Introduction

Description:

the Global Health Observatory (GHO) data repository under World Health Organization (WHO) keeps track of the health status as well as many other related factors for all countries The datasets are made available to public for the purpose of health data analysis. The dataset related to life expectancy, health factors for 193 countries has been collected from the same WHO data repository website and its corresponding economic data was collected from United Nation website. Among all categories of health-related factors only those critical factors were chosen which are more representative. It has been observed that in the past 15 years , there has been a huge development in health sector resulting in improvement of human mortality rates especially in the developing nations in comparison to the past 30 years. Therefore, in this project we have considered data from year 2000-2015 for 193 countries for further analysis.

Dataset:

The dataset is collected from kaggle :

<https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who>

Attributes:

Our dataset consist of 2820 rows and 22 columns.

**Columns description:**

country (Nominal) - the country in which the indicators are from (i.e. United States of America or Congo)

year (Ordinal) - the calendar year the indicators are from (ranging from 2000 to 2015)

status (Nominal) - whether a country is considered to be 'Developing' or 'Developed' by WHO standards

life\_expectancy (Ratio) - the life expectancy of people in years for a particular country and year

adult\_mortality (Ratio) - the adult mortality rate per 1000 population (i.e. number of people dying between 15 and 60 years per 1000 population);

infant\_deaths (Ratio) - number of infant deaths per 1000 population; similar to above, but for infants

alcohol (Ratio) - a country's alcohol consumption rate measured as liters of pure alcohol consumption per capita

percentage\_expenditure (Ratio) - expenditure on health as a percentage of Gross Domestic Product (gdp)

hepatitis\_b (Ratio) - number of 1 year olds with Hepatitis B immunization over all 1 year olds in population

measles (Ratio) - number of reported Measles cases per 1000 population

bmi (Interval/Ordinal) - average Body Mass Index (BMI) of a country's total population

under-five\_deaths (Ratio) - number of people under the age of five deaths per 1000 population

polio (Ratio) - number of 1 year olds with Polio immunization over the number of all 1 year olds in population

total\_expenditure (Ratio) - government expenditure on health as a percentage of total government expenditure

diphtheria (Ratio) - Diphtheria tetanus toxoid and pertussis (DTP3) immunization rate of 1 year olds

hiv/aids (Ratio) - deaths per 1000 live births caused by HIV/AIDS for people under 5; number of people under 5 who die due to

HIV/AIDS per 1000 births

gdp (Ratio) - Gross Domestic Product per capita

population (Ratio) - population of a country

thinness\_1-19\_years (Ratio) - rate of thinness among people aged 10-19 (Note: variable should be renamed to thinness\_10-19\_years to more accurately represent the variable)

thinness\_5-9\_years (Ratio) - rate of thinness among people aged 5-9

income\_composition\_of\_resources (Ratio) - Human Development Index in terms of income composition of resources (index ranging from 0 to 1)

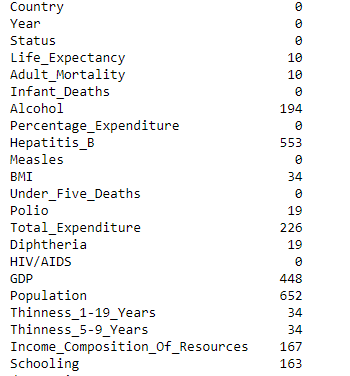
schooling (Ratio) - average number of years of schooling of a population

Tools:  
python , pandas(python data analysis library) , matplotlib and seaborn (python data visualization libraries)

Data Cleaning:

* Removing white spaces of column names.
* Renaming columns with appropriate name without spaces between
* Imputing nan values and dealing with some outliers

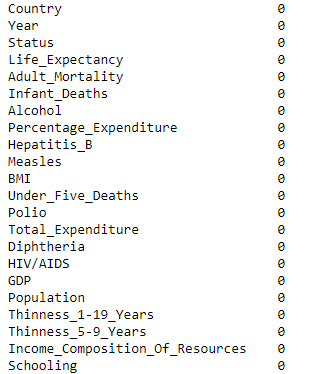
Data before cleaning



it looks like many columns containing null values, since this is time series data assorted by country, the best course of action would be to interpolate the data by country.

