

# **Project Documentation**

Project Report on  
**Predicting Life Expectancy using Machine Learning**  
Under  
**Remote Summer Internship Program 2020 by**  
**SmartInternz**

Project by:  
**Vedanti Vyawahare**

**BTECH 3rd year (CSE)**  
**Vishwakarma Institute of Technology, Pune**

Email: [verdantvyawahare23@gmail.com](mailto:verdantvyawahare23@gmail.com)

Mobile: +918411841969

**Project Scope, Schedule, Team & Deliverables:**

**Project Scope :** Good prognostication helps to determine the course of treatment and helps to anticipate the procurement of health care services and facilities, or more broadly: facilitates Advance Care Planning. So this problem statement is aimed at predicting Life Expectancy rate of a country given various features. It predicts the average lifetime of a human being and predicts on the basis of various factors like Regional variations, Economic Circumstances, Sex Differences, Mental Illnesses, Physical Illnesses, Education, Year of their birth and other demographic factors. So the end product will predict the future life expectancy of the person with the help of prior given appropriate matrix of features by the user like current 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.

**Project Summary :**

This project is to build a model while considering historical data from a period of 2000 to 2015 for all the countries.

The model trained in this project will be able to predict the average lifetime of a human being given some input factors .

With the help of this project any country can able to predict the expected lifetime of their countrymen and then accordingly take preventive measures to improve on their healthcare measures.

This will also help countries in improving a particular a particular field such as GDP,alcohol intake,etc which have a high impact on country's life expectancy.

**Project Team : Individual****Dataset Reference :**

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

# **Project Title: Predicting Life Expectancy using Machine Learning.**

## **1. Introduction :**

**1.1 Overview :** Life expectancy is a statistical measure of the average time a human being is expected to live. This project's aim is to predict the average life expectancy of a human being of a country which depends on various factors like Regional variations, Economic Circumstances, Sex Differences, Mental Illnesses, Physical Illnesses, Education, Year of their birth and other demographic factors.

**1.2 Purpose and working :** The purpose of this project is to predict the life expectancy of a person based on the inputs given by the user. In this project an UI is provided that is built using Node-Red application of IBM cloud in which the user will provide some values like year, adult mortality, alcohol, total expenditure, BMI, info regarding some diseases, thinness, schooling, etc. and the user will get the predicted life expectancy as output. The project uses a **Random Forest Regressor Machine Learning** model to predict the life expectancy which is deployed on IBM cloud using services like IBM Watson Studio and IBM Watson Machine Learning Service. The dataset used for the training of the model was downloaded from kaggle.com and Python is used to write the code for machine learning model.

