Project Name: Predicting Life Expectancy using Machine Learning

1. Project Summary:

A typical Regression Machine Learning project leverages historical data to predict insights into the future. This problem statement is aimed at predicting the 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 the 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.

Thus we will be developing an application using ML algorithm for the prediction of the life expectancy of such patients.

2. Project Requirements:

The project will help in determing the life expectancy rate on the basic of the dataset. The dataset will provide various details like kind of diseases leading to the death. thus with help of this information we will predict the life expectancy.

2.1 Functional Requirement:

Predicting life expectancy rate of a country.

2.2 Technical Requirements:

- Programming language - Python
- Machine learning Algorithms
- **The algorithms have been used are as follows**:
  Linear Regression
  Ridge Regression
  Lasso Regression
  Elastic Net Regression
  Linear Regression with Polynomic features
  Decision Tree Regression
  Random Forest Regression

2.3 Software Requirements:

- IBM cloud
- IBM Watson Studio
- Node-Red

3. Project Deliverables:

The application will help out in predicting the life expectancy on the basis of the information provided and which disease have the highest life expectancy rate can be predicted through the model.

4. Project Team:

This is a solo project.

5. Project Schedule:

The project is Scheduled for one month from 16 May 2020 to 15 June 2020.

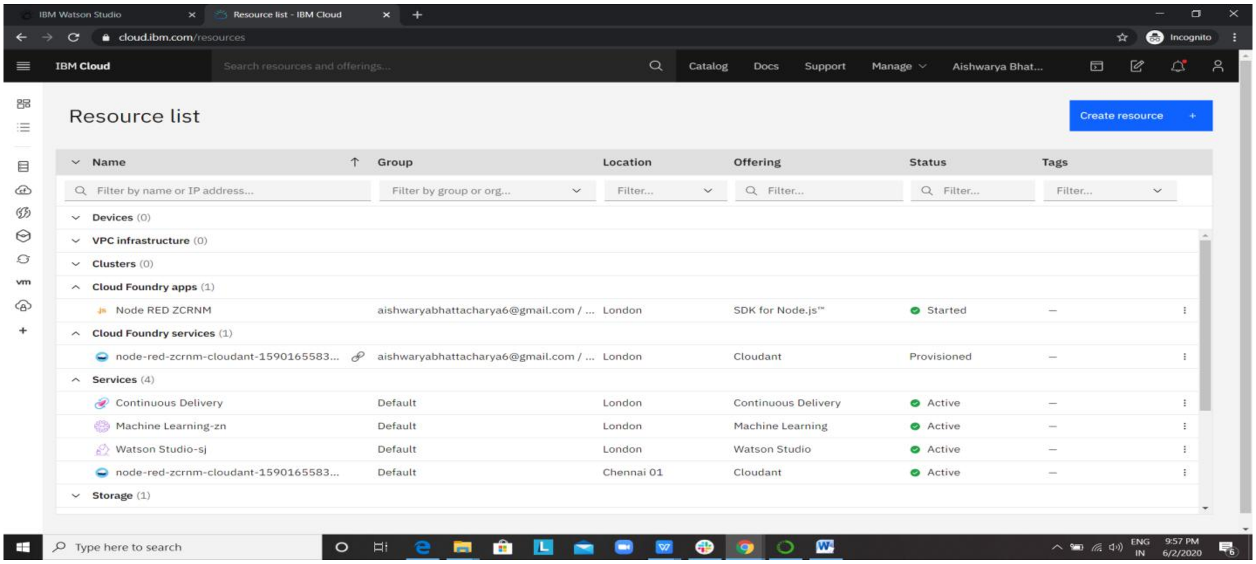
6. Phases in the Development:

1. Collecting the Dataset:

First most important thing for any project is collecting the data as per requirement of the model. Thus firstly we collect the data from the given source. For the project the dataset was "Life Expectancy".
Thus dataset was provided by the WHO in order for the analysis purpose. We have used this dataset for the prediction purpose.

