**PROJECT REPORT**

PREDICTING LIFE EXPECTANCY USING MACHINE LEARNING

BY

MILAN VISHNOI

For Completion of Internship

With

SMARTINTERNZ

Internship ID: Predicting Life Expectancy using Machine Learning - SB52231

Project ID: SPS\_PRO\_215

**CHAPTER 1**

**INTRODUCTION**

* 1. **OVERVIEW**

A typical Regression Machine Learning project leverages historical data to predict insights into the future. This problem statement is aimed at predicting Life Expectancy rate of a country given various features.

Life expectancy is a statistical measure of the average time a human being is expected to live, Life expectancy depends on various factors: Regional variations, Economic Circumstances, Sex Differences, Mental Illnesses, Physical Illnesses, Education, Year of their birth and other demographic factors. This problem statement provides a way to predict average life expectancy of people living in a country when various factors such as year, GDP, education, alcohol intake of people in the country, expenditure on healthcare system and some specific disease related deaths that happened in the country are given.

In my project, I have made Machine learning Model which can predict the average life expectancy of an average person living in a country by taking as input various features and statistical data related to that country. The data includes adult mortality, average alcohol consumption, percentage immunization of 1-year-olds for various diseases, government expenditure percentage on health of total government expenditure and average schooling in terms of number of years.

* 1. **PURPOSE**

The main purpose for building this project is to predict the life expectancy of a citizen living in country by using different statistical data of that country. We can also alter the data to check which parameter has the most impact on average life expectancy of a person living in a particular country and hence can focus more on improving that thing/factor of that country to increase the average life expectancy.

**CHAPTER 2**

**LITERATURE SURVEY**

**2.1 EXISTING PROBLEM**

Currently the existing problem is that the countries don’t know exactly where to invest in and what can be improved to improve the life expectancy of its citizens. We need to design a model with the help of Machine Learning which can tell the life expectancy of a nation by using the historical data related to it.

**2.2 PROPOSED SOLUTION**

As we have the data related to various countries and the life expectancy in those countries, we can leverage it for the purpose of training a model based on linear regression. We can use the AUTO AI service of IBM Watson Studio or can design our own model using Jupyter Notebook in IBM Watson Studio, both the services will use Watson Machine Learning Service and then we can deploy our model. Using Node-Red we can design UI for our project. If anyone wants to predict the life expectancy of a particular country then he/she needs to enter the inputs and the output will be displayed on the screen in terms of number of years.

**CHAPTER 3**

**THEORITICAL ANALYSIS**

**3.1 BLOCK DIAGRAM**

**Node Red UI**

**Watson Studio**

**Machine Learning**

**AUTO AI/ JUPYTER NOTEBOOK DEPLOYMENT MODEL**

**Input**

**Output**

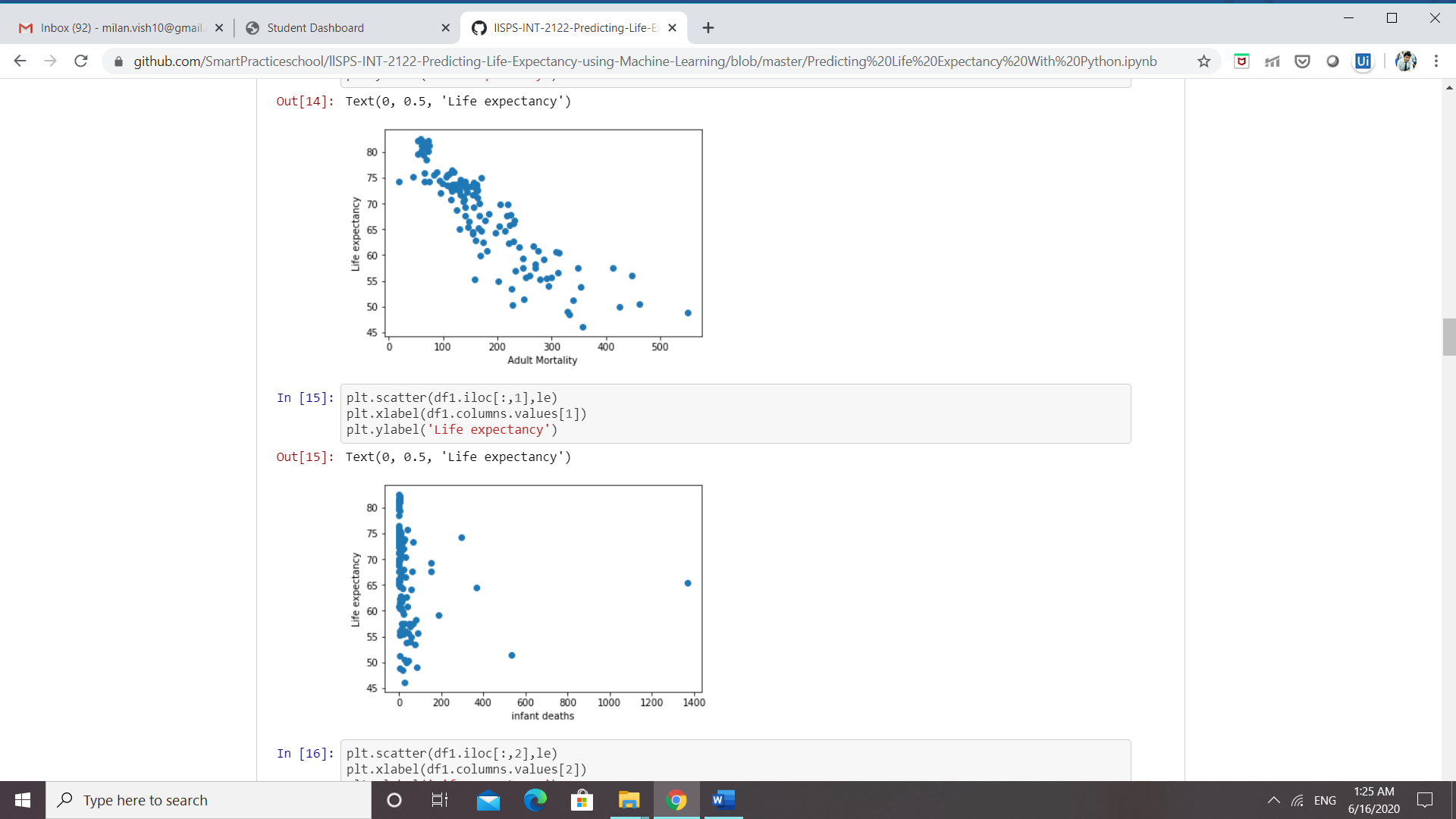
**3.2 SOFTWARE DESIGNING**

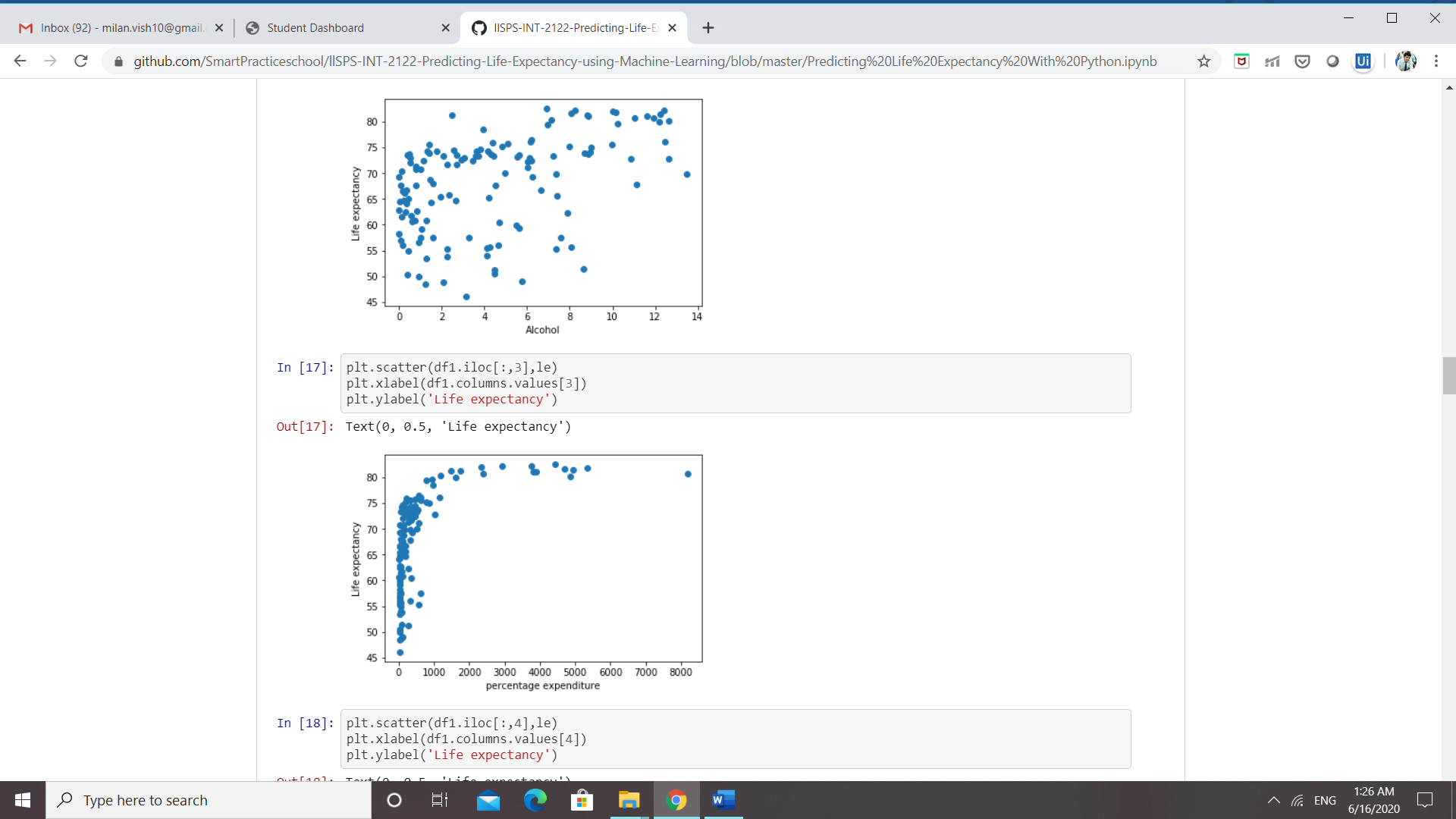
I have used IBM cloud for the purpose of implementation of my project. IBM Watson Studio along with its services is used. Node-Red is used for the purpose of designing the UI and interacting with the model. AUTO AI along with Watson Machine Learning Service can be used to get a model by just providing the data or Jupyter Notebook can be used for designing our own model.

