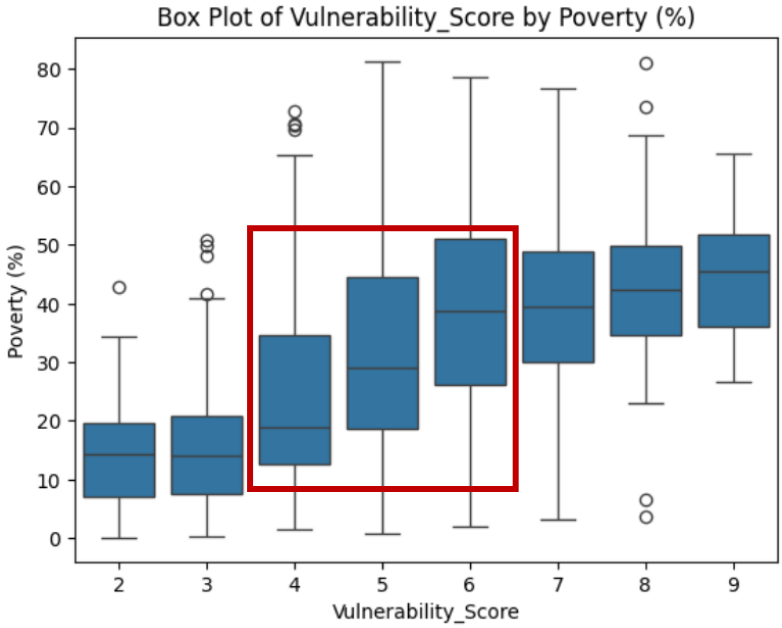
  1) Should government prioritize raising teachers' salaries and forming teachers than bulding new schools?

  2) Should customized learning and diversification of human teachers be the answer for development?



  The red line highlight the minimum in the range of poverty.

- **A vulnerability analysis trying to integrate most of the variables was made considering educational centres, illiteracy, discrimination, money laundry and poverty ratios**



As identified, there were 9 categories summing 1 if the district was considered in a lower third of performance regarding other districts. When we analyze how the frequency of being in a vulnerable state regarding different variables of analysis, we can observe that the most effect on the poverty mean lies when growing from 4 to 6 variables of vulnerability.