2. Setting up IBM Cloud Services:

For using the various Cloud services for the project development. One must first create an IBM Cloud account. Once the account is created we can access various services used for ML projects.



Resource list

Name	Group	Location	Offering	Status	Tags
Filter by name or IP address...	Filter by group or org...	Filter...	Filter...	Filter...	Filter...
Devices (0)					
VPC infrastructure (0)					
Clusters (0)					
Cloud Foundry apps (1)					
Node RED ZCRNM	aishwaryabhattacharya6@gmail.com / ...	London	SDK for Node.js™	Started	–
Cloud Foundry services (1)					
node-red-zcrnm-cloudant-1590165583...	aishwaryabhattacharya6@gmail.com / ...	London	Cloudant	Provisioned	–
Services (4)					
Continuous Delivery	Default	London	Continuous Delivery	Active	–
Machine Learning-zn	Default	London	Machine Learning	Active	–
Watson Studio-sj	Default	London	Watson Studio	Active	–
node-red-zcrnm-cloudant-1590165583...	Default	Chennai 01	Cloudant	Active	–
Storage (1)					

3. Creating a Watson Project:

Once the services required for the project are enabled we can go with for the creation of the project. Watson Studio allows us to create various project using different tools like Jupyter notebook, Auto AI, R Studio etc.

a) Configure the Watson studio:

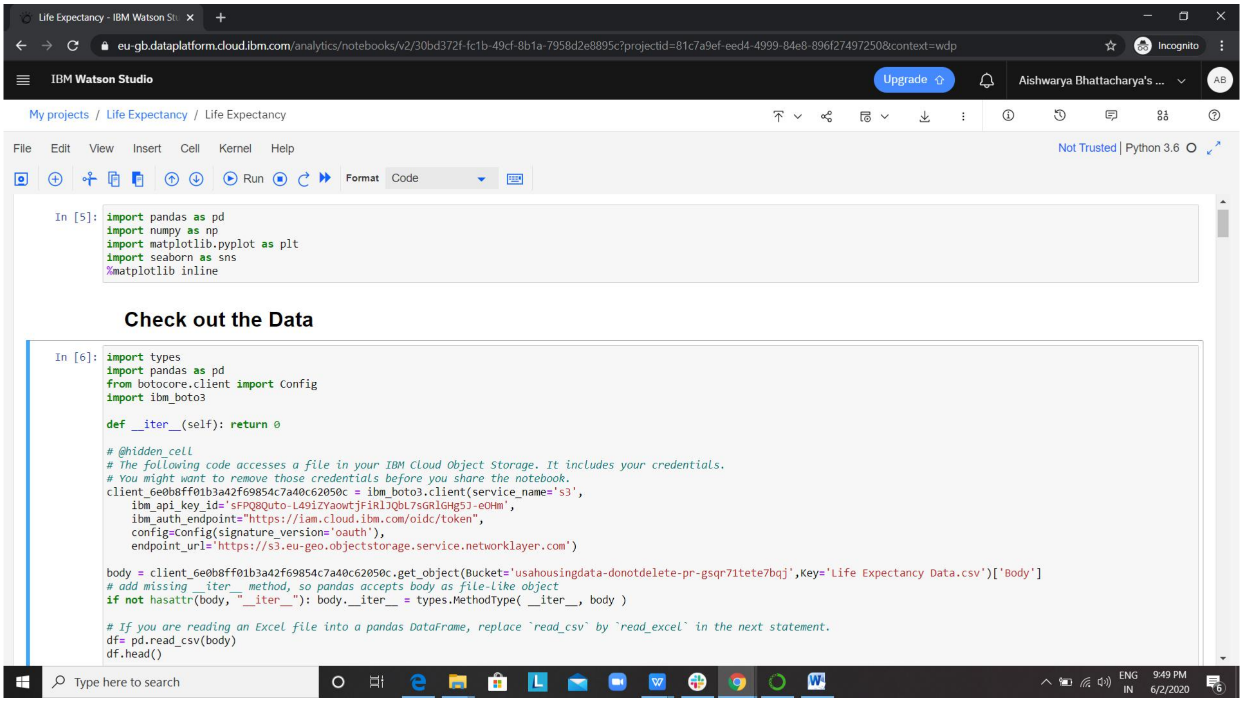
Once we are done with the creation of the watson project we can configure the various services associated with it. Also we can look for the various tools associated with it.