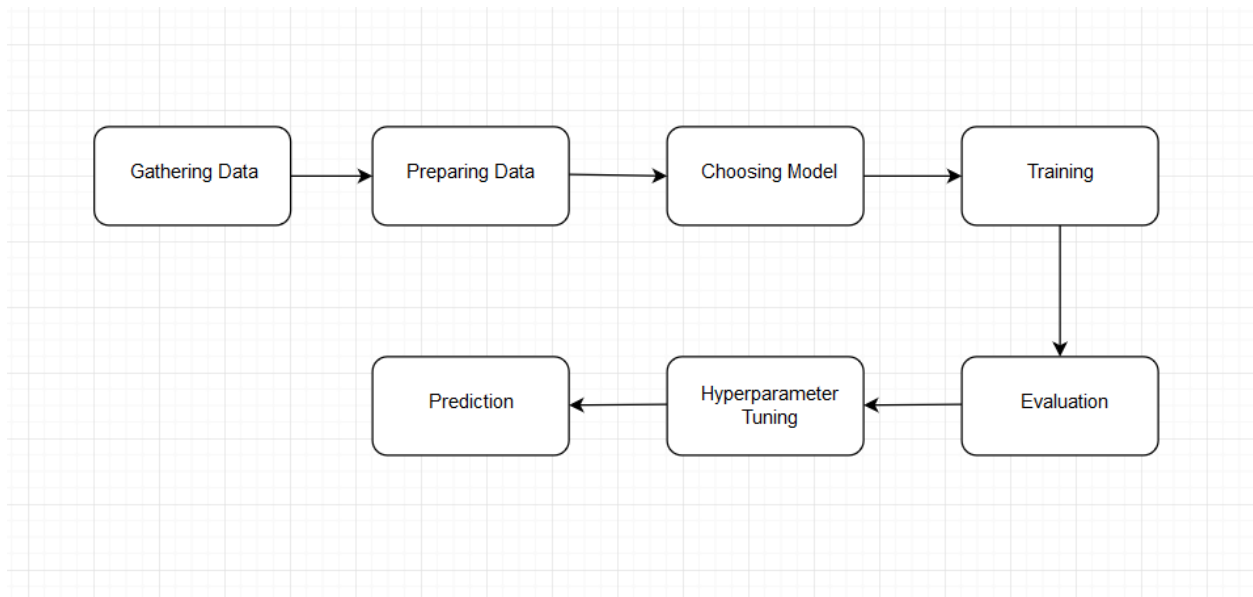
## **2. Literature Survey :**

**2.1 Existing problem :** Currently the life expectancy of a person is calculated on the basis of the already collected data and the life expectancy of a person in future coming years cannot be known because of lack of technology. Due to which many problems may arise because the people are not prepared to tackle the issues that may affect the life expectancy as they don't know whether a factor may increase or decrease the life expectancy.

**2.2 Proposed Solution :** If the life expectancy of a person can be predicted for the coming years then we can be aware of the factors that may affect the life expectancy either in a positive way or in a negative way and accordingly take necessary actions or precautions.

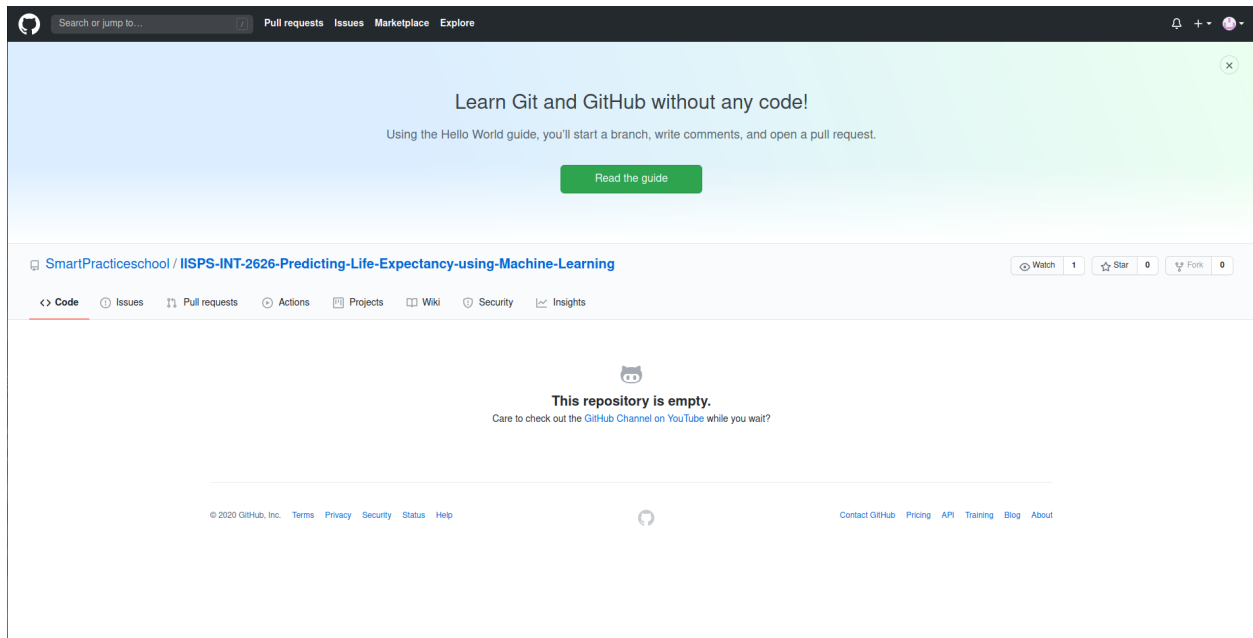
### 3. Theriotical Analysis :

