**SHRI VAISHNAV VIDYAPEETH VISHWAVIDYALAYA, INDORE**



**Project Report**

On

**Life Expectancy Data**

Submitted By

1. **Divyam Pithawa [19100BTCSAII05259]**
2. **Mradul Agrawal [19100BTCSAII05269]**
3. **Rudraksh Thorat [19100BTCSAII05280]**
4. **Sarthak Nahar [19100BTCSAII05283]**
5. **Shivam Sharma [19100BTCSAII05286]**

Under the Guidance of

**Mr. Sachin Garg**

**SHRI VAISHNAV INSTITUTE OF INFORMATION TECHNOLOGY**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**JULY 2021-JANUARY 2021**

**INTRODUCTION**

Although there have been lot of studies undertaken in the past on factors affecting life expectancy considering demographic variables, income composition and mortality rates. It was found that affect of immunization and human development index was not taken into account in the past. Also, some of the past research was done considering multiple linear regression based on data set of one year for all the countries. Hence, this gives motivation to resolve both the factors stated previously by formulating a regression model based on mixed effects model and multiple linear regression while considering data from a period of 2000 to 2015 for all the countries. Important immunization like Hepatitis B, Polio and Diphtheria will also be considered. In a nutshell, this study will focus on immunization factors, mortality factors, economic factors, social factors and other health related factors as well. Since the observations this dataset are based on different countries, it will be easier for a country to determine the predicting factor which is contributing to lower value of life expectancy. This will help in suggesting a country which area should be given importance in order to efficiently improve the life expectancy of its population.

**Content**

The project relies on accuracy of data. 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 data-sets are made available to public for the purpose of health data analysis. The data-set 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. The individual data files have been merged together into a single data-set. On initial visual inspection of the data showed some missing values. As the data-sets were from WHO, we found no evident errors. Missing data was handled in R software by using Missmap command. The result indicated that most of the missing data was for population, Hepatitis B and GDP. The missing data were from less known countries like Vanuatu, Tonga, Togo, Cabo Verde etc. Finding all data for these countries was difficult and hence, it was decided that we exclude these countries from the final model data-set. The final merged file(final dataset) consists of 22 Columns and 2938 rows which meant 20 predicting variables. All predicting variables was then divided into several broad categories:​Immunization related factors, Mortality factors, Economical factors and Social factors.

**Acknowledgements**

The data was collected from WHO and United Nations website with the help of Deeksha Russell and Duan Wang.

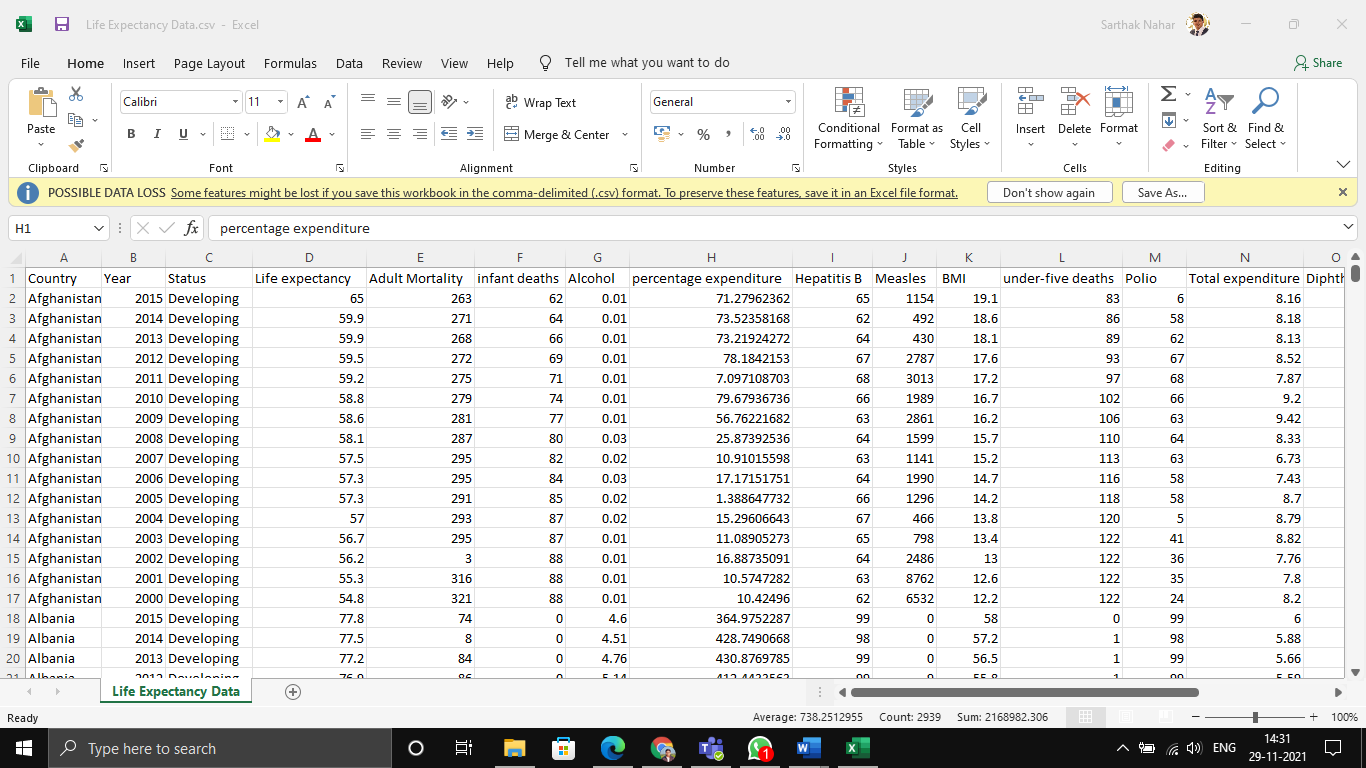
**Public Dataset available at below website**

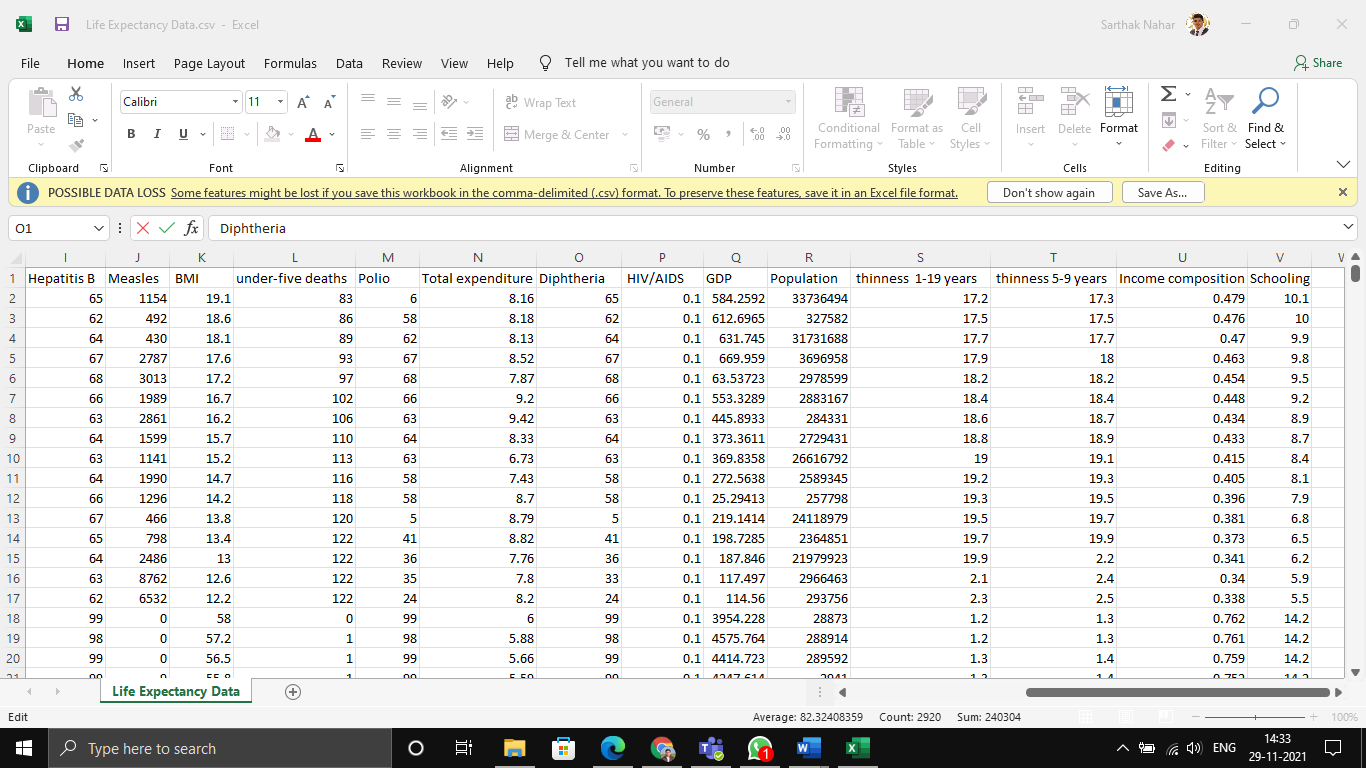
<https://www.kaggle.com/kumarajarshi/life-expectancy-who>

**Tools Used**

Jupyter Notebook.

**DataSet (Screenshot) : -**





**About Columns of Dataset: -**

1. Country - Country
2. Year - Year
3. Status - Developed or Developing status
4. Life expectancy - Life Expectancy in age
5. Adult Mortality - Adult Mortality Rates of both sexes (probability of dying between 15 and 60 years per 1000 population)
6. Infant Deaths - Number of Infant Deaths per 1000 population
7. Alcohol - Alcohol, recorded per capita (15+) consumption (in litres of pure alcohol)
8. Percentage Expenditure - Expenditure on health as a percentage of Gross Domestic Product per capita(%)
9. Hepatitis B - Hepatitis B (HepB) immunization coverage among 1-year-olds (%)
10. Measles - Number of reported cases per 1000 population
11. BMI - Average Body Mass Index of entire population
12. Under-Five Deaths - Number of under-five deaths per 1000 population
13. Polio - Polio (Pol3) immunization coverage among 1-year-olds (%)
14. Total Expenditure - General government expenditure on health as a percentage of total government expenditure (%)
15. Diphtheria - Diphtheria tetanus toxoid and pertussis (DTP3) immunization coverage among 1-year-olds (%)
16. HIV/AIDS - Deaths per 1 000 live births HIV/AIDS (0-4 years)
17. GDP - Gross Domestic Product per capita (in USD)
18. Population - Population of the country
19. Thinness 1-19 Years - Prevalence of thinness among children and adolescents for Age 10 to 19 (% )
20. Thinness 5-9 Years - Prevalence of thinness among children for Age 5 to 9(%)
21. Income Composition of Resources - Human Development Index in terms of income composition of resources (index ranging from 0 to 1)
22. Schooling - Number of years of Schooling(years)

**Problem Statement are: -**

Que 1. Finding top 10 Countries with maximum number of Life Expectancy.

Que 2. Find the Correlation between GDP and Life Expectancy.

Que 3. Which Disease affect Life Expectancy the most.

Que 4. Which Country's Life Expectancy increased the most in 15 years.

Que 5. Top Five Country with Highest Adult Mortality.

Que 6. Draw a Graph b/w Income Composition and Population.

Que 7. In which Year the Total Expenditure is Highest in India.

Que 8. What are the predicting variables that actually affecting the Life Expectancy?

Que 9. Should a country having a lower life expectancy value (<65) increase its healthcare expenditure in order to improve its average lifespan?

Que 10. How does Infant and Adult mortality rates affect Life Expectancy?

Que 11. Does Life Expectancy have positive or negative correlation with Status of the Country, drinking alcohol, BMI, GDP etc.

Que 12. What is the impact of schooling on the lifespan of humans?

Que 13. Do densely populated countries tend to have lower life expectancy?

Que 14. Draw a Line Graph of Year wise GDP of India.

Que 15. Which Country infant die more often.

Que 16. In which Year the Under Five Death is Highest.

Que 17. Draw a Graph between Year and Schooling.

Que 18. Country which consumed most alcohol.

Que 19. Countries whose government shows awareness about their health condition (more expenditure on health issues).

Que 20. In which year India shows healthiest Body Mass Index (BMI).

Que 21. Top 5 countries whose infant mortality rate is lowest.

Que 22. Countries whose literacy rate is best.

Que 23. Countries with lowest literacy rate.

Que 24. In which Year the Total Expenditure is Lowest in India.

Que 25. In which year the Income Composition of Resources is Highest in India.

Que 26. How much the GDP has increased in the last 15 years in India.

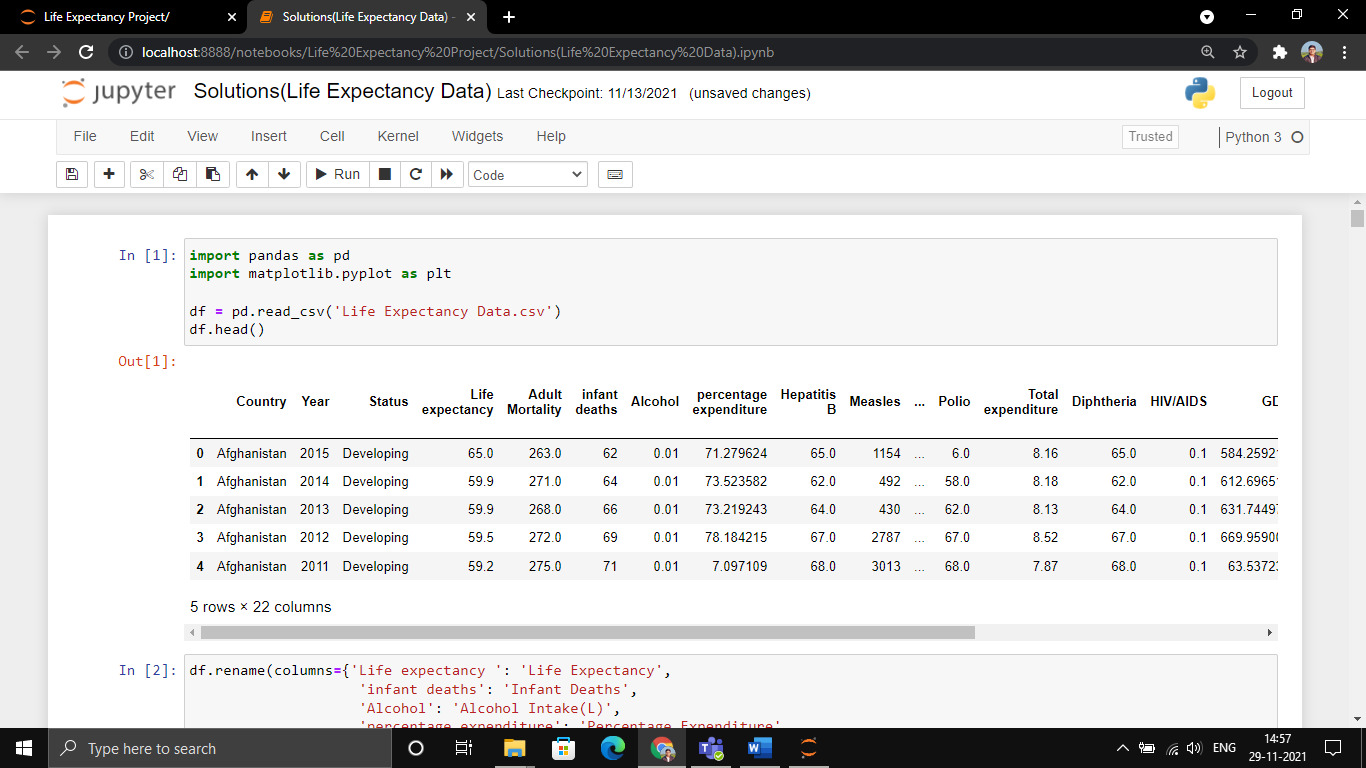
Que 27. How many countries are Developed till the year 2015.

Que 28. List of Countries which has Developed in the last 15 Years.