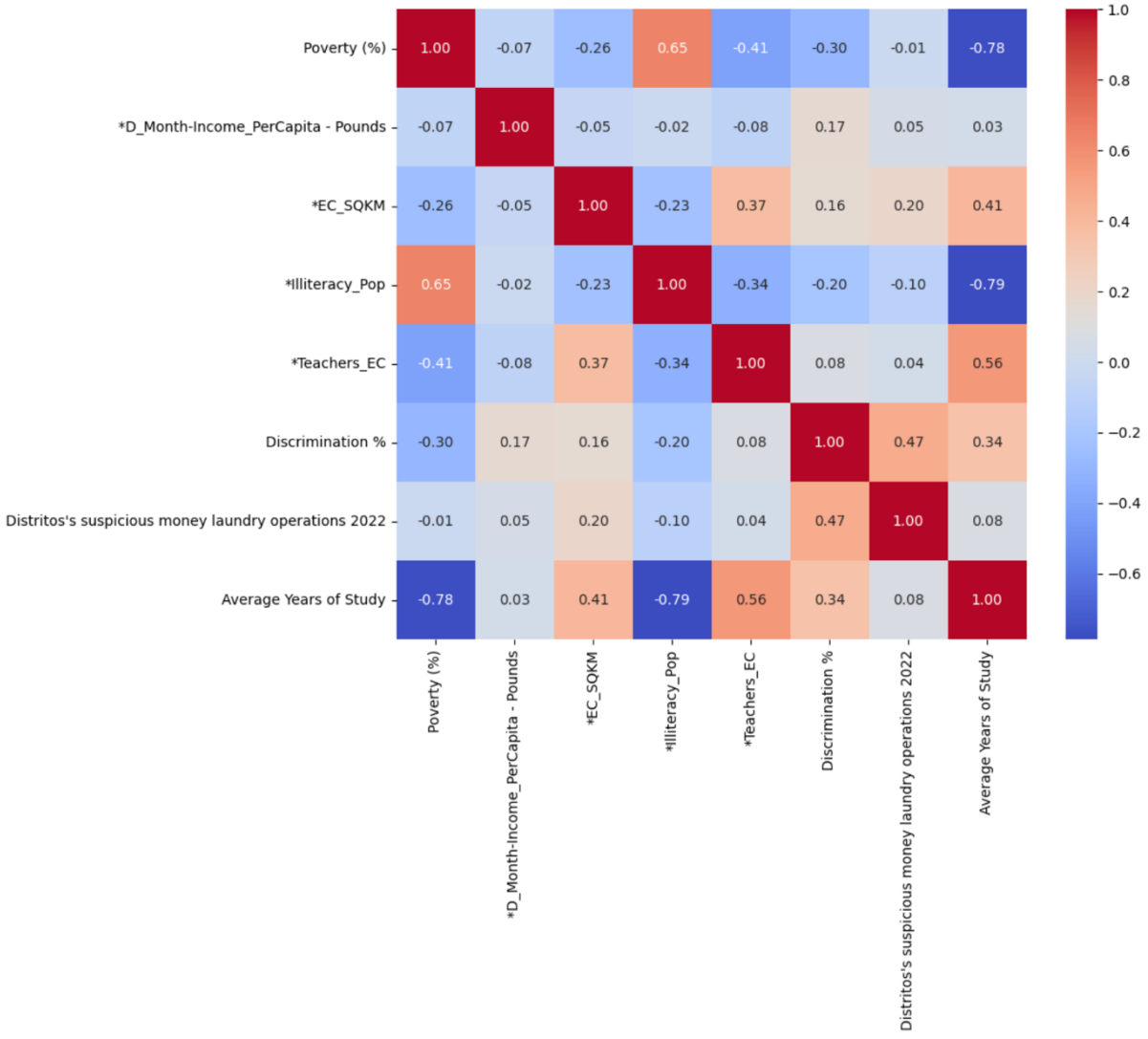
Therefore, it is an interesting hypothesis that by tackling districts with a median level of vulnerability, we perhaps could have the most impact on poverty.

- **Regarding a correlation matrix analysis for all the variables of analysis**

  We identify here that the highest correlation towards **poverty is the Average Years of study** with a -0.78 correlation (if we ignore correlation between years of study and illiteracy which naturally have a direct correlation).

  Also, a very -0.41 correlation between the number of teachers per education centre and poverty still hold an interesting value, thus reinforces the importance to delve **on customization of learning**.

  On the other hand, other variables of interest show to be how money laundry and discrimination is interestingly correlated.