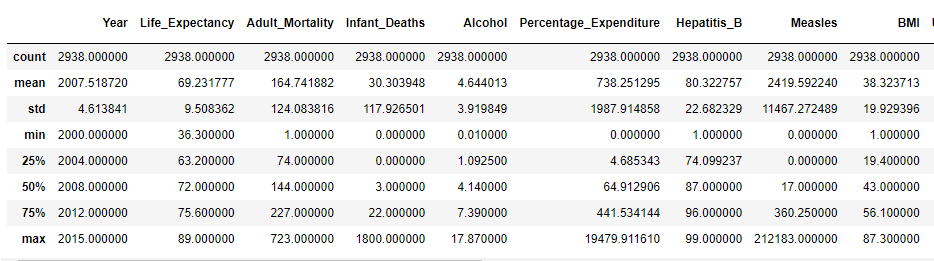
However, when attempting to interpolate by country it doesn't fill in any values as the countries' data for all the null values are null for each year, so we will impute the data by year and we will impute by each year's mean value.

After cleaning by imputing mean values:

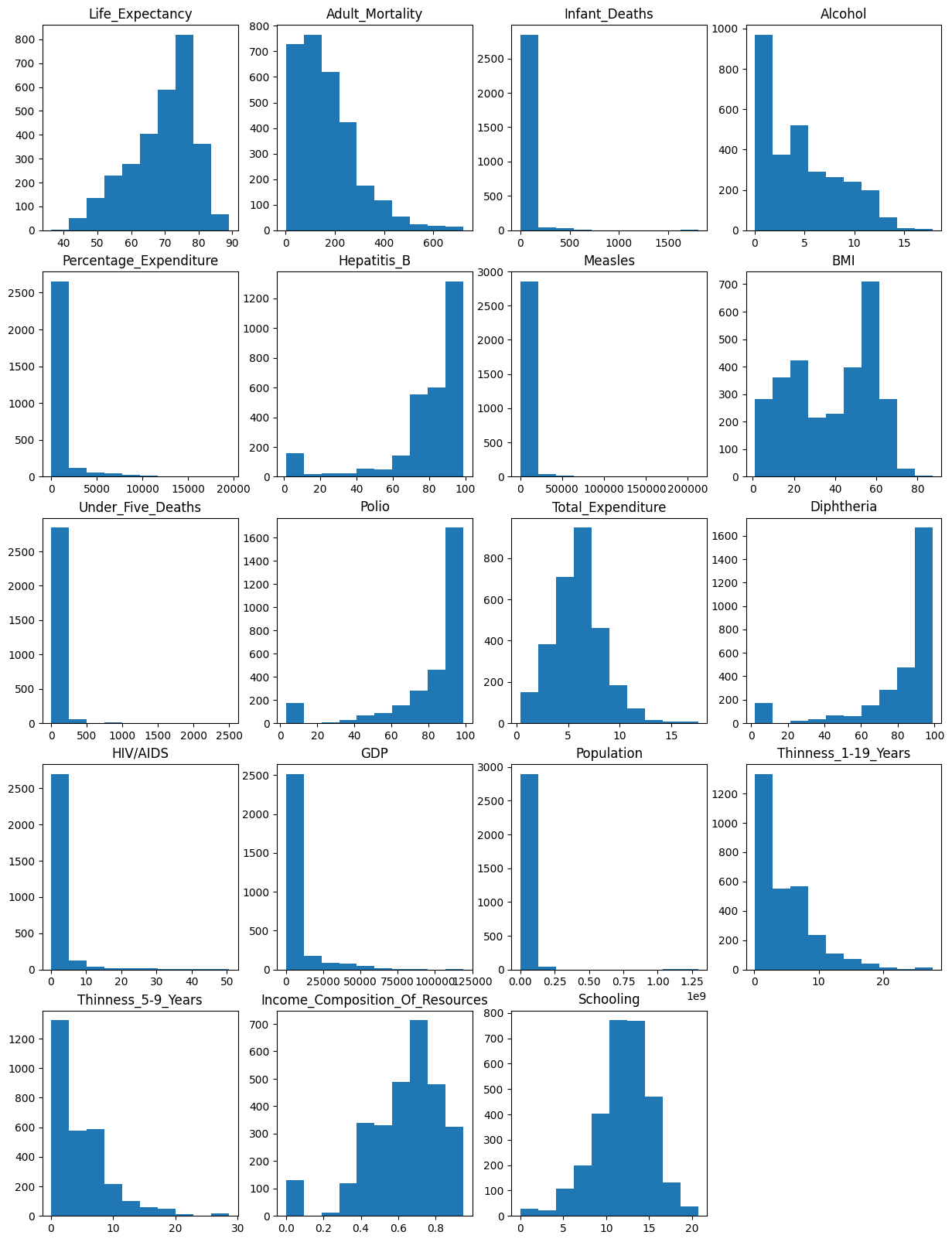


Descriptive analysis

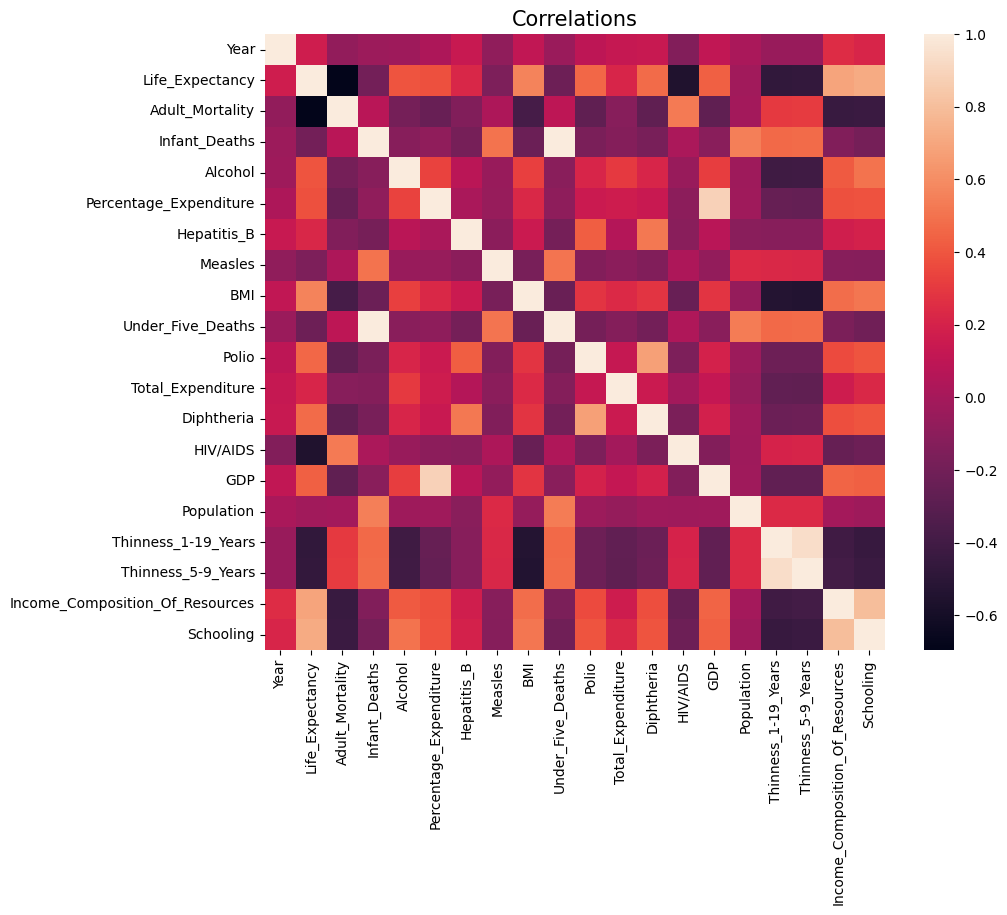
Descriptive table with top features :