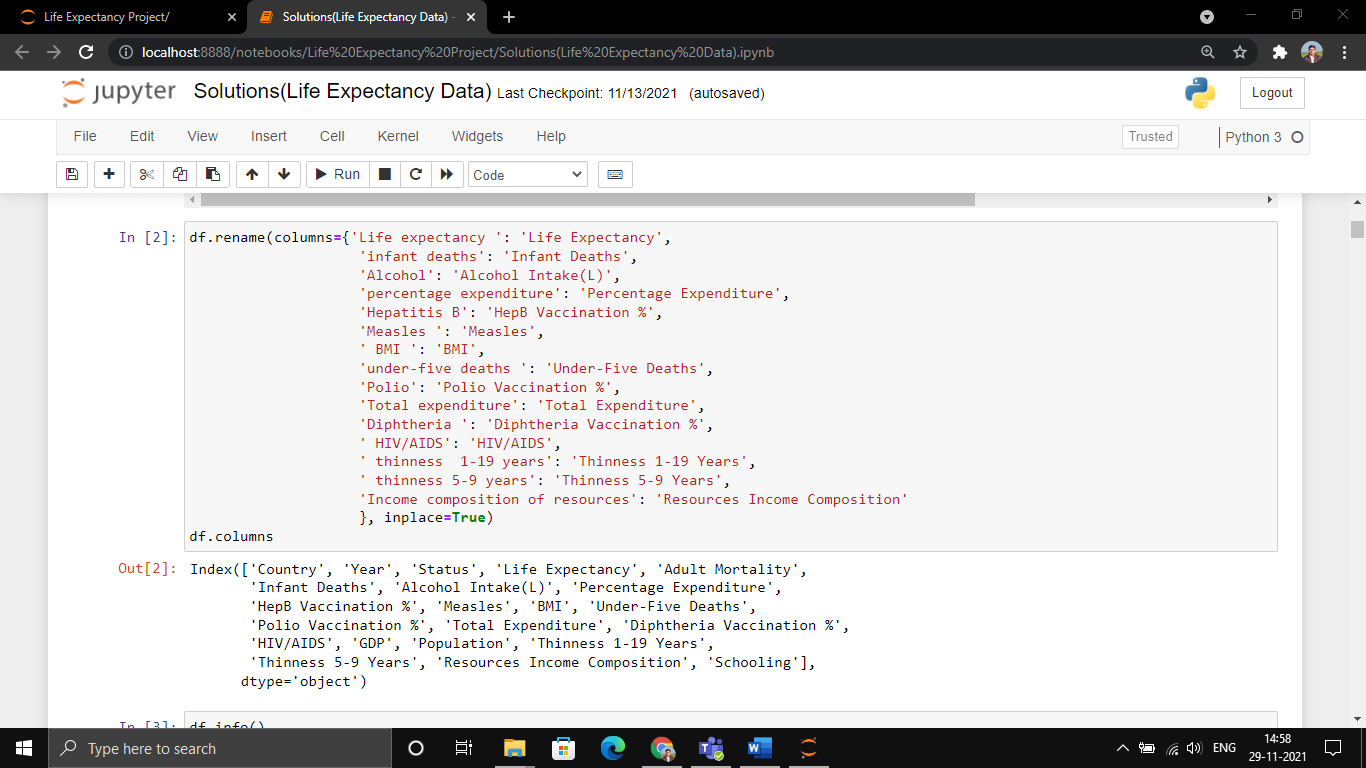
Que 29. Draw a Graph of Year wise Schooling in India.

Que 30. Draw a Graph of Life Expectancy of different Countries.

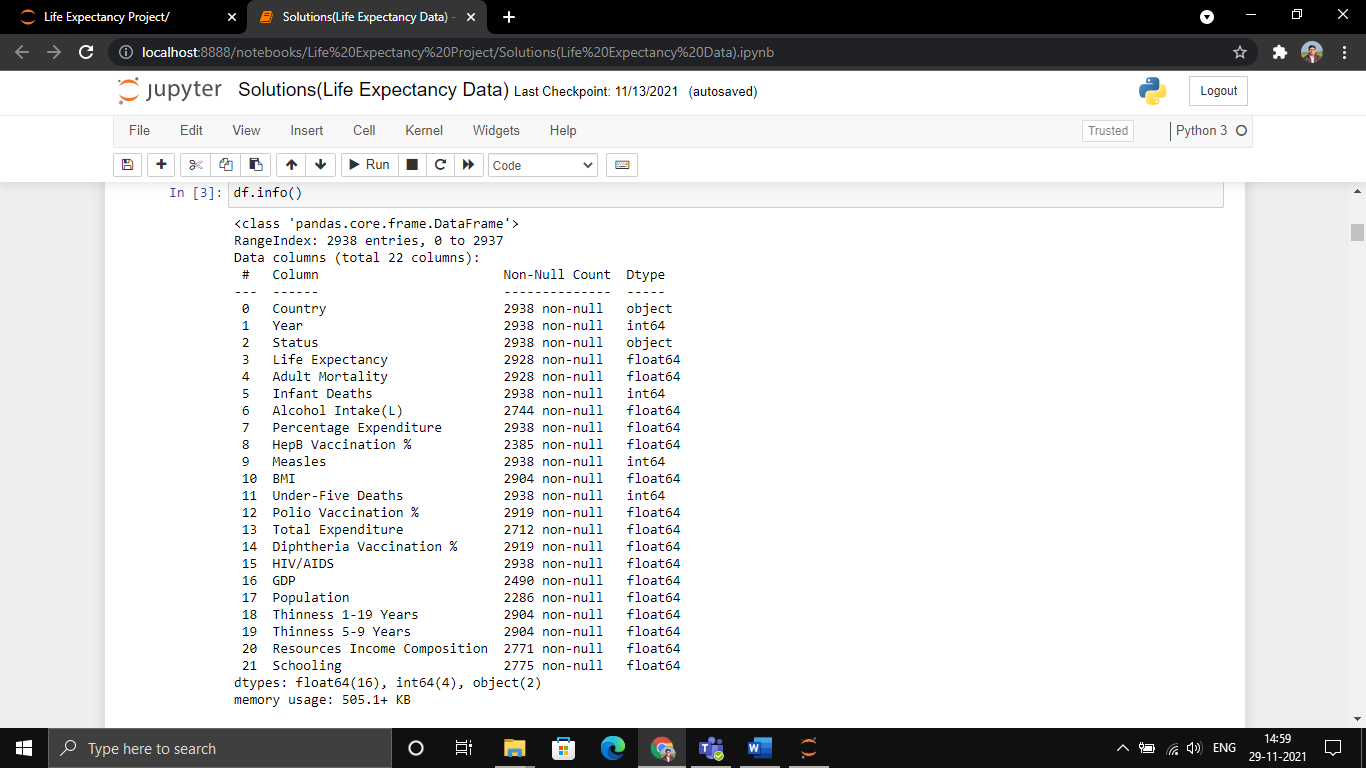
**Step 1. Importing Data set to a Python Code**



**Step 2. Renaming Columns of a Data Set**

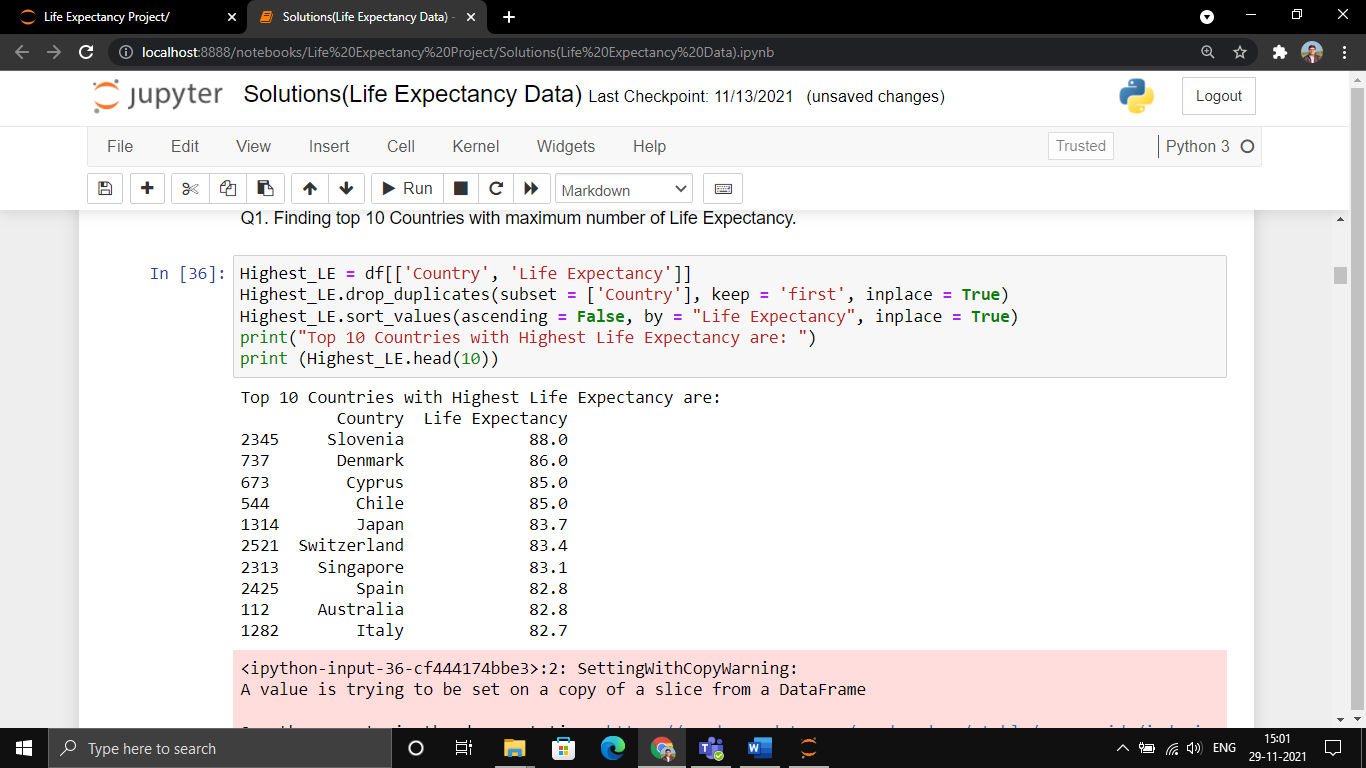


**Step 3. Finding info about Data Set**

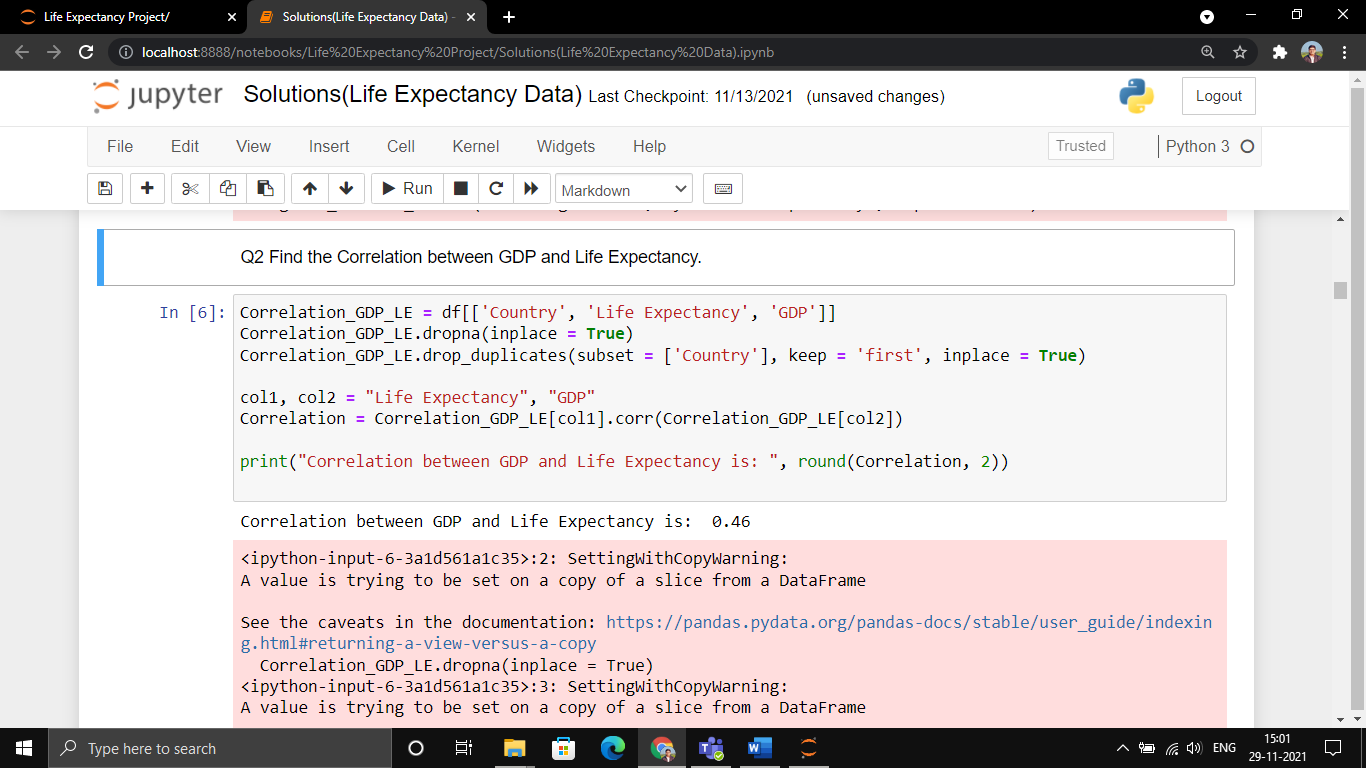


**Solutions of Problem Statement: -**

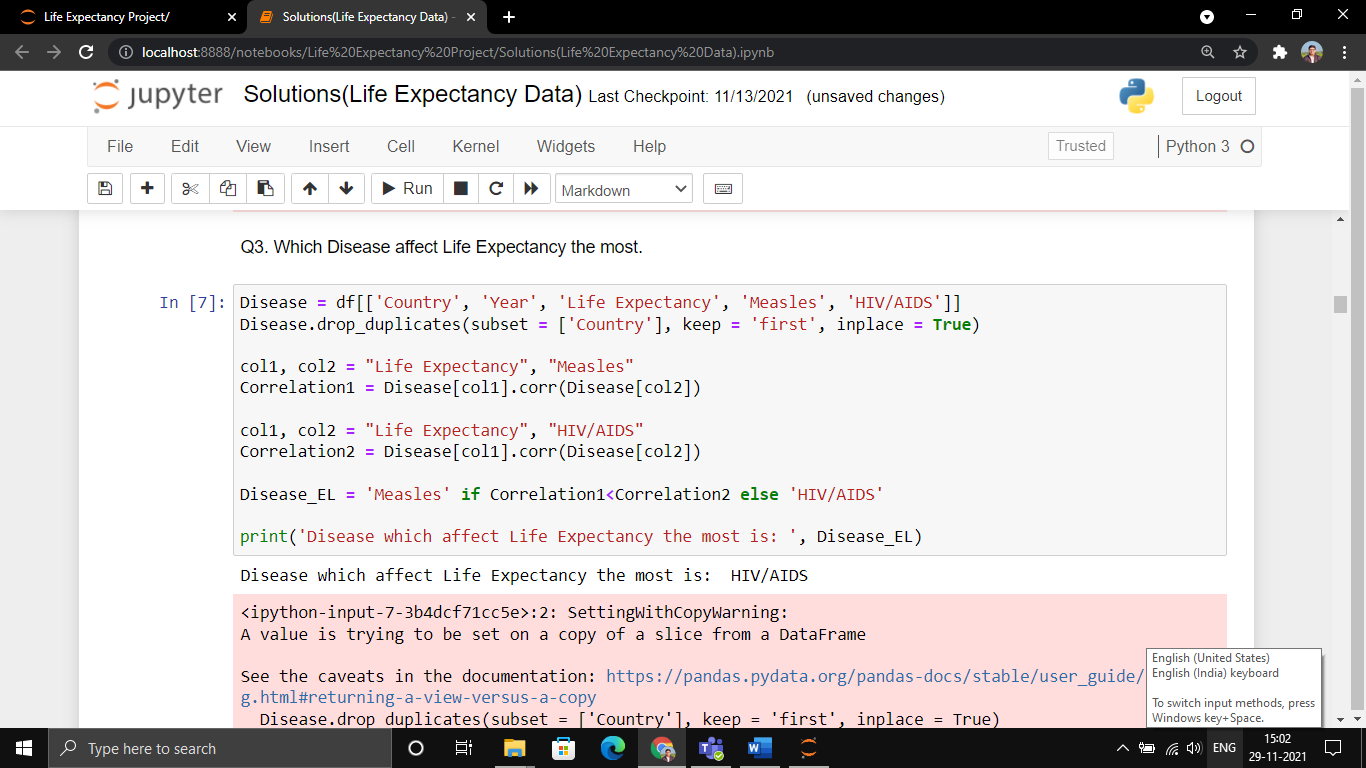
Que 1. Finding top 10 Countries with maximum number of Life Expectancy.