4. Creating Machine Learning Services:

As we are creating the Machine Learning Model for the prediction of the Life Expectancy we must create the Machine learning services in IBM cloud which will help in building up the model.

a) Create Jupyter Notebook and Import Dataset:

Firstly in the project we need to add the Jupyter Notebook(It is the platform for developing the model and actual implementation. Once the Jupyter notebook is created we must import the data. The data set is Inserted to code In pandas data frame.


Life Expectancy - IBM Watson St
eu-gb.dataplatform.cloud.ibm.com/analytics/notebooks/v2/30bd372f-fc1b-49cf-8b1a-7958d2e8895c?projectid=81c7a9ef-eed4-4999-84e8-896f27497250&context=wdp
Incognito
IBM Watson Studio
Upgrade
Aishwarya Bhattacharya's ...
AB
My projects / Life Expectancy / Life Expectancy
File Edit View Insert Cell Kernel Help
Not Trusted | Python 3.6
Run
Format Code
Type here to search
ENG IN
9:49 PM 6/2/2020

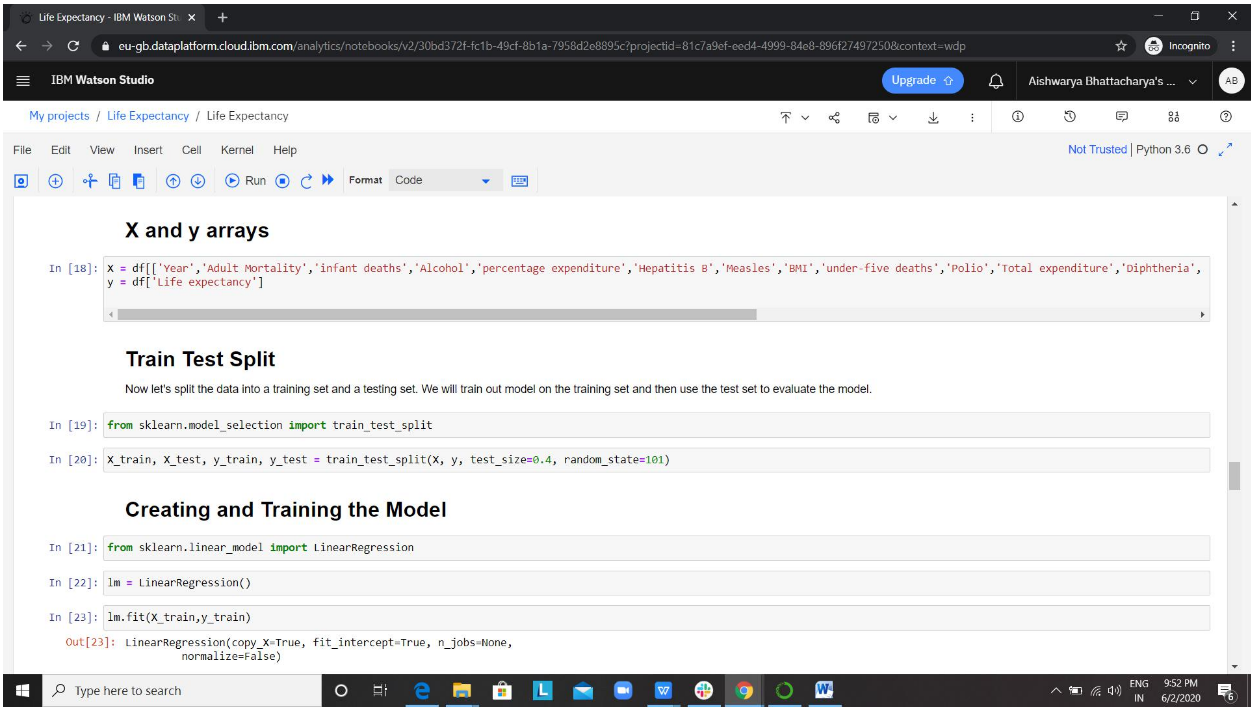
```
In [5]: import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        import seaborn as sns
        %matplotlib inline
```