plotting distribution of all continous variables

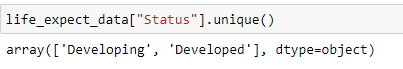


Correlation heat map between all the features with each other

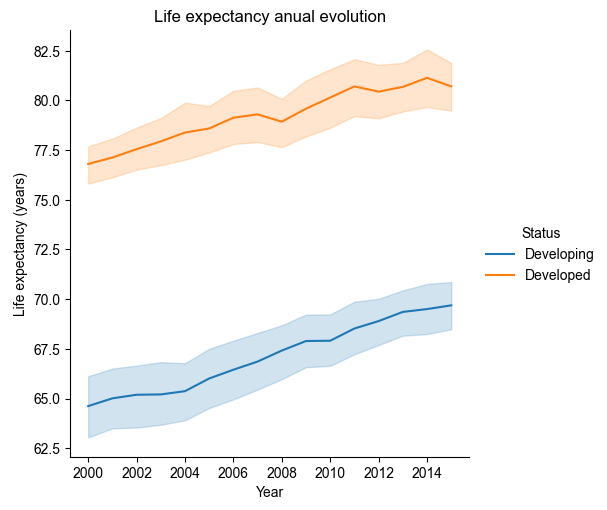


Exploratory analysis

during our exploratory analysis we analyze our data with comparing between the developing and developed country to get a lot of interesting insights.



The line graph shows life expectancy per year for developing and developed countries



Result:

We can see that the rate of life expectancy increases in both developing and developed countries,

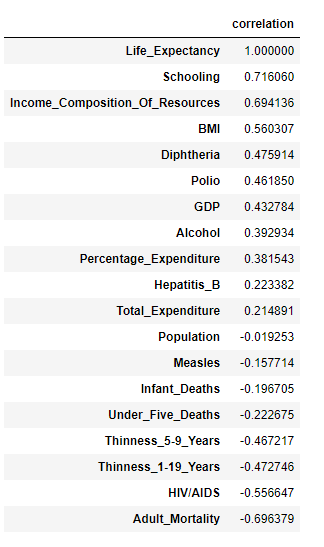
But developed countries starting with more years than developing countries.

Hypothesis Questions:

1. How does Infant and Adult mortality rates affect life expectancy?
2. Does Life Expectancy have positive or negative relationship with drinking alcohol ?
3. does schooling have positive or negative impact on the life expectancy?

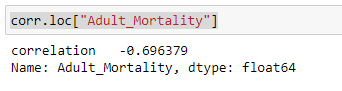
In order to continue investigating our questions we will compute the correlation between every feature with life expectancy .

Result:

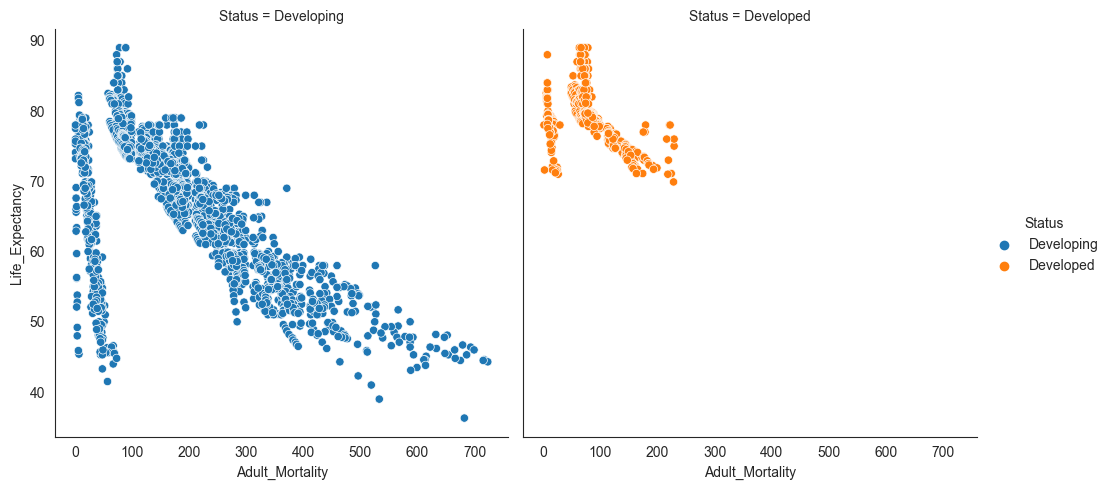


1. How does Infant and Adult mortality rates affect life expectancy?

**Adult mortality**



Scatter plot between adult mortality and life expectancy , we can see that as adult mortality increases life expectancy decreases In both developed and developing countries.

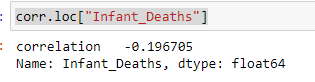


Result:

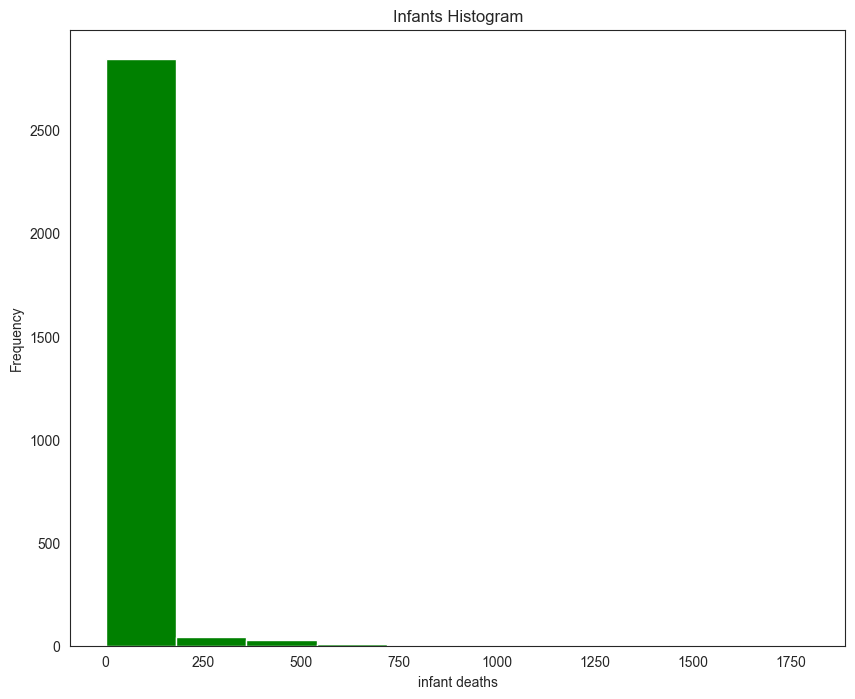
we can see that Adult\_mortality has strong negative correlation with life expectancy.