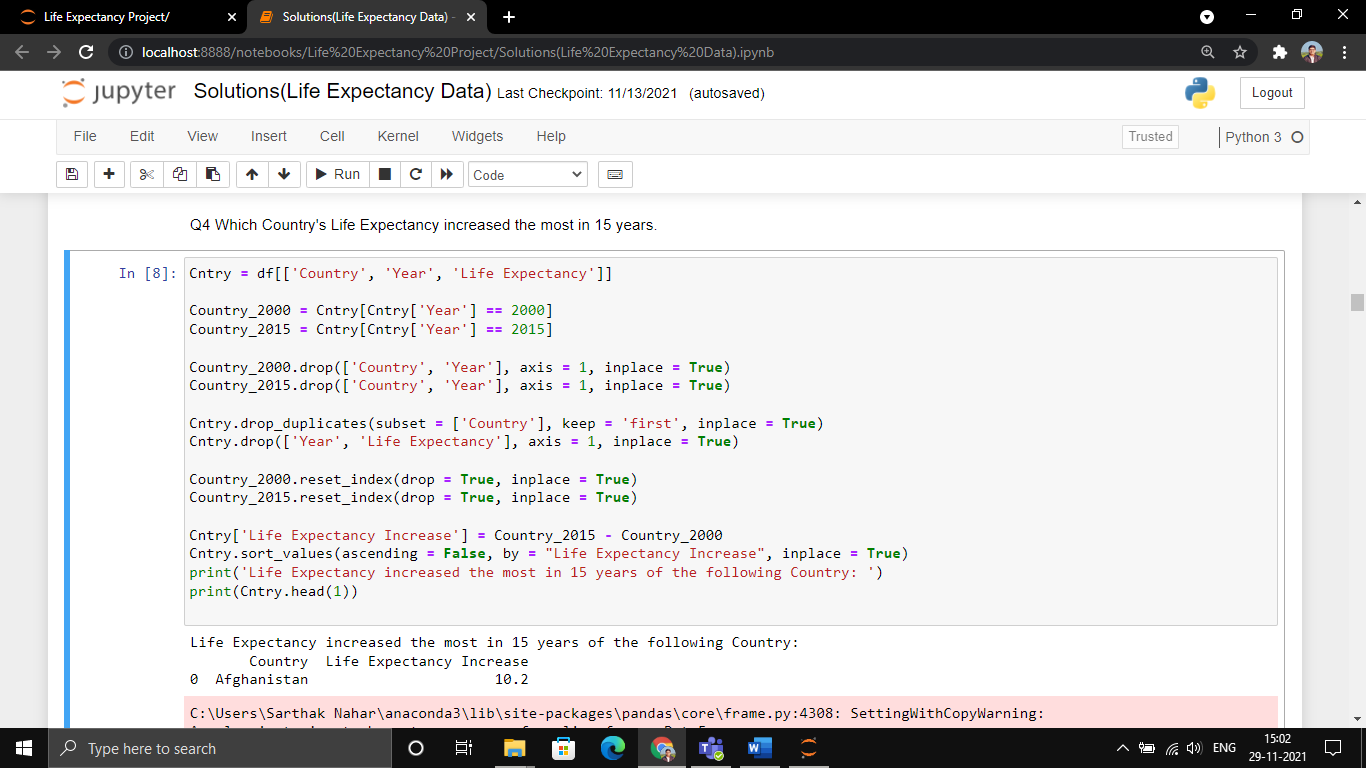
Que 2. Find the Correlation between GDP and Life Expectancy.



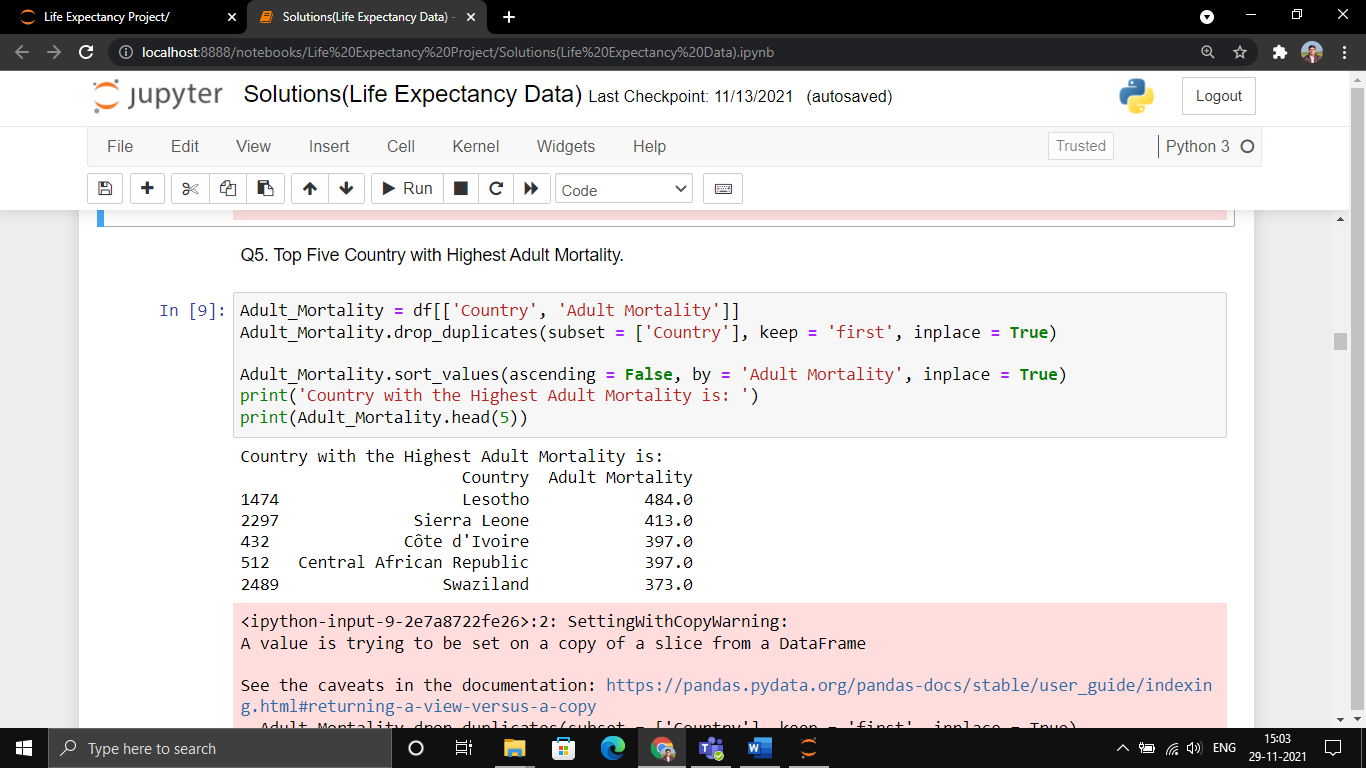
Que 3. Which Disease affect Life Expectancy the most.



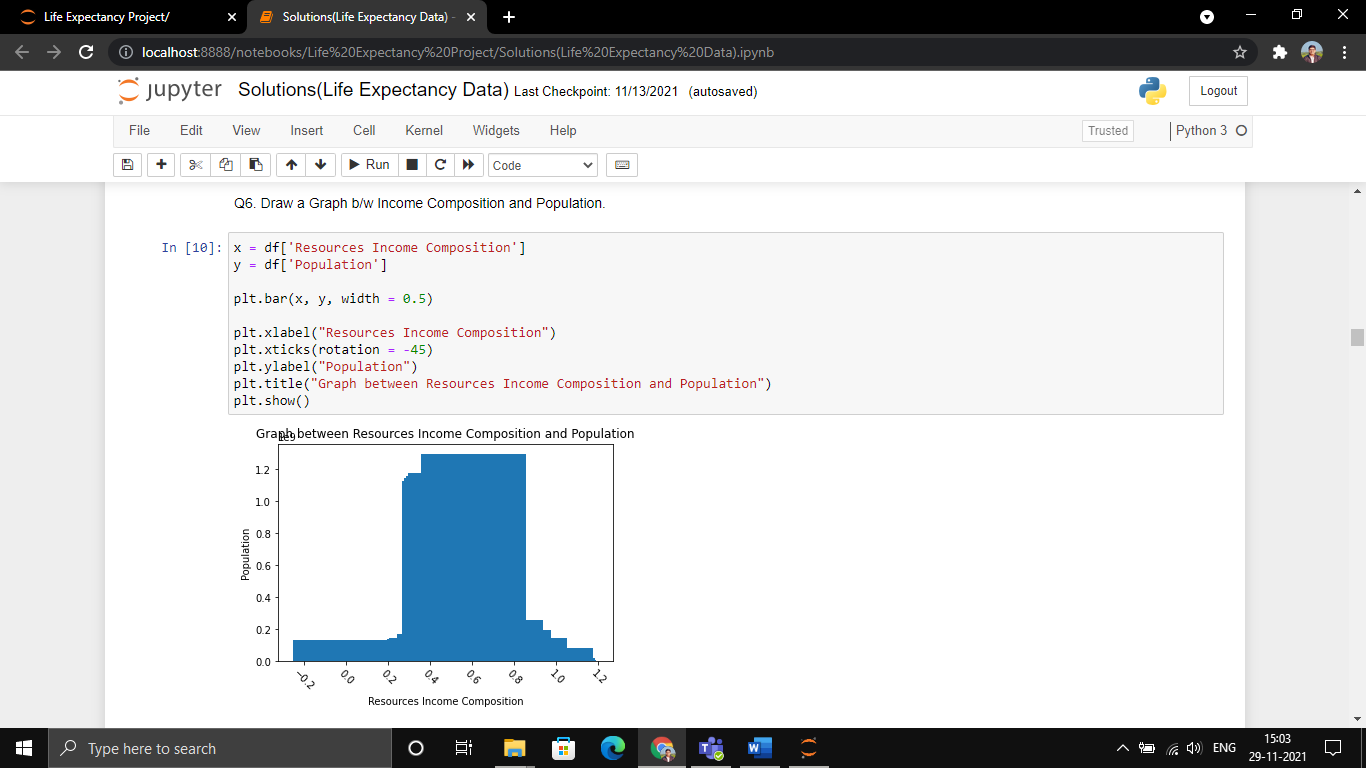
Que 4. Which Country's Life Expectancy increased the most in 15 years.



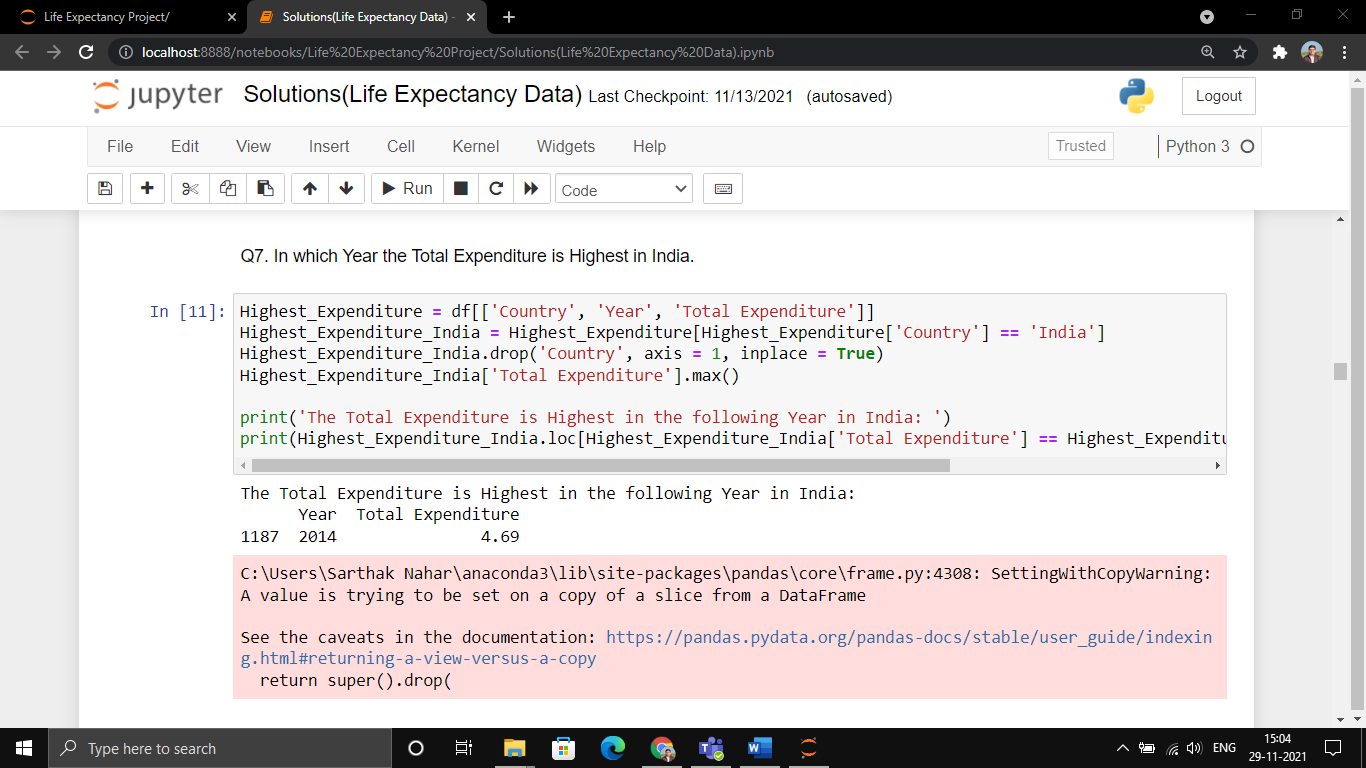
Que 5. Top Five Country with Highest Adult Mortality.



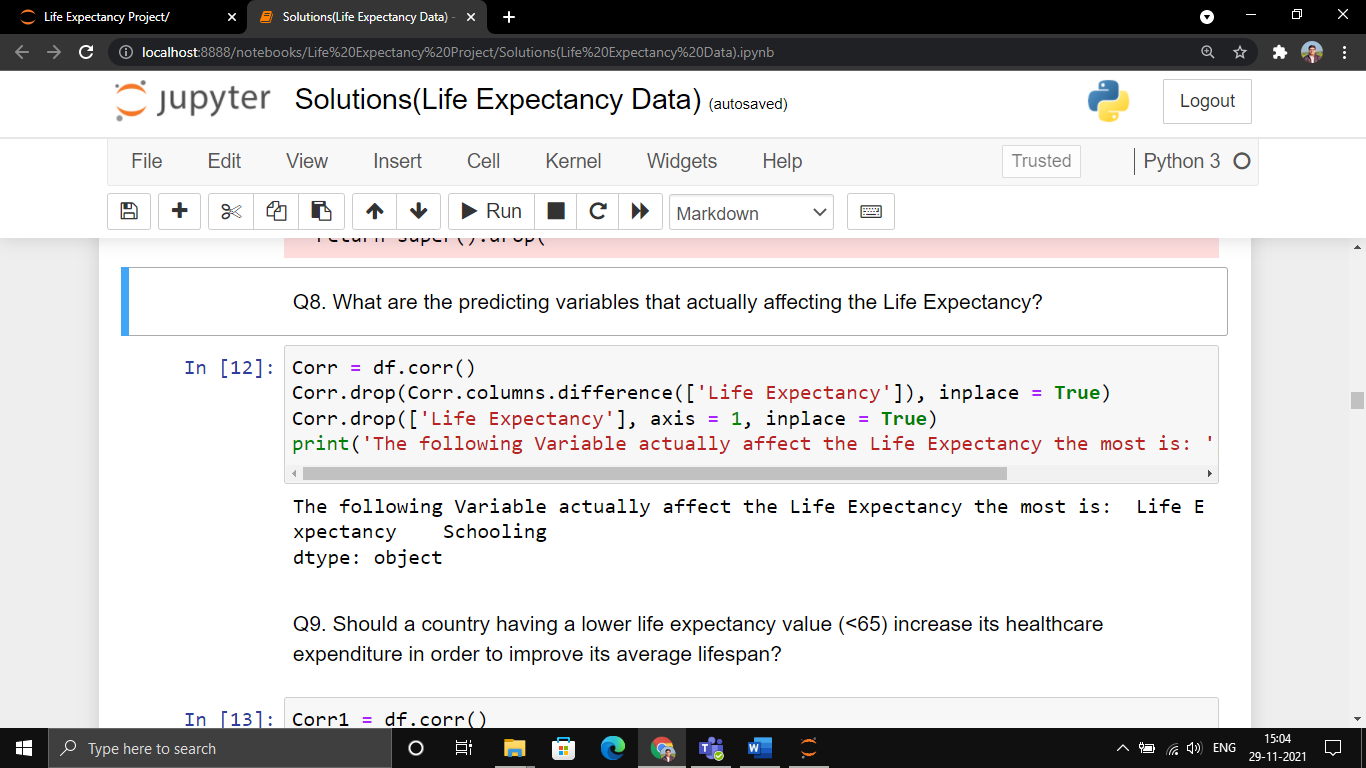
Que 6. Draw a Graph b/w Income Composition and Population.