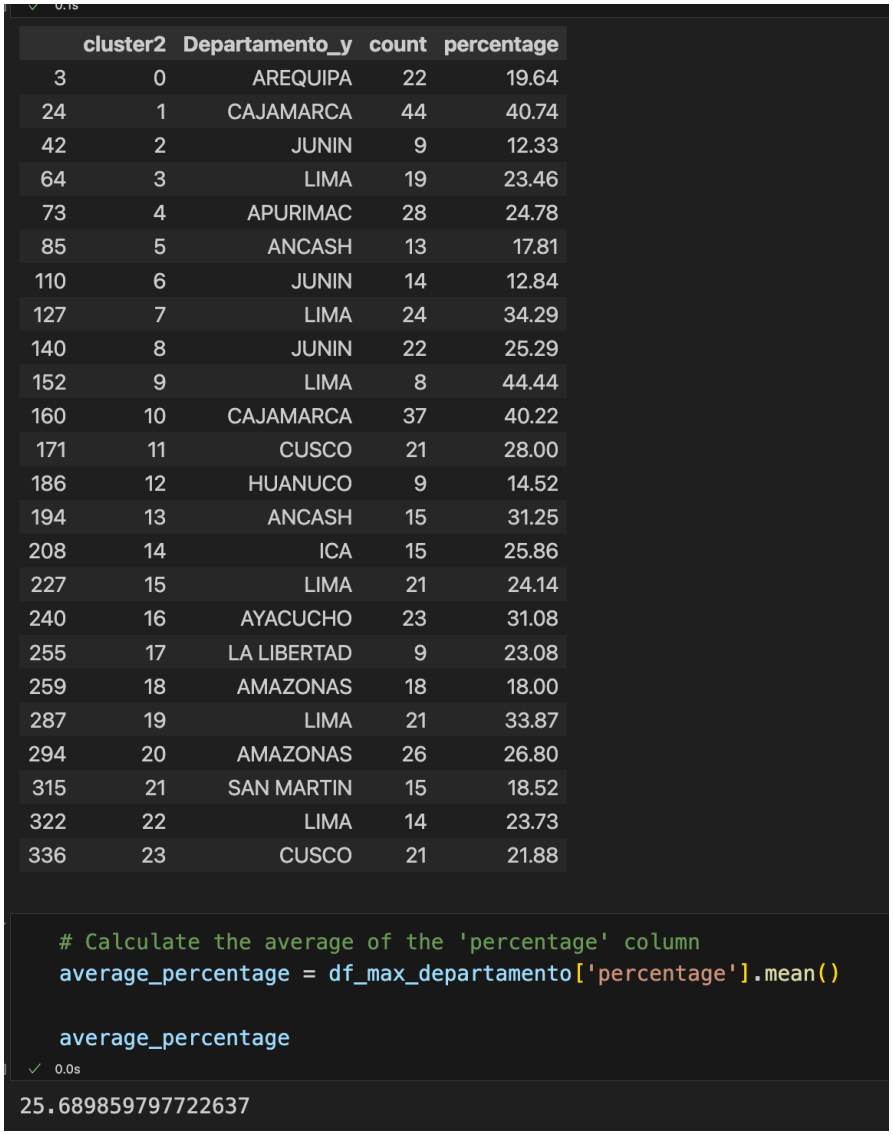


- **Regarding Kmeans analysis for clustering considering the better performing variables: poverty, years of study, teachers per educational centre, illiteracy and vulnerability score**

  After kmean execution as an unsupervised learning algorithm library, the model was asked to generate 24 new clusters. Currently, Peru is already geographically classified in 24 clusters based on geography and they are referenced as for any decision to make withing the government.

  Nonetheless, I believe we should ask ourselves, is geography the real and most urgent way to analyze smaller districts and to take decisions? Or should we cluster them in a different way based on urgency?

  Therefore, when running the analysis, we this imaginative new 24 clusters, just a 25% coincidence with previous geographical clusters was found. Therefore, this could be the start of starting to question the way we classify countries, even more now that we are not ruled only by a physical world but also by technology where geography is less relevant.



	cluster2	Departamento_y	count	percentage
3	0	AREQUIPA	22	19.64
24	1	CAJAMARCA	44	40.74
42	2	JUNIN	9	12.33
64	3	LIMA	19	23.46
73	4	APURIMAC	28	24.78
85	5	ANCASH	13	17.81
110	6	JUNIN	14	12.84
127	7	LIMA	24	34.29
140	8	JUNIN	22	25.29
152	9	LIMA	8	44.44
160	10	CAJAMARCA	37	40.22
171	11	CUSCO	21	28.00
186	12	HUANUCO	9	14.52
194	13	ANCASH	15	31.25
208	14	ICA	15	25.86
227	15	LIMA	21	24.14
240	16	AYACUCHO	23	31.08
255	17	LA LIBERTAD	9	23.08
259	18	AMAZONAS	18	18.00
287	19	LIMA	21	33.87
294	20	AMAZONAS	26	26.80
315	21	SAN MARTIN	15	18.52
322	22	LIMA	14	23.73
336	23	CUSCO	21	21.88