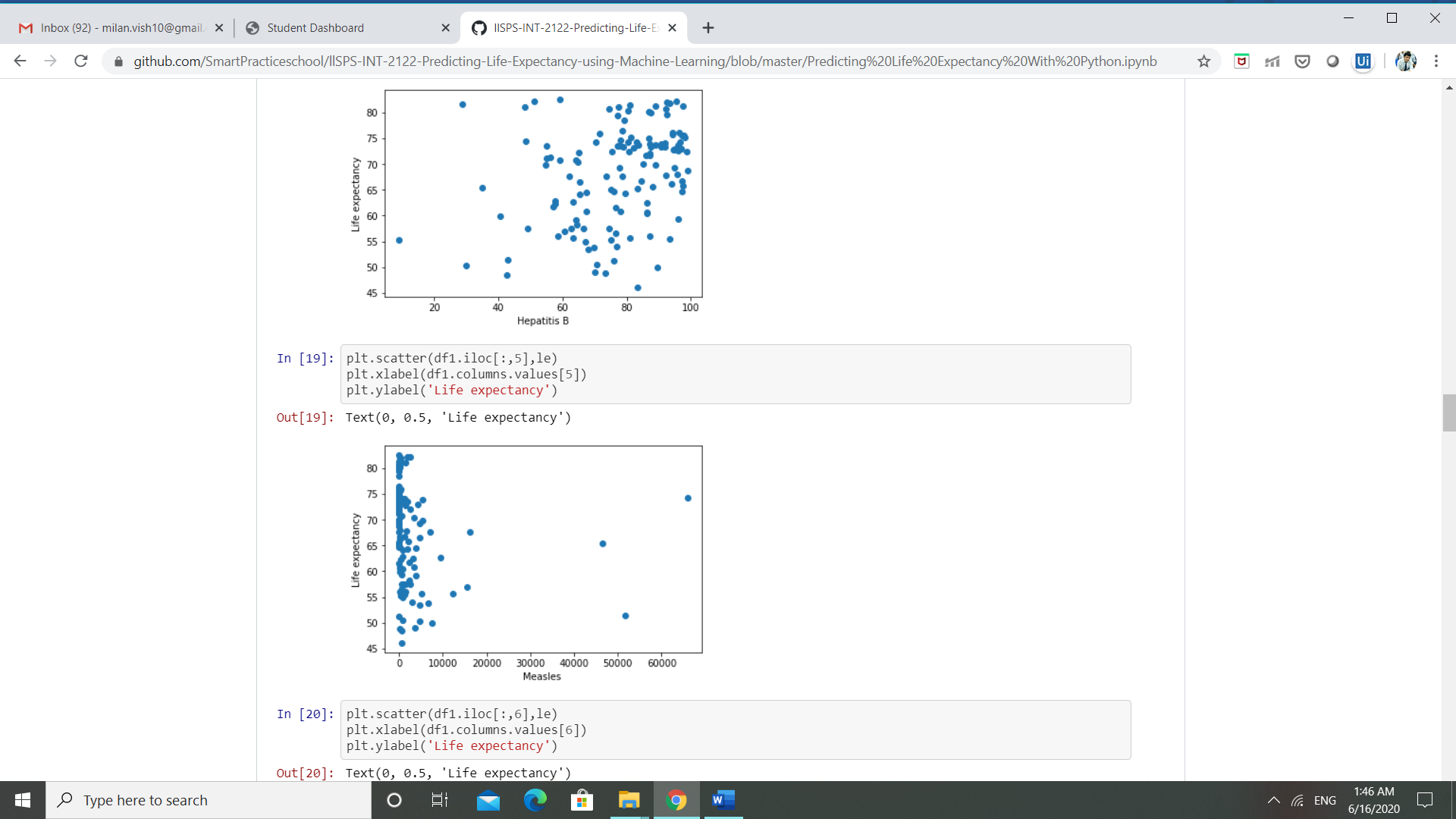
**CHAPTER 4**

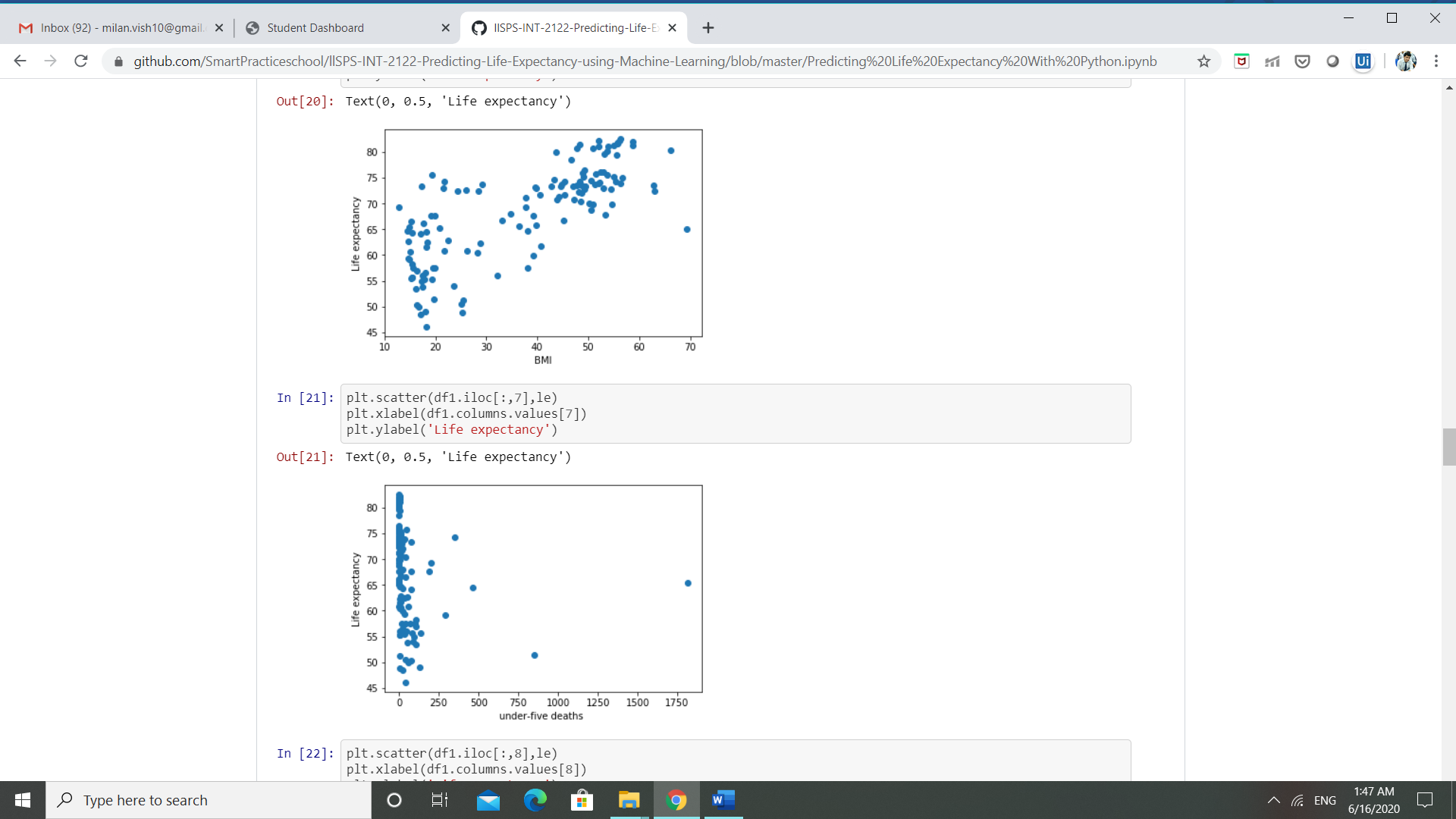
**EXPERIMENTAL INVESTIGATIONS**

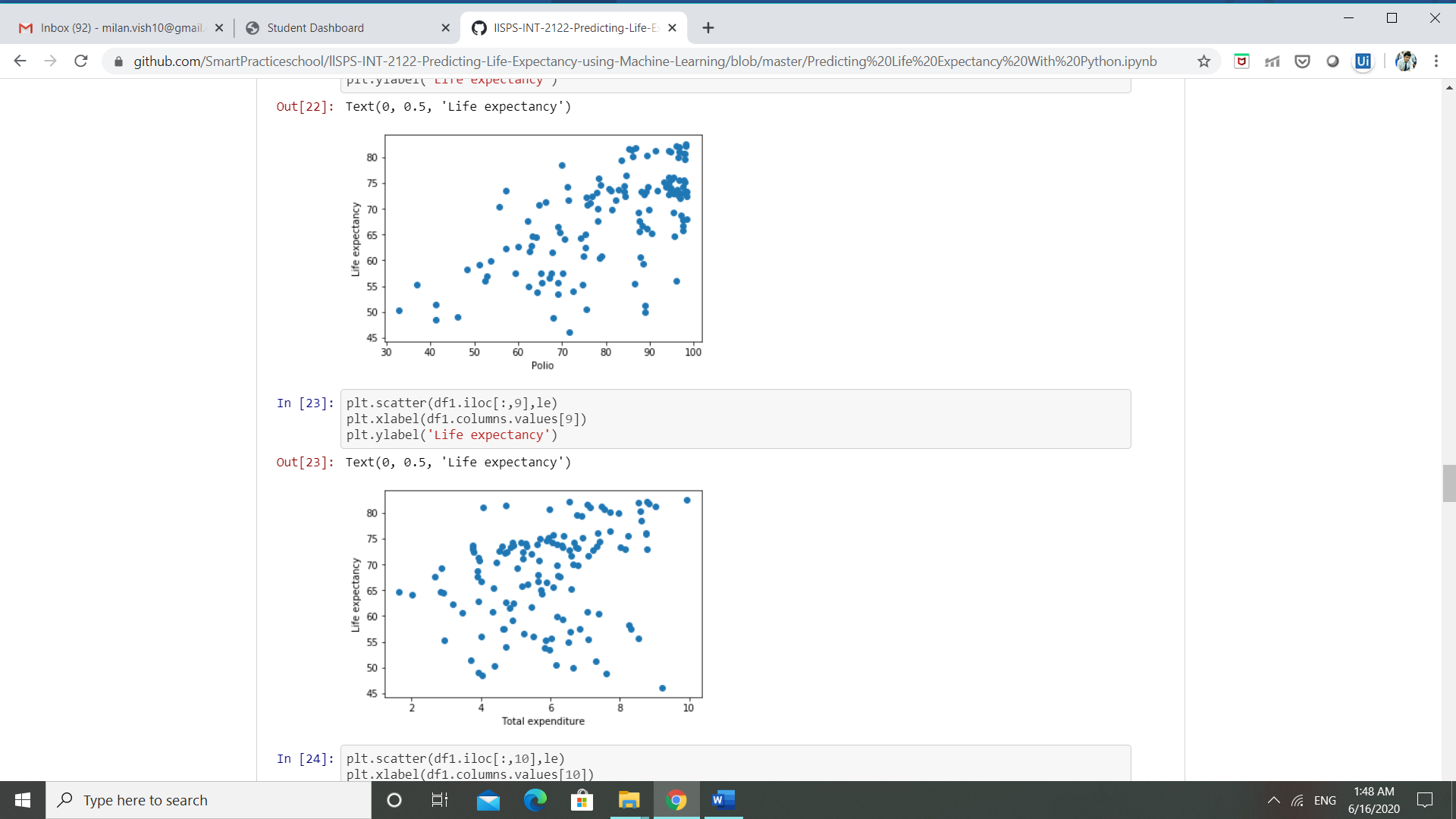
On drawing various charts against the Life Expectancy column in the dataset, we can know which factor is influencing the Life Expectancy most and can only keep those columns in our dataframe. Following are the scatter plot of Life Expectancy with other columns provided in the dataset. Firstly we have to performance pre-processing on data by doing average of values country-wise and removing the Nan values.