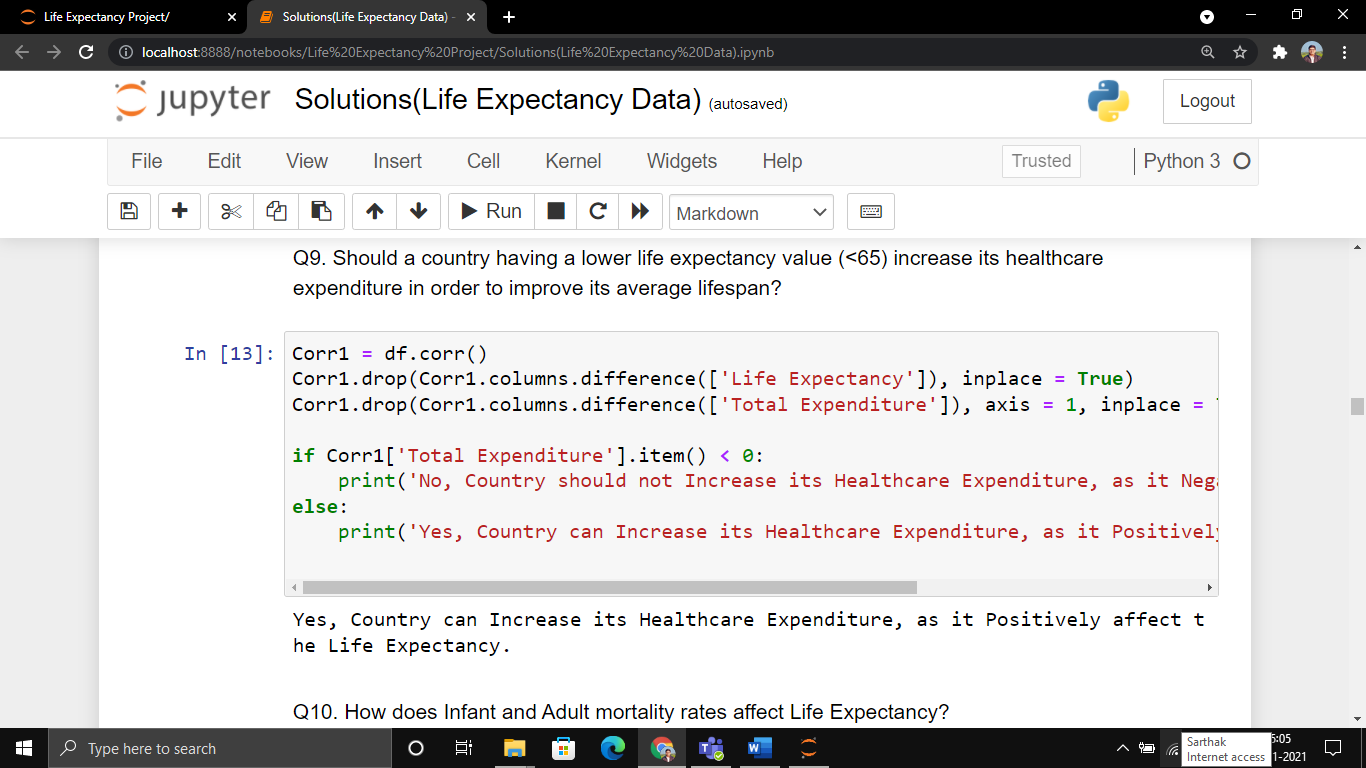
Que 7. In which Year the Total Expenditure is Highest in India.



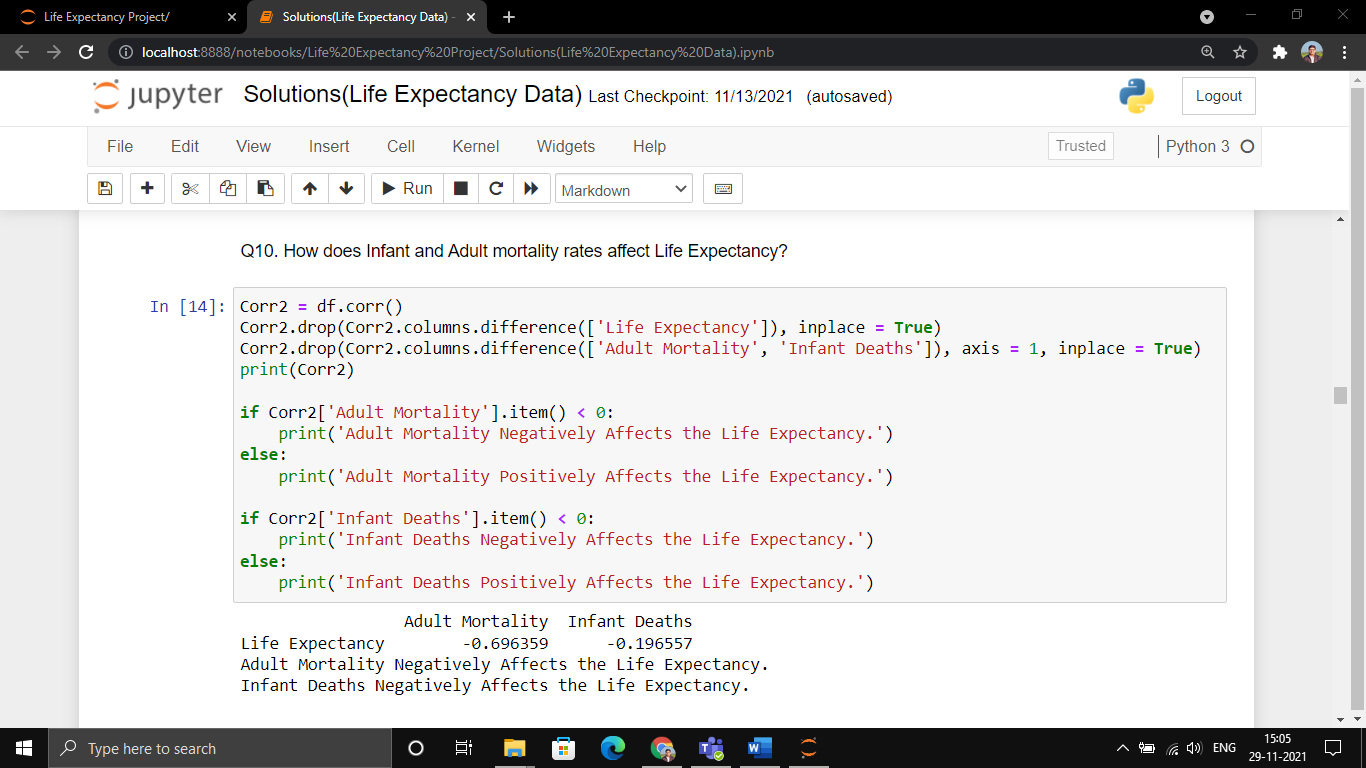
Que 8. What are the predicting variables that actually affecting the Life Expectancy?



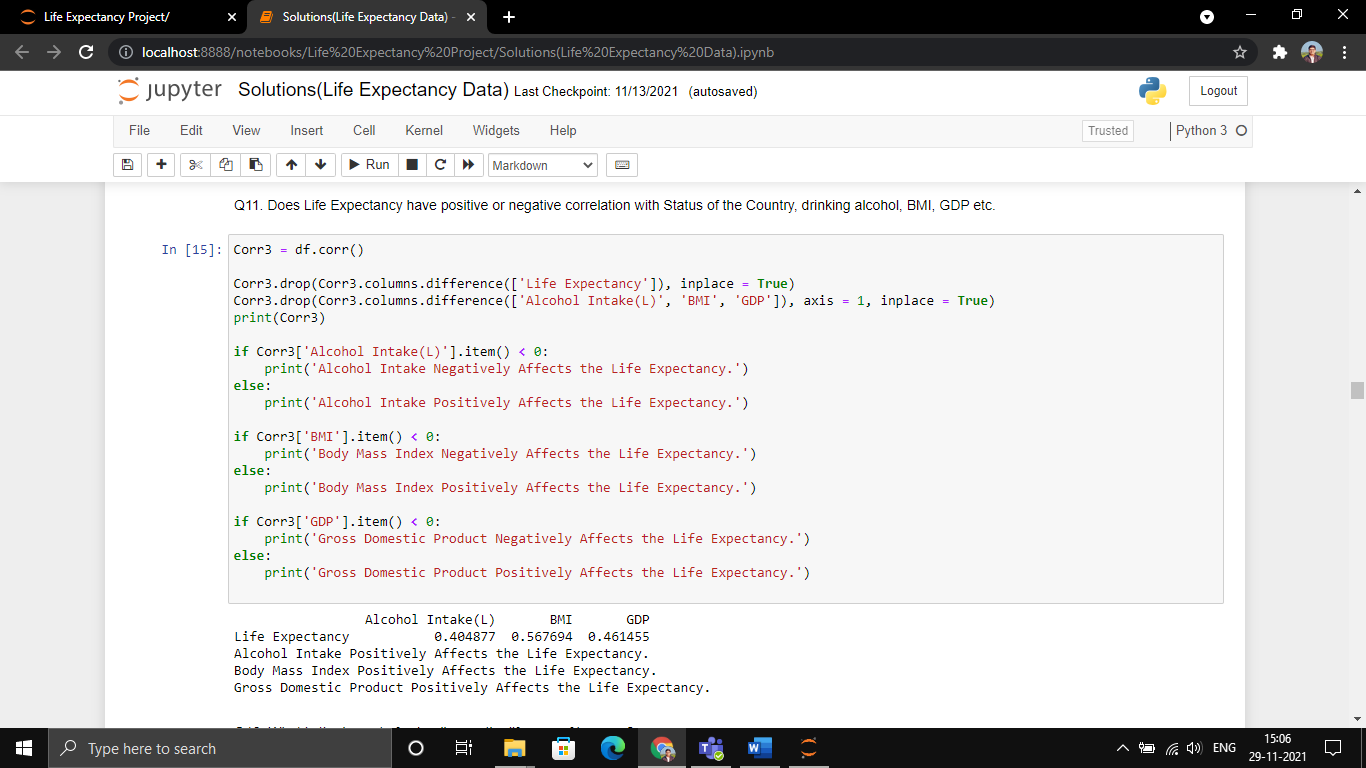
Que 9. Should a country having a lower life expectancy value (<65) increase its healthcare expenditure in order to improve its average lifespan?



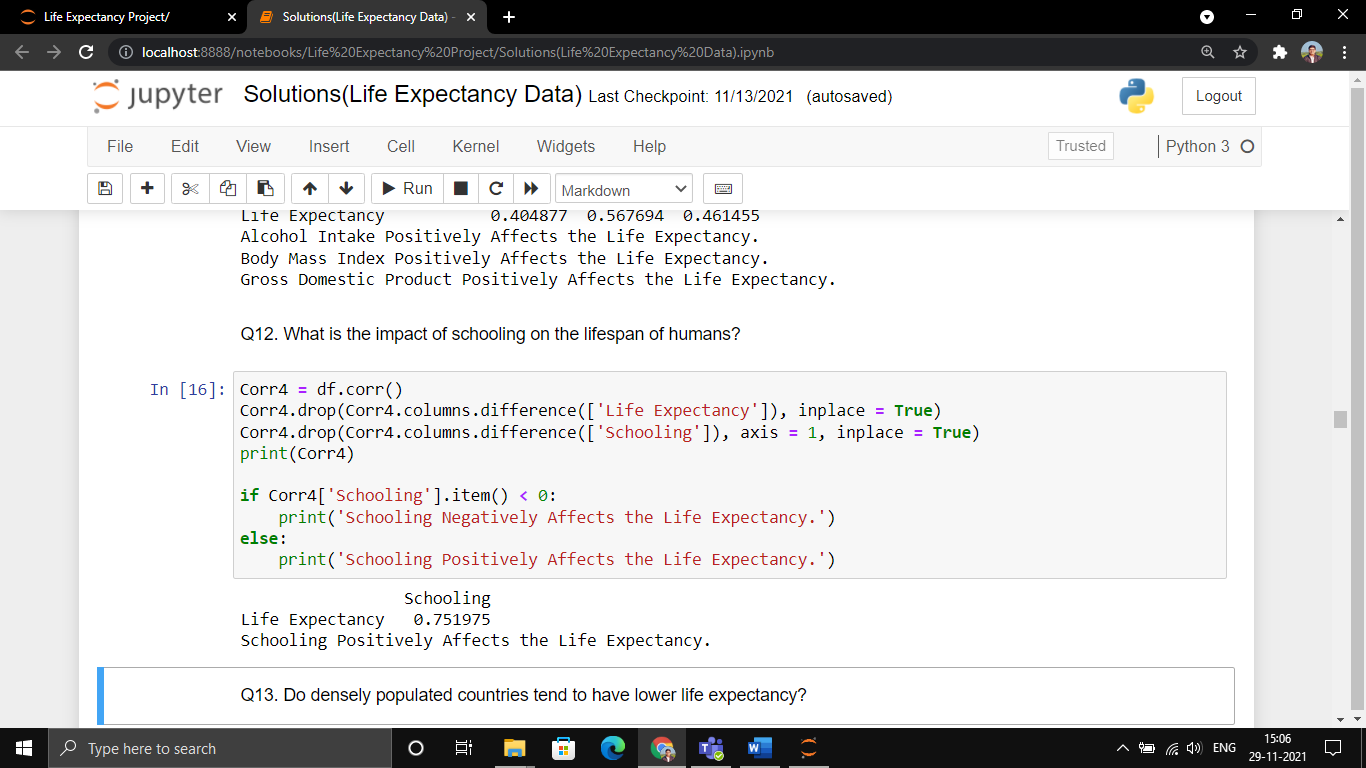
Que 10. How does Infant and Adult mortality rates affect Life Expectancy?



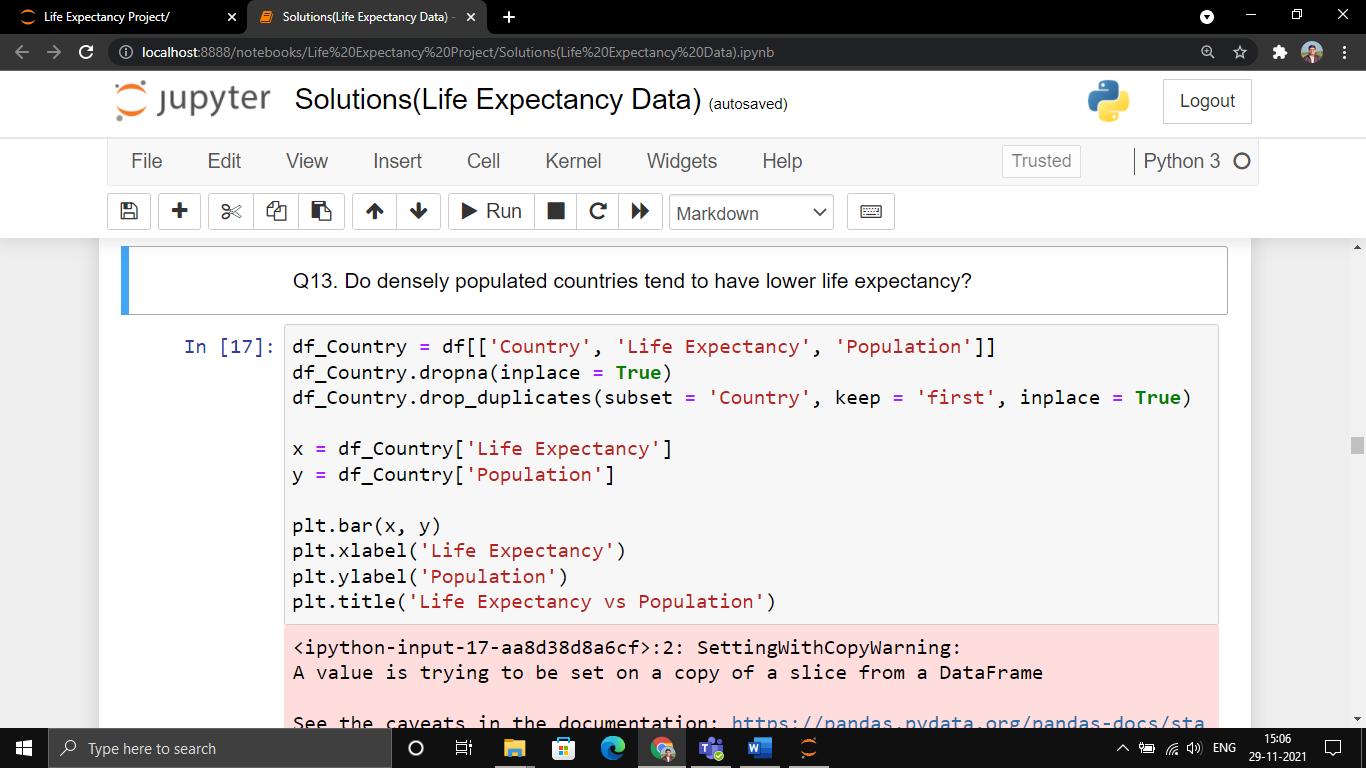
Que 11. Does Life Expectancy have positive or negative correlation with Status of the Country, drinking alcohol, BMI, GDP etc.