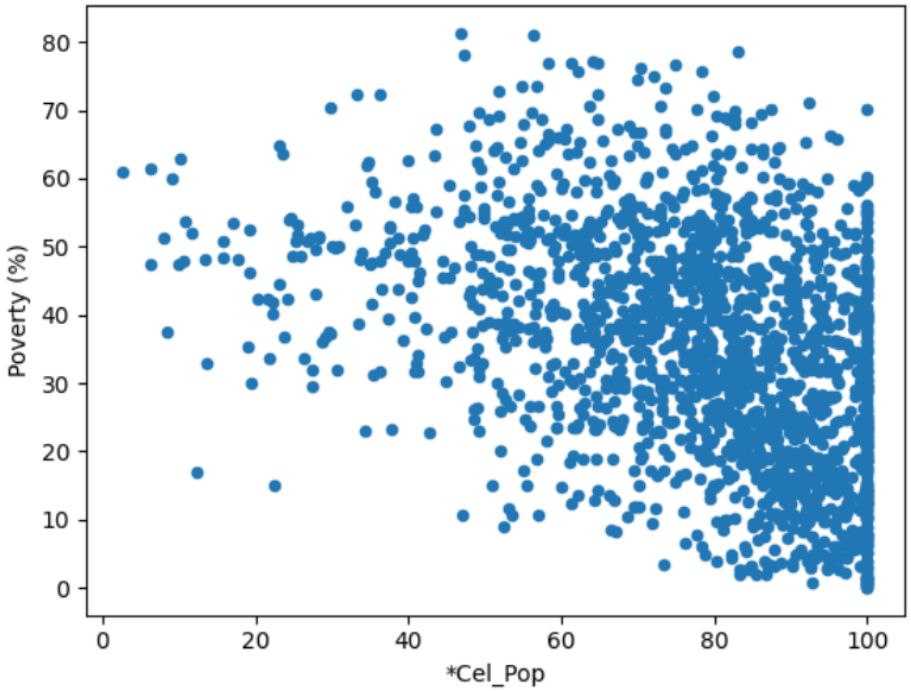
```
# Calculate the average of the 'percentage' column
average_percentage = df_max_departamento['percentage'].mean()

average_percentage
```

0.0s

```
25.689859797722637
```

- **Analyzing connectivity and Poverty**

  While connectivity has the potential to contribute to poverty alleviation and development goals, the actual outcomes are complex and require careful consideration of the mechanisms through which internet technologies influence various social and economic dimensions. (Viecens, 2017)



Even there seems to be a trend correlation between poverty and the population that has a cellphone, data shows that possession of cellphones is already very concentrated towards almost everyone having one. Therefore, there is a potential as a technological tool to access even vulnerable populations.

In addition, allowing more accessibility to children in their houses to education is a great alternative. Data reveals that around 20% of students in rural areas have to walk for more than an hours to arrive to their schools (ProExpansion, 2023).

6. Conclusions

Main Take Aways:

1. Quality of education isn't solely determined by the number of educational centers per inhabitant, highlighting complexities in urban areas.

2. There's a **concentration of teachers per educational center**, posing concerns about educational reliance and the qualifications of teachers.

3. Government earnings per capita show a negative correlation with poverty, indicating potential issues like **corruption or inequality**.

4. **Prioritizing teacher** training over building new schools might be more effective in poverty alleviation.

5. Focusing on districts with **moderate vulnerability could yield significant poverty reduction**.

6. Educational investment and **customization of learning approaches** are crucial in poverty alleviation efforts.

7. Alternative **clustering methods beyond geography** could offer better insights for policymaking.

8. Cellphone ownership presents an opportunity to **leverage technology in reaching vulnerable populations and addressing poverty**.