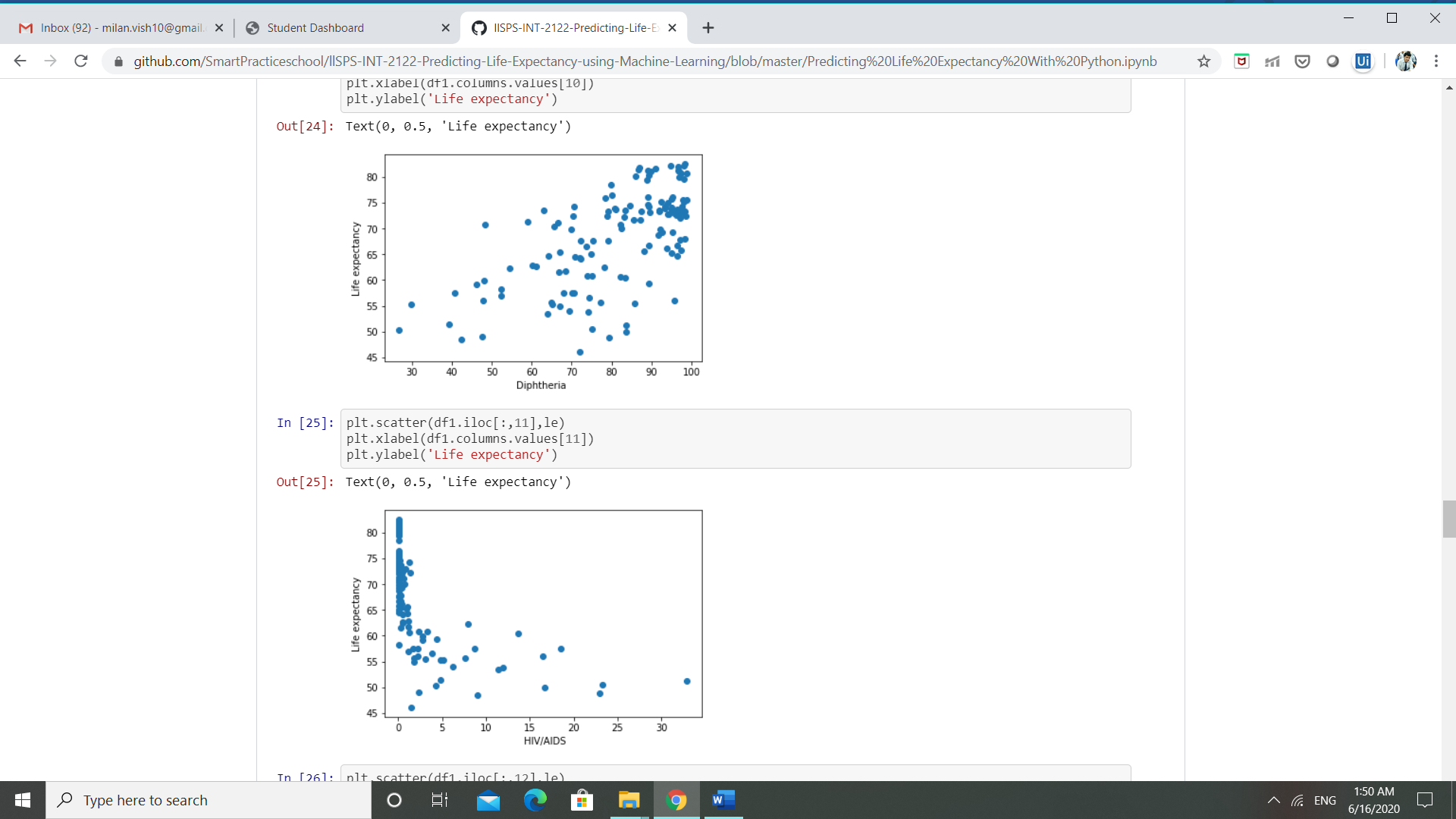


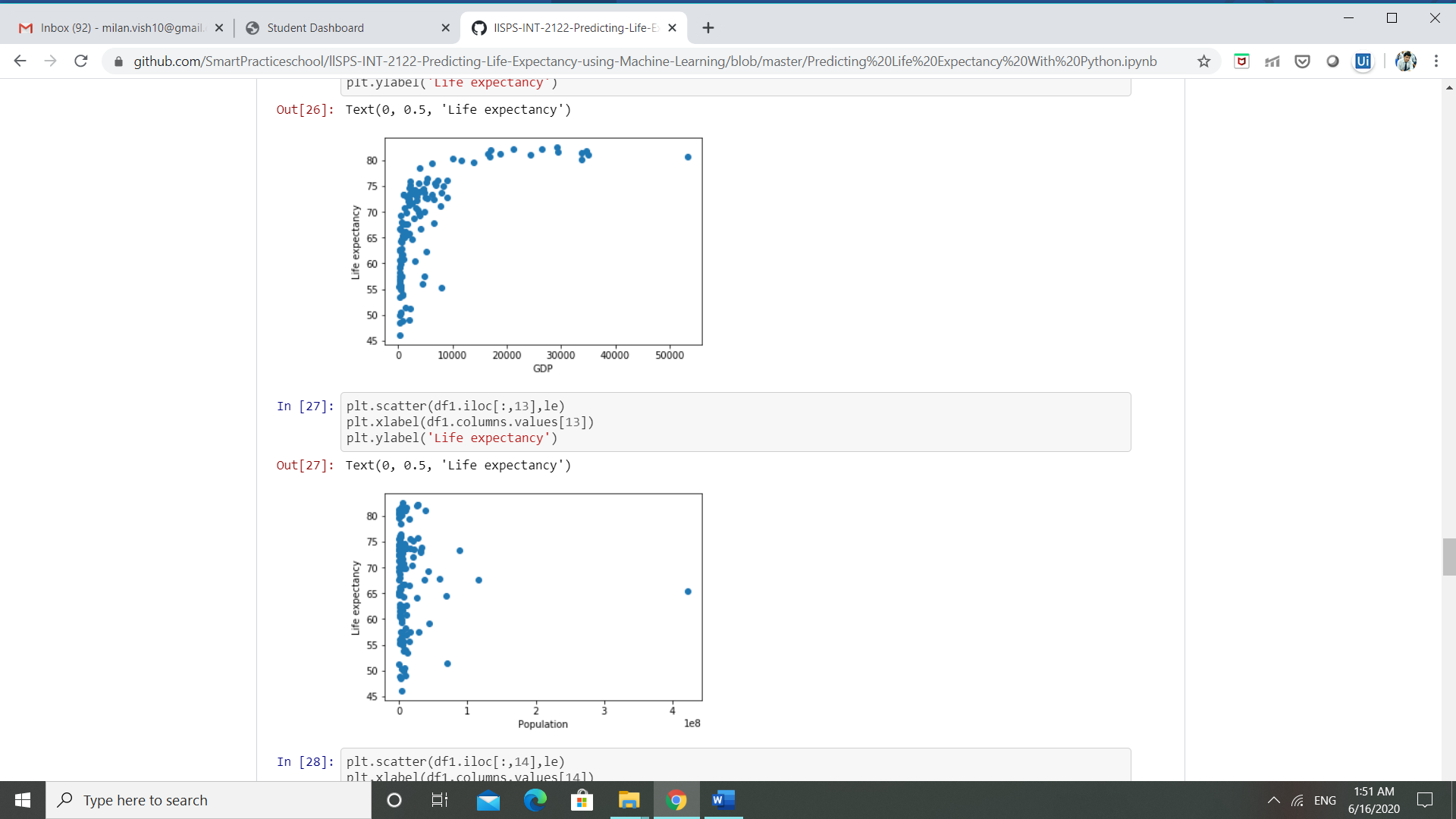


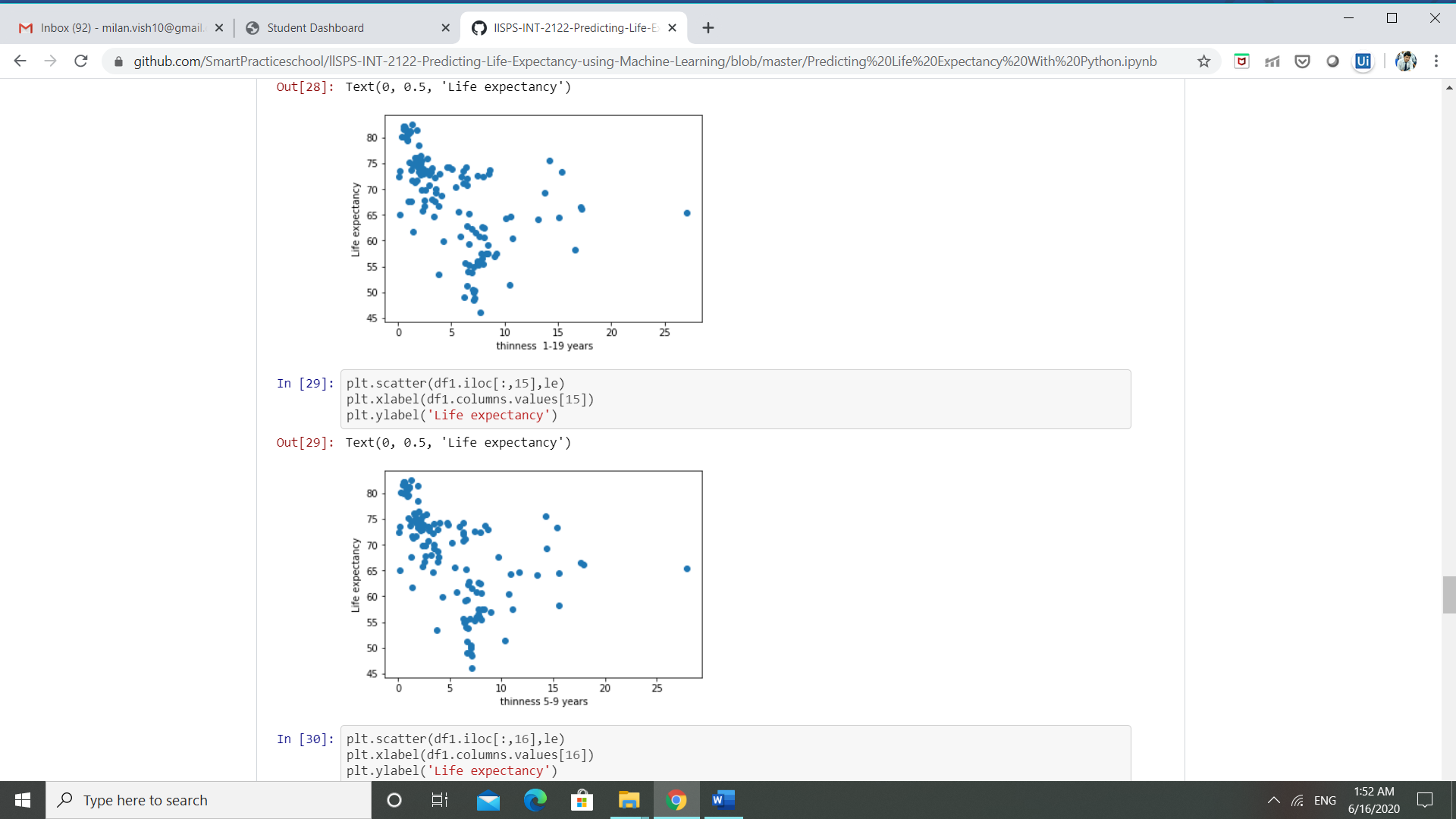


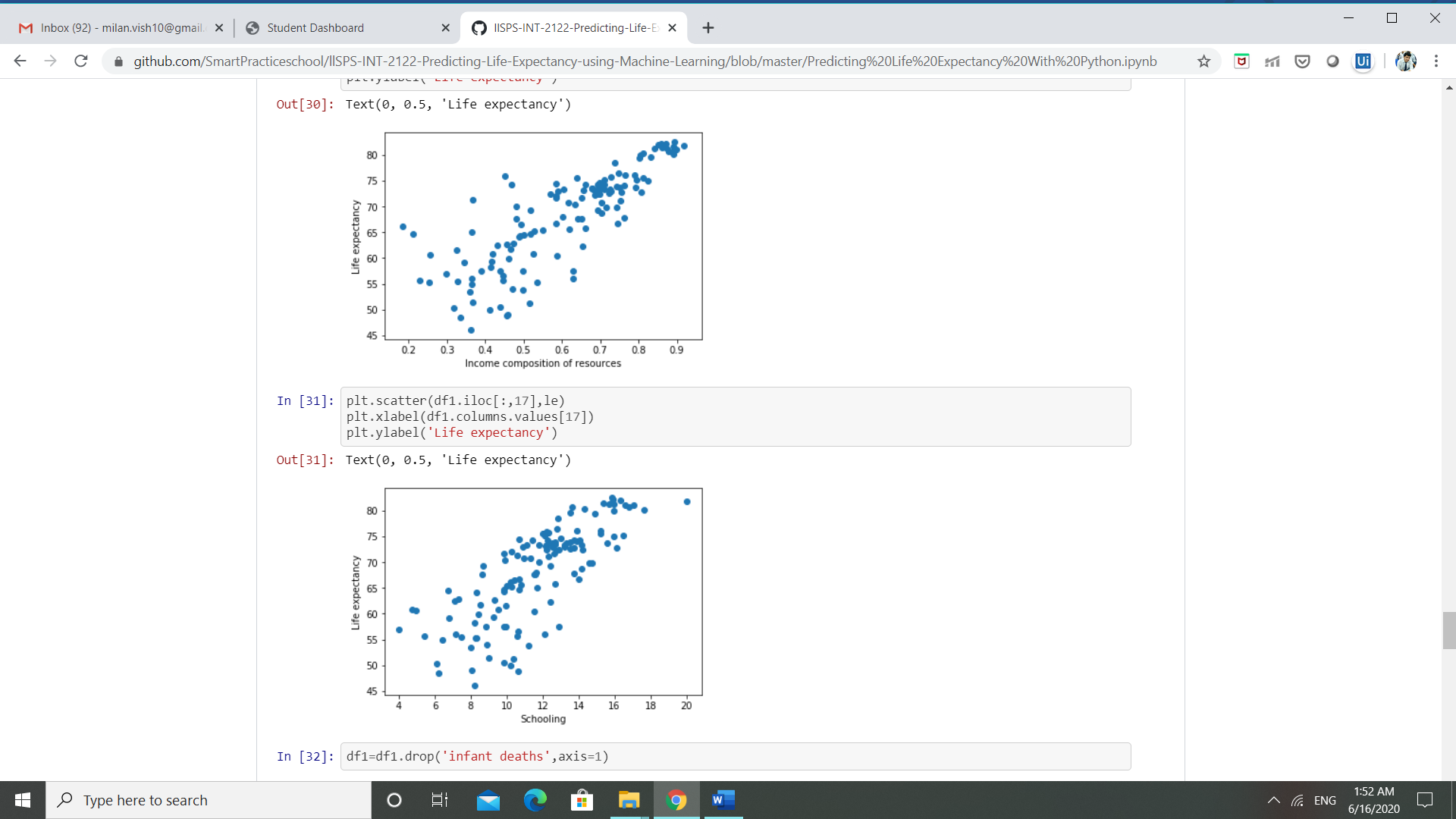




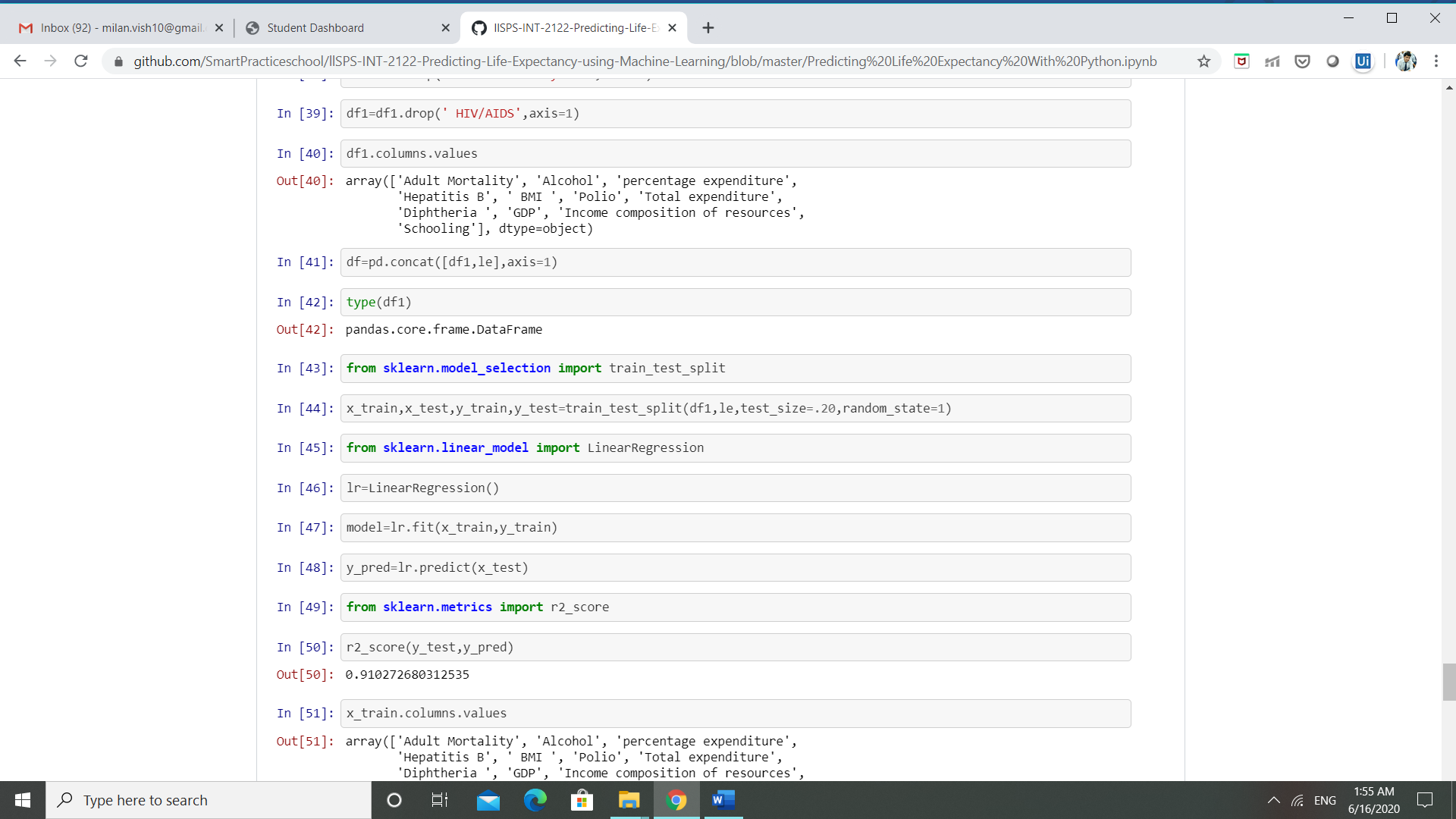








So, after plotting different charts we came to the conclusion that the following factors have the most impact on the life expectancy.



We have removed all the columns that were not that much correlated with life expectancy.

