*Predicting Life Expectancy problem using machine learning*

**Project Scope:**

The project tries to create a model based on data provided by the World Health Organization (WHO) to evaluate the life expectancy for different countries in years. The data offers a time frame from 2000 to 2015. The data originates from here: https://www.kaggle.com/kumarajarshi/life-expectancy-who/data

            The output algorithms have been used to test if they can maintain their accuracy in predicting the life expectancy for data they haven’t been trained. I am going to use Regression algorithms to form the model.

**Schedule:**

The duration of this project is one month. I have to submit this project within 5th of July, 2020.

**Team:**

I, Bratati Ghosh am going to do this project with the help of smartinternz's mentor. I am a 3rd year B.Tech (IT) student from Government College Of Engineering & Ceramic Technology.

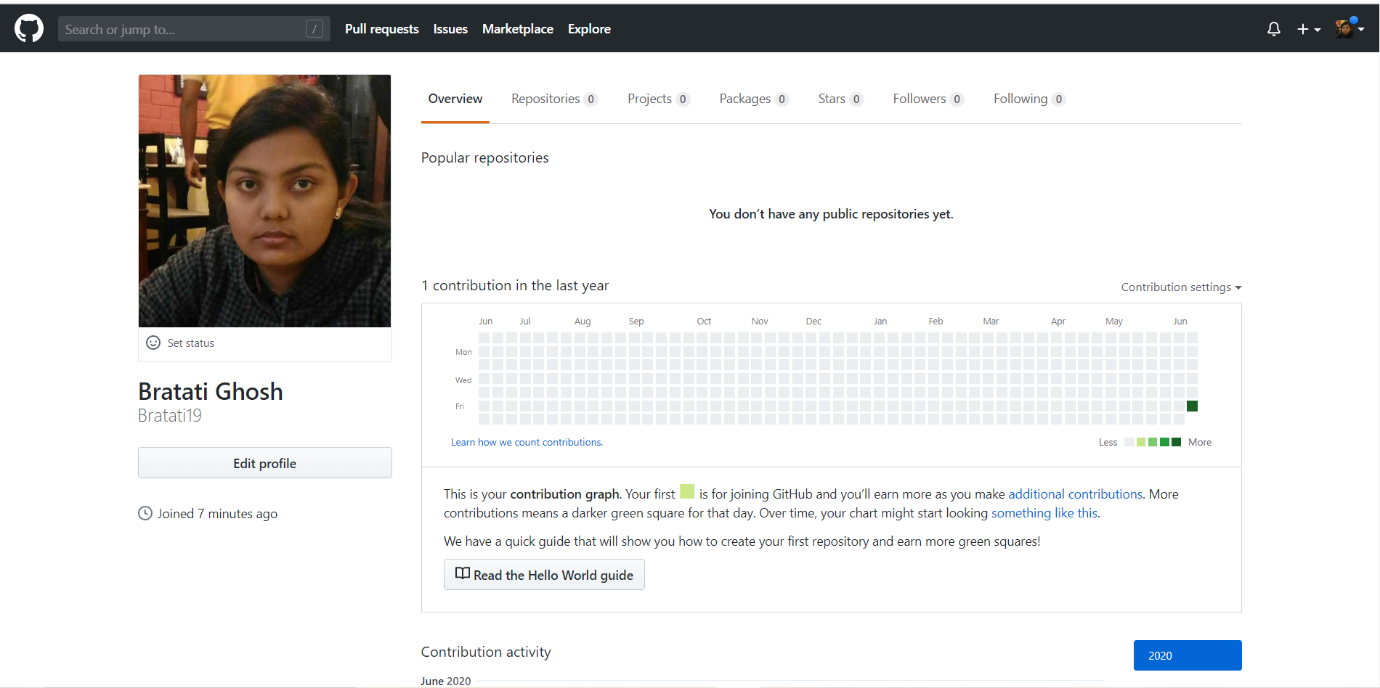
**Deliverables:**

At the end of the project, the best regression model which will give the best accuracy and least error will be our output model. The final model will be able to predict the future life expectancy based on new data set and also the project will tell us which features actually affect life expectancy and which features do not.

**Setting up the development environment:**

**1)Creating a github account:** I have created a github account to post all the details of my project there.

The screen shot of my github account is given below.



**2) What is github?**

Github is a web-based platform used for version control. Git simplifies the process of working with other people and makes it easy to collaborate on projects. Team members can work on files and easily merge their changes in with the master branch of the project.

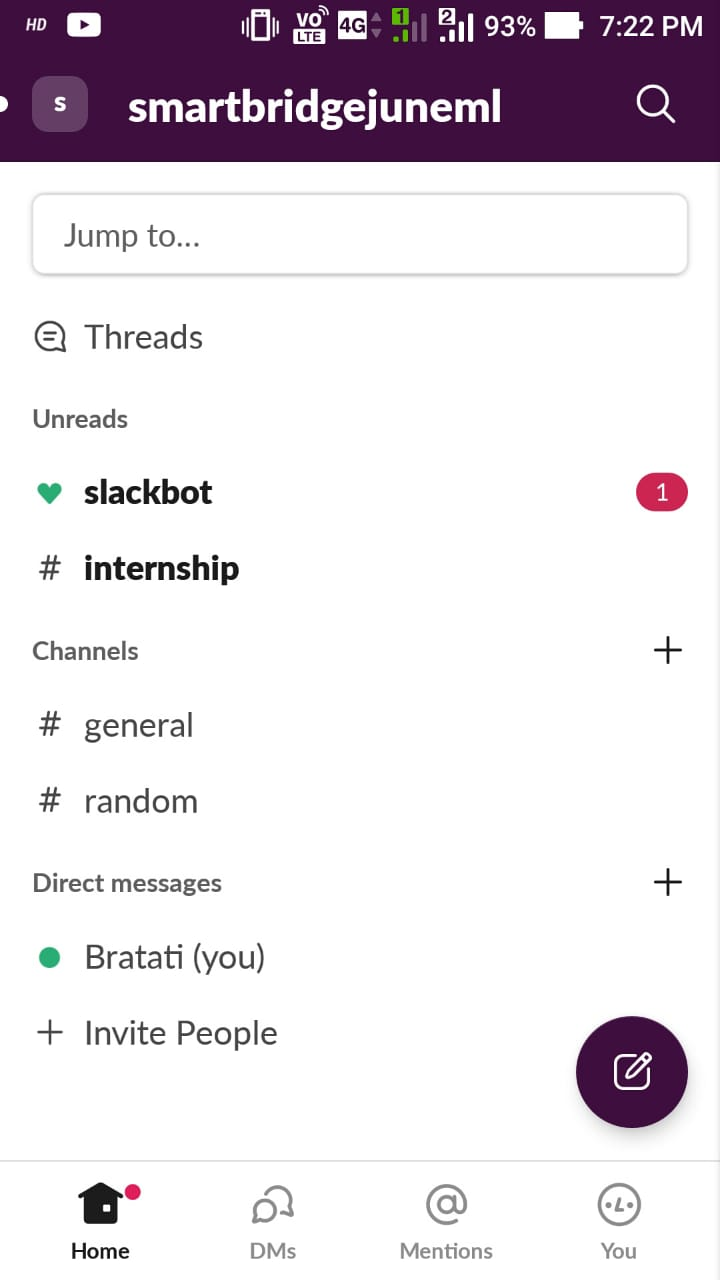
**3)What is slack?**

Slack is a collaboration hub that can replace email to help you and your team work together seamlessly. It’s designed to support the way people naturally work together, so you can collaborate with people online as efficiently as you do face-to-face.

**4)Installation of slack and creation of a slack account:**

I have installed slack and created an account using my email and a password and joined the workspace of smartbridge's internship of machine learning.

I have given the screen shot under this.



**5)Working with document writer:**

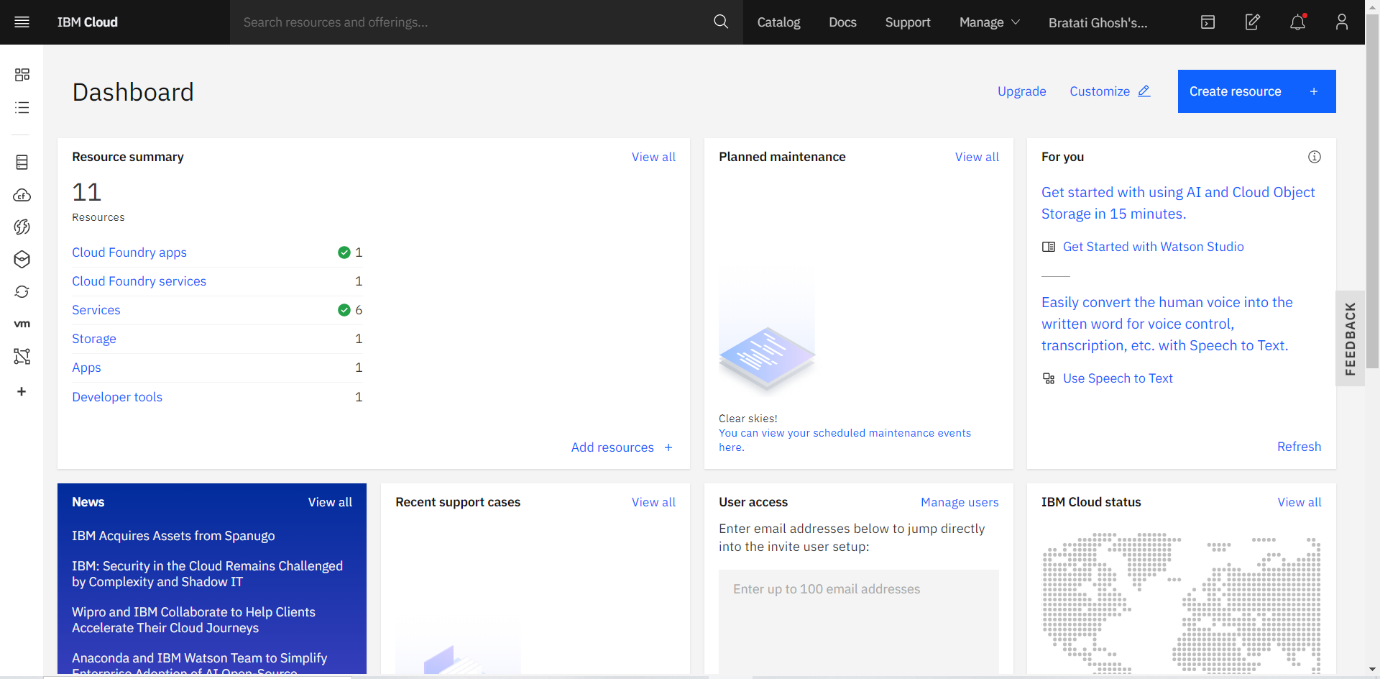
For the documentation of my project I am using zoho writer as provided by the smartbridge platform.

**Exploring IBM cloud platform:**

**1) Creating IBM cloud account:**

To use the IBM cloud services I have created an IBM cloud account using smartinternz

email id for free.

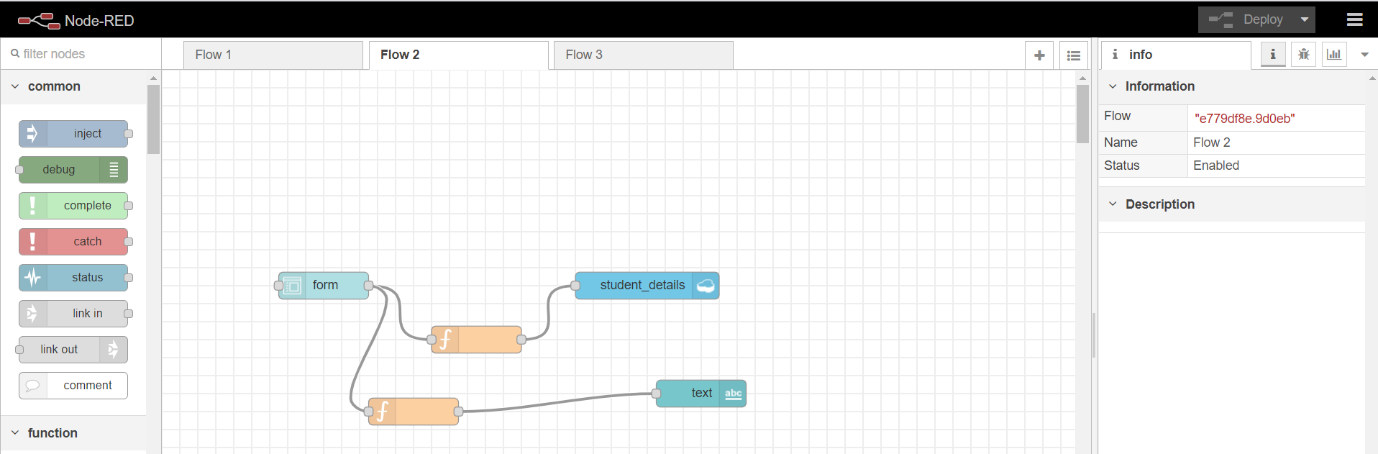


**2)Creating a node-red starter application:**

Node-RED is a programming tool for wiring together hardware devices, APIs and online services. Primarily, it is a visual tool designed for the Internet of Things, but it can also be used for other applications to very quickly assemble flows of various services.