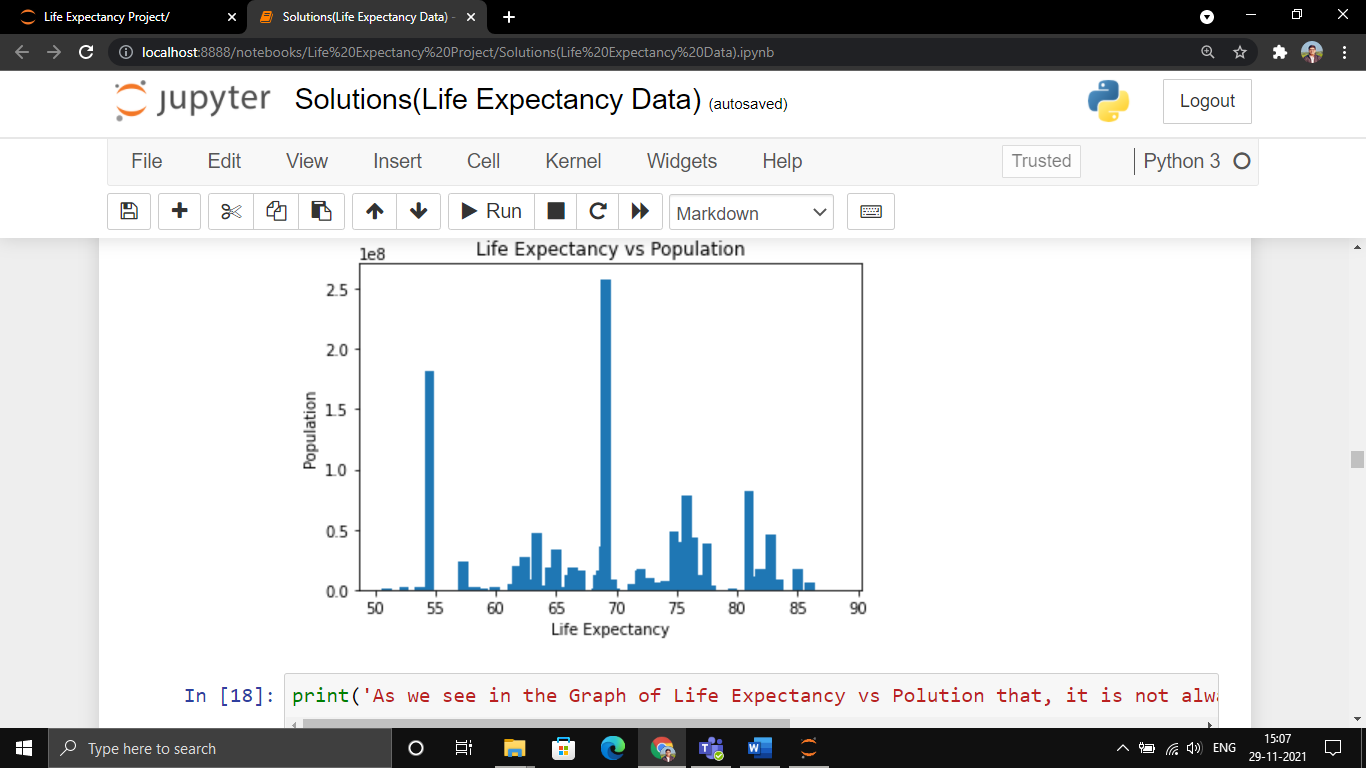


Que 12. What is the impact of schooling on the lifespan of humans?

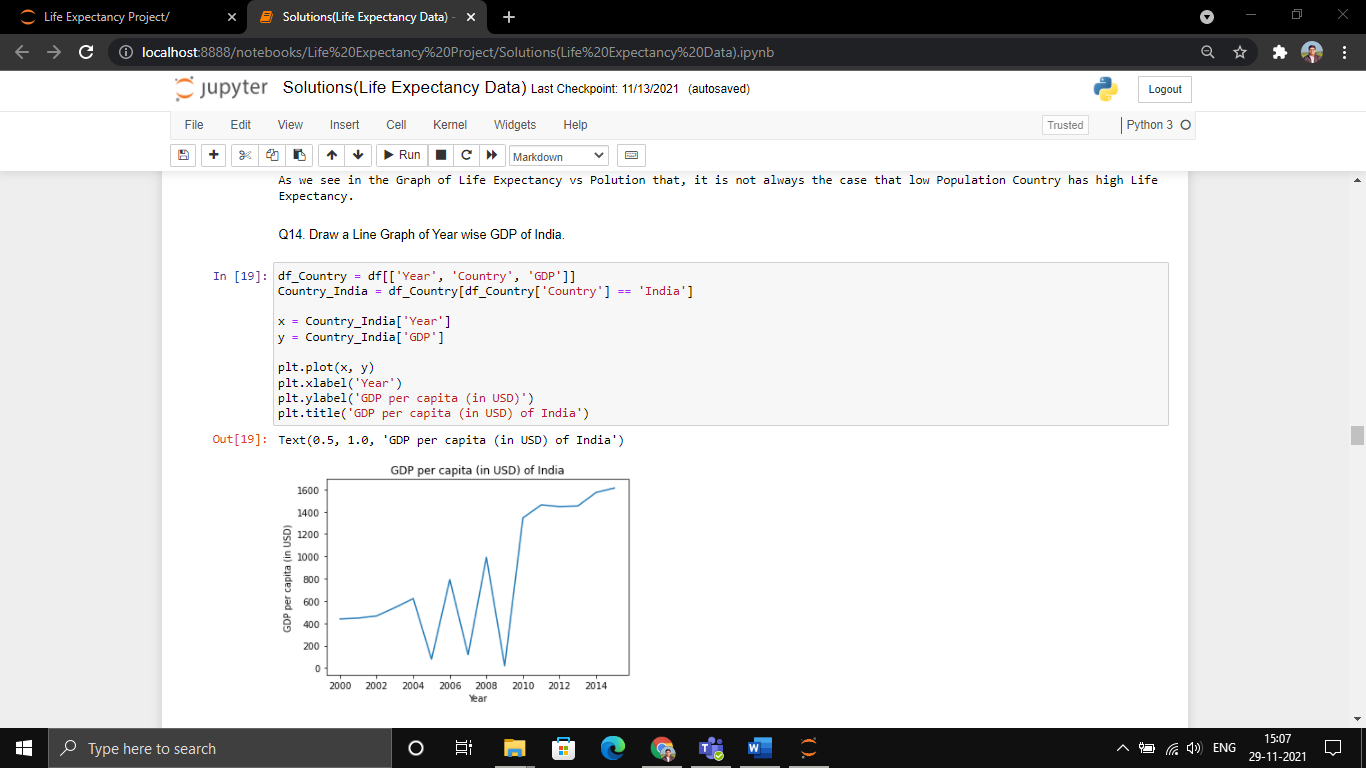


Que 13. Do densely populated countries tend to have lower life expectancy?

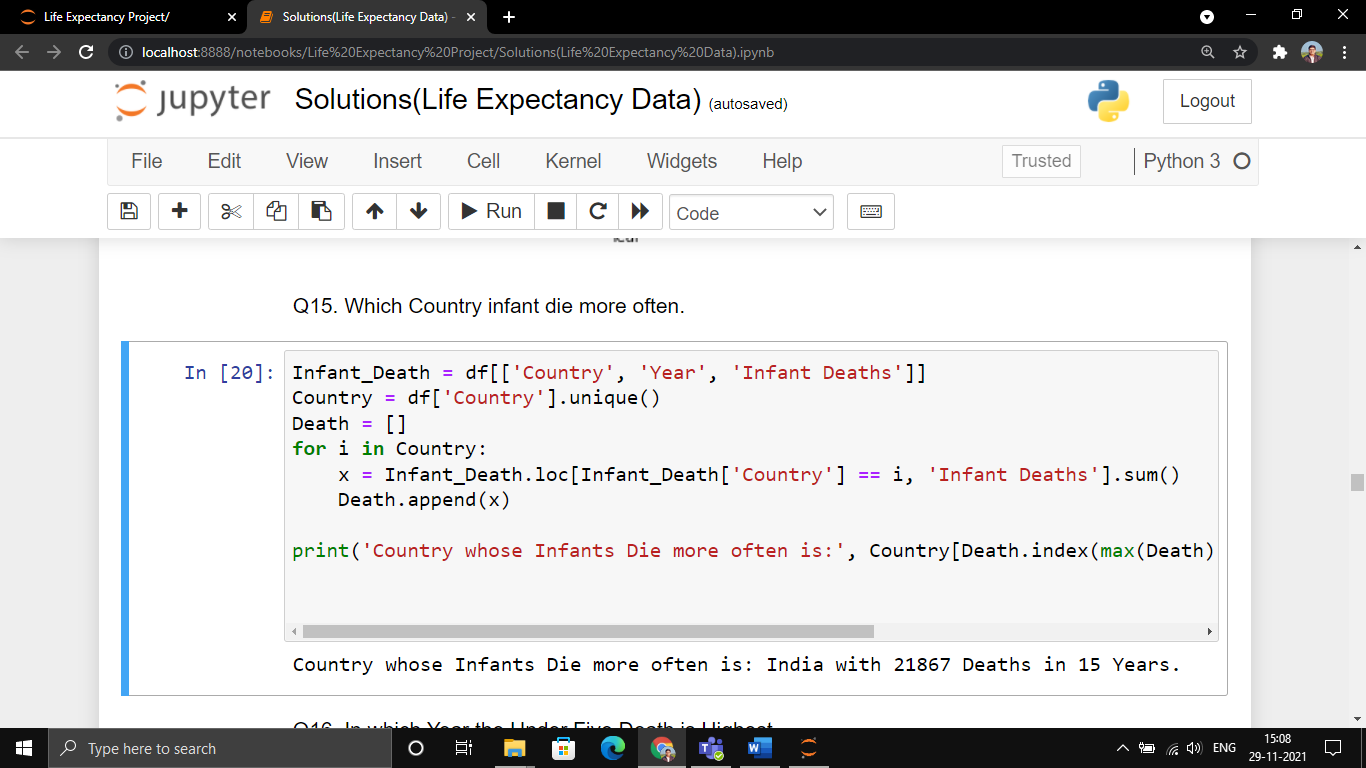




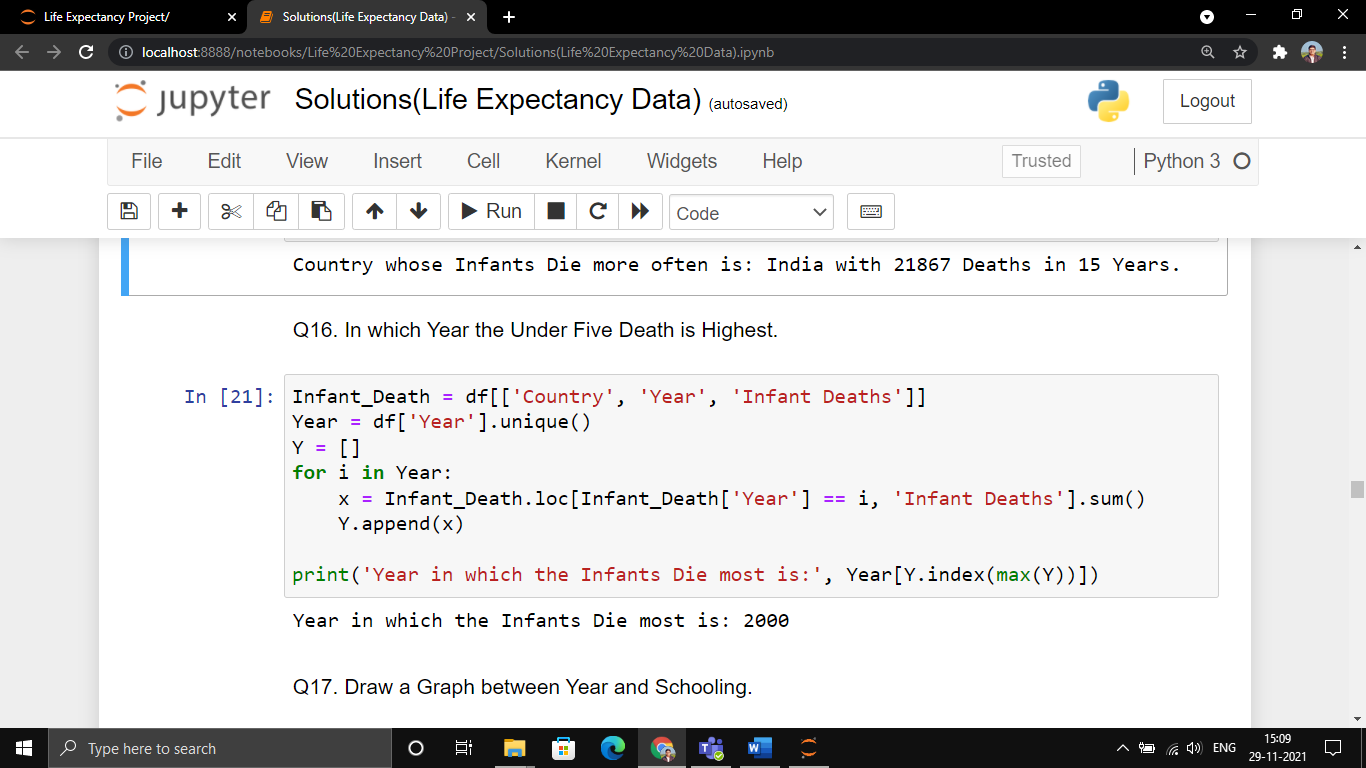
Que 14. Draw a Line Graph of Year wise GDP of India.