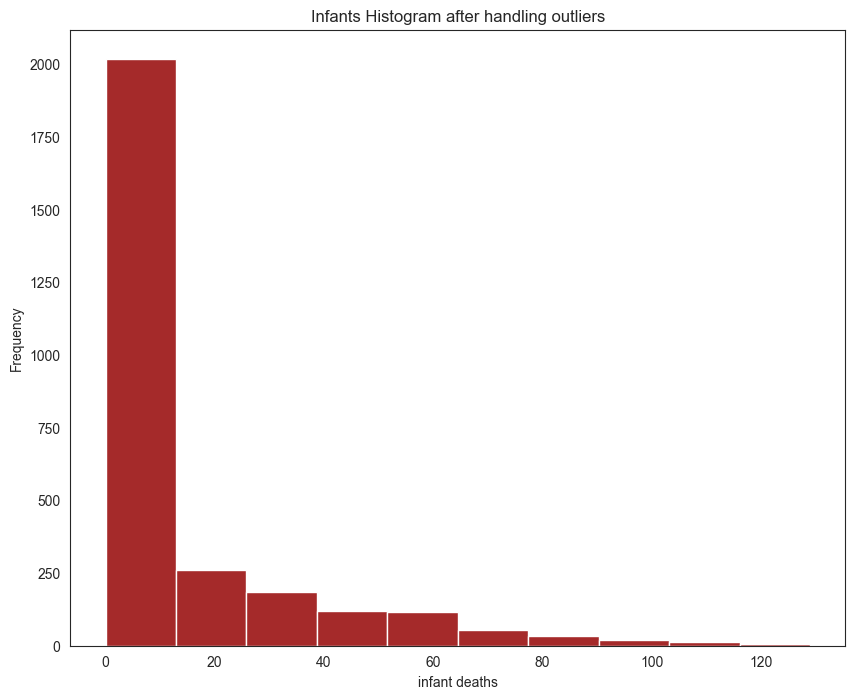
**Infant Deaths**

Infant deaths has kind of weak correlation with the life expectancy but maybe this correlation is misleading as a result of outliers

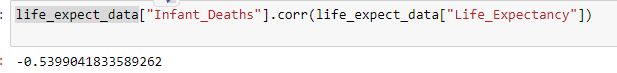


we can see from our histogram that most of our data lies between 0 to 100





checking the correlation again

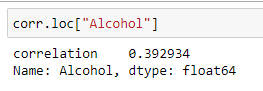


Result:

Huge difference , that means our assumption was right of existing of outliers misleading information.

We conclude that infant deaths affect negatively the life expectancy.

### 2. Does Life Expectancy have positive or negative relationship with drinking alcohol ?

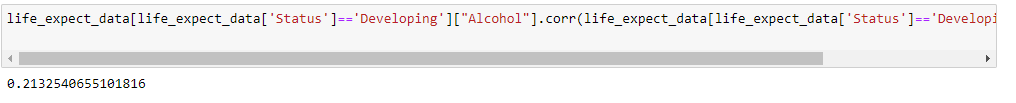


The correlation above showed 0.39 positive correlation of between Alcohol and life expectancy, we have doubt that maybe we have misleading information so we will test the correlation when separating the data by developing and developed countries to see the result.

Scatter plot between alcohol in developing countries and life Expectancy



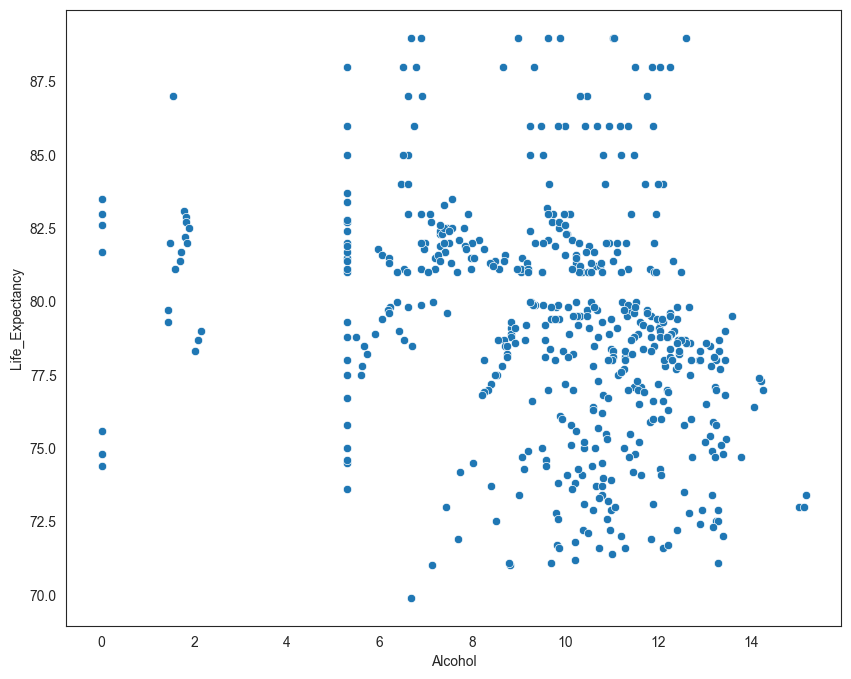
Correlation between developing countries' life expectancy and alcohol