In conclusion, the evaluation underscores the **intricate relationship between education, socioeconomic factors, and policy decisions in the fight against poverty** in Peru. It urges a new approach, emphasizing the importance of prioritizing teacher training, addressing educational gaps, and targeting interventions based on vulnerability rather than solely geographical considerations.

Moreover, the **potential of technology, particularly cellphone ownership, offers a promising avenue for reaching marginalized communities, while reducing kids risks and time to walk to schools**. By adopting innovative strategies and reevaluating traditional approaches, Peru stands to make meaningful strides towards lasting poverty alleviation and inclusive development.

III. DATA SCIENCE METHODOLOGY, QUALITY AND ETHICS

9. Data Exploration, Evaluating and Cleansing

- **Data Integration**: Utilizing a combination of Python programming and Excel functions like INDEX and MATCH, approximately 20 datasets were merged to create a comprehensive dataset for analysis.

- **Population Discrepancies**: Discrepancies between rural, urban, and total population datasets were identified and addressed to ensure accuracy and consistency across the board.

- **Teacher Data Standardization**: Standardization efforts were undertaken to rectify inconsistencies such as the use of "-" instead of "0" or blanks in the secondary teacher dataset. In cases where primary teacher numbers were close to zero, it was assumed that the quantity of qualified teachers also approached zero based on average data.

- **Budget and Financial Data Normalization**: Budget, income, and expense datasets were normalized to ensure uniformity in unit measures, all values being scaled to millions. Additionally, discrepancies in district names across different datasets were reconciled to facilitate accurate data matching.

- **Money Laundering Dataset Corrections**: Abbreviated department names in the money laundering dataset were manually adjusted to ensure consistency. Furthermore, in the absence of district-level data, region-level data was used as a proxy, acknowledging limitations in available sources.

- **Cellphone and Internet Access Data Conversion**: Datasets pertaining to cellphone and internet access, initially structured by household, were converted to reflect individual access, considering data from Datum with a conversion factor of 3.94. (DATUM, 2016)

- **Geospatial Data Processing**: To ascertain the area of each district, a GeoJSON file was utilized and tabulated for clarity using www.geojson.io. Subsequently, the file was transformed into a CSV dataset, and area values were scaled based on comparison with real-world Wikipedia values, necessitating the adjustment of SHAPE_AREA by a factor of 10,000 for accurate representation.

By meticulously addressing these data exploration, evaluation, and cleansing challenges, the resulting dataset offers a robust foundation for subsequent analysis, ensuring reliability and integrity in the findings derived from the research.

10. Data Quality and Ethics (bias, ethics, limitations, considerations)

- **Bias/Limitation:** While we were able to calculate district surface areas in square kilometers, the lack of specific topographical data poses a limitation. For instance, the time taken to traverse a kilometer can vary significantly depending on terrain features like mountains. We included an analysis of agricultural typology, assuming that districts with more complex topography might prioritize agriculture less. However, this assumption is purely for academic discussion and should not be directly applied to assess relationships.

- **Limitation:** Some cases in the data reveal more educational centers than expected teachers, suggesting a potential discrepancy in the reported number of teachers.

- **Limitation**: The conclusions drawn from relation analysis may suggest directions for further investigation, but they do not establish causality. Therefore, any correlation suggestion or regression analyses conducted are based on hypothetical and academic foundations.

- **Limitation**: The ratio of teachers to inhabitants assumes a similar percentage of the population attends school, which may not always hold true.

- **Limitation/Bias**: The "**black box**" concept should be considered after running cluster with the SKLearn library, since it is an unsupervised algorithm, there is no full clear understanding how clusters were made. Therefore, it is relevant to review characteristics of each cluster to delve on understanding and be aware that further training with bigger datasets should be looked for. Additionally, bias must be supervised:

    a) **Aggregation bias** through clustering my oversee some relevant districts that have behavior or extreme vulnerability. Therefore, it is advised as the continuity of this project to look for outliers (Jayavel, 2024)
    b) **Evaluation bias** since urgency is only assessed based on the currently available and selected variables, however, further other social variables should be eventually considered if the project aims to extend. (Jayavel, 2024)

- **Verification**: Continuation of data collection efforts could improve reliability, although for academic purposes, we rely primarily on government sources.