Check out the Data

```
In [6]: import types
        import pandas as pd
        from botocore.client import Config
        import ibm_boto3

        def __iter__(self): return 0

        # @hidden_cell
        # The following code accesses a file in your IBM Cloud Object Storage. It includes your credentials.
        # You might want to remove those credentials before you share the notebook.
        client_6e0b8ff01b3a42f69854c7a40c62050c = ibm_boto3.client(service_name='s3',
            ibm_api_key_id='sFPQ8Quto-L49iZYaowtjFiRlJQbL7sGRlGHg5J-eOHm',
            ibm_auth_endpoint="https://iam.cloud.ibm.com/oidc/token",
            config=Config(signature_version='oauth'),
            endpoint_url='https://s3.eu-geo.objectstorage.service.networklayer.com')

        body = client_6e0b8ff01b3a42f69854c7a40c62050c.get_object(Bucket='usahousingdata-donotdelete-pr-gsqr71tete7bqj',Key='Life Expectancy Data.csv')['Body']
        # add missing __iter__ method, so pandas accepts body as file-like object
        if not hasattr(body, "__iter__"): body.__iter__ = types.MethodType( __iter__, body )

        # If you are reading an Excel file into a pandas DataFrame, replace `read_csv` by `read_excel` in the next statement.
        df= pd.read_csv(body)
        df.head()
```

b) Choose the appropriate model for prediction:

We can use any model for the the prediction person and with the help of it you can train and test the dataset.
For the project I have been choosing the **Random Forest Regression Model** for the development purpose.


Life Expectancy - IBM Watson Stu
eu-gb.dataplatform.cloud.ibm.com/analytics/notebooks/v2/30bd372f-fc1b-49cf-8b1a-7958d2e8895c?projectid=81c7a9ef-eed4-4999-84e8-896f27497250&context=wdp
Incognito
IBM Watson Studio
Upgrade
Aishwarya Bhattacharya's ...
My projects / Life Expectancy / Life Expectancy
File Edit View Insert Cell Kernel Help
Not Trusted | Python 3.6
Run
Format Code
Type here to search
ENG IN
9:52 PM 6/2/2020

X and y arrays

```
In [18]: X = df[['Year','Adult Mortality','infant deaths','Alcohol','percentage expenditure','Hepatitis B','Measles','BMI','under-five deaths','Polio','Total expenditure','Diphtheria',
         y = df['Life expectancy']
```

Train Test Split

Now let's split the data into a training set and a testing set. We will train out model on the training set and then use the test set to evaluate the model.