**CHAPTER 5**

**FLOWCHART**

Start

Input the details about the country

Process the input and return the output

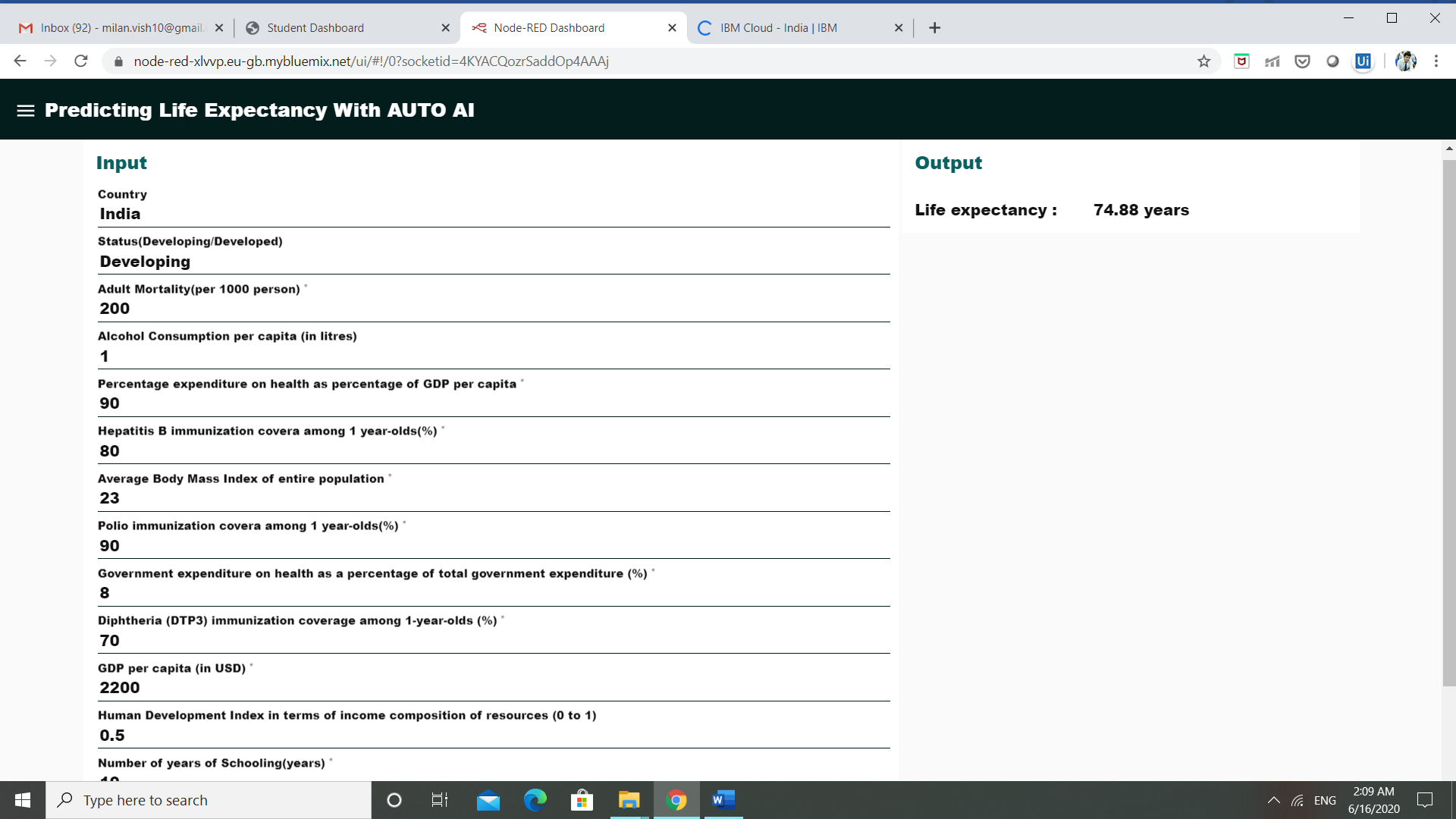
Display the output (years)

STOP

**CHAPTER 6**

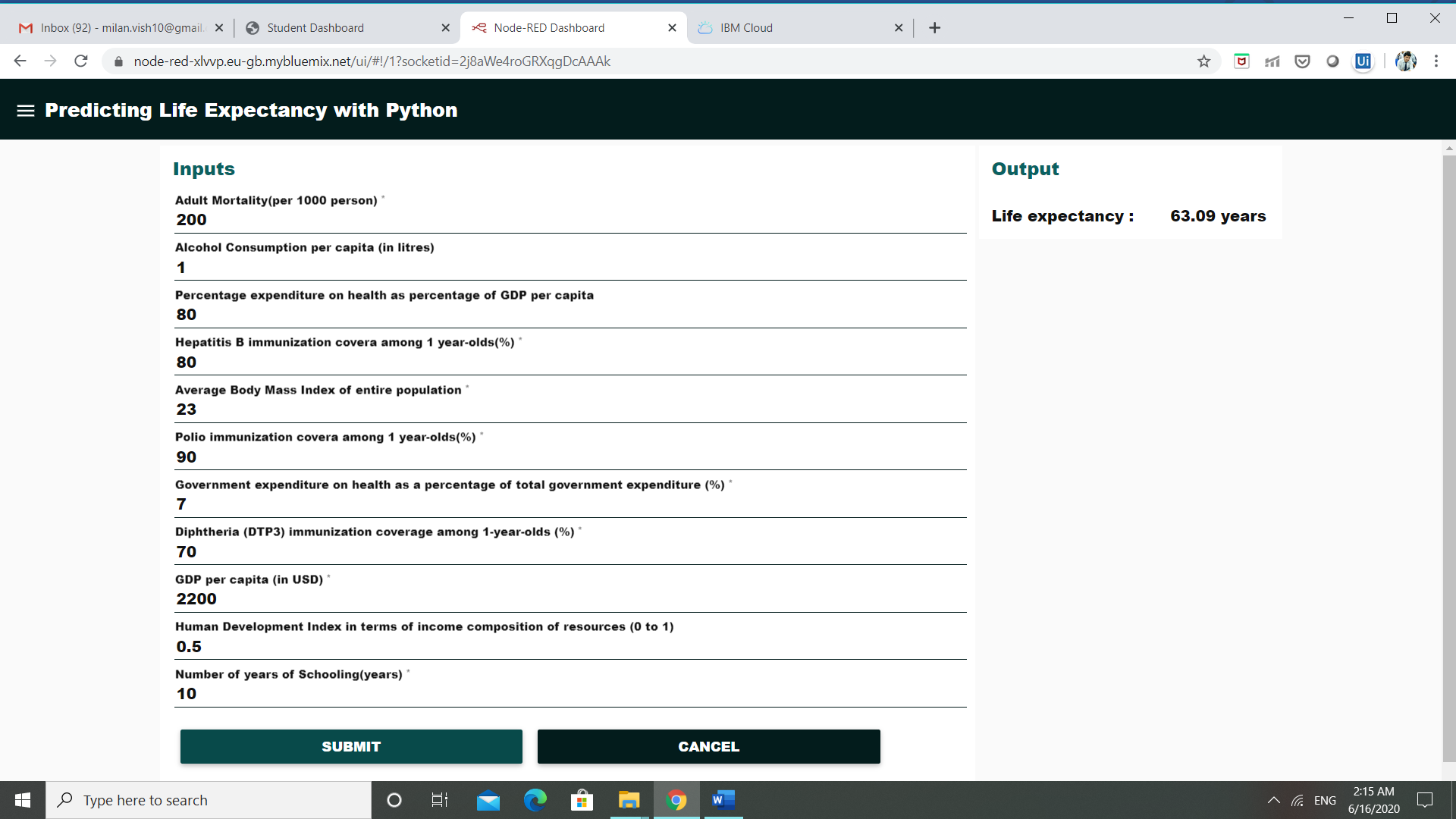
**RESULT**

The following are the snapshots of results:



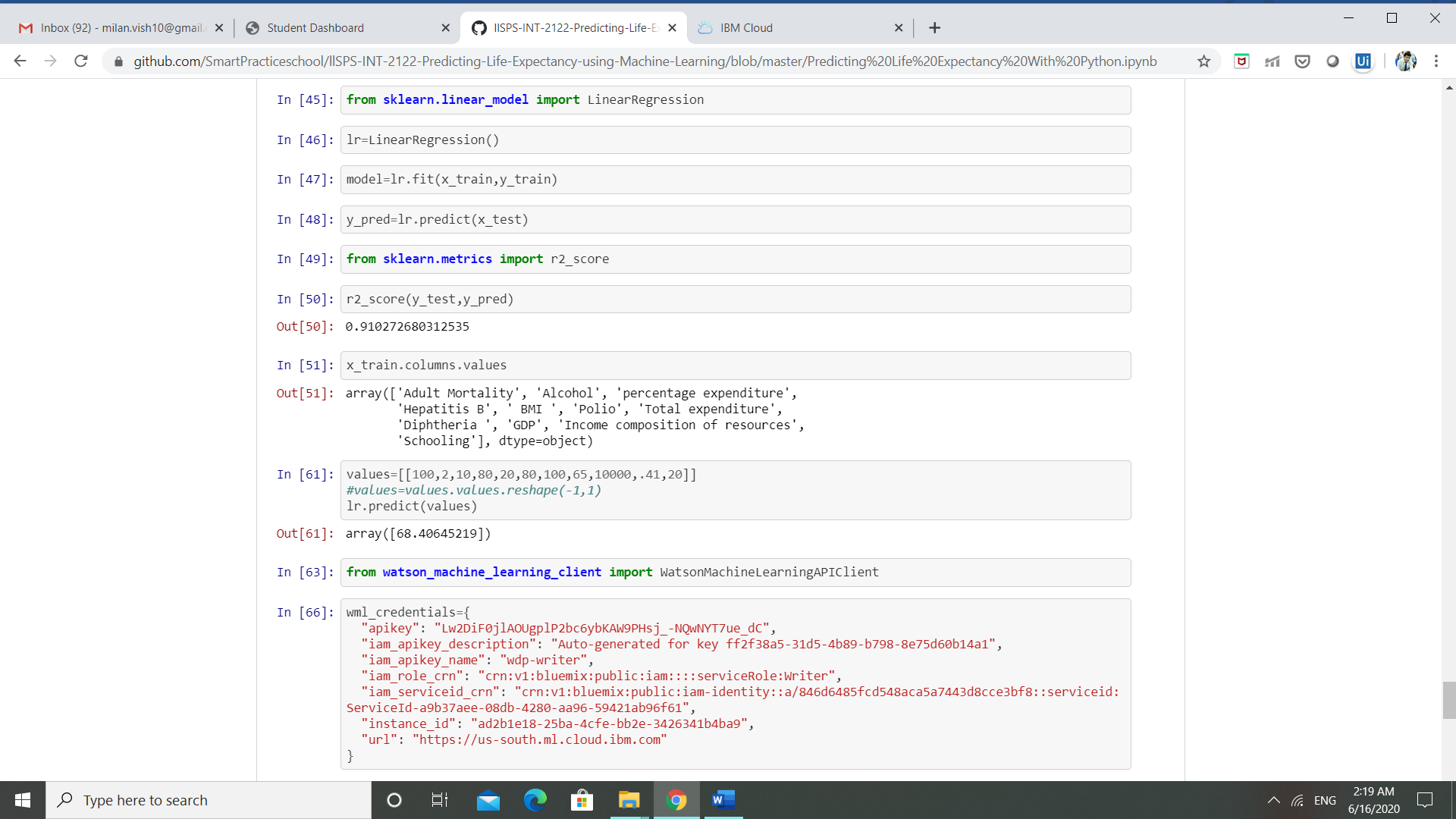
This is the snapshot of the result that we got when we used the model trained by Auto AI.