#### 3.1 Block Diagram :

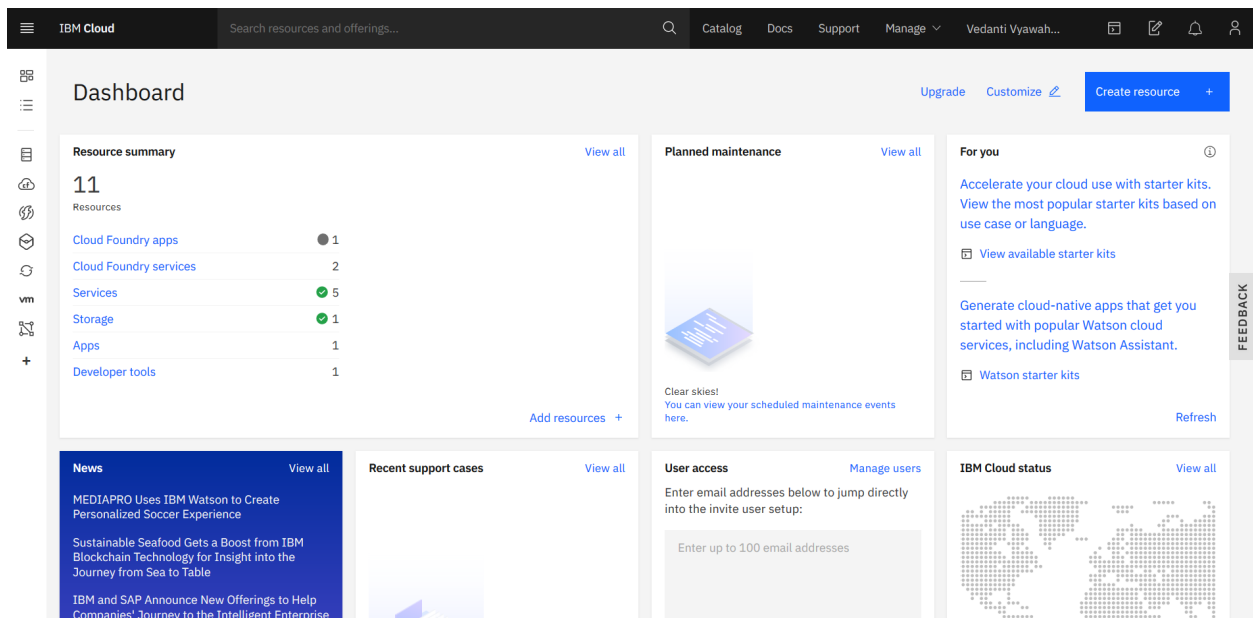


## 3.2 Software Designing : The software requirements of this project are:

Git Hub account -



IBM account -



# Watson Machine Learning-

IBM Cloud

Search resources and offerings...

Q

Catalog

Docs

Support

Manage

Vedanti Vyawahare...

Resource list /

WatsonMachineLearning Active Add tags

Details


Actions...

Manage

Service credentials

Plan

Connections



Watson Machine Learning

Welcome! Get Started with Watson Machine Learning in Watson Studio.

Access in Watson Studio

Documentation

Community

FEEDBACK

# Jupyter Notebook-

IBM Watson Studio

Upgrade

Vedanti Vyawahare's Acco...

VY

My projects / ML1 / prediction

File Edit View Insert Cell Kernel Help

Not Trusted | Python 3.6 with Spark

In [30]:

```
import types
import pandas as pd
from botocore.client import Config
import 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 = boto3.client(service_name='s3',
                      aws_access_key_id='71a503vfoN1VU8AqMF-ad3qNZVjwvTlpMyntX931k',
                      aws_secret_access_key='...',
                      config=Config(signature_version='oauth'),
                      endpoint_url='https://s3-api.us-geo.objectstorage.service.networklayer.com')

body = client.get_object(Bucket='ml1-donotdelete-pr-wmk3amdc18llcz', Key='life_prediction.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)

df_data_1 = pd.read_csv(body)
df_data_1.head()
```

Out[30]:

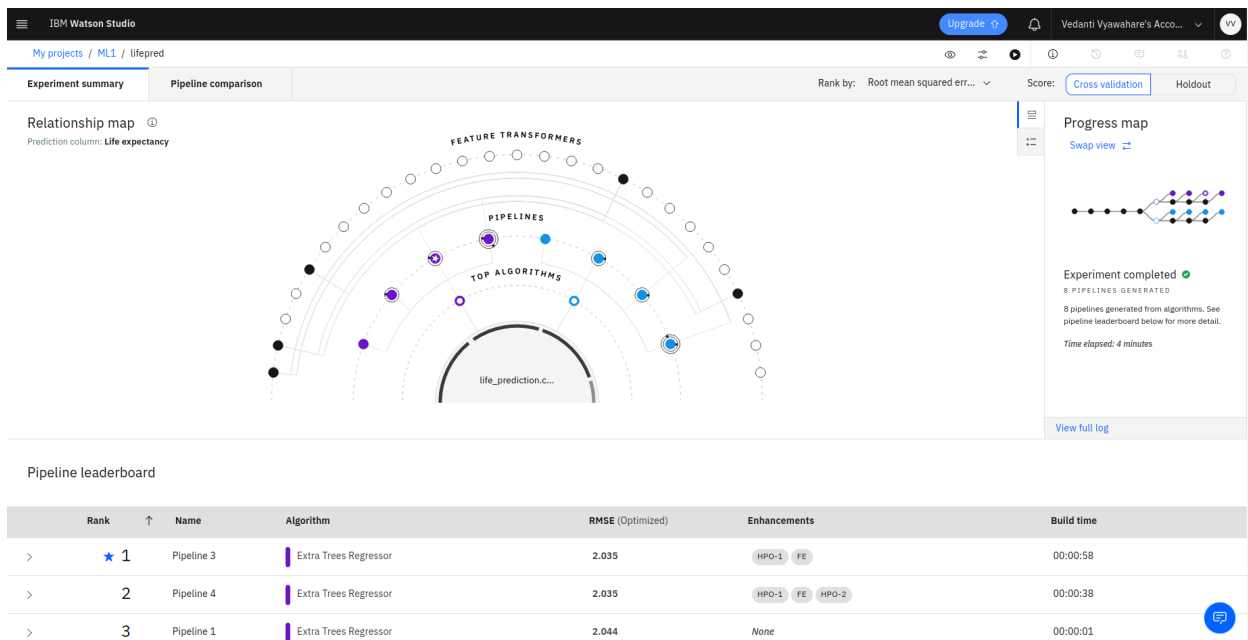
	Country	Status	Year	Schooling	Adult Mortality	Infant deaths	Alcohol	percentage expenditure	Hepatitis B	Measles ...	Polio	Total expenditure	Diphtheria	HIV/AIDS	GDP	Population	thinness 1-19 years	thinness 5-9 years	Income composition of resources	Life expectancy	
0	Alghanistan	Developing	2015	10.1	263.0	62	0.01	71.279624	65.0	1154	...	6.0	8.16	65.0	0.1	584.259210	33736494.0	17.2	17.3	0.479	65.0
1	Alghanistan	Developing	2014	10.0	271.0	64	0.01	73.523582	62.0	492	...	58.0	8.18	62.0	0.1	612.696514	327562.0	17.5	17.5	0.476	59.9
2	Alghanistan	Developing	2013	9.9	268.0	66	0.01	73.219243	64.0	430	...	62.0	8.13	64.0	0.1	631.744976	31731688.0	17.7	17.7	0.470	59.9
3	Alghanistan	Developing	2012	9.8	272.0	69	0.01	78.184215	67.0	2787	...	67.0	8.52	67.0	0.1	669.959000	3696958.0	17.9	18.0	0.463	59.5
4	Alghanistan	Developing	2011	9.5	275.0	71	0.01	7.097109	68.0	3013	...	68.0	7.87	68.0	0.1	63.537231	2978599.0	18.2	18.2	0.454	59.2