I have created a basic node - red starter application using inject and debug node.

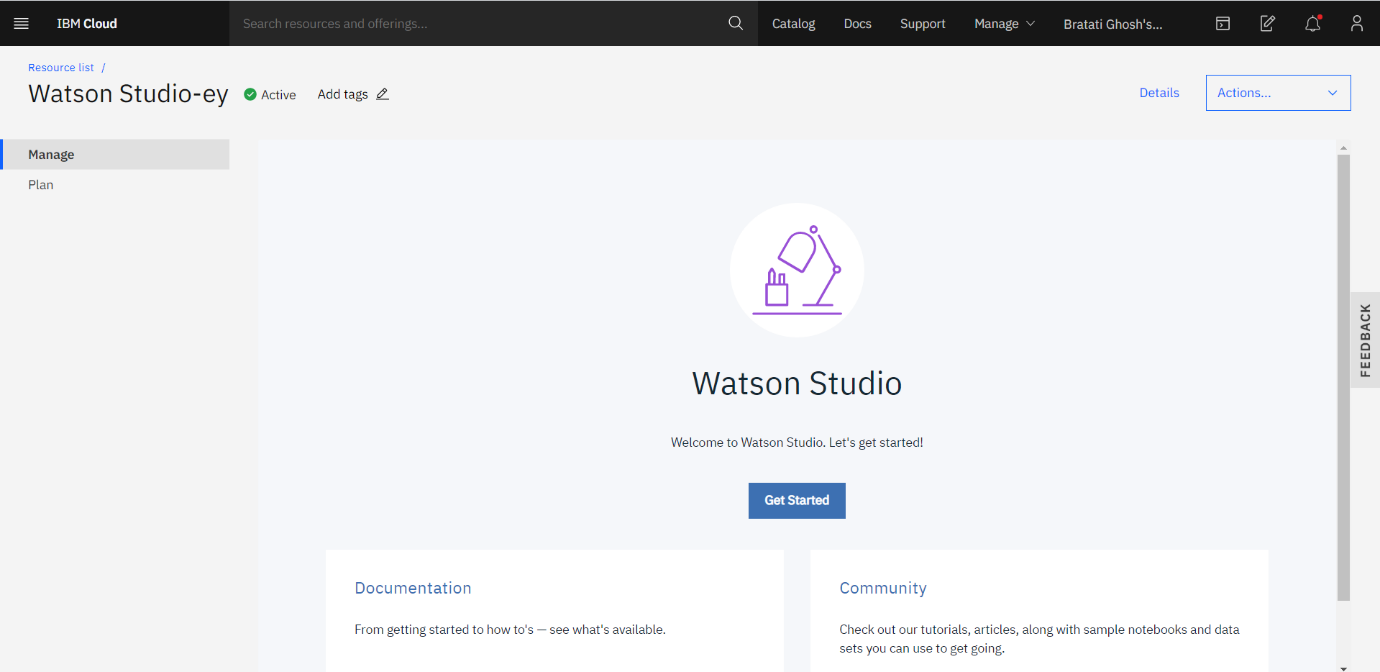
The screen shot is given below.



**1) Exploring Watson Machine Learning:**

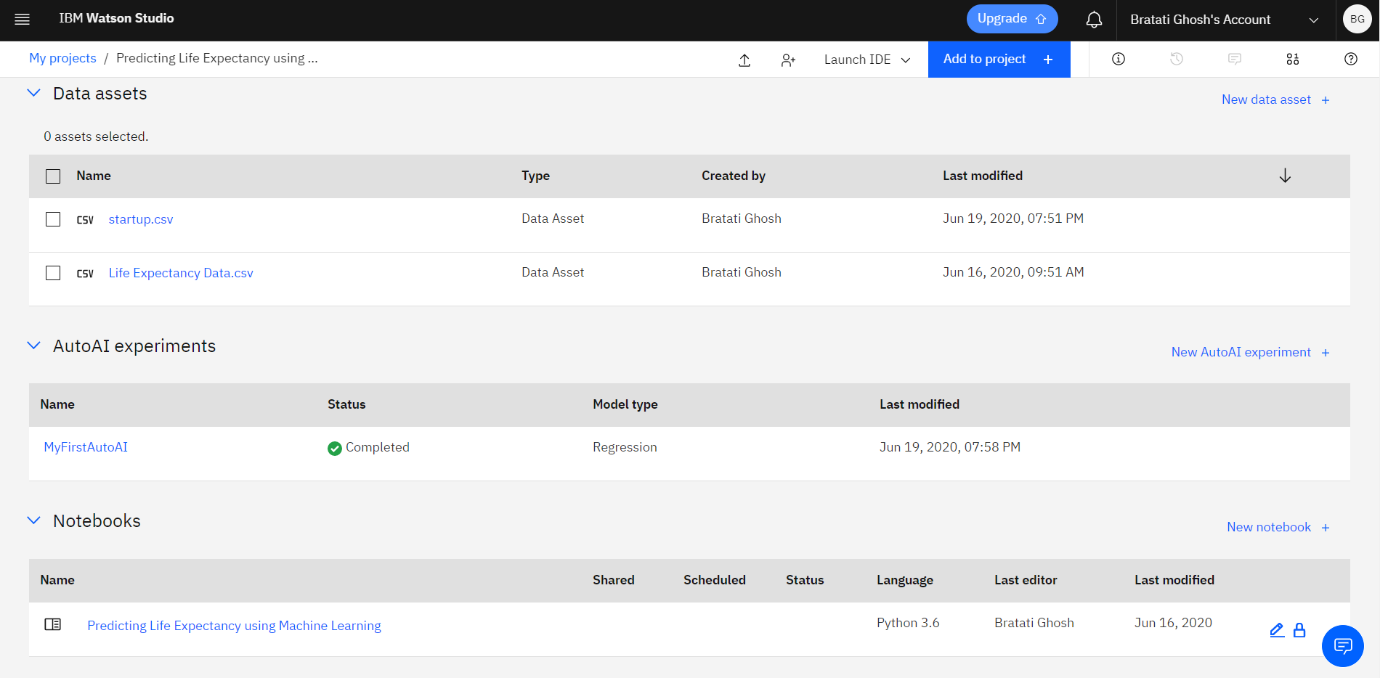
Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

I have explored IBM Watson Machine Learning service. The screen shot is given below.



2) Exploring IBM watson usecases:

I have explored different watson studio services like machine learning, notebook, AutoAI etc. The screen shot is given below



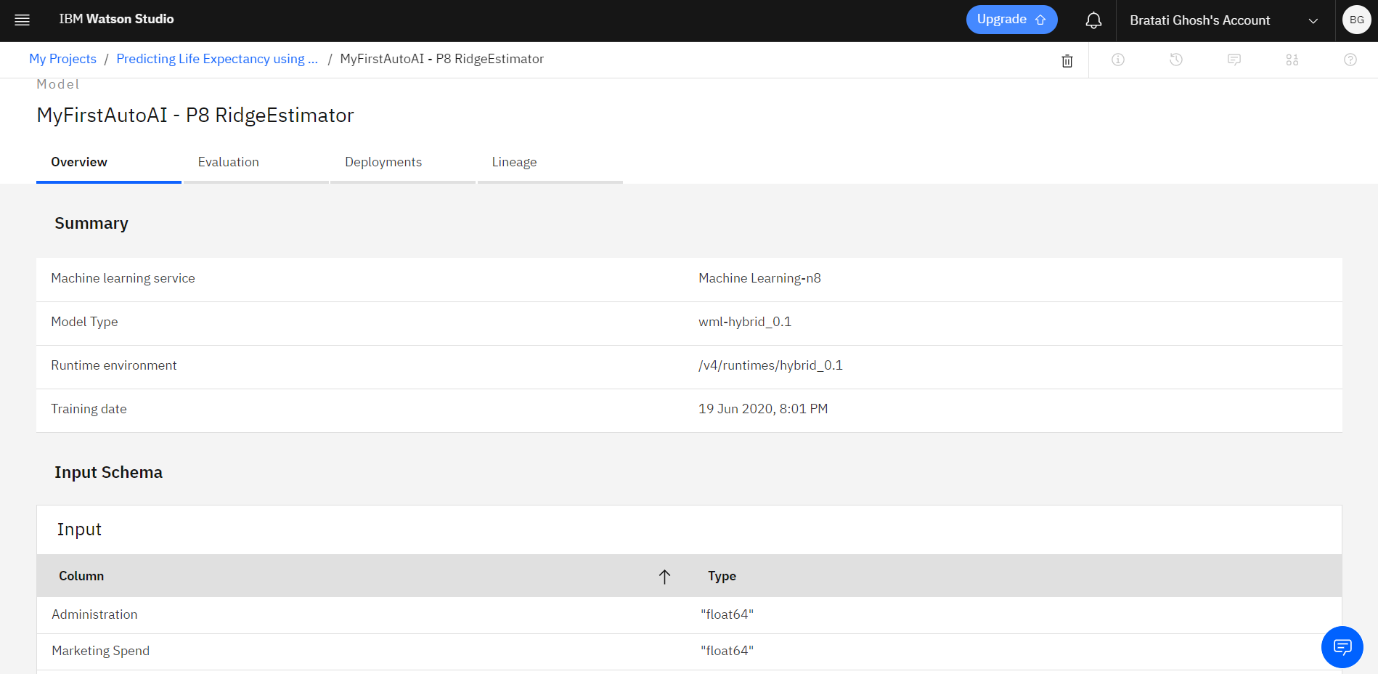
**Introduction to Watson Studio:**

**1)Building my own ML model in watson studio:**

I have built my ML model in watson studio using watson notebook. My model was built using the algorithm polynomial linear regression. It was the prediction of wind energy generation using weather report.

**2)Automating my ML model:**

Using Watson studio auto ai I have automated my ML model.



**Collecting the dataset for the project:**

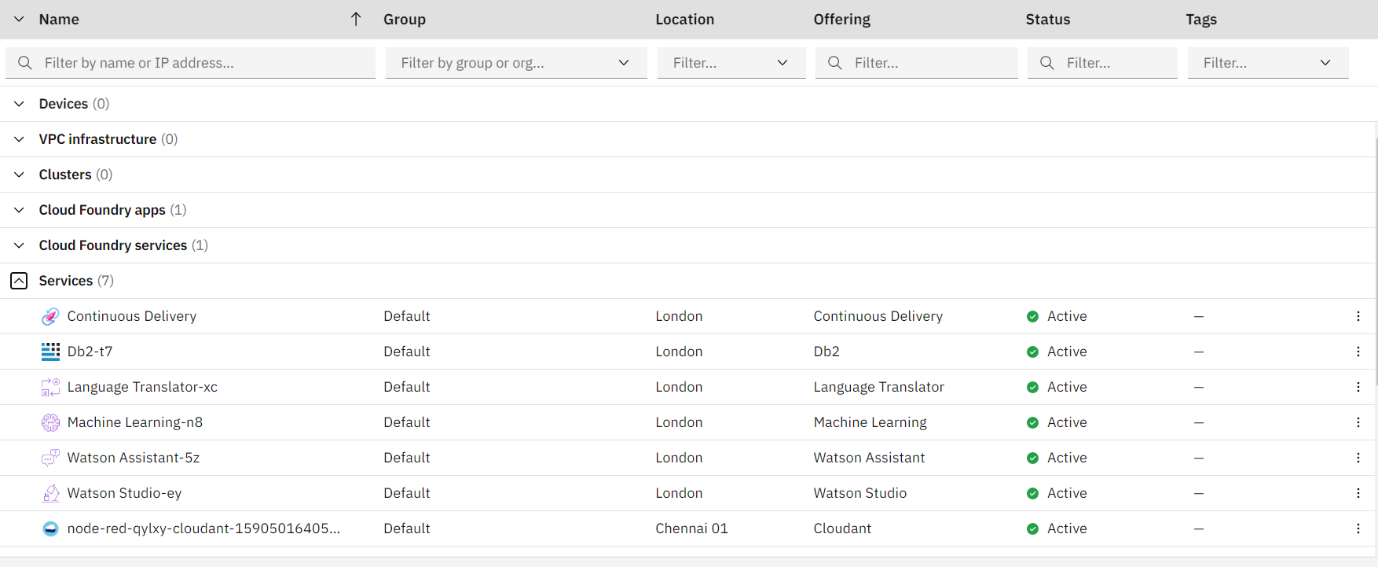
I have collected the data from kaggle.com. The data is given by WHO(World Health Organization). The data set is consists of 22 columns and 2938 rows.

The url of my dataset is : <https://www.kaggle.com/kumarajarshi/life-expectancy-who>

**Creating Necessery IBM cloud services:**

I have created necessery IBM cloud services which will help me to build my project further.