Now lets look at the results that we go using that model which I have designed by myself using Jupyter Notebook.

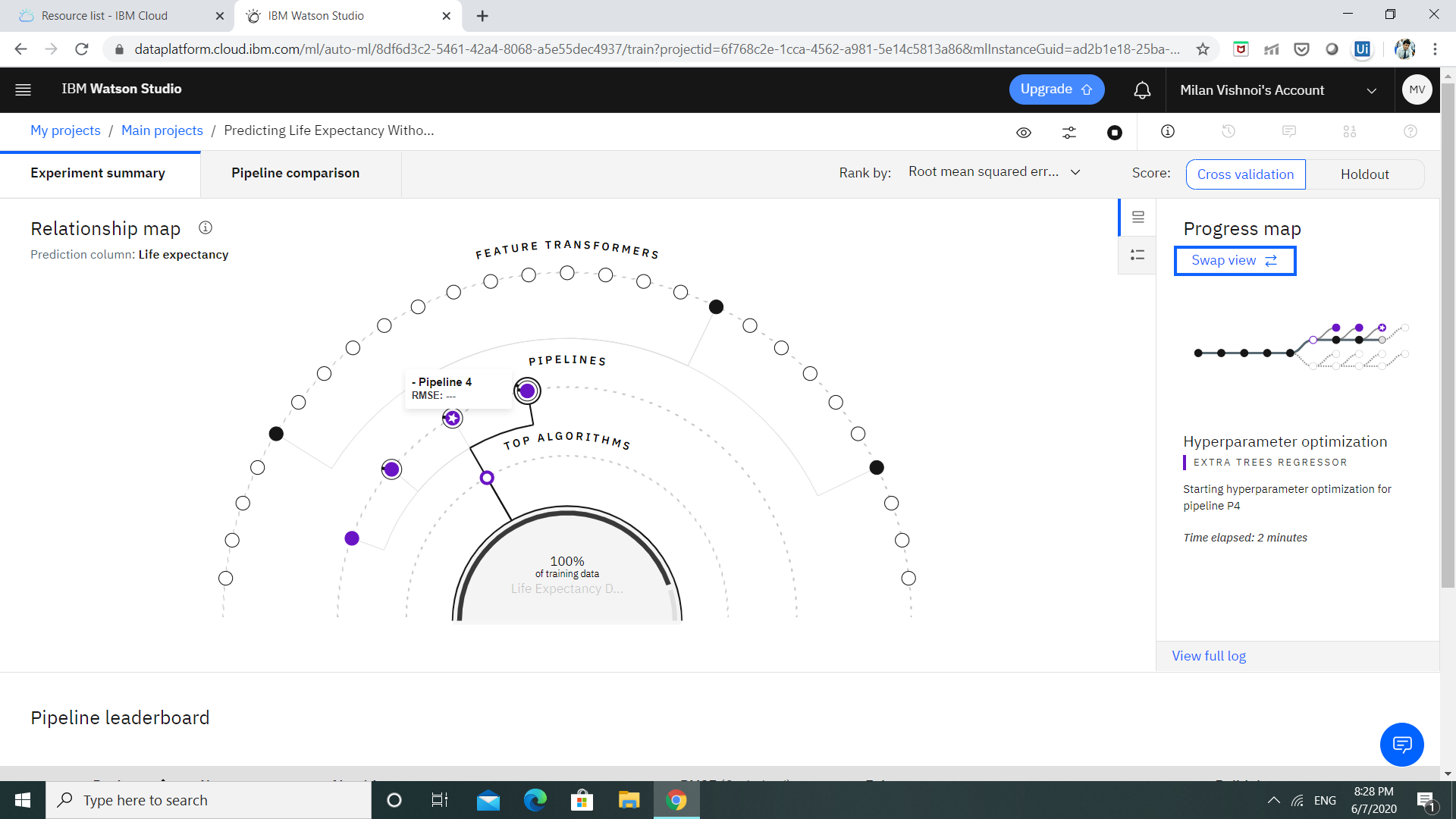


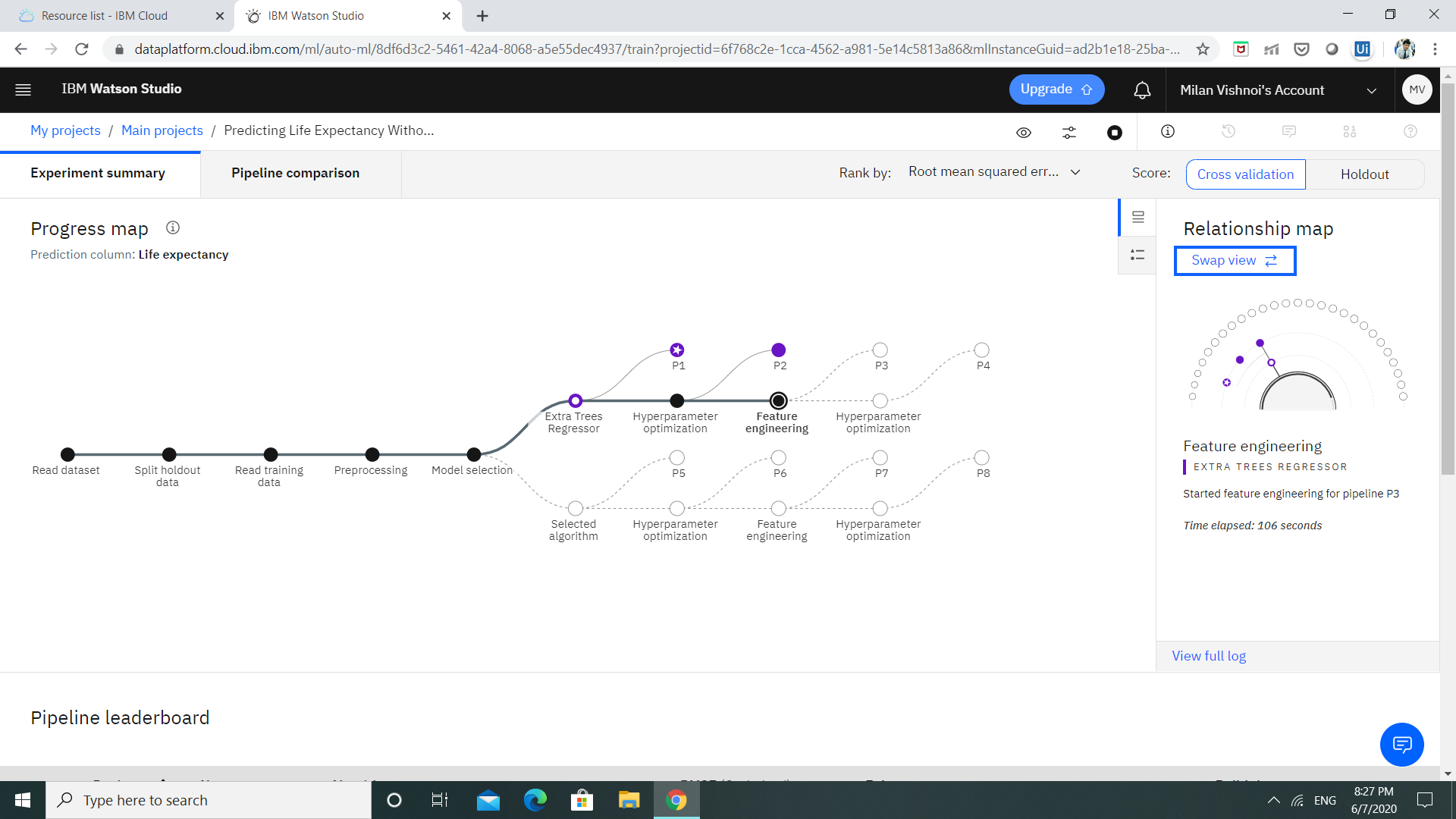
These are the results of the model that I manually designed using Jupyter Notebook.