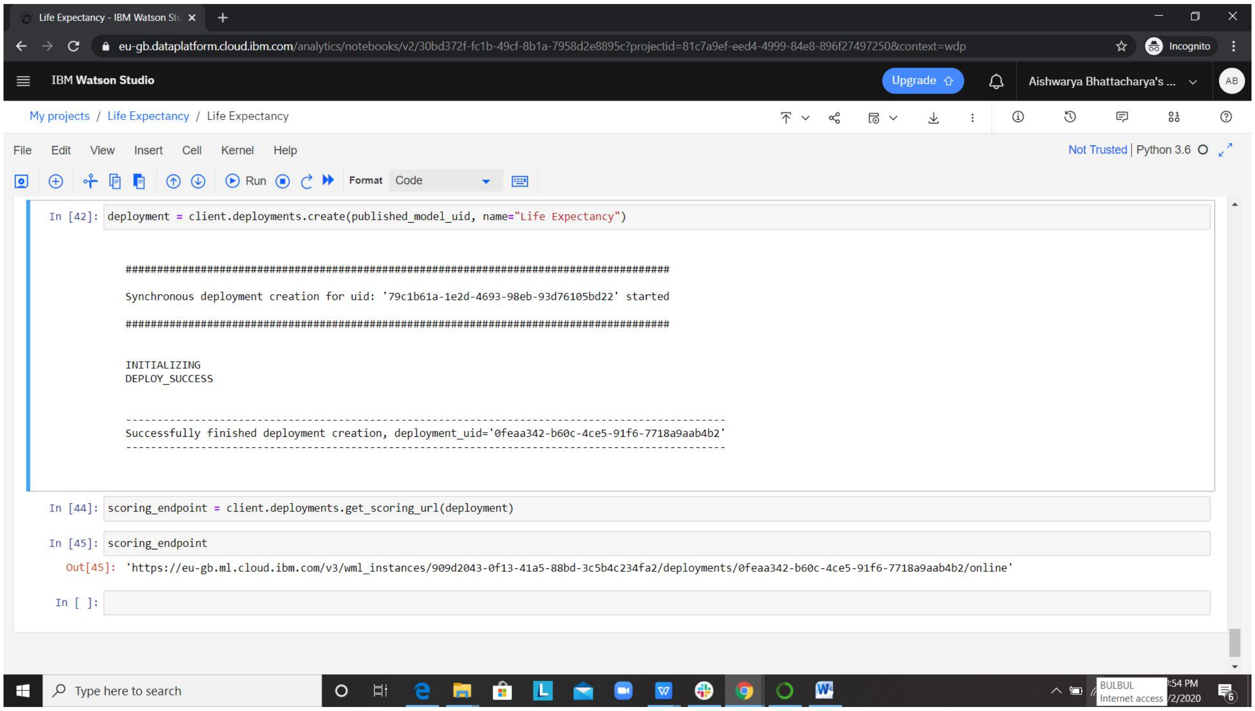
```
In [19]: from sklearn.model_selection import train_test_split

In [20]: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.4, random_state=101)
```

Creating and Training the Model

```
In [21]: from sklearn.linear_model import LinearRegression

In [22]: lm = LinearRegression()

In [23]: lm.fit(X_train,y_train)

Out[23]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None,
                  normalize=False)
```

c) Deployment of Model:

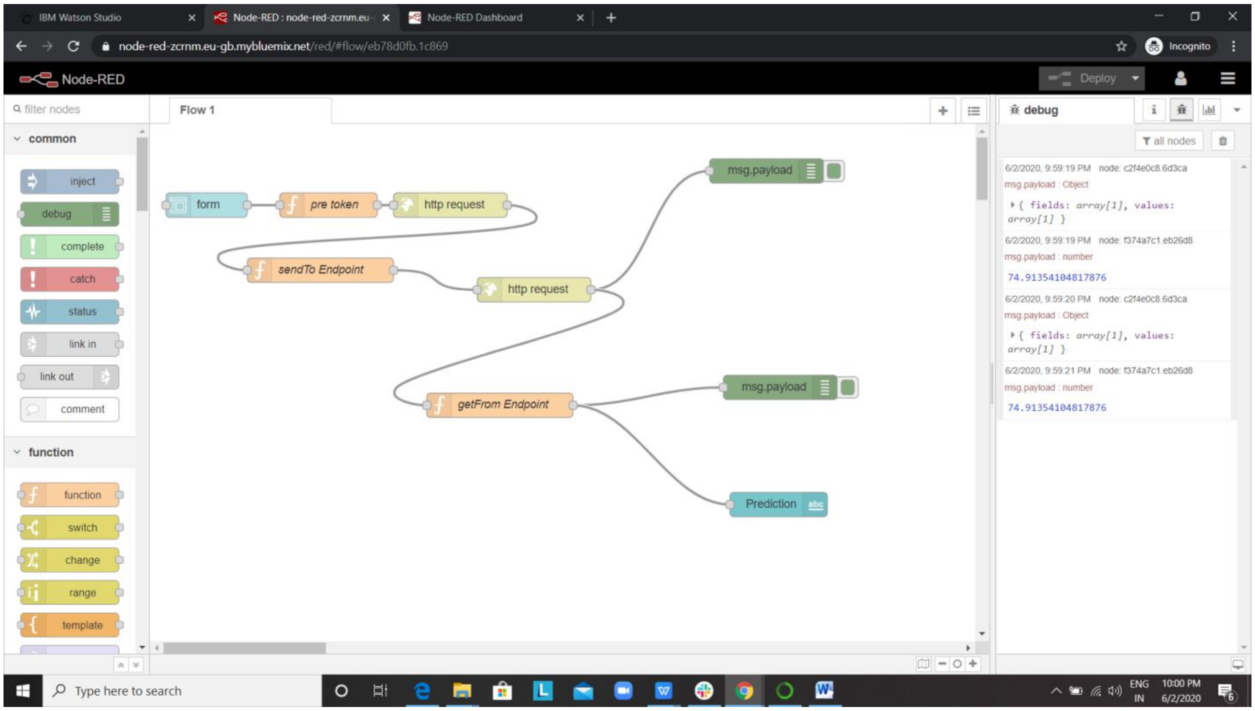
Once we're done with building the model ,we must deploy the model.
The deployed model will be stored in IBM Cloud Storage.


Life Expectancy - IBM Watson Stu
eu-gb.dataplatform.cloud.ibm.com/analytics/notebooks/v2/30bd372f-fc1b-49cf-8b1a-7958d2e8895c?projectid=81c7a9ef-eed4-4999-84e8-896f27497250&context=wdp
Incognito
IBM Watson Studio
Upgrade
Aishwarya Bhattacharya's ...
My projects / Life Expectancy / Life Expectancy
File Edit View Insert Cell Kernel Help
Not Trusted | Python 3.6
Run
Format Code
Type here to search
BULBUL Internet access
:54 PM /2/2020