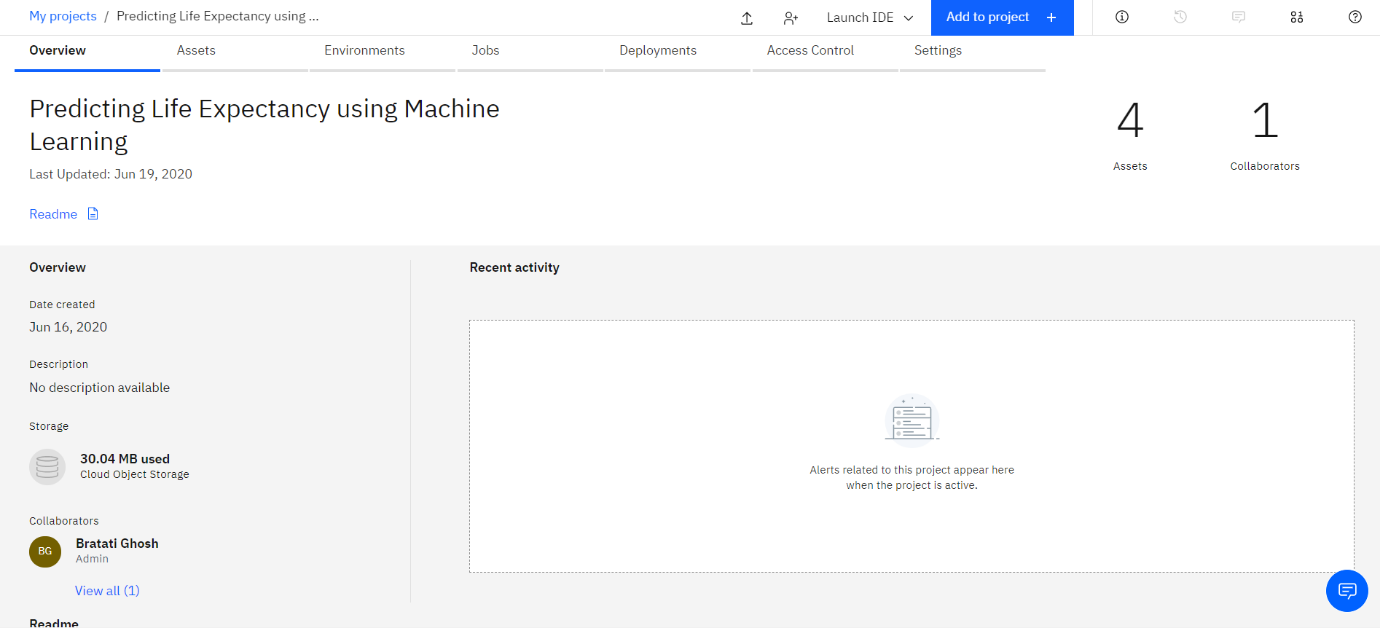
The screen shot of created services is given below:



**Creating a watson studio project:**

I have created a watson studion project to solve the given problem. The name of my project is : "[Prediction of life expectancy using Machine Learning](https://eu-gb.dataplatform.cloud.ibm.com/projects/261c10c0-e452-4205-b69f-5461ff7902c3?context=wdp)"

The screen shot of my created project is given below:

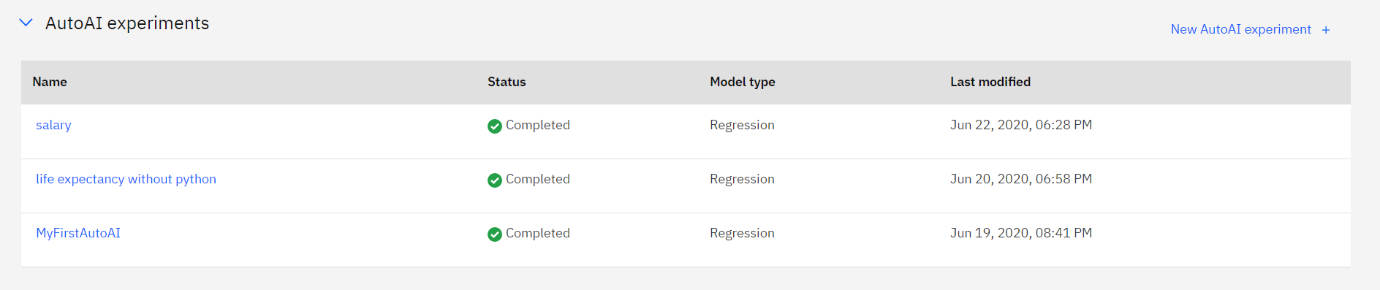


**Creating Machine Learning Service:**

I have created a Watson Machine Learning service for my project.

**Automate ML model:**

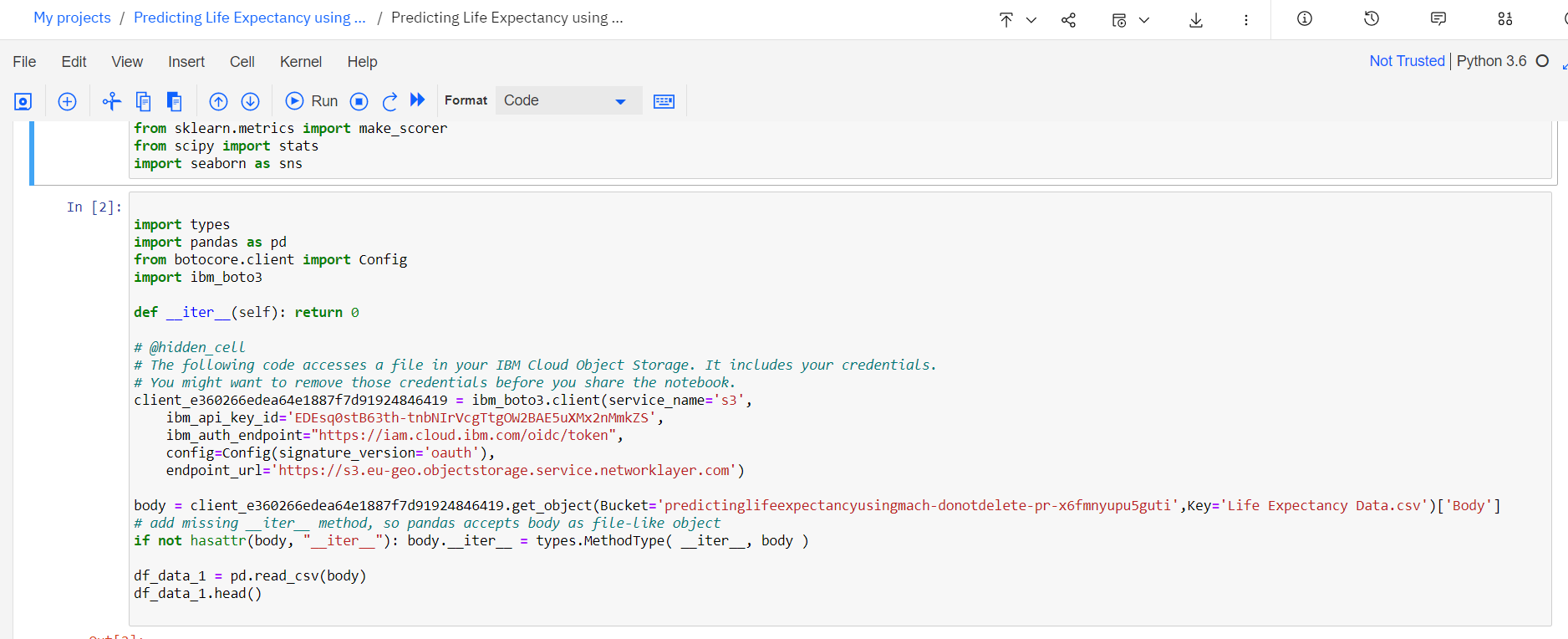
I have created a ml model using auto AI technique.The screen shot of my work is given below:



**Creating a Jupyter Notebook in Watson studio and importing the data:**

I have created a jupyter notebook and imported the data set in IBM cloud.

The screen shots are given below:



[Predicting Life Expectancy With Python](https://smartinternz.com/Student/workspace/2864#collapse5)

**1)Collecting the dataset for the project:**

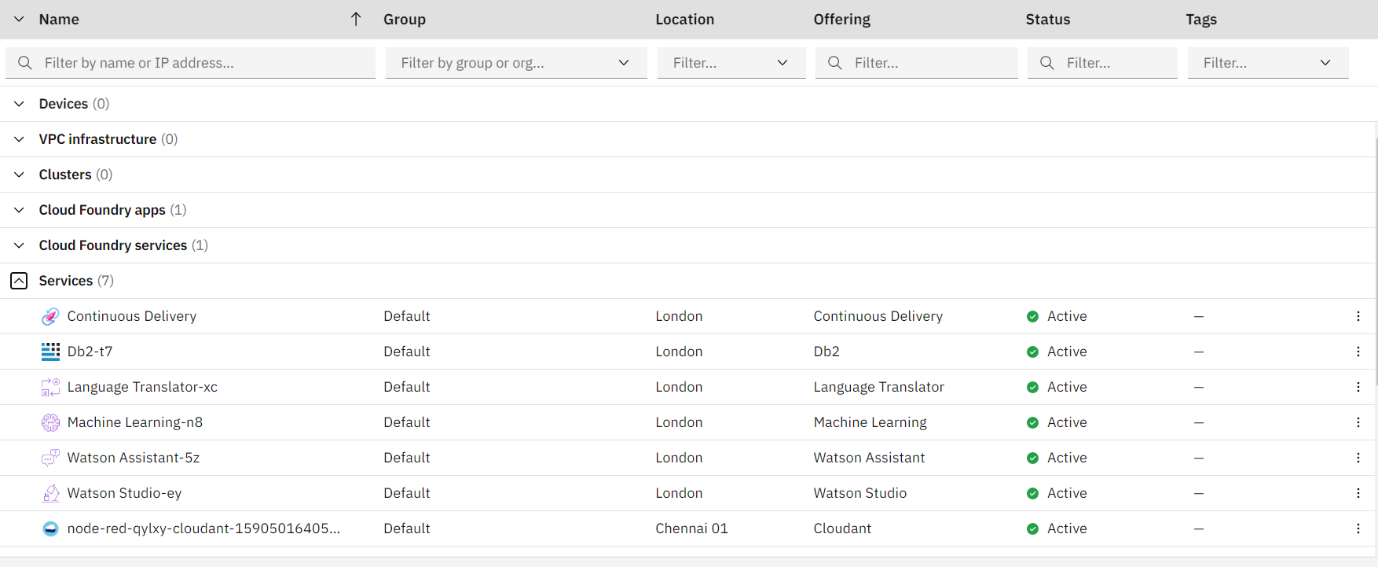
I have collected the data from kaggle.com. The data is given by WHO(World Health Organization). The data set is consists of 22 columns and 2938 rows.

The url of my dataset is : <https://www.kaggle.com/kumarajarshi/life-expectancy-who>

**2)Creating Necessery IBM cloud services:**

I have created necessery IBM cloud services which will help me to build my project further.