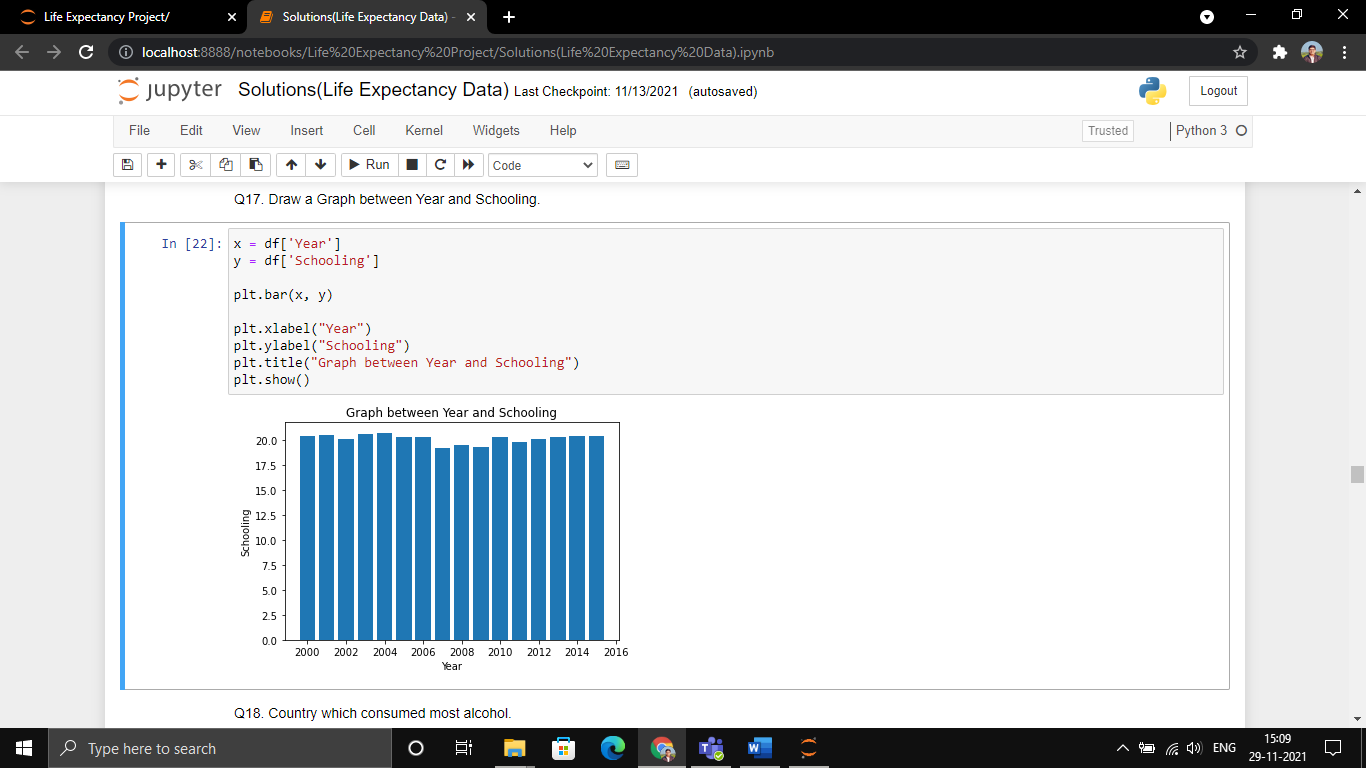
Que 15. Which Country infant die more often.



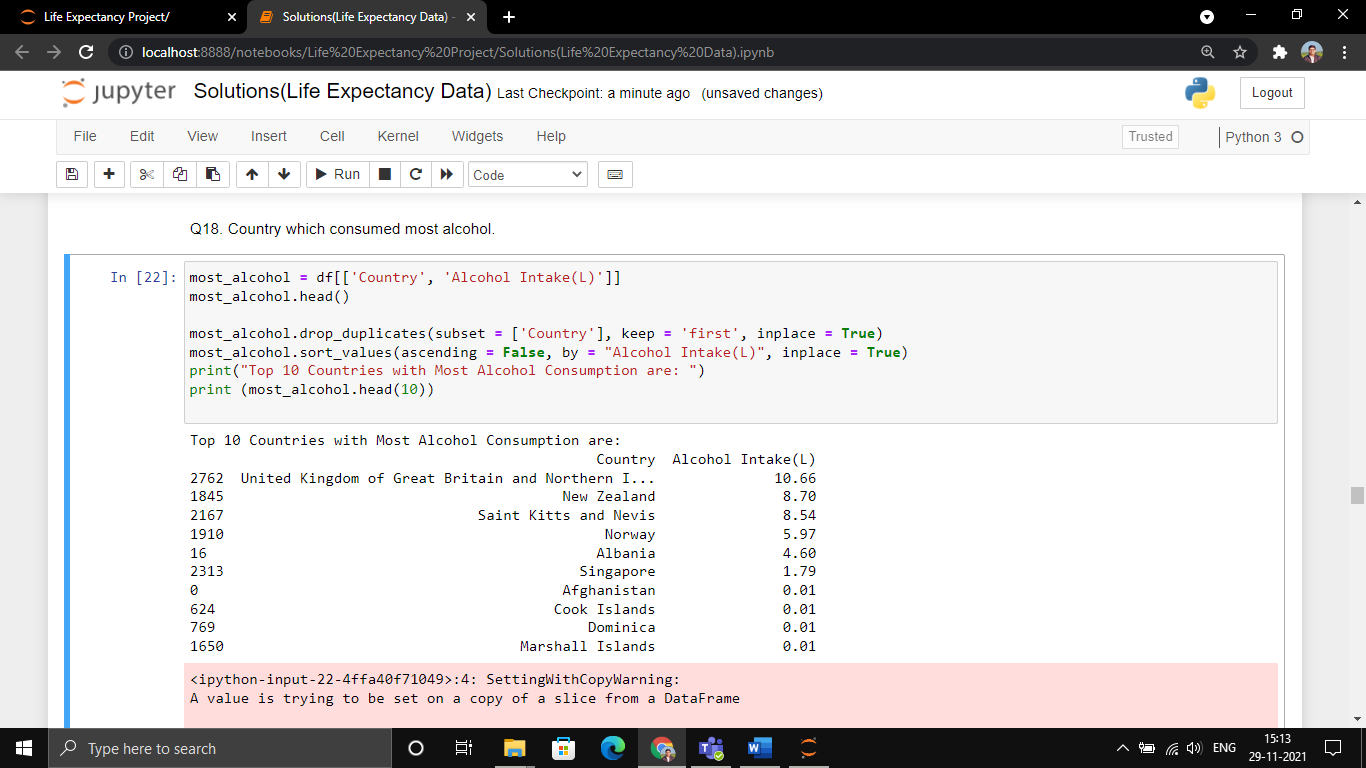
Que 16. In which Year the Under Five Death is Highest.



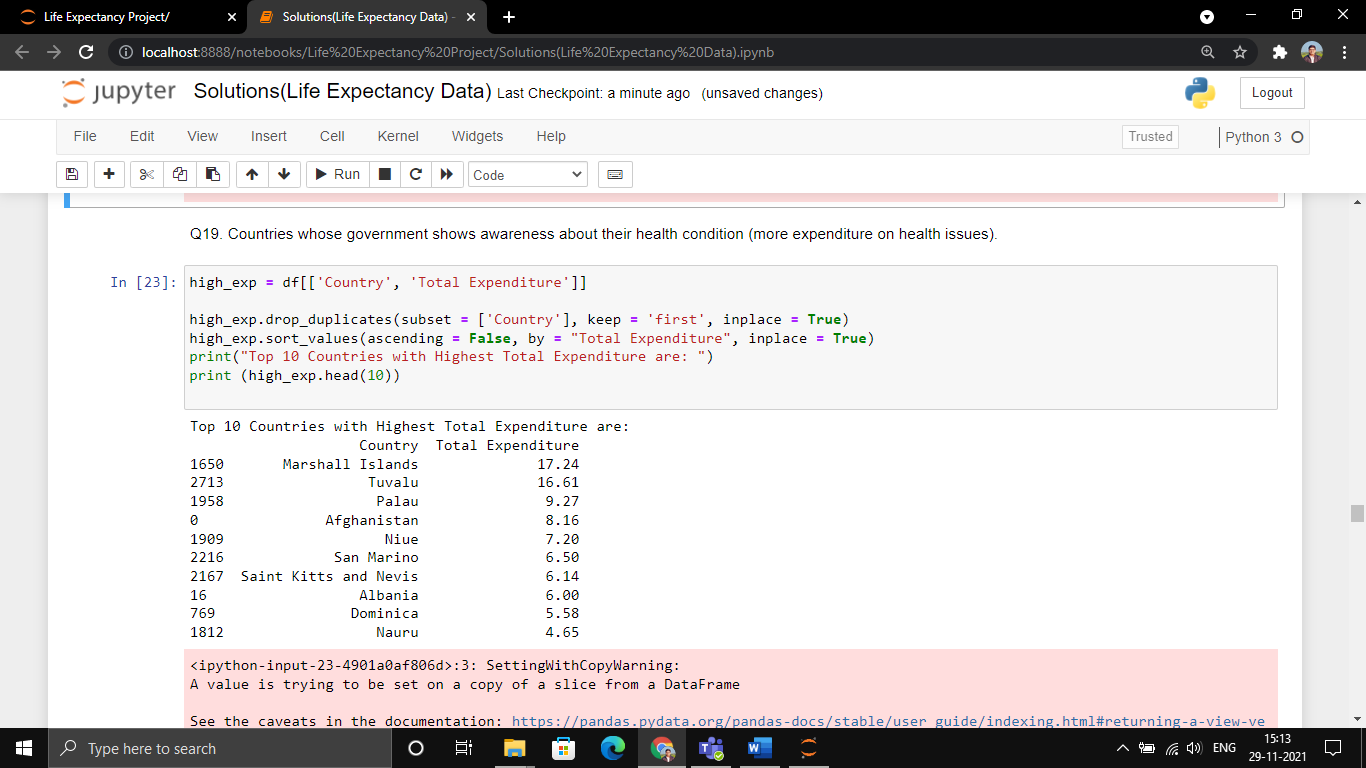
Que 17. Draw a Graph between Year and Schooling.



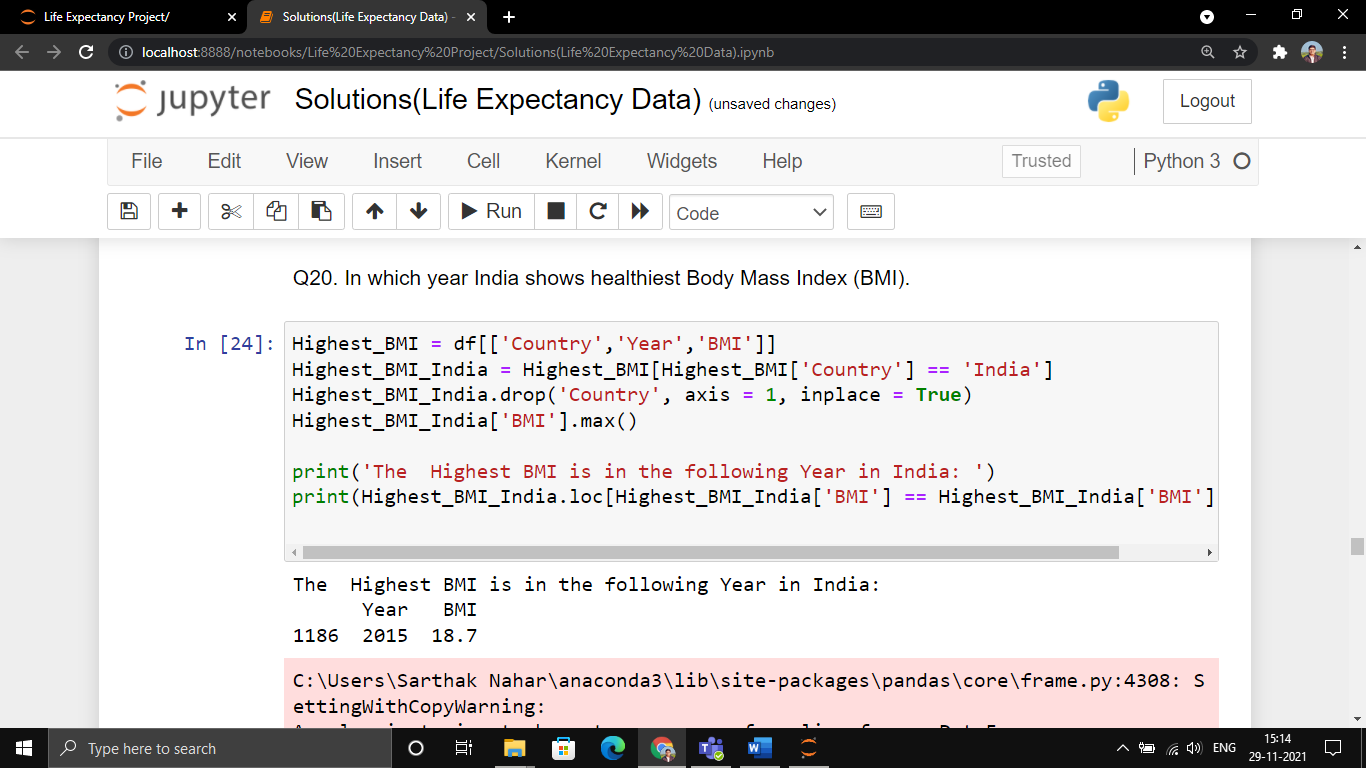
Que 18. Country which consumed most alcohol.



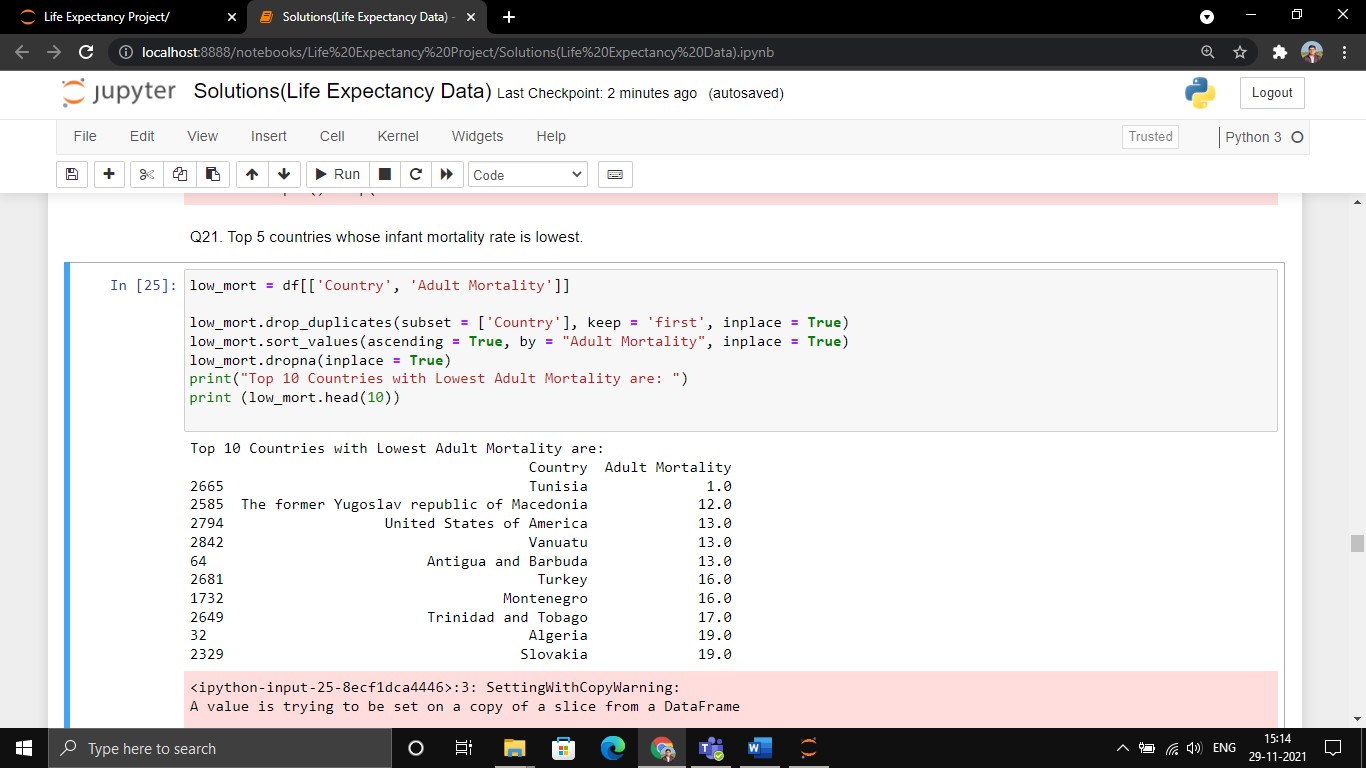
Que 19. Countries whose government shows awareness about their health condition (more expenditure on health issues).