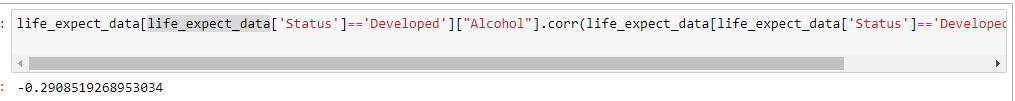
Result:

in developing countries we can see weak correlation between alcohol and life expectancy.

Scatter plot between alcohol in developed countries and life Expectancy



Correlation between developing countries' life expectancy and alcohol

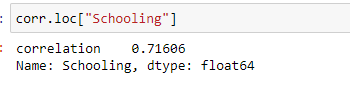


Result:

in developed country we can see a weak negative correlation between alcohol and life expectancy.

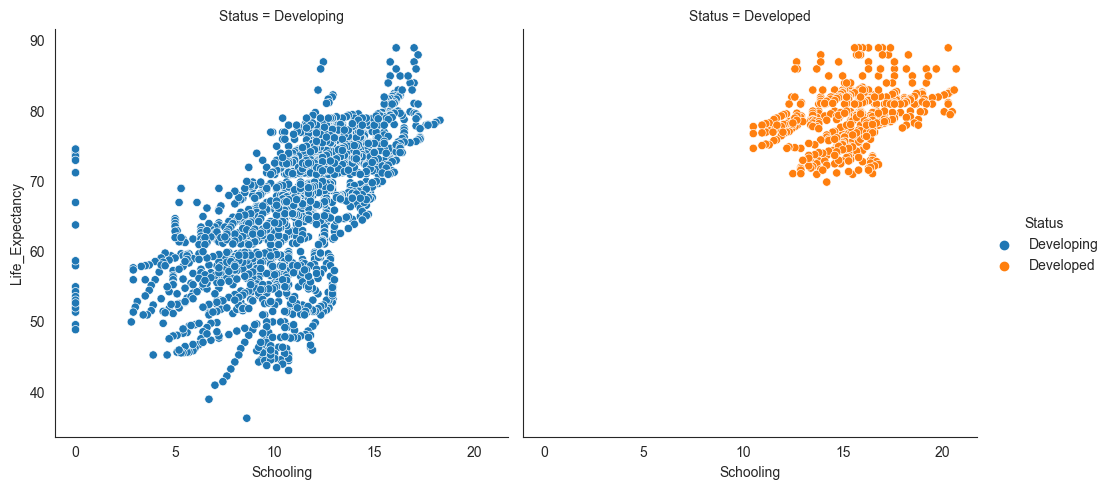
3. Does schooling have positive or negative impact on the life expectancy?

Correlation between Schooling and life expectancy



we have a strong positive correlation between schooling and life expectancy .

scatter plot between schooling and life expectancy we can see that as schooling increases the life expectancy increases In both developed and developing countries.



From the above graph:

Obviously Schooling has a positive impact on life expectancy.

Discussions:

* Life expectancy is increasing over the last 15 years in the developed and developing countries
* Adult deaths in developing countries is much higher than developed countries lead to lower life expectancy , even so adult death rates have negative effect on life expectancy on both types of countries.
* Infant early deaths affect the life expectancy.
* Alcohol has a negative effect on life expectancy in the developed countries.
* Schooling has very positive impact on life expectancy in both developed and developing countries

Conclusion:

This project analyzed the different aspect that effect the life expectancy in different point of views and with different variable.

We discussed the influence between the developing and developed countries with variables affecting it like adult and infant death, alcohol , schooling effects on the life expectancy.