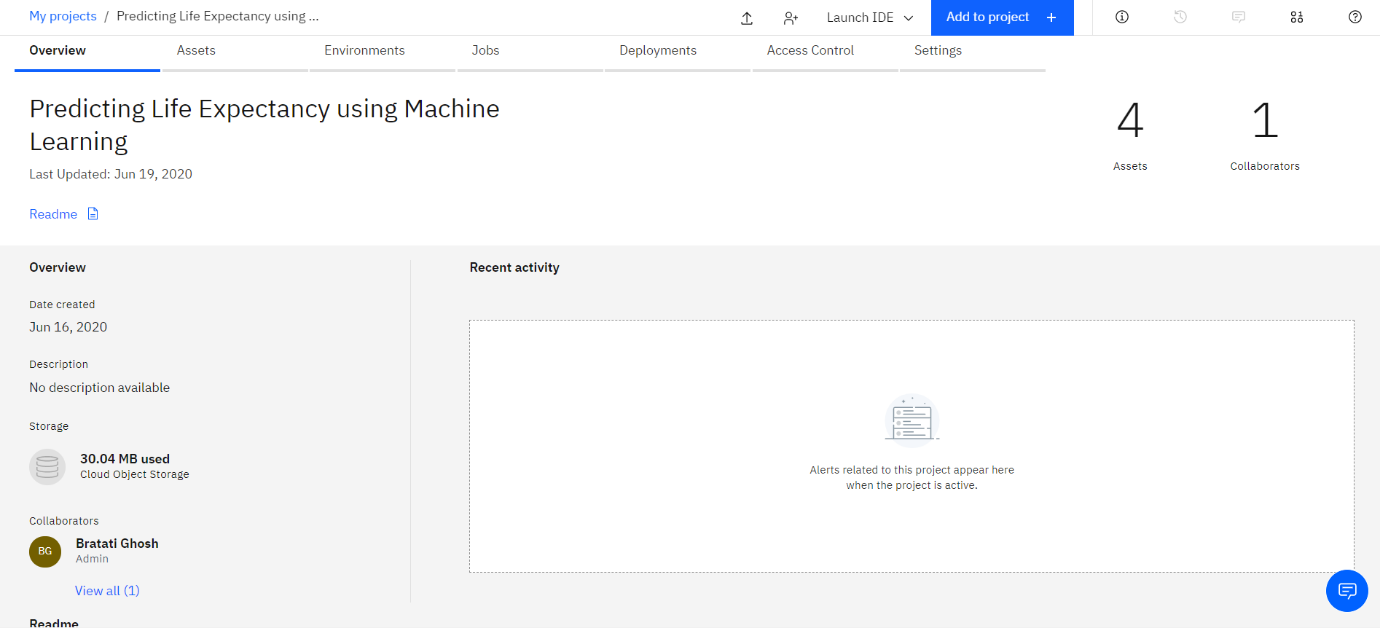
The screen shot of created services is given below:



**3)Creating a watson studio project:**

I have created a watson studion project to solve the given problem. The name of my project is : "[Prediction of life expectancy using Machine Learning](https://eu-gb.dataplatform.cloud.ibm.com/projects/261c10c0-e452-4205-b69f-5461ff7902c3?context=wdp)"

The screen shot of my created project is given below:



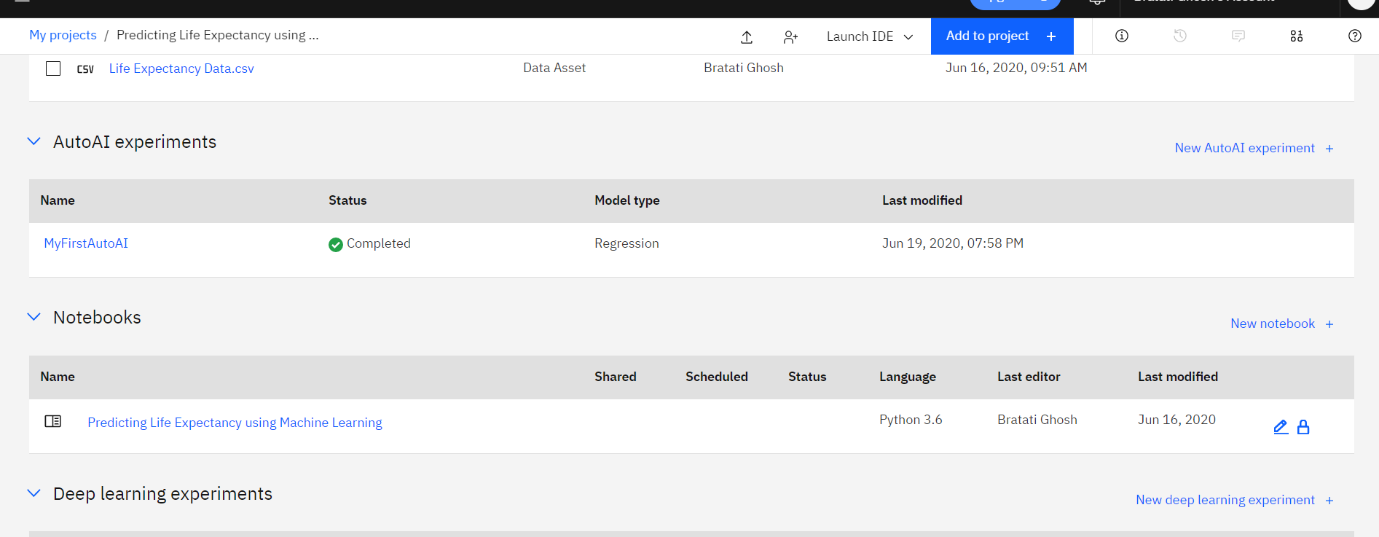
**4)Creating Machine Learning Service:**

I have created a Watson Machine Learning service for my project.

**5)Creating a Jupyter Notebook in Watson studio and importing the data:**

I have created a jupyter notebook and imported the data set in IBM cloud.

The screen shots are given below:



**6)Building a Machine Learning Model and creating Endpoints for Node-red Integration:**

Using Linear Regression I have built a Machine Learning model and created end-points for Node-red intregration.

**Linear Regression:**

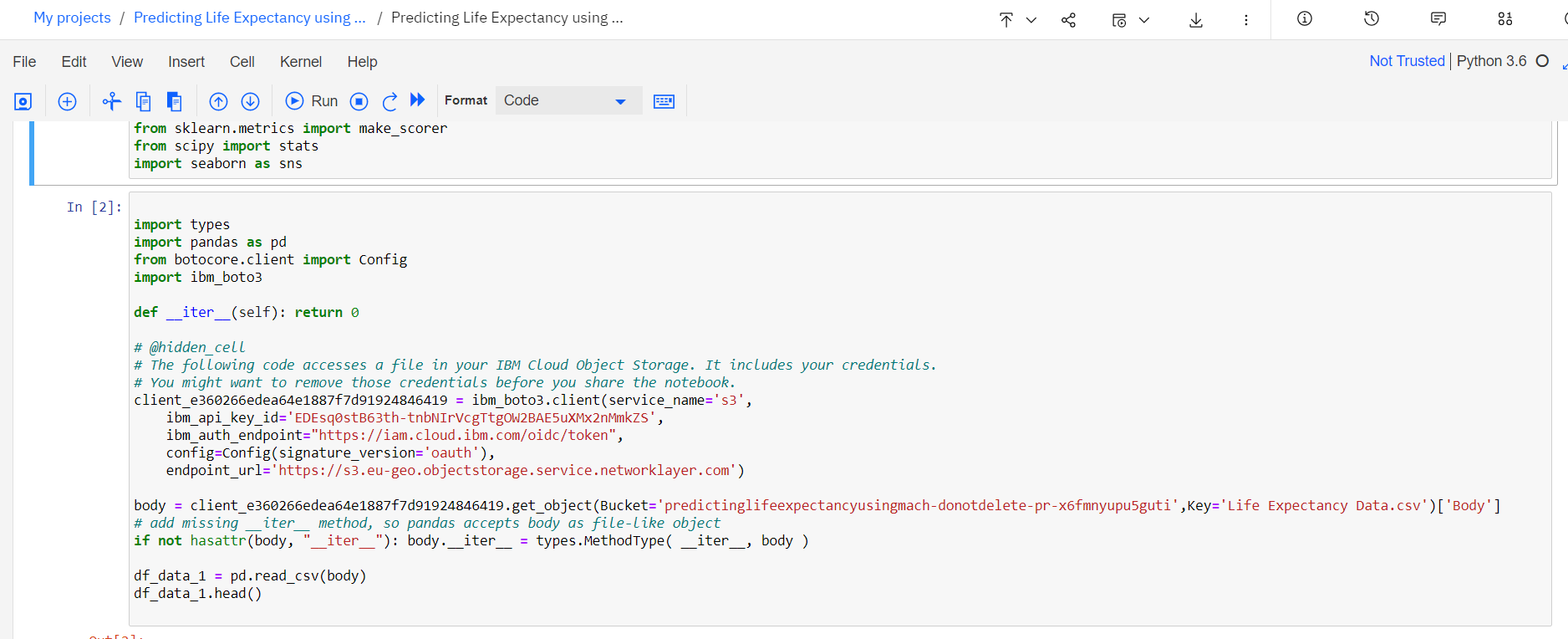
Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable. For example, a modeler might want to relate the weights of individuals to their heights using a linear regression model.

Before attempting to fit a linear model to observed data, a modeler should first determine whether or not there is a relationship between the variables of interest. This does not necessarily imply that one variable *causes* the other (for example, higher SAT scores do not *cause* higher college grades), but that there is some significant association between the two variables. A scatterplot can be a helpful tool in determining the strength of the relationship between two variables. If there appears to be no association between the proposed explanatory and dependent variables (i.e., the scatterplot does not indicate any increasing or decreasing trends), then fitting a linear regression model to the data probably will not provide a useful model. A valuable numerical measure of association between two variables is the correlation co-efficient, which is a value between -1 and 1 indicating the strength of the association of the observed data for the two variables.

A linear regression line has an equation of the form ***Y = a + bX***, where ***X*** is the explanatory variable and ***Y*** is the dependent variable. The slope of the line is ***b***, and

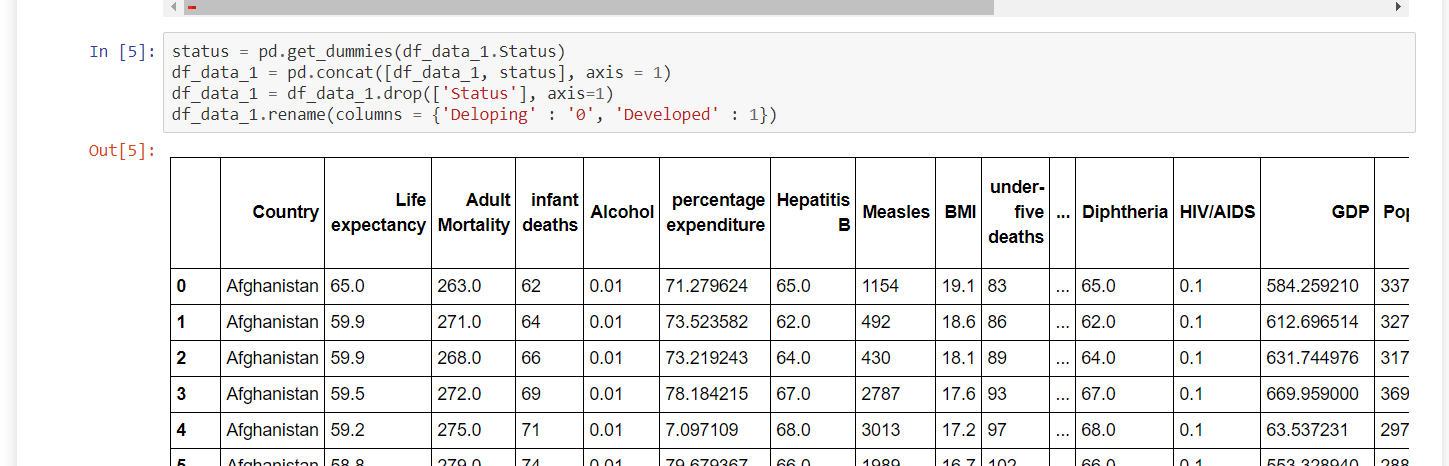
a is the intercept (the The screen shot of my code is given below:

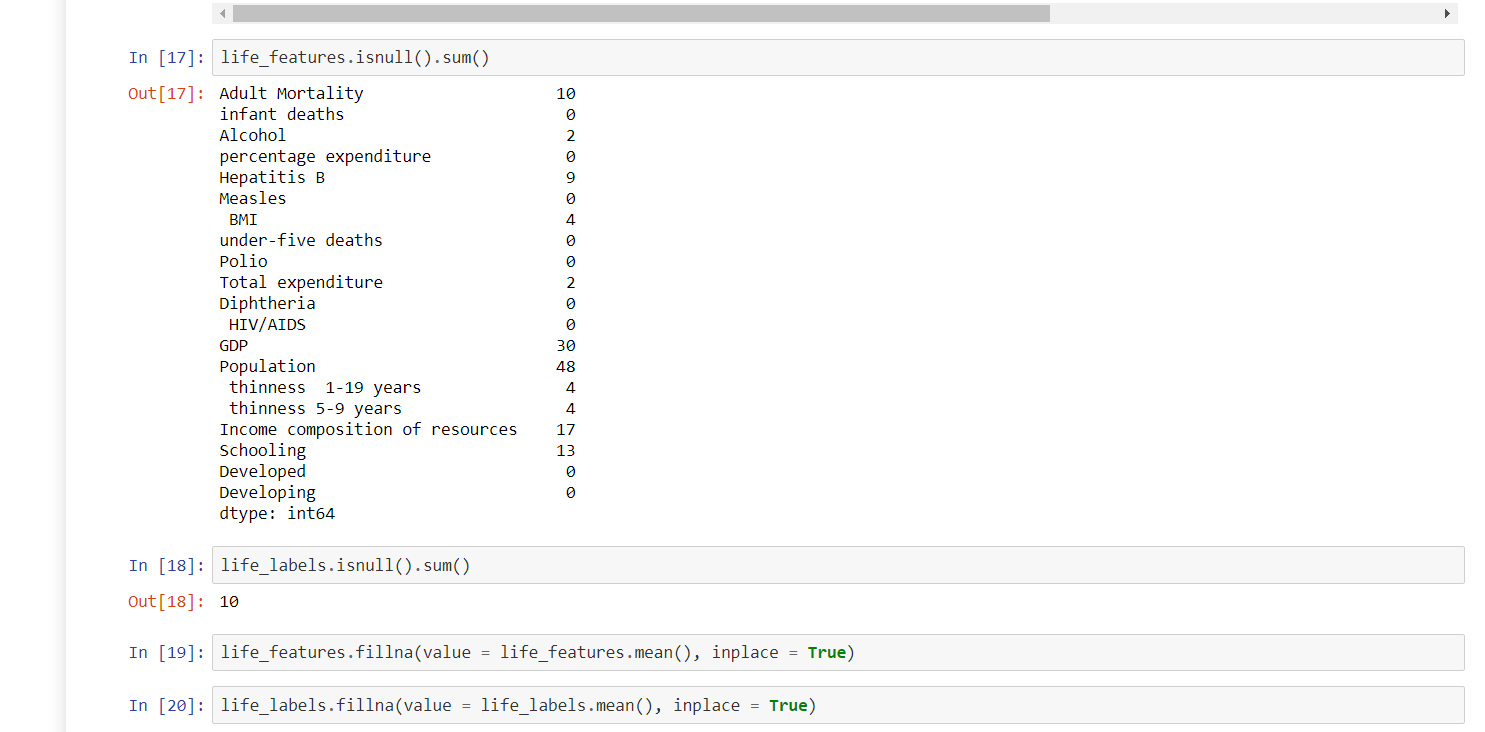
1)Importing Libraries and dataset into a dataframe:



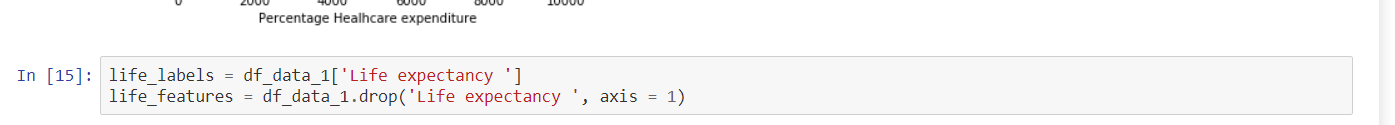


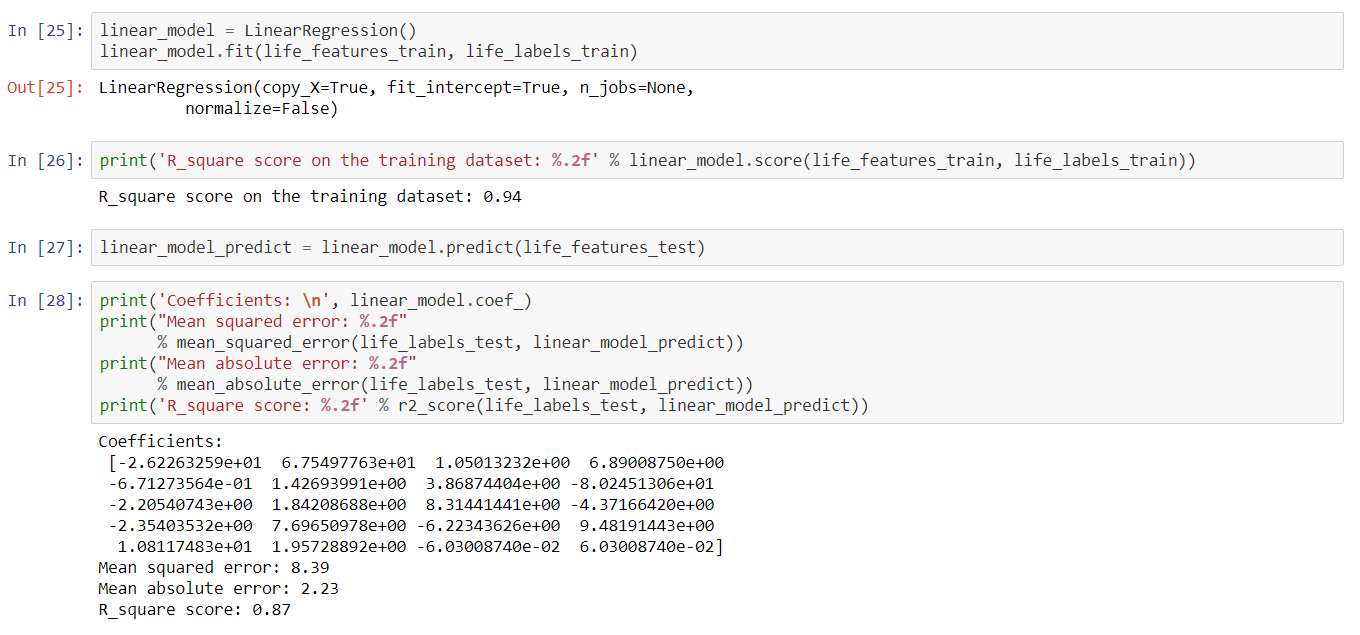
2) Handling missing/null values:





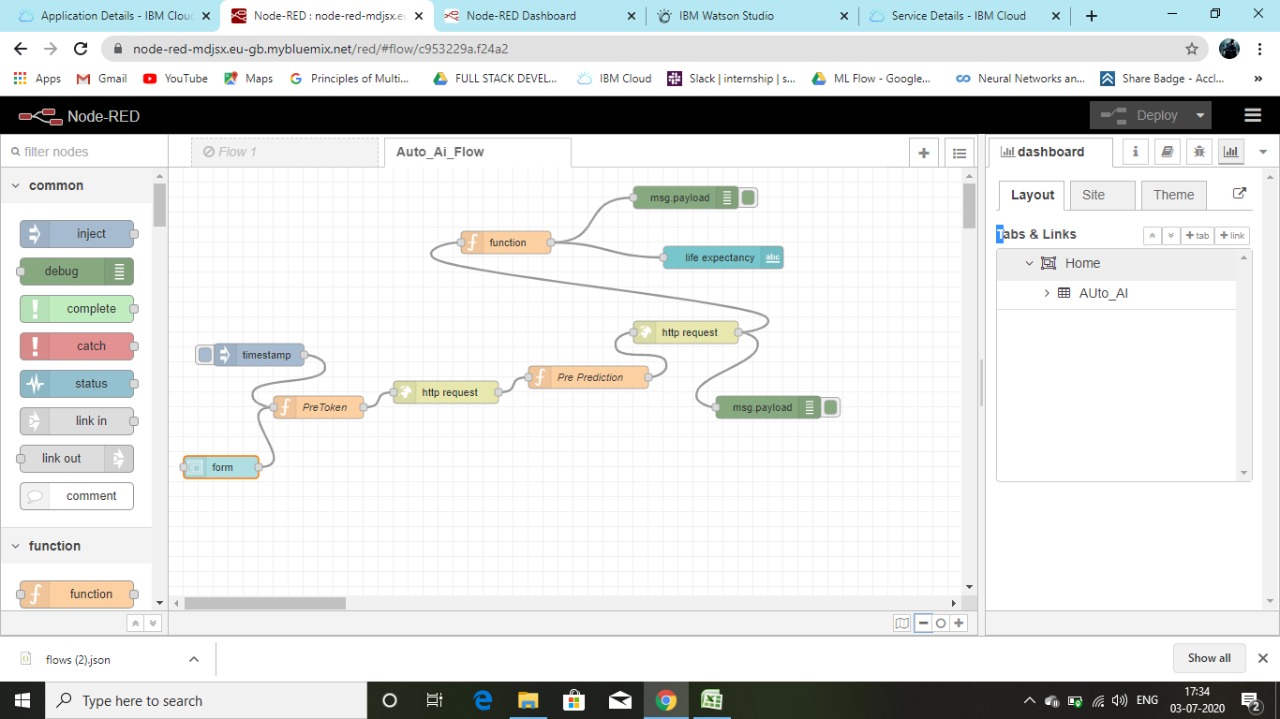
3)splitting dataset and model building:

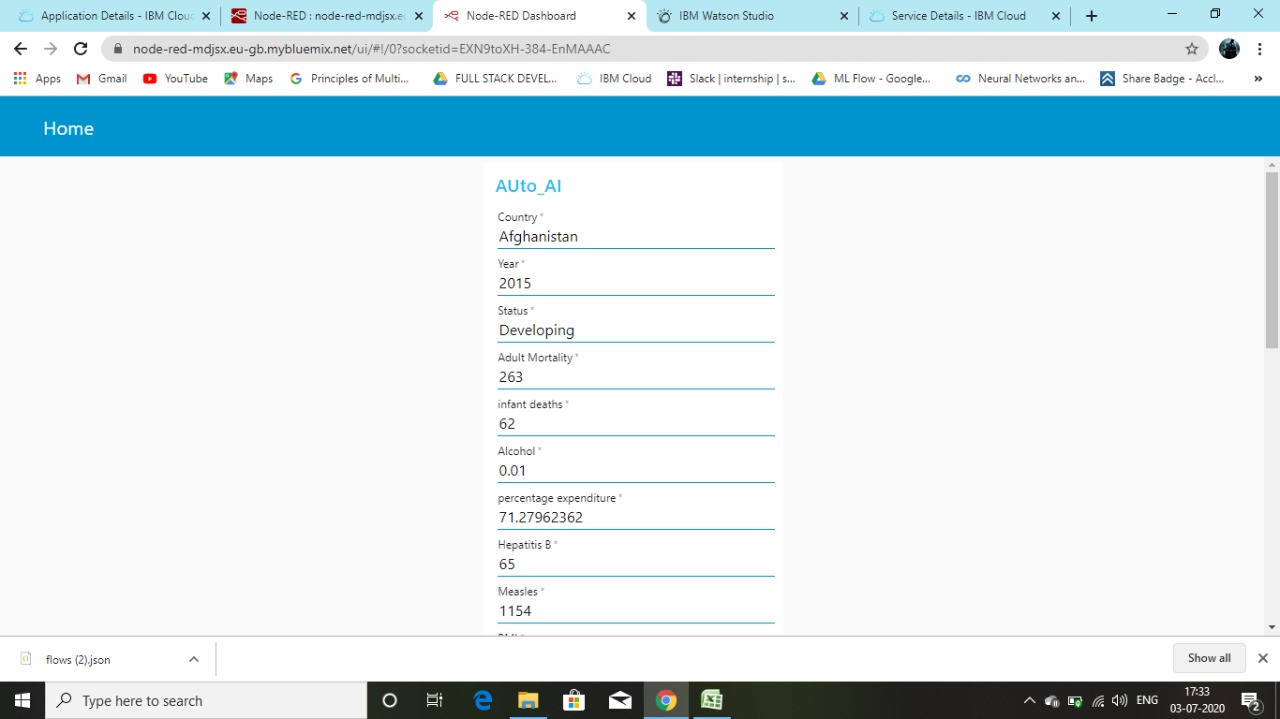


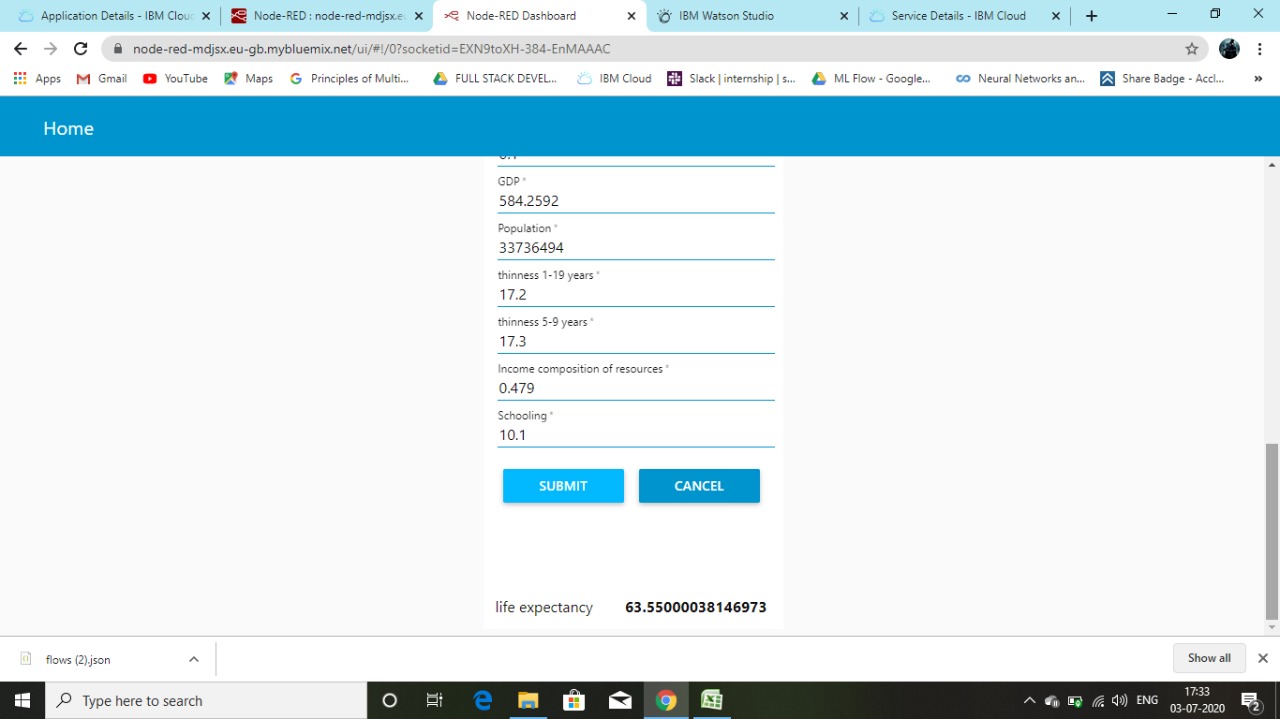


7)node-red representation:

**[{"id":"7429c9e1.3161d8","type":"ui\_form","z":"c953229a.f24a2","name":"","label":"","group":"4e357c19.5bad54","order":1,"width":0,"height":0,"options":[{"label":"Country","value":"a","type":"text","required":true,"rows":null},{"label":"Year","value":"b","type":"number","required":true,"rows":null},{"label":"Status","value":"c","type":"text","required":true,"rows":null},{"label":"Adult Mortality","value":"d","type":"number","required":true,"rows":null},{"label":"infant deaths","value":"e","type":"number","required":true,"rows":null},{"label":"Alcohol","value":"f","type":"number","required":true,"rows":null},{"label":"percentage expenditure","value":"g","type":"number","required":true,"rows":null},{"label":"Hepatitis B","value":"h","type":"number","required":true,"rows":null},{"label":"Measles ","value":"i","type":"number","required":true,"rows":null},{"label":" BMI ","value":"j","type":"number","required":true,"rows":null},{"label":"under-five deaths ","value":"k","type":"number","required":true,"rows":null},{"label":"Polio","value":"l","type":"number","required":true,"rows":null},{"label":"Total expenditure","value":"m","type":"number","required":true,"rows":null},{"label":"Diphtheria ","value":"n","type":"number","required":true,"rows":null},{"label":" HIV/AIDS","value":"o","type":"number","required":true,"rows":null},{"label":"GDP","value":"p","type":"number","required":true,"rows":null},{"label":"Population","value":"q","type":"number","required":true,"rows":null},{"label":" thinness  1-19 years","value":"r","type":"number","required":true,"rows":null},{"label":"thinness 5-9 years","value":"s","type":"number","required":true,"rows":null},{"label":"Income composition of resources","value":"t","type":"number","required":true,"rows":null},{"label":"Schooling","value":"u","type":"number","required":true,"rows":null}],"formValue":{"a":"","b":"","c":"","d":"","e":"","f":"","g":"","h":"","i":"","j":"","k":"","l":"","m":"","n":"","o":"","p":"","q":"","r":"","s":"","t":"","u":""},"payload":"","submit":"submit","cancel":"cancel","topic":"","x":70,"y":400,"wires":[["8ee77c66.5145"]]},{"id":"4e357c19.5bad54","type":"ui\_group","z":"","name":"AUto\_AI","tab":"61ef45cd.b0c28c","order":2,"disp":true,"width":"6","collapse":false},{"id":"61ef45cd.b0c28c","type":"ui\_tab","z":"","name":"Home","icon":"dashboard","disabled":false,"hidden":false}]**







**8) Using Streamlit library:**

**app.py**

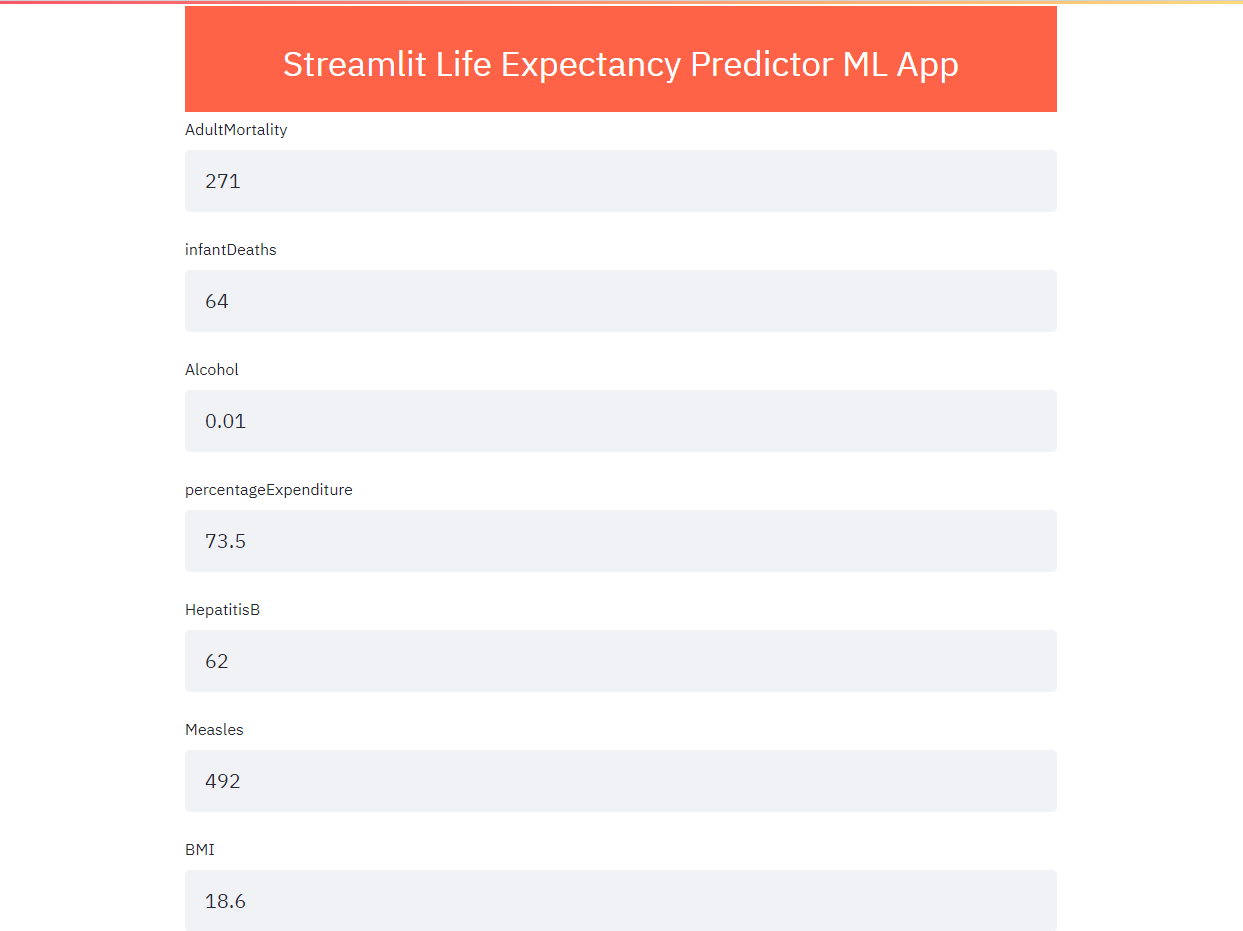
**import numpy as np**  
**import pickle**  
**import pandas as pd**  
**#from flasgger import Swagger**  
**import streamlit as st**   
  
**from PIL import Image**  
  
**#app=Flask(\_\_name\_\_)**  
**#Swagger(app)**  
  
**model=pickle.load(open('model.pkl','rb'))**  
  
**#@app.route('/')**  
**def welcome():**  
**return "Welcome All"**  
  
**#@app.route('/predict',methods=["Get"])**  
**def predict\_note\_authentication(AdultMortality,infantDeaths,Alcohol,percentageExpenditure,HepatitisB,Measles,BMI,underFiveDeaths,Polio,TotalExpenditure,Diphtheria,HIV\_AIDS,GDP,Population,thinness\_1\_19\_years,thinness\_5\_9\_years,Income\_composition\_of\_resources,Schooling):**  
**input=np.array([[AdultMortality,infantDeaths,Alcohol,percentageExpenditure,HepatitisB,Measles,BMI,underFiveDeaths,Polio,TotalExpenditure,Diphtheria,HIV\_AIDS,GDP,Population,thinness\_1\_19\_years,thinness\_5\_9\_years,Income\_composition\_of\_resources,Schooling]]).astype(np.float64)**  
  
**prediction=model.predict(input)**  
**print((prediction))**  
**return(prediction)**  
  
  
  
**def main():**  
**st.title("Life Expectancy Prediction")**  
**html\_temp = """**  
**<div style="background-color:tomato;padding:10px">**  
**<h2 style="color:white;text-align:center;">Streamlit Life Expectancy Predictor ML App </h2>**  
**</div>**  
**"""**  
**st.markdown(html\_temp,unsafe\_allow\_html=True)**  
**AdultMortality = st.text\_input("AdultMortality","Type Here")**  
**infantDeaths= st.text\_input("infantDeaths","Type Here")**  
**Alcohol= st.text\_input("Alcohol","Type Here")**  
**percentageExpenditure = st.text\_input("percentageExpenditure","Type Here")**  
**HepatitisB = st.text\_input("HepatitisB","Type Here")**  
**Measles = st.text\_input("Measles","Type Here")**  
**BMI = st.text\_input("BMI","Type Here")**  
**underFiveDeaths = st.text\_input("underFiveDeaths","Type Here")**  
**Polio = st.text\_input("Polio","Type Here")**  
**TotalExpenditure = st.text\_input("TotalExpenditure","Type Here")**  
**Diphtheria= st.text\_input("Diphtheria","Type Here")**  
**HIV\_AIDS = st.text\_input("HIV\_AIDS","Type Here")**  
**GDP = st.text\_input("GDP","Type Here")**  
**Population = st.text\_input("Population","Type Here")**  
**thinness\_1\_19\_years = st.text\_input("thinness\_1\_19\_years","Type Here")**  
**thinness\_5\_9\_years = st.text\_input("thinness\_5\_9\_years","Type Here")**  
**Income\_composition\_of\_resources = st.text\_input("Income\_composition\_of\_resources","Type Here")**  
**Schooling = st.text\_input("Schooling","Type Here")**  
**result=""**  
**if st.button("Predict"):**  
**result=predict\_note\_authentication(AdultMortality,infantDeaths,Alcohol,percentageExpenditure,HepatitisB,Measles,BMI,underFiveDeaths,Polio,TotalExpenditure,Diphtheria,HIV\_AIDS,GDP,Population,thinness\_1\_19\_years,thinness\_5\_9\_years,Income\_composition\_of\_resources,Schooling)**  
  
**st.success('The output is {}'.format(result))**  
**if st.button("About"):**  
**st.text("Lets LEarn")**  
**st.text("Built with Streamlit")**