The R2 score for the model is:



The snapshots of AUTO AI model being trained are as follows:





**CHAPTER 7**

**APPLICATIONS**

The applications of this project can be found in improving the healthcare sector and also in planning and also in making strategy regarding various things. It can also be used to analyse which factor has how much impact on overall life expectancy of the citizens of a nation and then the work can be done to improve the life expectancy.

**CHAPTER 8**

**CONCLUSION**

By observing various graphs and also the output data we can conclude that some factors influence the life expectancy more than the others. Among such are adult mortality, immunization of the citizens against various diseases, schooling, alcohol consumption etc.

**CHAPTER 9**

**FUTURE SCOPE**

As more and more data will be added, we can see improvement in the level if prediction and can design even more accurate machine learning models which can more accurately predict the life expectancy of citizens of a country.

**CHAPTER 10**

**BIBILOGRAPHY**

* <https://developer.ibm.com/tutorials/how-to-create-a-node-red-starter-application/>
* <https://cloud.ibm.com/docs/overview?topic=overview-whatis-platform>
* <https://bookdown.org/caoying4work/watsonstudio-workshop/jn.html>
* <https://developer.ibm.com/tutorials/watson-studio-auto-ai/>
* <https://www.kaggle.com/kumarajarshi/life-expectancy-who>
* <https://bookdown.org/caoying4work/watsonstudio-workshop/jn.html#deploy-model-as-web-service>
* <https://bookdown.org/caoying4work/watsonstudio-workshop/auto.html#add-asset-as-auto-ai>

**APPENDIX**

1. **Source code**

**import** **pandas** **as** **pd**

**import** **numpy** **as** **np**

**import** **matplotlib.pyplot** **as** **plt**

df=pd.read\_csv("https://raw.githubusercontent.com/SmartPracticeschool/llSPS-INT-2122-Predicting-Life-Expectancy-using-Machine Learning/master/dataset/Life**%20E**xpectancy%20Data.csv")

df.head()

df=df.groupby("Country").mean()

df.head()

df=df.drop('Year',axis=1)

df.isna().count()

df=df.dropna()

df.head()

df1=df1.drop('Measles ',axis=1)

df1=df1.drop('under-five deaths ',axis=1)

df1=df1.drop('Population',axis=1)

df1=df1.drop(' thinness 1-19 years',axis=1)In [38]:

df1=df1.drop(' thinness 5-9 years',axis=1)

df1=df1.drop(' HIV/AIDS',axis=1)

df1.columns.values

df=pd.concat([df1,le],axis=1)

df1=df1.drop('Measles ',axis=1)

**from** **sklearn.model\_selection** **import** train\_test\_split

x\_train,x\_test,y\_train,y\_test=train\_test\_split(df1,le,test\_size=.20,random\_state=1)

**from** **sklearn.linear\_model** **import** LinearRegression

lr=LinearRegression()

model=lr.fit(x\_train,y\_train)

y\_pred=lr.predict(x\_test)

**from** **sklearn.metrics** **import** r2\_score

r2\_score(y\_test,y\_pred)

**from** **watson\_machine\_learning\_client** **import** WatsonMachineLearningAPIClient

In [66]:

wml\_credentials={

"apikey": "xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx",

"iam\_apikey\_description": "Auto-generated for key xxxxxxxxxxxxxxxxxxxxxxx",

"iam\_apikey\_name": "wdp-writer",

"iam\_role\_crn": "crn:v1:bluemix:public:iam::::serviceRole:Writer",

"iam\_serviceid\_crn": "crn:v1:bluemix:public:iam-identity::a/846d6485fcd548aca5a7443d8cce3bf8::serviceid:ServiceId-xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx",

"instance\_id": "xxxxxxxxxxxxxxxxxxxxxxxxxxxxx",

"url": "https://us-south.ml.cloud.ibm.com"

}

In [67]:

client=WatsonMachineLearningAPIClient(wml\_credentials)

In [69]:

model\_props={

client.repository.ModelMetaNames.AUTHOR\_NAME : "",

client.repository.ModelMetaNames.AUTHOR\_EMAIL : " ",

client.repository.ModelMetaNames.NAME:"Life Expectancy Predcitor"

}

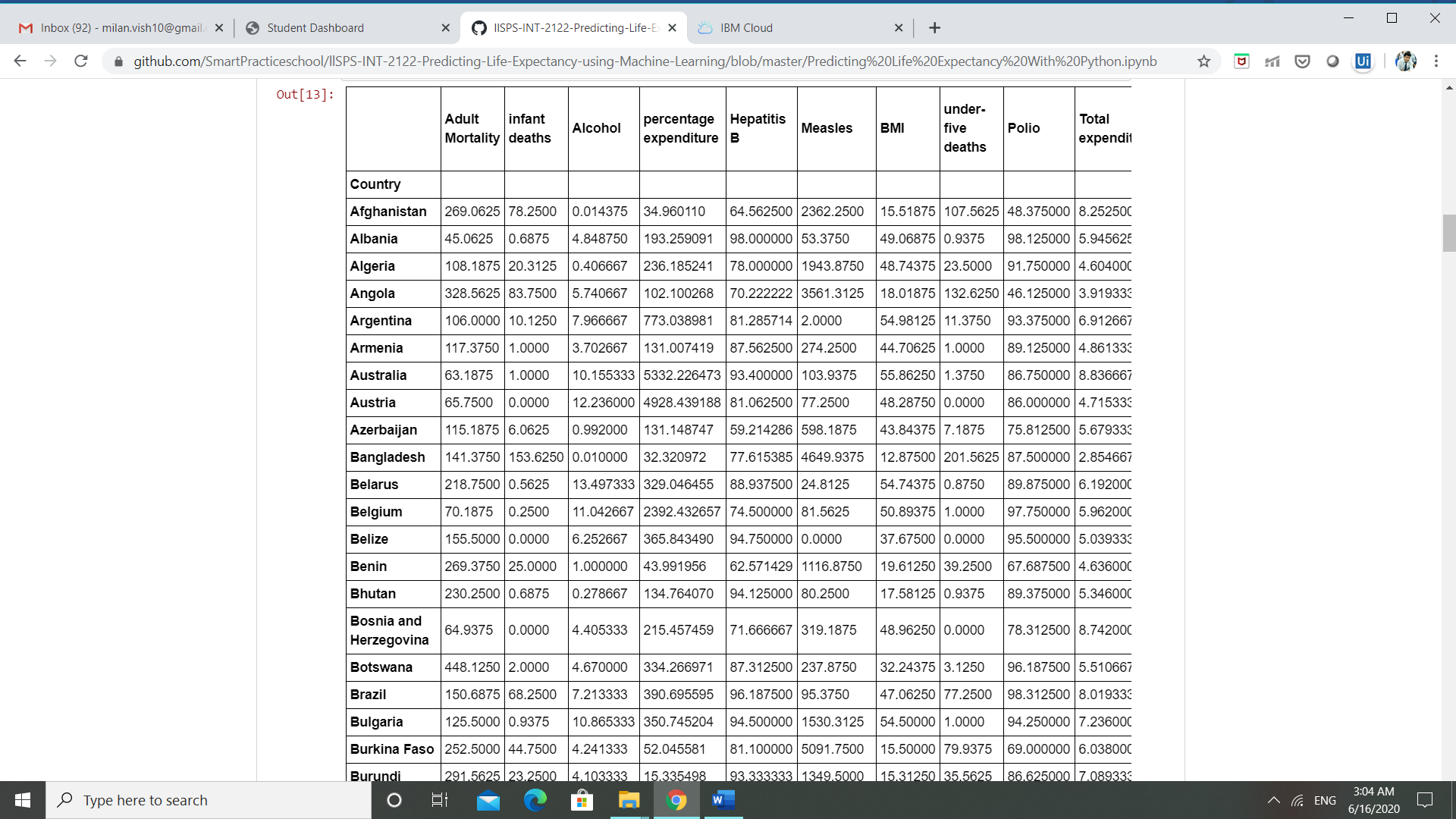
model\_artifacts=client.repository.store\_model(lr,meta\_props=model\_props)

guid=client.repository.get\_model\_uid(model\_artifacts)

deploy=client.deployments.create(guid,name="Predicting Life Expectancy With Python")

deploy[‘metadata’][‘entity’][‘scoring\_url’]

1. **Dataset**



**Link to the dataset:**

<https://github.com/SmartPracticeschool/llSPS-INT-2122-Predicting-Life-Expectancy-using-Machine-Learning/blob/master/dataset/Life%20Expectancy%20Data.csv>