- **Ethics**: While a tele-educational system is proposed as a potential solution, post-pandemic statistics reveal a gender gap in literacy, with families prioritizing education for sons over daughters using a single cellphone. (Burneo, 2022)

- **Ethics and Risk**: The analysis of underserved education districts must be handled carefully to avoid potential **misuse of data**. For instance, private education companies could exploit districts with above-average income but classified as underserved (Jayavel, 2024). For instance, a private "education" company with no compromise on real education, but with profit, could sell bad quality education programs to district's authorities that in our dataset reflect above than average income but are still classified as underserved regarding educational resources

11. Data Visualization / UX Data Visualization and Website

- **Objectives**:
    - Be able to highlight main finding in a strategically visual focused way to point out the right pains and urgency points to tackle
    - Be structurally disruptive to gain attention and engagement
    - Be aesthetically pleasing to gain attention and engagement
- **Website:**
    - Main: https://santiagowonsiu.wixsite.com/peru-district-analys
    - Data Analysis / Data Visualization: https://santiagowonsiu.wixsite.com/peru-district-analys/blank-1
    - Data Methodology, quality and ethics: https://santiagowonsiu.wixsite.com/peru-district-analys/about-3
    - Project proposal: https://santiagowonsiu.wixsite.com/peru-district-analys/about-4

12. Future additional ideas to complement anlaysis

- **Data scrapping** on Facebook groups on comments (while also extracting district of the comment issuer) to analyze
    - grammar and orthography errors through an LLM model to add on a new dataset on performance / education quality
    - discrimination / hate speech
    - potential political corruption cases
- **Deep Learning Neural Networks** for improved regressions (MLP) Estimated Effects (i.e., Regression analysis AND Neural Network MLP) on the society/district performance variables

- **Web Scrapping**: Wikipedia web scrapping for height on the districts to get a new variable on vulnerability

- **Run further analysis on outliers** on the data to look for urgent in need districts and avoid the aggregation bias from the analyzed clusters

IV. TOWARDS A DIGITAL EDUCATION PROPOSAL

13. Context

With the aims to revolutionize the education sector in Peru by addressing the complexities highlighted in the analysis. It leverages technology to provide customized learning solutions, prioritize teacher training, and bridge the educational gap in vulnerable districts.

14. Challenge and Idea Proposal:

Drawing an analogy to fine-tuning complex deep learning models, consider the well-established education system as already functioning. Despite the availability of high-quality educational content on platforms like YouTube, there's still a need for a supportive backbone. In this analogy, less-educated teachers in rural areas can serve as the vital link, fine-tuning the learning process for students to achieve more effective outcomes.

Why?

- **Addressing Educational Inequality:** can provide access to quality education resources and opportunities to students in underserved areas, helping bridge the gap between urban and rural educational disparities. By offering online learning modules and resources, it can empower students who lack access to traditional educational infrastructure.

- **Enhancing Teacher Quality**: Prioritizing teacher training over building new schools, as suggested in the project idea, can lead to improved teaching standards and educational outcomes. Also, it can offer professional development courses and resources for teachers, enhancing their skills and qualifications to deliver better education to students.