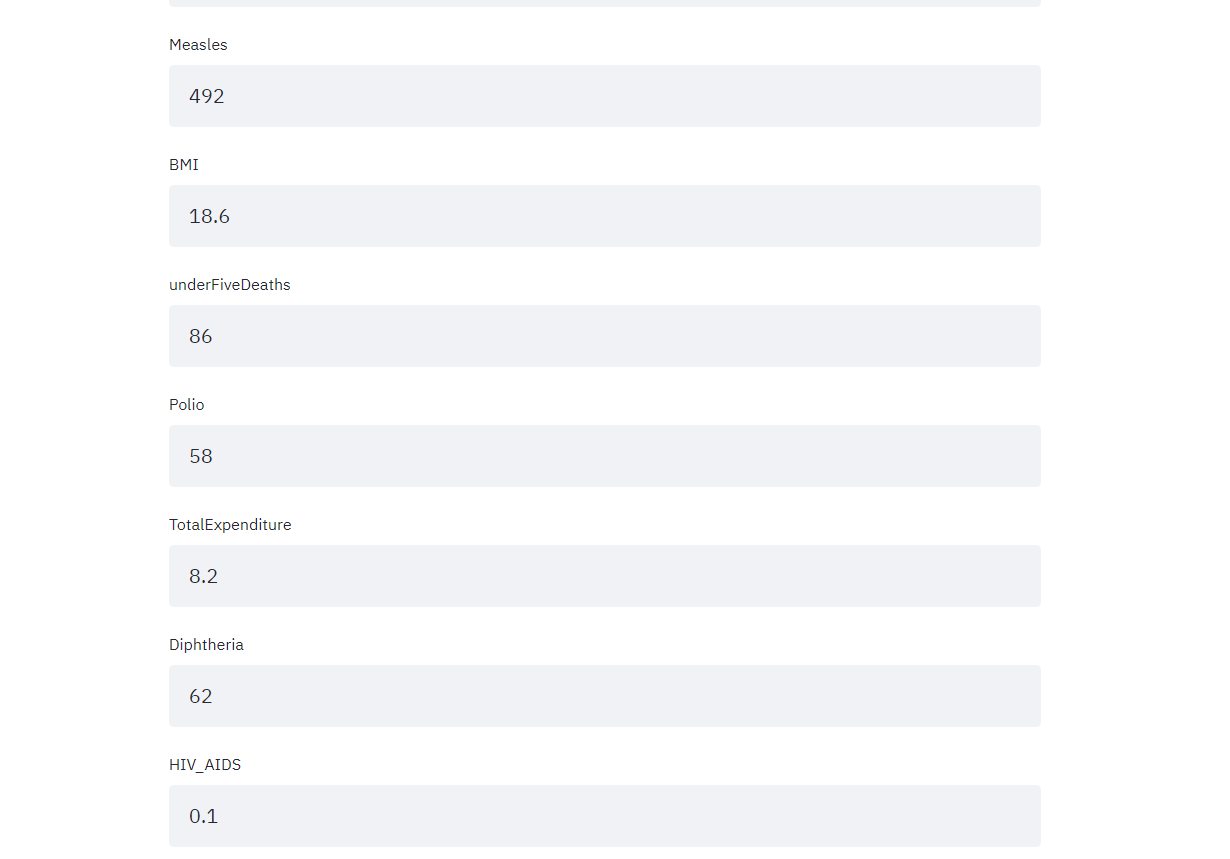
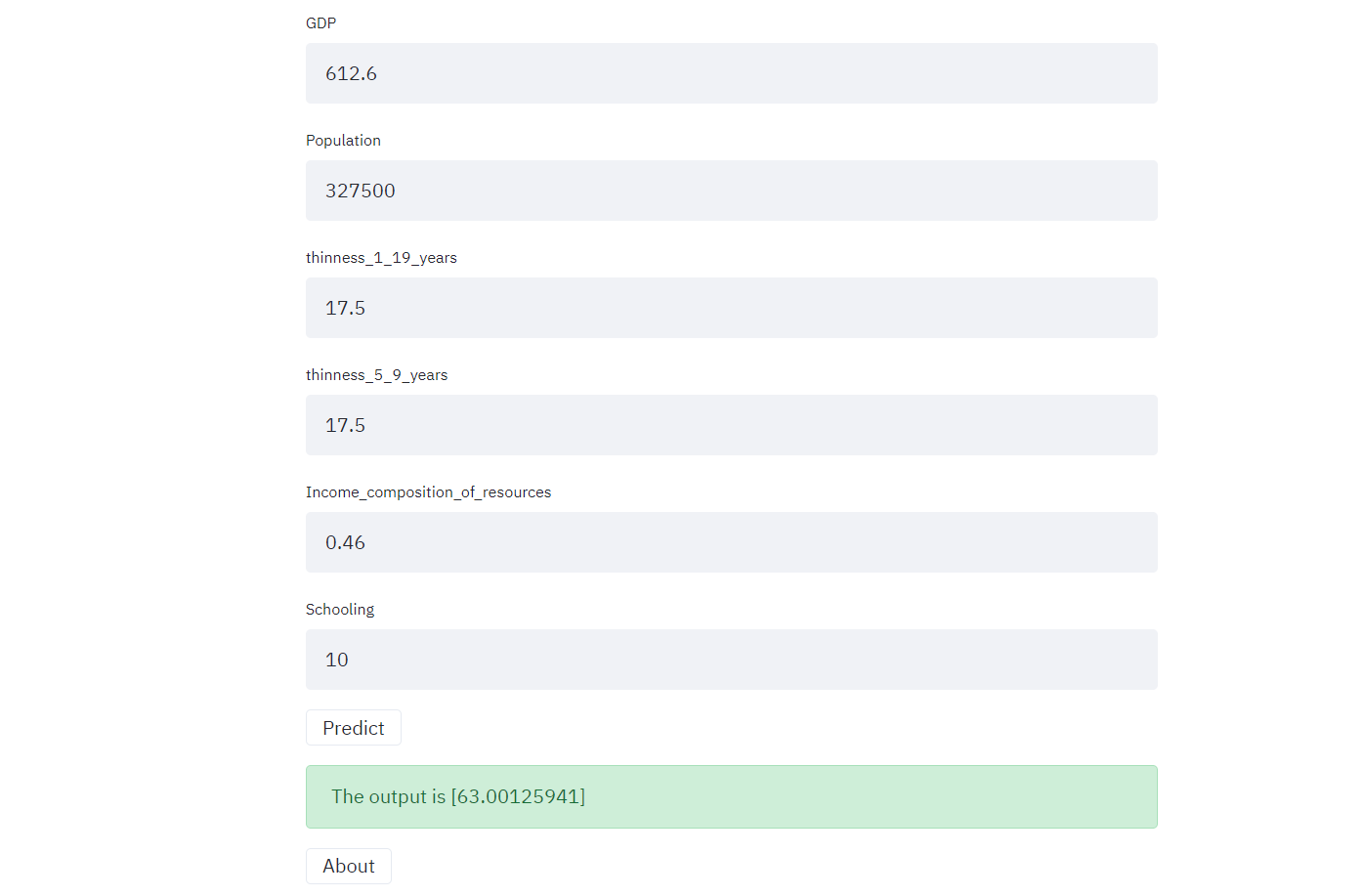
**if \_\_name\_\_=='\_\_main\_\_':**  
**main()**

**test.py**

import numpy as np  
import pandas as pd  
import matplotlib.pyplot as plt # ploting , visualization  
import seaborn as sns # ploting  
from sklearn import model\_selection #scikit learn  
from sklearn import linear\_model  
from sklearn import metrics  
from sklearn import preprocessing  
from sklearn import utils  
from sklearn import feature\_selection  
import warnings  
import pickle  
warnings.filterwarnings("ignore")  
df= pd.read\_csv("e:/C/datasets\_12603\_17232\_Life Expectancy Data.csv")  
df.head()  
df.isnull().sum()  
df=df.dropna()  
X=df.drop(["Year","Country","Status","Life expectancy "],axis=1)  
Y=df['Life expectancy ']  
Y=Y.astype('int')  
X=X.astype('int')  
from sklearn.model\_selection import train\_test\_split  
Xtrain, Xtest, Ytrain, Ytest = train\_test\_split(X, Y, test\_size=0.25, random\_state=42)  
l=linear\_model.LinearRegression()  
l.fit(Xtrain,Ytrain)  
pickle.dump(l,open('model.pkl','wb'))  
model=pickle.load(open('model.pkl','rbs'))

The screenshot of the output is given below:





[Predicting Life Expectancy Without Python](https://smartinternz.com/Student/workspace/2864#collapse6)

**1)Collecting the dataset for the project:**

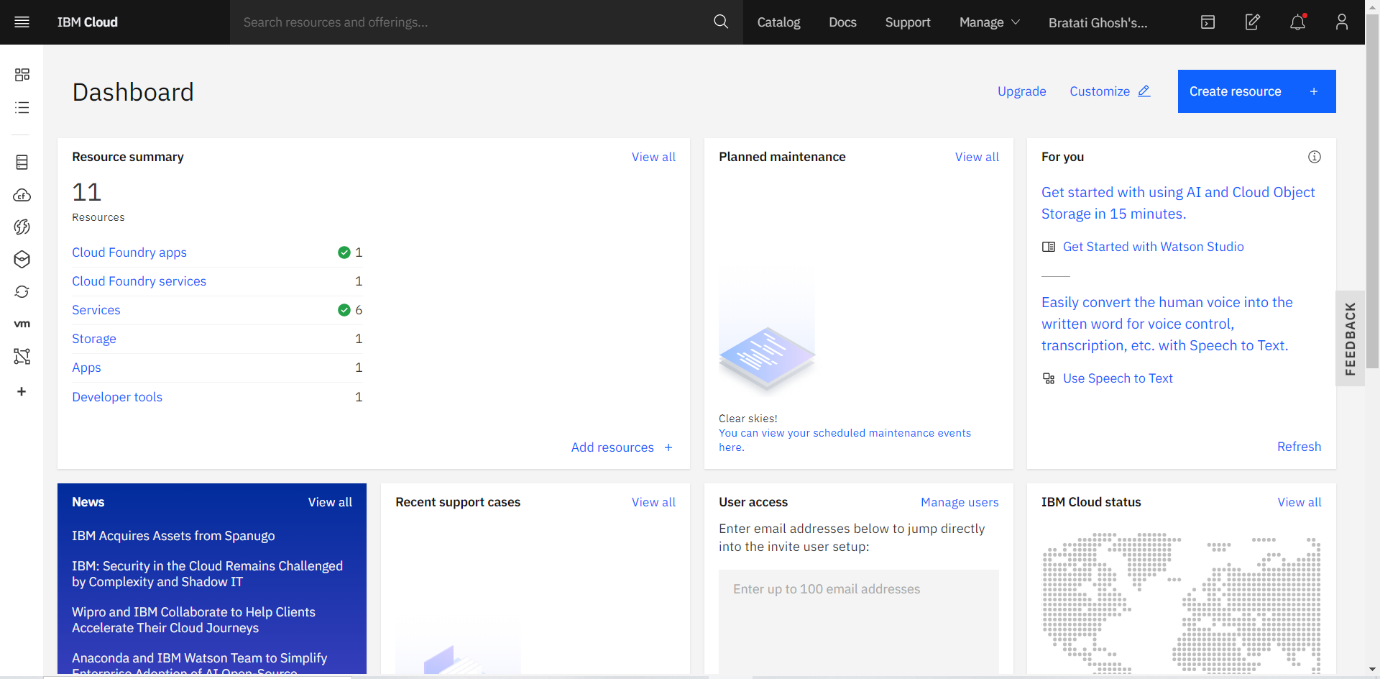
I have collected the data from kaggle.com. The data is given by WHO(World Health Organization). The data set is consists of 22 columns and 2938 rows.

The url of my dataset is : <https://www.kaggle.com/kumarajarshi/life-expectancy-who>

**2)Creating Necessery IBM cloud services:**

I have created necessery IBM cloud services which will help me to build my project further.

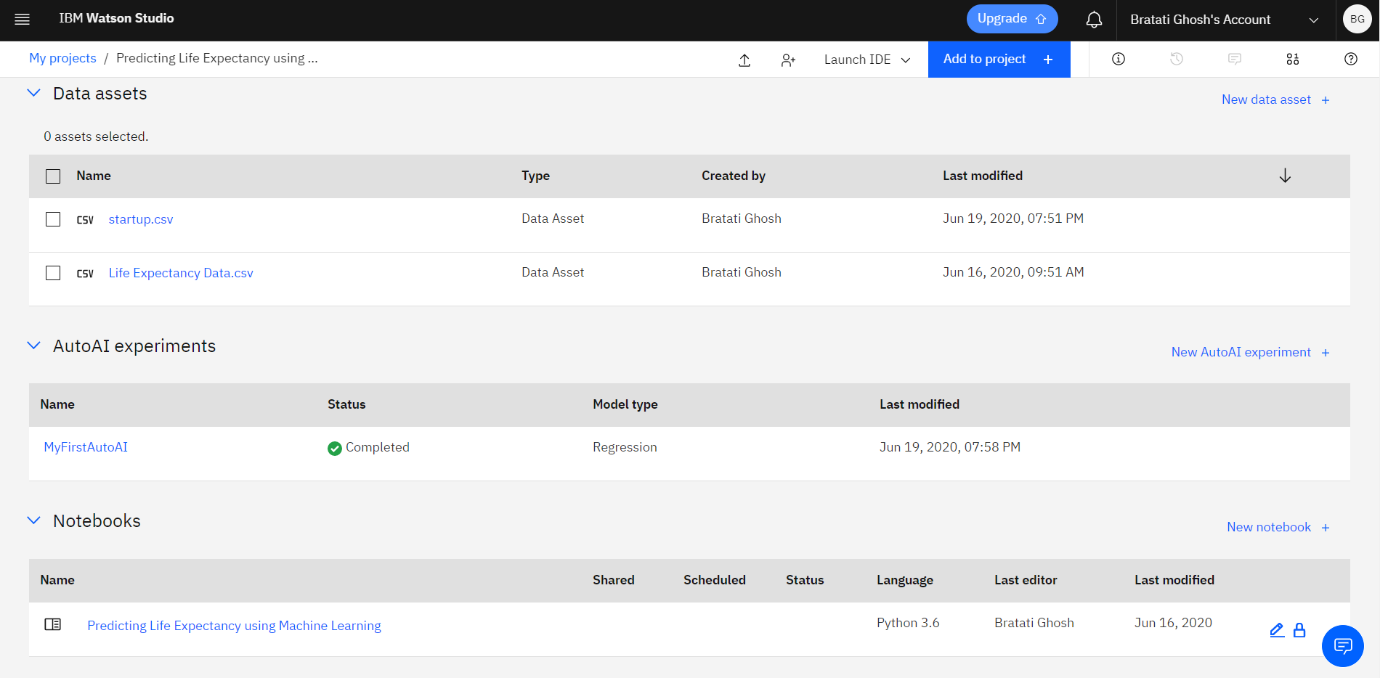
The screen shot of created services is given below:



**3)Creating a watson studio project:**

I have created a watson studion project to solve the given problem. The name of my project is : "[Prediction of life expectancy using Machine Learning](https://eu-gb.dataplatform.cloud.ibm.com/projects/261c10c0-e452-4205-b69f-5461ff7902c3?context=wdp)"