5 rows x 22 columns

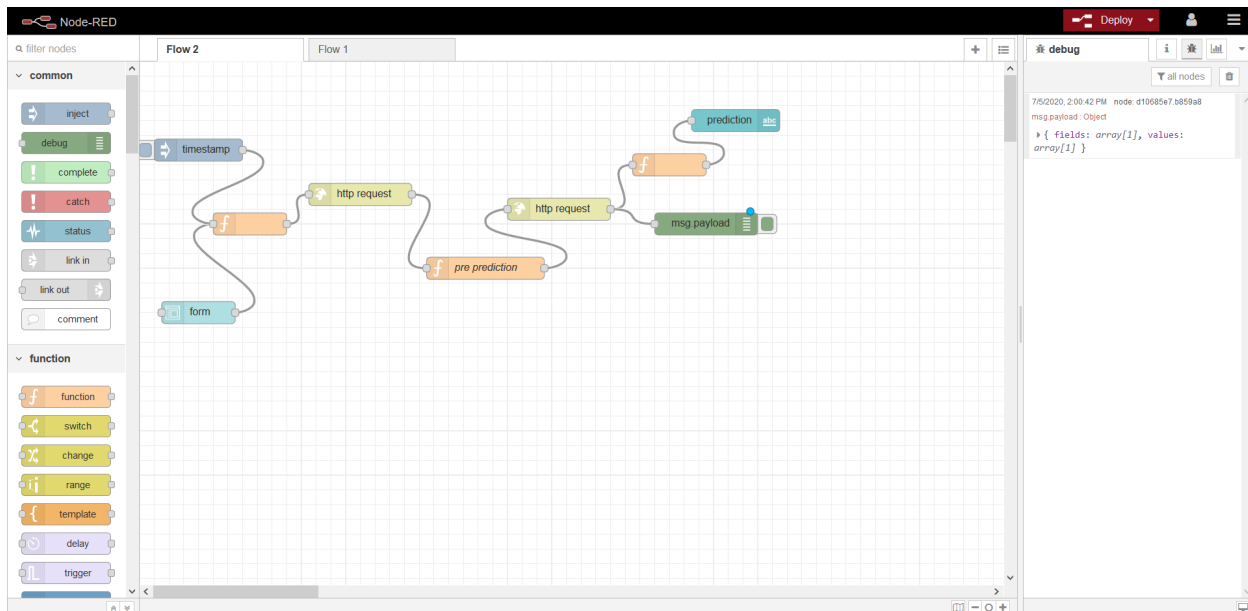
In [31]:

```
#df_data_1 = df_data_1.replace(0, null)
# fill missing values with mean column values
df_data_1.fillna(df_data_1.mean(), inplace=True)
# count the number of NaN values in each column
print(df_data_1.isnull().sum())
```

## Auto AI -



## Node-Red -



**4.Experimental Investigation :** This prediction is for continuous value. So algorithm can be used for it are regression algorithms such as linear regression, random forest regression. By predicting the values by linear regression, accuracy got nearly 92% and by random forest accuracy got nearly 99%. So random forest is better approach than linear regression.

Features used of predictions are as follows :