- **Fighting Poverty**: Focusing on districts with moderate vulnerability and providing tailored educational interventions can contribute to poverty reduction. By equipping students with the knowledge and skills they need to succeed

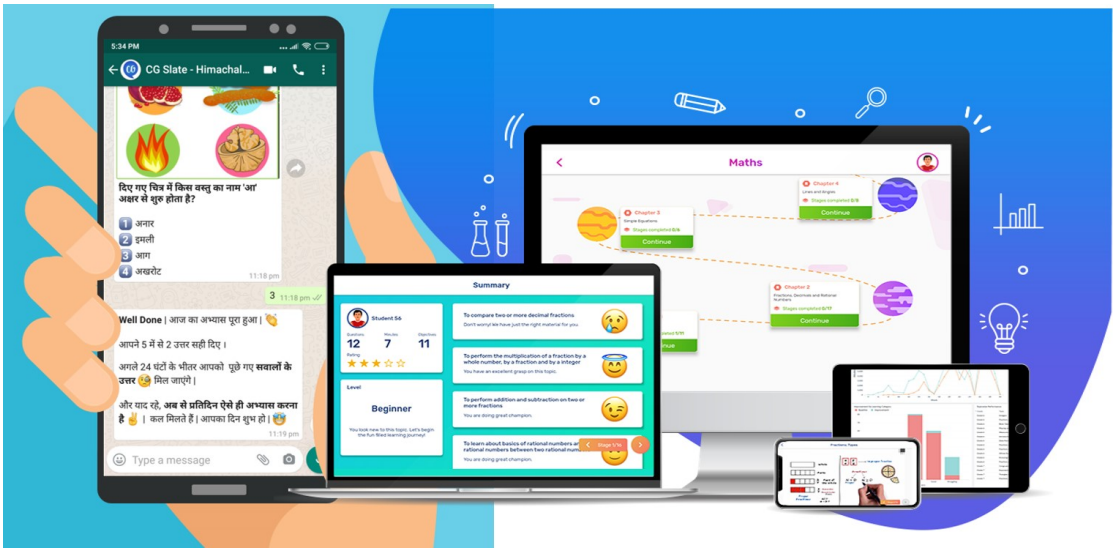
academically and professionally, the project can empower individuals to break the cycle of poverty and achieve socio-economic mobility.

- **Other societal benefits and alternatives**: Education combats corruption by using the curriculum as a tool for knowledge acquisition and behavior modification. By focusing on the affective domain, education can instill values, ethics, and a sense of responsibility in individuals, equipping them to make informed decisions and resist engaging in corrupt practices. This approach helps individuals understand the consequences of corruption, the need for change, and empowers them to make ethical choices, ultimately contributing to a corruption-free society. (Olofu, 2021)

15. Benchmark Analysis – Opportunity for Value Proposal

- **ConveGenius**: is an Indian EdTech comprehensive educational platform designed to revolutionize learning experiences for students across various levels. Through its innovative approach, ConveGenius aims to make education more accessible, engaging, and effective for learners, particularly in underserved communities.

  Some exciting business model features include learning through Whatsapp. This could be inspiring for the current project since several districts in Peru might have limited internet coverage (e.g., hard to stream videos) and Whatsapp can be very efficient in that scenario.



https://convegenius.com/

- **Synthesis:** is Space X's education for kids arm that is the integration of cutting-edge technology and pedagogical strategies to create engaging and personalized learning environments. By harnessing the power of artificial

intelligence, virtual reality, gamification, and other emerging technologies, the initiative aims to make learning more interactive, immersive, and effective.

https://www.synthesis.com/

- **Duolingo**: “When technologist Luis von Ahn was building the popular language-learning platform Duolingo, he faced a big problem: Could an app designed to teach you something ever compete with addictive platforms like Instagram and TikTok? He explains how Duolingo harnesses the psychological techniques of social media and mobile games to get you excited to learn — all while spreading access to education across the world.” ((Ahn, 2023)

https://www.ted.com/talks/luis_von_ahn_how_to_make_learning_as_addictive_as_social_media

16. Key Model Features Ideas:

- **Customized Learning Platform**: Develop a comprehensive online learning platform that offers tailored educational content based on individual student needs and learning styles. Utilize AI and machine learning algorithms to adapt content and pace according to student progress.