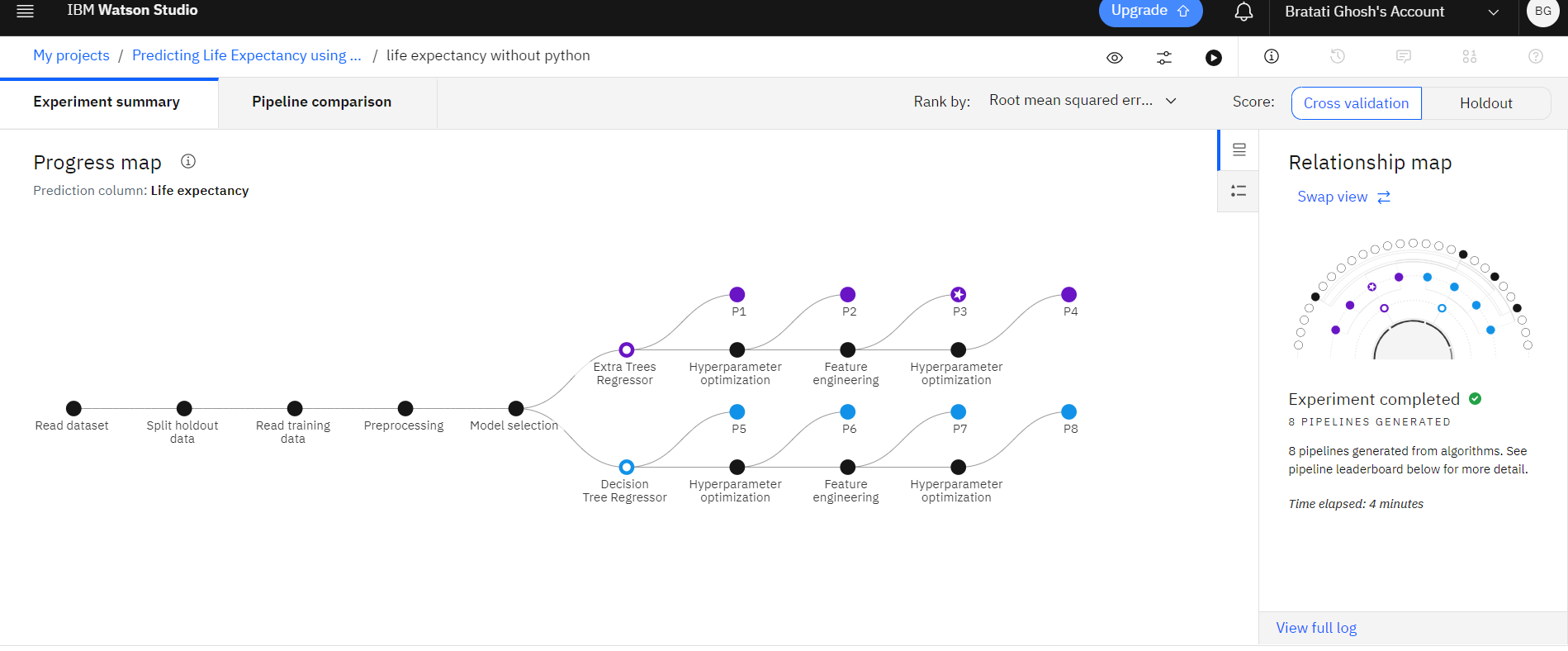
The screen shot of my created project is given below:

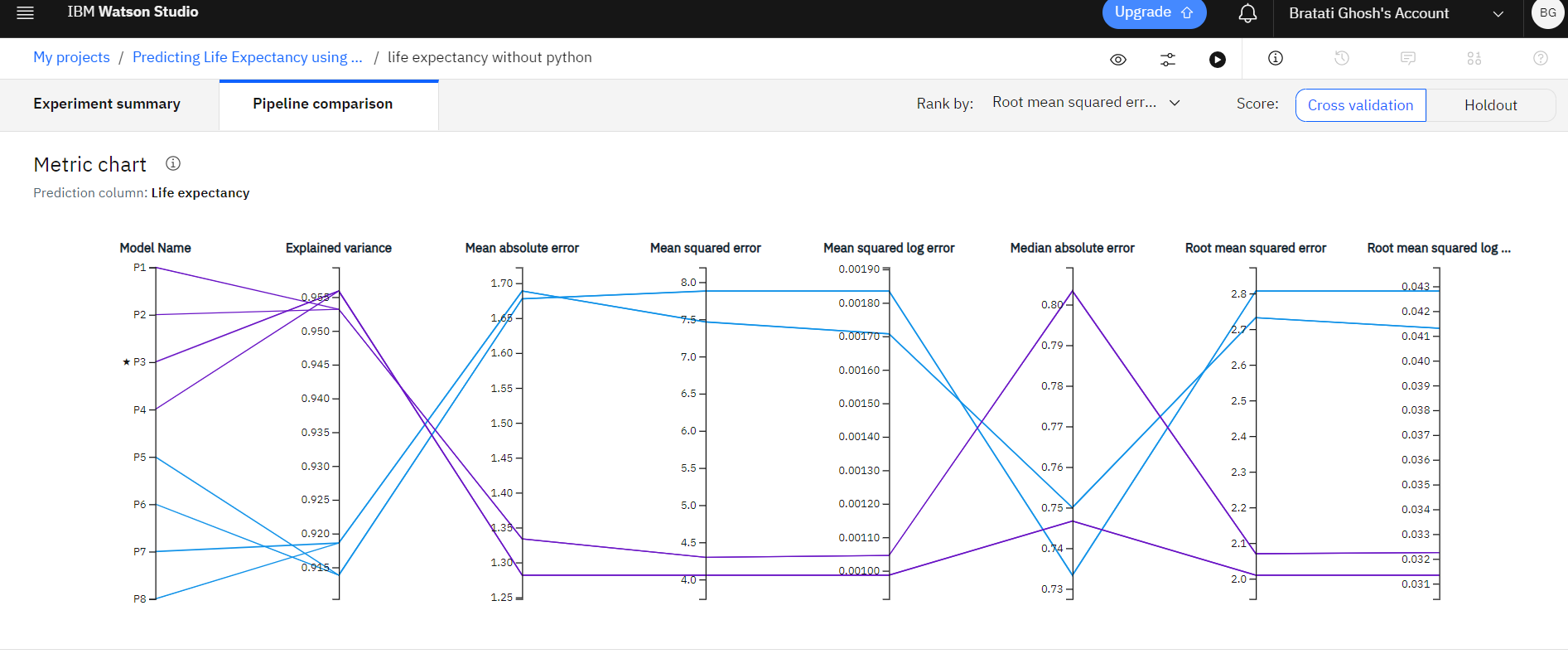


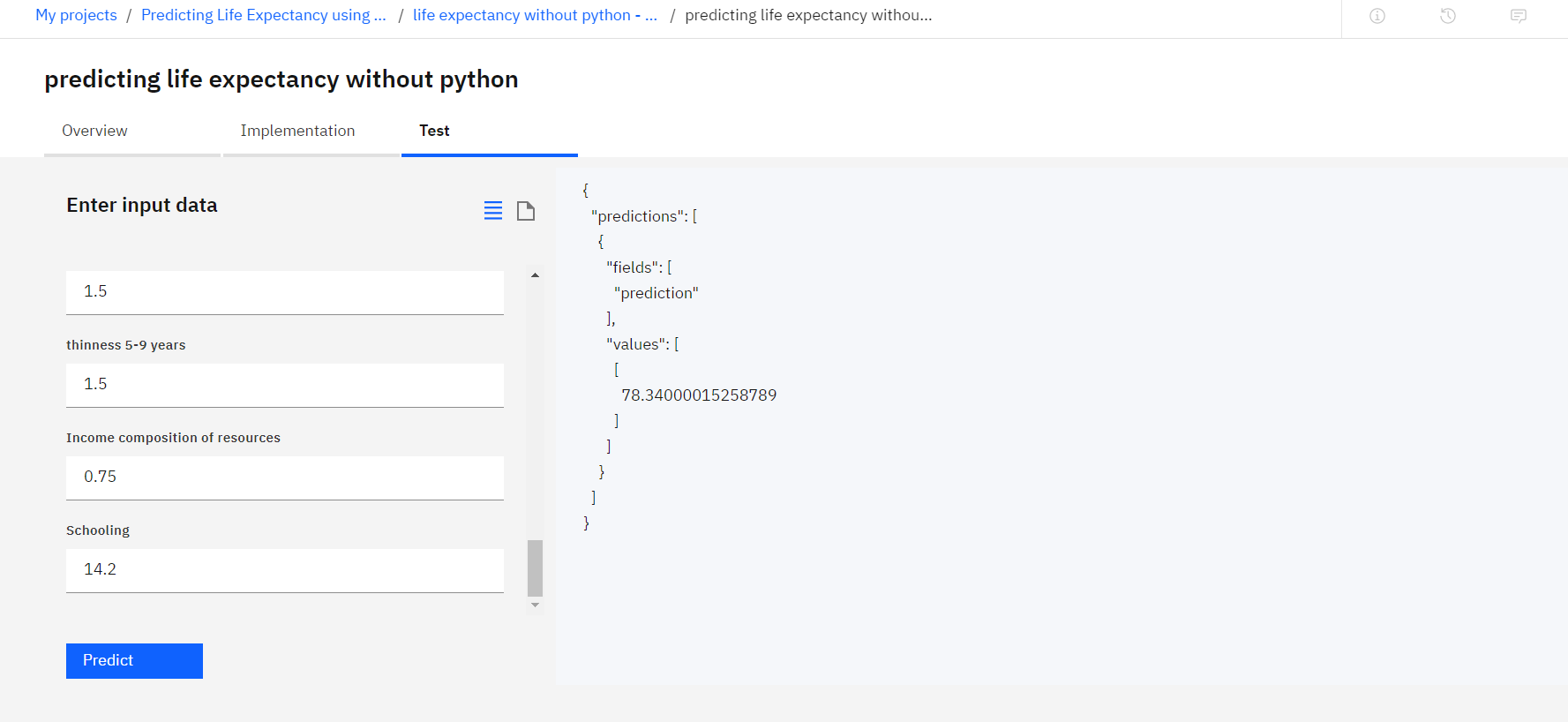
**4)Creating Machine Learning Service:**

I have created a Watson Machine Learning service for my project.

**5)Importing dataset and creating AutoAi Experiment:**







**Node-red Flow:**

**[{"id":"7429c9e1.3161d8","type":"ui\_form","z":"c953229a.f24a2","name":"","label":"","group":"4e357c19.5bad54","order":1,"width":0,"height":0,"options":[{"label":"Country","value":"a","type":"text","required":true,"rows":null},{"label":"Year","value":"b","type":"number","required":true,"rows":null},{"label":"Status","value":"c","type":"text","required":true,"rows":null},{"label":"Adult Mortality","value":"d","type":"number","required":true,"rows":null},{"label":"infant deaths","value":"e","type":"number","required":true,"rows":null},{"label":"Alcohol","value":"f","type":"number","required":true,"rows":null},{"label":"percentage expenditure","value":"g","type":"number","required":true,"rows":null},{"label":"Hepatitis B","value":"h","type":"number","required":true,"rows":null},{"label":"Measles ","value":"i","type":"number","required":true,"rows":null},{"label":" BMI ","value":"j","type":"number","required":true,"rows":null},{"label":"under-five deaths ","value":"k","type":"number","required":true,"rows":null},{"label":"Polio","value":"l","type":"number","required":true,"rows":null},{"label":"Total expenditure","value":"m","type":"number","required":true,"rows":null},{"label":"Diphtheria ","value":"n","type":"number","required":true,"rows":null},{"label":" HIV/AIDS","value":"o","type":"number","required":true,"rows":null},{"label":"GDP","value":"p","type":"number","required":true,"rows":null},{"label":"Population","value":"q","type":"number","required":true,"rows":null},{"label":" thinness  1-19 years","value":"r","type":"number","required":true,"rows":null},{"label":"thinness 5-9 years","value":"s","type":"number","required":true,"rows":null},{"label":"Income composition of resources","value":"t","type":"number","required":true,"rows":null},{"label":"Schooling","value":"u","type":"number","required":true,"rows":null}],"formValue":{"a":"","b":"","c":"","d":"","e":"","f":"","g":"","h":"","i":"","j":"","k":"","l":"","m":"","n":"","o":"","p":"","q":"","r":"","s":"","t":"","u":""},"payload":"","submit":"submit","cancel":"cancel","topic":"","x":70,"y":400,"wires":[["8ee77c66.5145"]]},{"id":"4e357c19.5bad54","type":"ui\_group","z":"","name":"AUto\_AI","tab":"61ef45cd.b0c28c","order":2,"disp":true,"width":"6","collapse":false},{"id":"61ef45cd.b0c28c","type":"ui\_tab","z":"","name":"Home","icon":"dashboard","disabled":false,"hidden":false}]**