Country

Status

Year

Schooling

Adult Mortality

infant deaths

Alcohol

percentage expenditure

Hepatitis B

Measles

BMI

under-five deaths

Polio

Total expenditure

Diphtheria

HIV/AIDS

GDP

Population

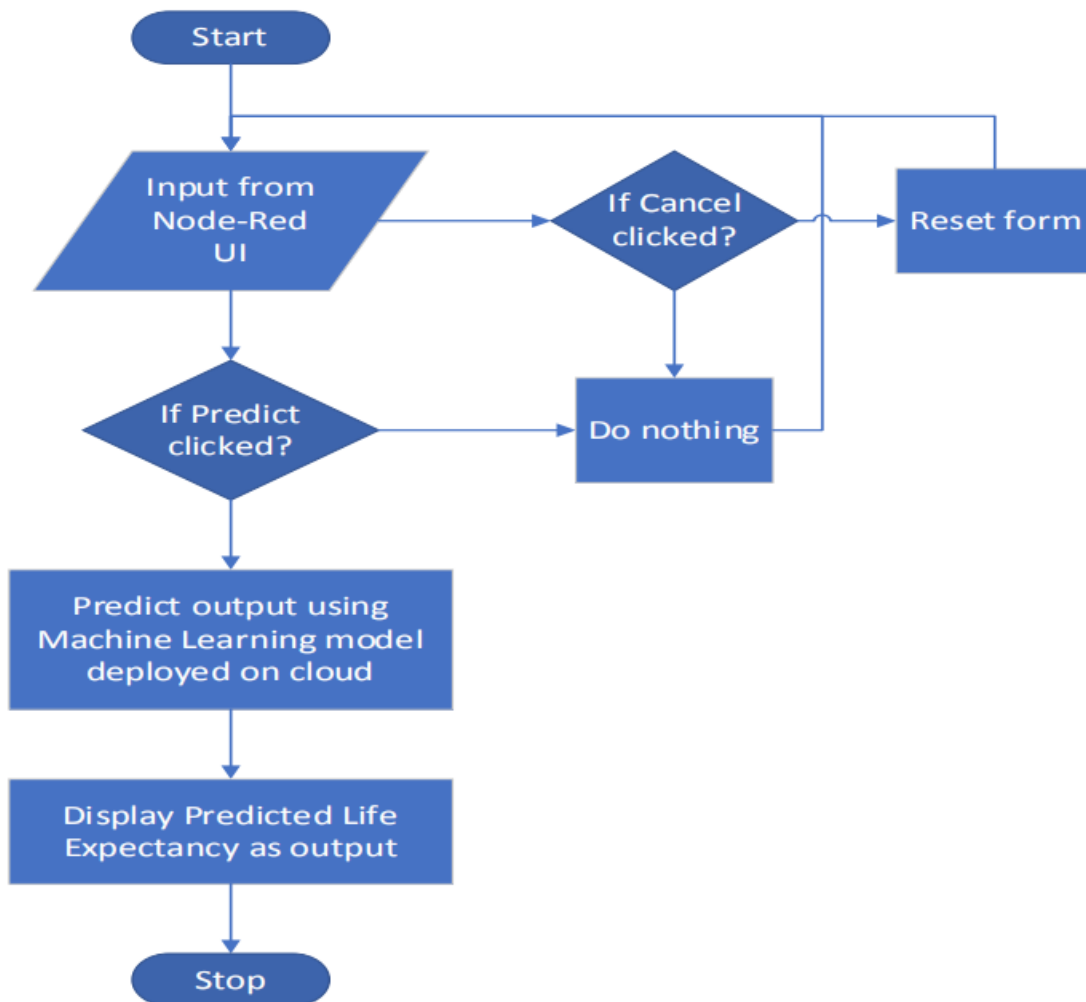
thinness 1-19 years

thinness 5-9 years

Income composition of resources



## 5.Flowchart :



**6. Result :** The result of the project is the predicted life expectancy of a person based on the provided inputs.

## **7. Advantages and Disadvantages :**

### **Advantages :**

1. Life expectancy can be estimated at any age, e.g. life expectancy at 65 years. Gives more weight to deaths at younger ages. Life expectancy has been used nationally to monitor health inequalities.
2. It is adjusted for age profile of a population. Can be monitored annually. Has been used nationally as a proxy indicator to monitor progress towards national health inequalities target.
3. One can look at specific causes of death, e.g. cancer, stroke.

### **Disadvantages :**

1. At smaller geographies, due to smaller numbers more subject to random variation year on year. To mitigate this, pooled years can be used but may limit trend analysis.
2. At smaller geographies may be influenced by nursing homes in the area.
3. Accuracy is not 100% percent.

## **8. Applications :**

- i) Insurance
- ii) Tax
- iii) Pharmaceutical Industry

**9. Conclusion :** Prognostication of life expectancy is difficult for humans. Research shows that machine learning technique offers a feasible and promising approach to predicting life expectancy. The research has potential for real-life applications, such as supporting timely recognition of the right moment to start Advance Care Planning.

**10. Future Scope :** The machine learning model used for the project can be trained well using more and accurate data and more appropriate and trained model can be used to increase the accuracy and get more accurate result.

## **11. Bibilography :**

i) <https://en.wikipedia.org/wiki/Prediction>

ii) <https://towardsdatascience.com/workflow-of-a-machine-learning-project-ec1dba419b94>

iii) <https://www.kaggle.com/>

iv) <https://www.geeksforgeeks.org/random-forest-regression-in-python/>

## **APPENDIX**

### **A. Source Code :**

[https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/9b8a1d9f-6f7f-41ed-9e9f-61d7dea91a2a/view?access\\_token=7eb91dcea95540ba746adc8fcf19c542accbb1eaaef3380404125b2b87d4ea7a](https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/9b8a1d9f-6f7f-41ed-9e9f-61d7dea91a2a/view?access_token=7eb91dcea95540ba746adc8fcf19c542accbb1eaaef3380404125b2b87d4ea7a)

### **B. Node-red App :**

<https://node-red-xslem.mybluemix.net/ui/>

**THANK YOU !!**