- **Teacher Training Program**: Establish a robust teacher training program focusing on modern teaching methodologies, digital literacy, and subject mastery. Provide ongoing support and resources to enhance teaching effectiveness.

- **Poverty Alleviation Initiatives**: Collaborate with government agencies and NGOs to implement targeted interventions in districts with moderate vulnerability. Offer scholarships, educational resources, and mentorship programs to empower students and break the cycle of poverty.

- **Alternative Clustering Methodologies**: Develop advanced data analytics tools to identify clusters of districts based on socioeconomic factors, educational performance, and resource allocation. Use these insights to inform policy decisions and resource allocation for maximum impact.

- **Mobile Learning App**: Launch a mobile learning app accessible to students across Peru, especially those in remote areas with limited access to traditional educational facilities. The app provides interactive lessons, quizzes, and educational games to supplement classroom learning.

17. Data Ethics on the Project to consider

- **Privacy and consent**: Since this would be a digital solution, data might be collected and consent must be given, also acknowledging that data leaks from already vulnerable populations can be very marginally harmful

- **Data bias and fairness**: When developing the educational curricula, globalization and standardization of education can suppress other diversities such as input into education on minorities or specific cultural inputs through education that shouldn't be lost on every single district in Peru

18. KPIs on Ethics of the project

- **Impact on learning** (be sure to measure the right "learning goals") perhaps the learning outcomes are different across different communities

- **Effective inclusion / tailoring** on districts' and regions' culture into the education process

V. Sketchbooks' Reference Glossary

Workbook_1.ipynb

VI. Bibliography / Dataset references

INEI. (2024). *Sistema de Informacion Distrital para la Gestion Publica*. Retrieved from https://estadist.inei.gob.pe/report

Sanchez, J. (2016). Retrieved from https://github.com/juaneladio/peru-geojson/tree/master

Topazio, D. (2022). *HiH Agricultural Typologies Dataset - Peru*. Retrieved from https://data.apps.fao.org/catalog/dataset/hih-peru-agricultural-typology-dataset

INEI. (2022). Retrieved from https://datacrim.inei.gob.pe/panel/mapa

Fhima, F. (2023). Retrieved from How does corruption affect sustainable development? A threshold non-linear analysis: https://www.sciencedirect.com/science/article/abs/pii/S0313592623000401

Olofu, M. A. (2021). *Building a corruption-free society in Nigeria through emphasis on the affective domain in Basic Education curriculum.* Retrieved from https://scholar.google.com/scholar?hl=es&as_sdt=0%2C5&q=corruption+as+the+main+affective+towards+development&btnG=#:~:text=Building%20a%20corruption%2Dfree%20society%20in%20Nigeria%20through%20emphasis%20on%20the%20affective%20domain%20in%20Basic%20Educat

Viecens, H. G. (2017). *Connected for Development? Theory and evidenceabout the impact of Internet technologies onpoverty alleviation.*

Jayavel, D. K. (2024). Critical Data Representation and Analysis Week 2.

Comercio, E. (2020). Retrieved from https://elcomercio.pe/ecdata/mas-de-130-mil-maestros-de-colegios-ejercen-sin-un-titulo-profesional-ecdata-noticia/

DATUM. (2016). Retrieved from https://www.datum.com.pe/new_web_files/files/pdf/nuevas_dinamicas_en_las_familias_peruanas.pdf

ProExpansion. (2023). Retrieved from https://proexpansion.com/en/articles/420-escuelas-rurales-en-el-peru-cuanto-tiempo-les-toma-a-los-ninos-trasladarse-a-sus-escuelas

Burneo, J. (2022). Retrieved from https://southernvoice.org/wp-content/uploads/2022/06/Brechas-genero-Peru-Barrantes-et-al-2022.pdf

Ahn, L. V. (2023). Retrieved from https://www.ted.com/talks/luis_von_ahn_how_to_make_learning_as_addictive_as_social_media