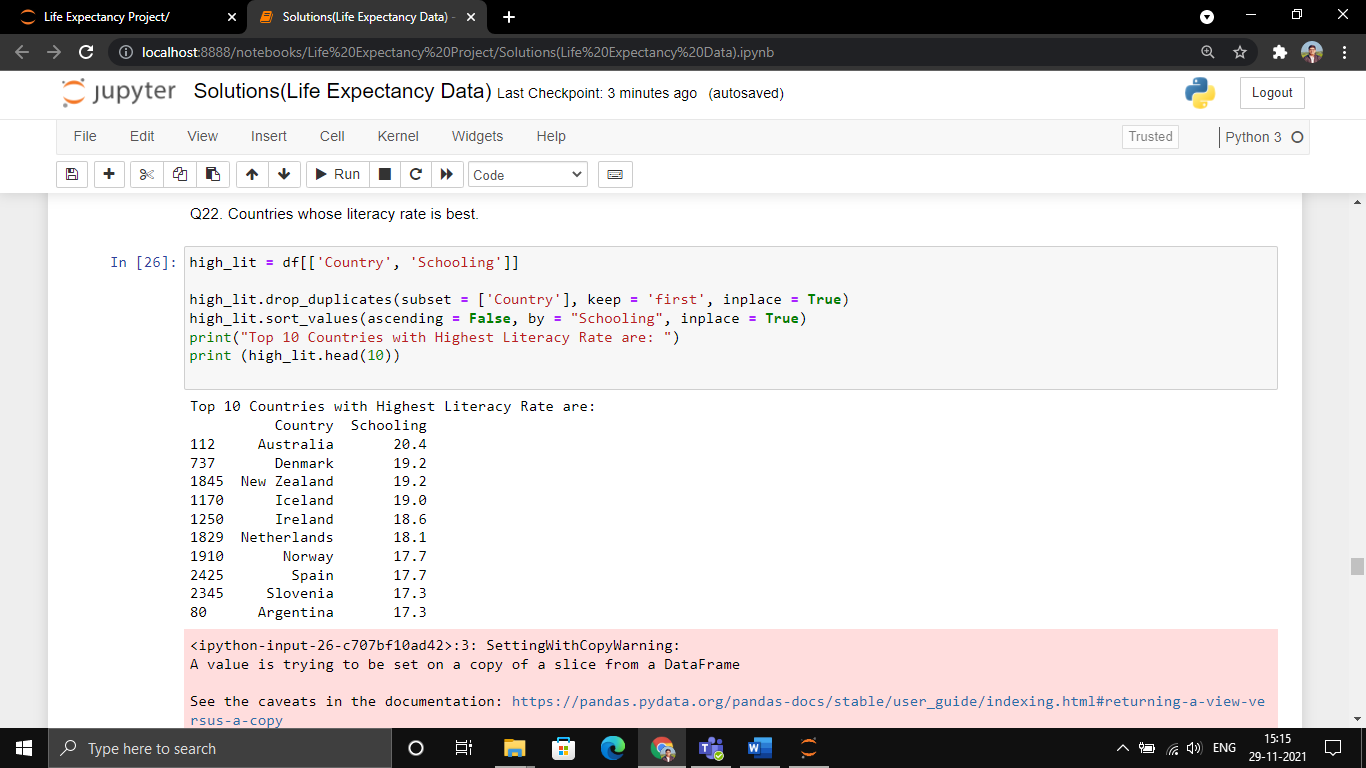
Que 20. In which year India shows healthiest Body Mass Index (BMI).



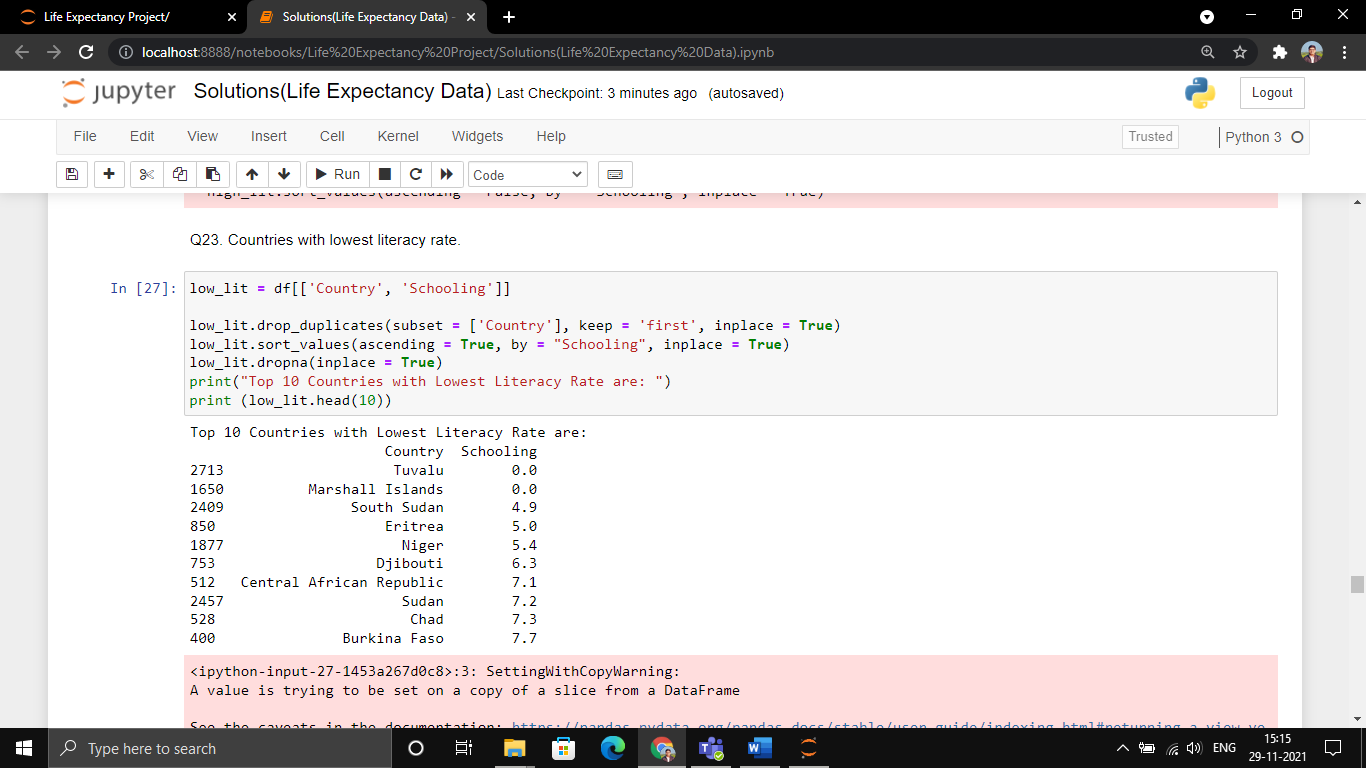
Que 21. Top 5 countries whose infant mortality rate is lowest.



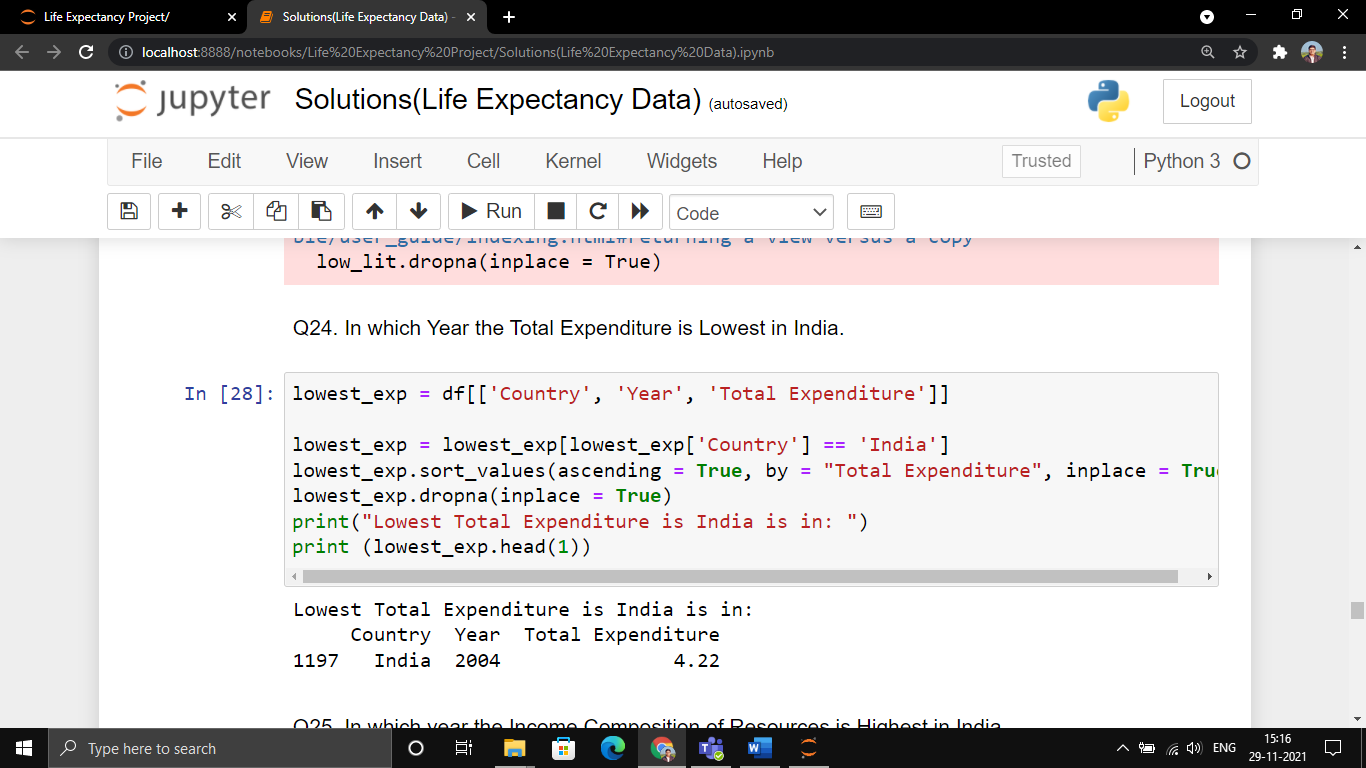
Que 22. Countries whose literacy rate is best.



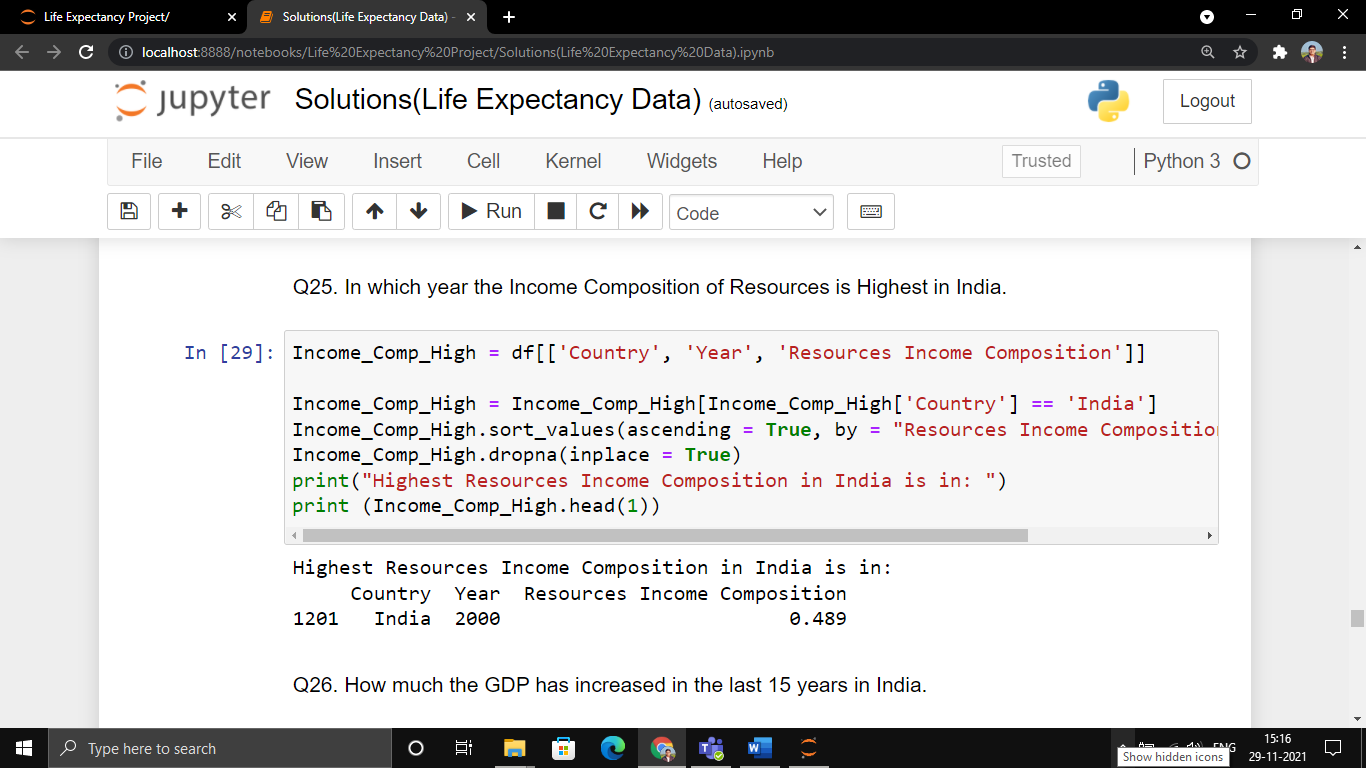
Que 23. Countries with lowest literacy rate.



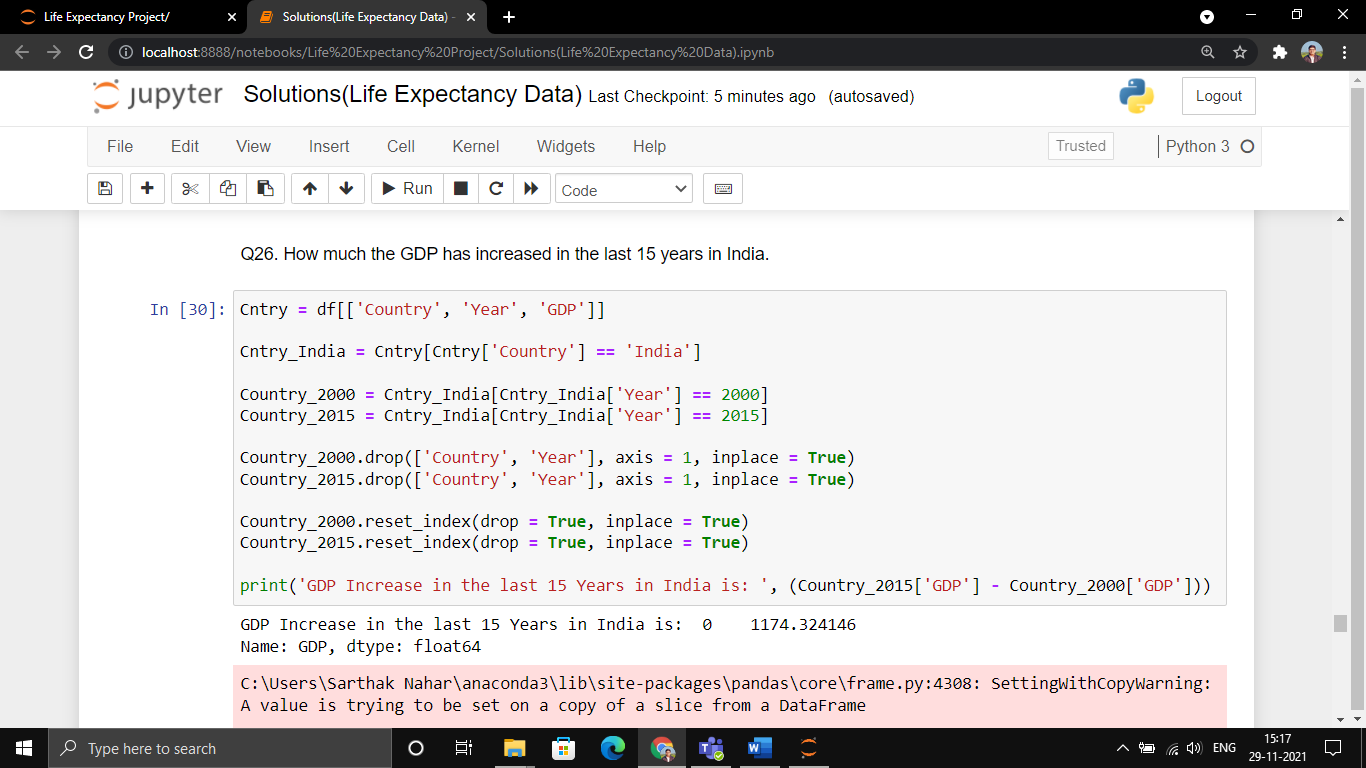
Que 24. In which Year the Total Expenditure is Lowest in India.