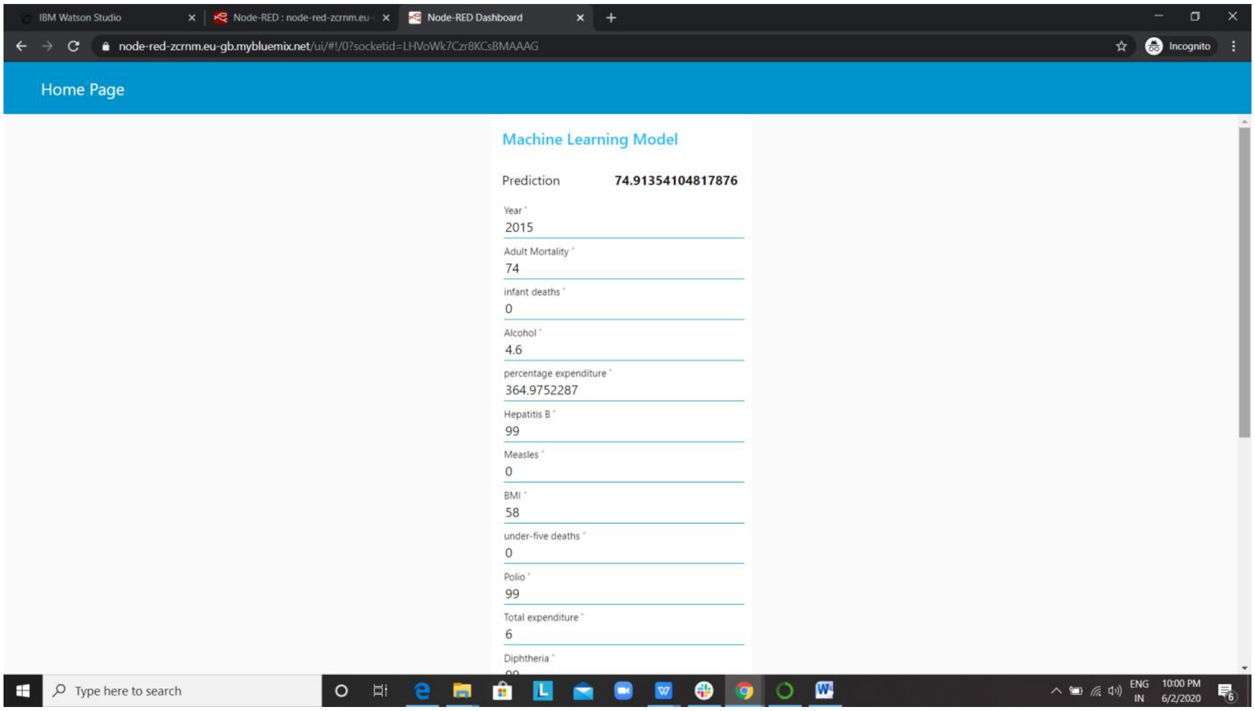
```
In [42]: deployment = client.deployments.create(published_model_uid, name="Life Expectancy")


           ################################################################################

           Synchronous deployment creation for uid: '79c1b61a-1e2d-4693-98eb-93d76105bd22' started

           ################################################################################


           INITIALIZING
           DEPLOY_SUCCESS


           ----------------------------------------------------------------------------------------
           Successfully finished deployment creation, deployment_uid='0feaa342-b60c-4ce5-91f6-7718a9aab4b2'
           ----------------------------------------------------------------------------------------

In [44]: scoring_endpoint = client.deployments.get_scoring_url(deployment)

In [45]: scoring_endpoint

Out[45]: 'https://eu-gb.ml.cloud.ibm.com/v3/wml_instances/909d2043-0f13-41a5-88bd-3c5b4c234fa2/deployments/0feaa342-b60c-4ce5-91f6-7718a9aab4b2/online'

In [ ]:
```

5. Create a Node-red Flow:

Once the model is deployed we can create the node red flow to create an API for the model . Thus API will act as an front end to the model. From here we get the Life Expectancy prediction data.



```
6/2/2020, 9:59:19 PM   node: c2f4e0c8.6d3ca
msg.payload : Object
▸{ fields: array[1], values:
array[1] }
6/2/2020, 9:59:19 PM   node: f374a7c1.eb26d8
msg.payload : number
74.91354104817876
6/2/2020, 9:59:20 PM   node: c2f4e0c8.6d3ca
msg.payload : Object
▸{ fields: array[1], values:
array[1] }
6/2/2020, 9:59:21 PM   node: f374a7c1.eb26d8
msg.payload : number
74.91354104817876
```


IBM Watson Studio
Node-RED : node-red-zcrnm.eu-
Node-RED Dashboard
node-red-zcrnm.eu-gb.mybluemix.net/ui/#!/0?socketid=LHVoWk7Czr8KCsBMAAAG
Incognito
Home Page
Type here to search
ENG
IN
10:00 PM
6/2/2020

Machine Learning Model

Prediction 74.91354104817876

Year *
2015

Adult Mortality *
74

infant deaths *
0

Alcohol *
4.6

percentage expenditure *
364.9752287

Hepatitis B *
99

Measles *
0

BMI *
58

under-five deaths *
0

Polio *
99

Total expenditure *
6

Diphtheria *

7. Conclusion:

From the project we conclude the Prediction of life expectancy
from the given dataset by working on Watson studio as the back end and Node Red Flow as the front end.