Plan of Action:

Use pandas profiling to determine the data and the quality of the fields

<https://pypi.org/project/pandas-profiling/>

<https://towardsdatascience.com/exploratory-data-analysis-with-pandas-profiling-de3aae2ddff3>

It will help analyze de distribution

correlation of the attributes

distribution of the variables ( cind123 )

The output of the exploratory data analysis, I can see the distribution and correlation of all the attributes. Datatypes of the attributes

**You the eda report ( exploratory data analysis ) HTML or PDF**

Better the HTML

We need to go through that report, we neet to match the questions against the question on my abstract. The proposed technicques are applicale

**Who used this techniques before**

I care about the process itself

Analize the questions, because the questions against techniques

Why answering of the questions is worth, who will be stakeholders is this important

=

Find if the questions are being answering by one technique

Correlated with each other

Nlt analyze the text, text analytics

Beside data analytics

**Groups of information by column:**

* Population
* Fertility
* Mortality
* Migration

**Dependant variable:**

Population size

**Independent variables:**

* Fields within the groups Fertility, Mortality, Migration
* Time: in years
  + Demographics: such total population, male and female population, median age, natural change births minus deaths
* Economic factors to add ( from other datasets )
* Environmental factors ( to add from other datasets )
* Policy and governance ( to add regulations etc )
* Education and health care ( to add from other datasets )
* Migration( already included )

**Data collected types of data :**

* Most of the data is numeric not many categorical attributes

**Questions:**

**What are the influential predictors ( independent ) that can clasify the dependant variable )**

**with only find correlations**

Which factors affect the population size in predictive models

The question is how to set up the algorithm to work all the variables against 2 fields ( country, year )

Multiple linear regression

For each country I have the collective information from years 1950 to 2021

Feature selector, which is the most important predictors

Filter based: multicolinearity correlations

wrappers :

embedded techniques or hibrids

stratify clustering based on country

fuature selection technique which factors are more important

check with library

study reduction, backward elimination

<https://ourworldindata.org/future-population-growth>

* [The UN projects that the global population will increase from a population of around 8 billion in 2022 to 10.4 billion by the end of the century. By that time, the UN projects, fast global population growth will come to an end.](https://ourworldindata.org/future-population-growth#global-population-growth)
* [Beneath the global level, there are of course big differences between different world regions and countries. While in some regions the world population will likely grow rapidly for the coming decades, other regions will continue to see declining population numbers.](https://ourworldindata.org/future-population-growth#population-growth-by-world-region)
* [Global population growth is determined by the number of births and deaths. Improving health is increasing the size of the population as it is decreasing mortality. The countervailing trend is falling fertility rates – the trend of couples having fewer children is what brought rapid population growth to an end in many countries already, and what will bring an end to rapid population growth globally.](https://ourworldindata.org/future-population-growth#projections-of-the-drivers-of-population-growth)
* [The global population growth rate has already slowed down considerably: it reached its peak at over 2% in the 1960s and has been falling since.](https://ourworldindata.org/future-population-growth#the-population-growth-rate)
* [The UN projections for the global population growth rates, which have been produced since the 1950s, have a good track record in projecting the size of the global population.](https://ourworldindata.org/future-population-growth#how-accurate-have-past-population-projections-been)
* [While the UN projections are most widely known, there are other very-carefully-produced projections. The demographers of WC-IIASA model what will happen according to different scenarios and make clear that the population growth rate tomorrow depends on what we do today. Rapid progress in getting children – and especially girls – into schools will result in a much smaller global population.](https://ourworldindata.org/future-population-growth#progress-in-global-education-now-matters-for-the-size-of-the-global-population-in-the-21st-century)
* [The biggest disagreement between different projections is concerning the future of Africa. While the UN projects an almost 3-fold increase of the population of Africa, other researchers find a much smaller increase more likely.](https://ourworldindata.org/future-population-growth#whether-the-world-population-will-reach-10-billion-will-likely-depend-on-africa)

<https://medium.com/@aisharm13/world-population-data-analysis-and-model-for-prediction-34287391d665>

regression model to predict the population at any moment

https://www.sciencedaily.com/releases/2020/07/200715150444.htm