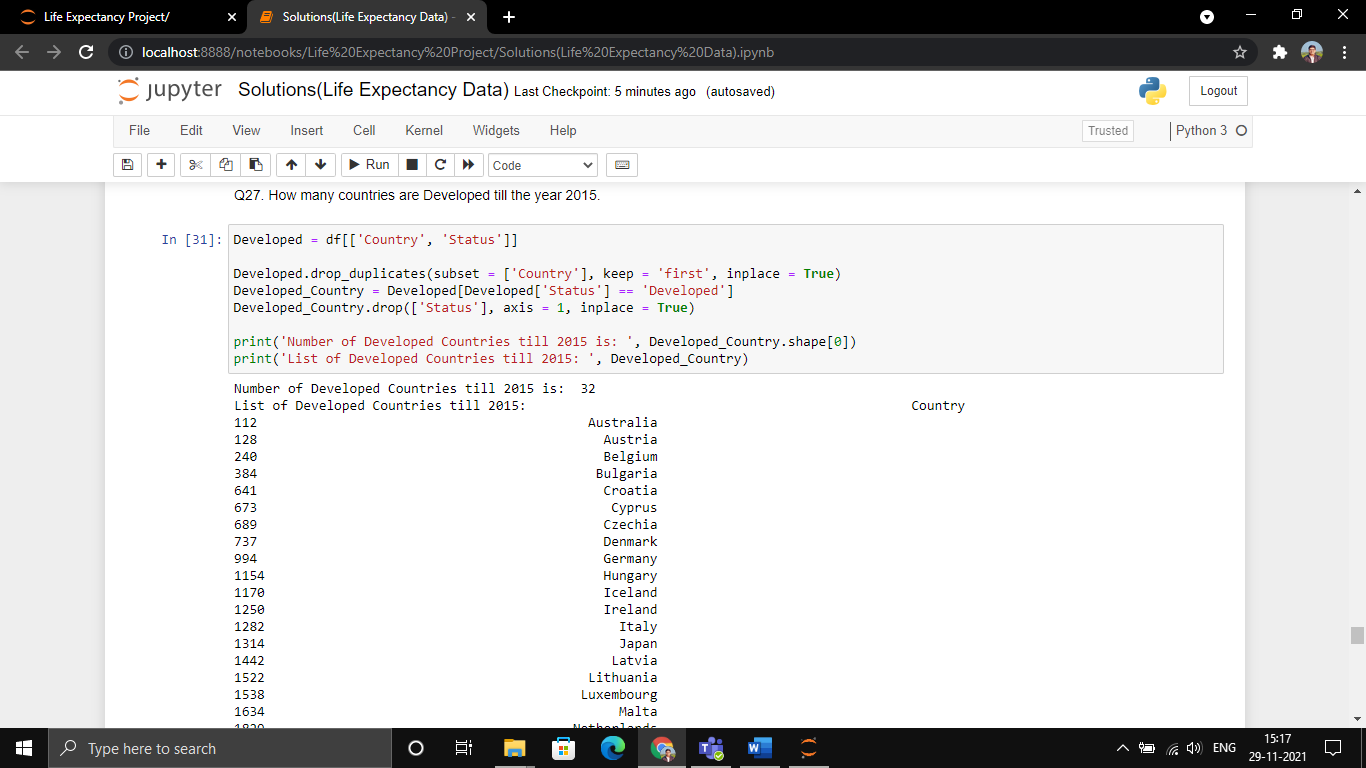
Que 25. In which year the Income Composition of Resources is Highest in India.

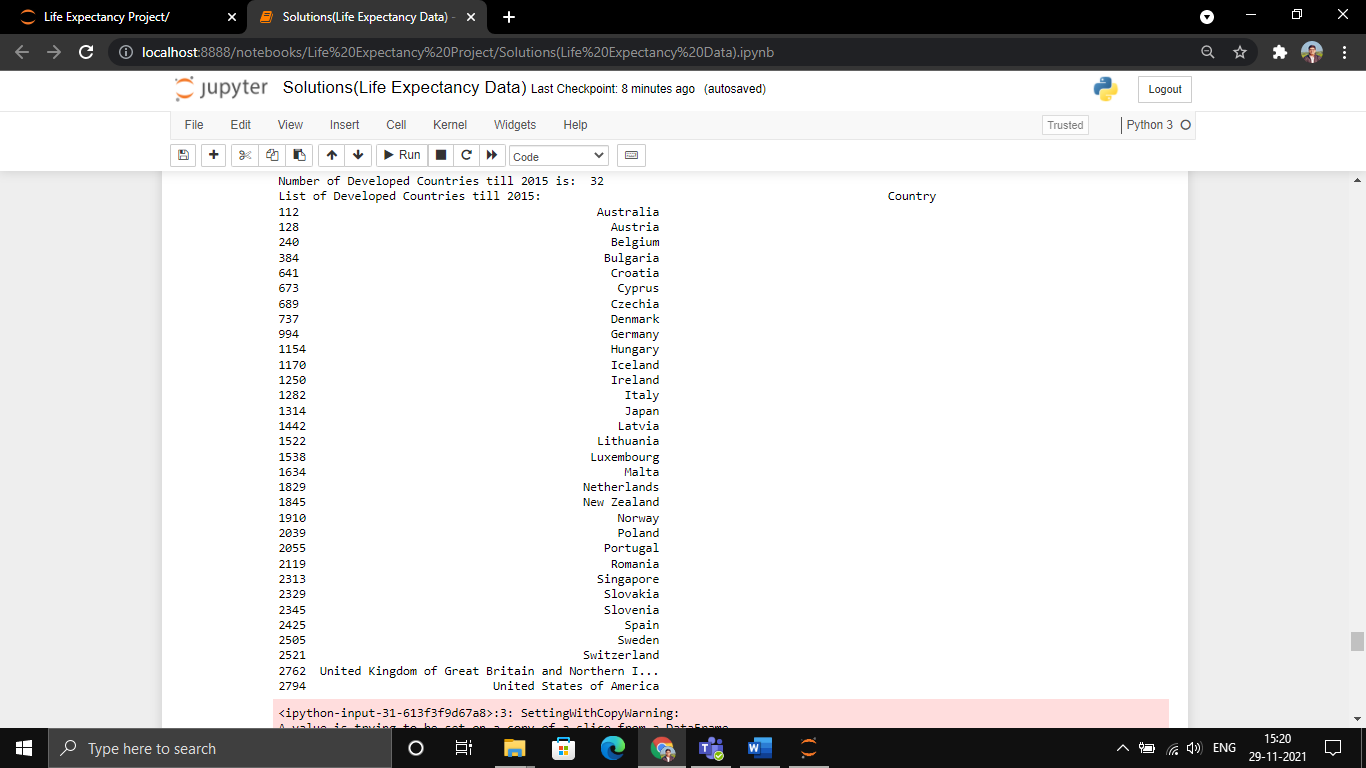


Que 26. How much the GDP has increased in the last 15 years in India.

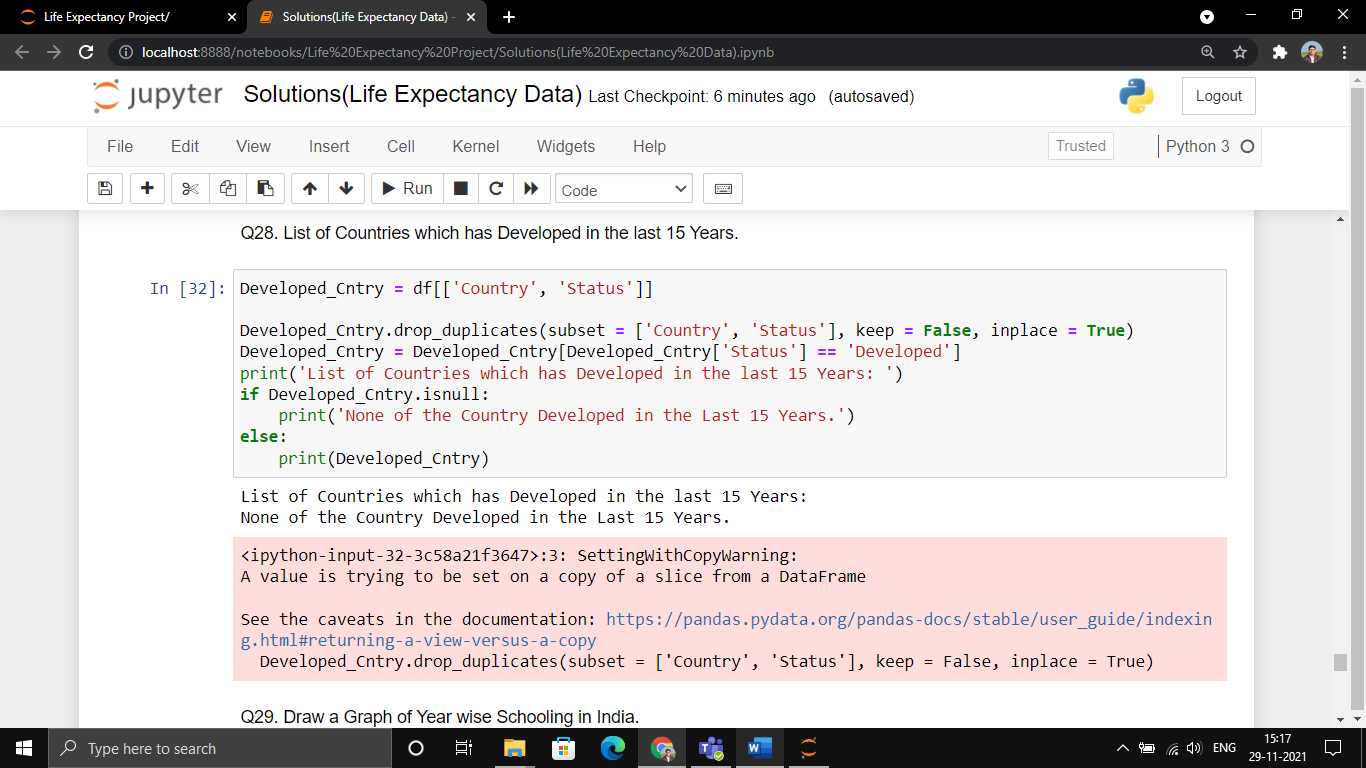


Que 27. How many countries are Developed till the year 2015.

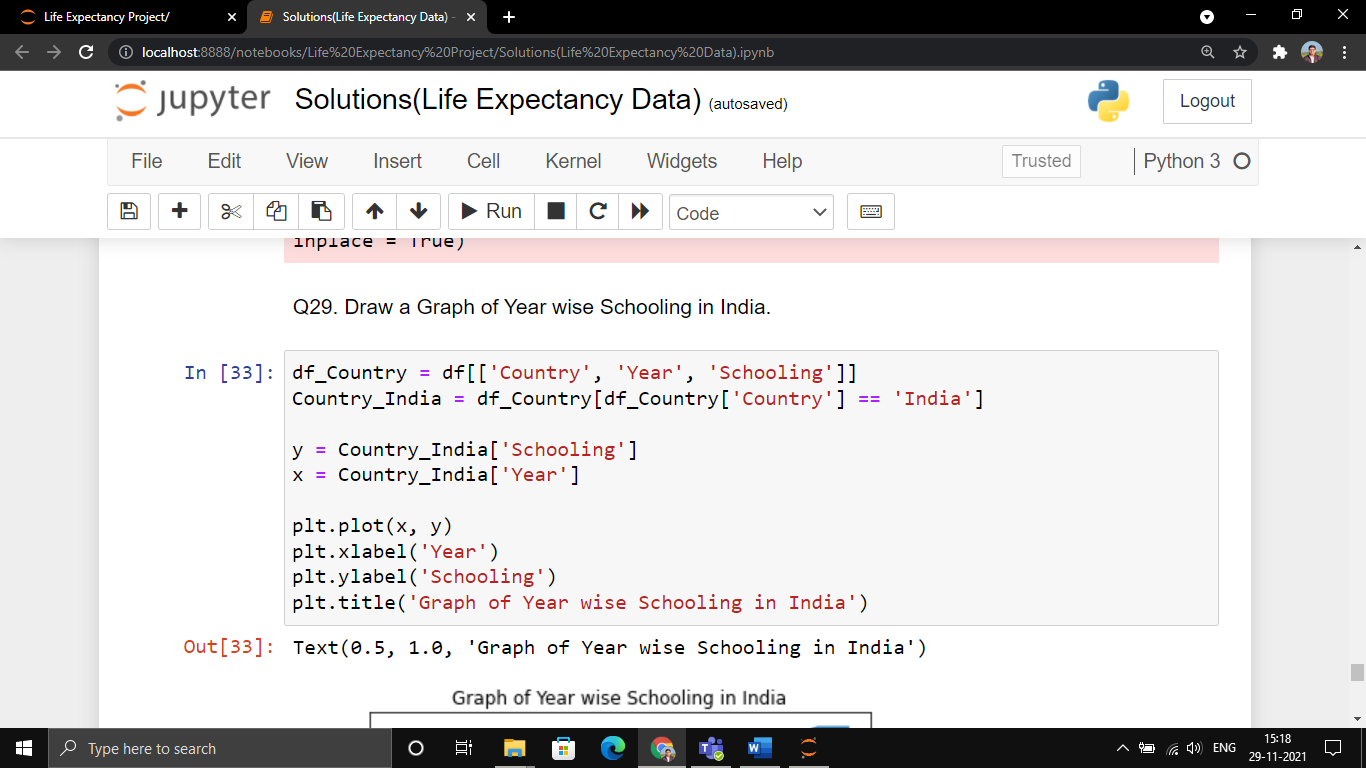


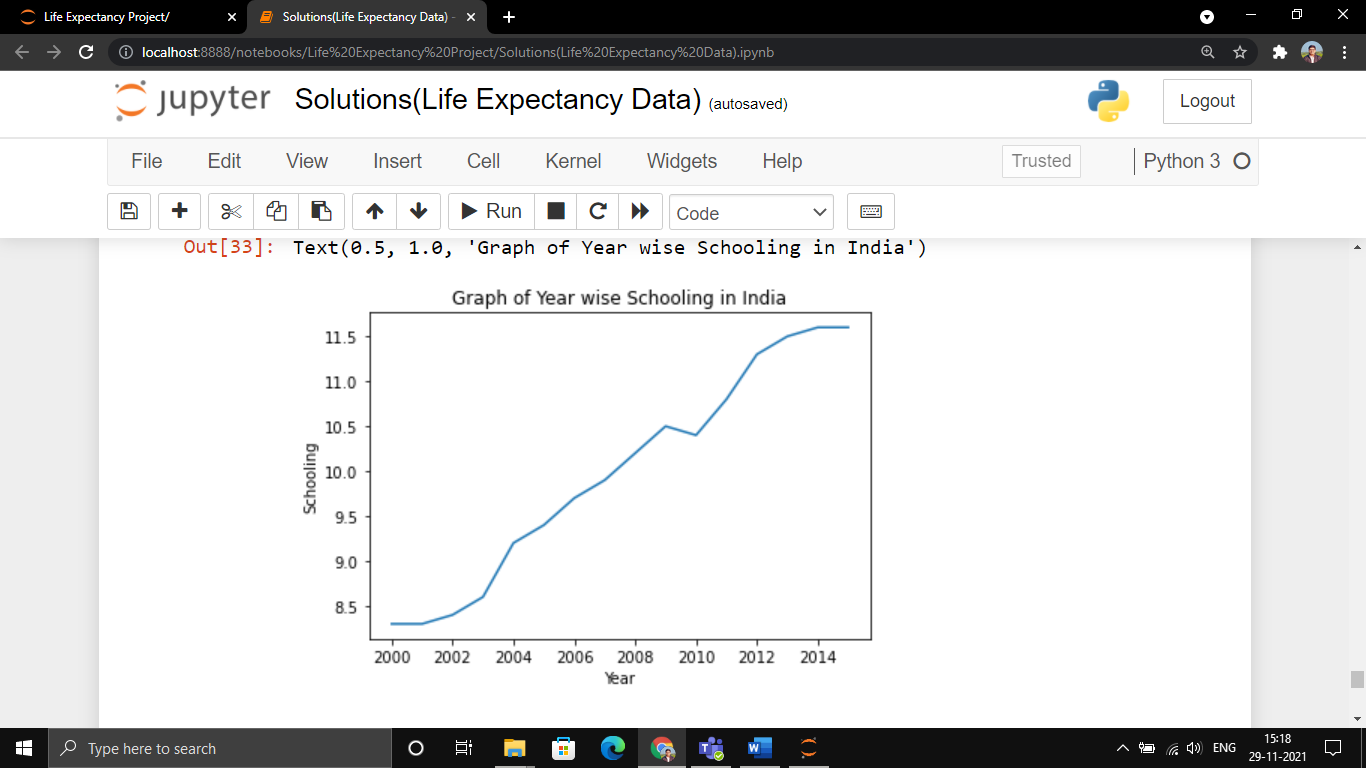


Que 28. List of Countries which has Developed in the last 15 Years.



Que 29. Draw a Graph of Year wise Schooling in India.





Que 30. Draw a Graph of Life Expectancy of different Countries.

