

Laboratorio 4

Despliegue de modelos de ML mediante uso de API's

Integrantes:

- Daniel Alejandro Ángel – 201911345
- Jairo Adolfo Céspedes Plata - 201912008

Link del repositorio de GitHub:

<https://github.com/danr844/Laboratorio4-BI-Heroku.git>

Escenarios de prueba:

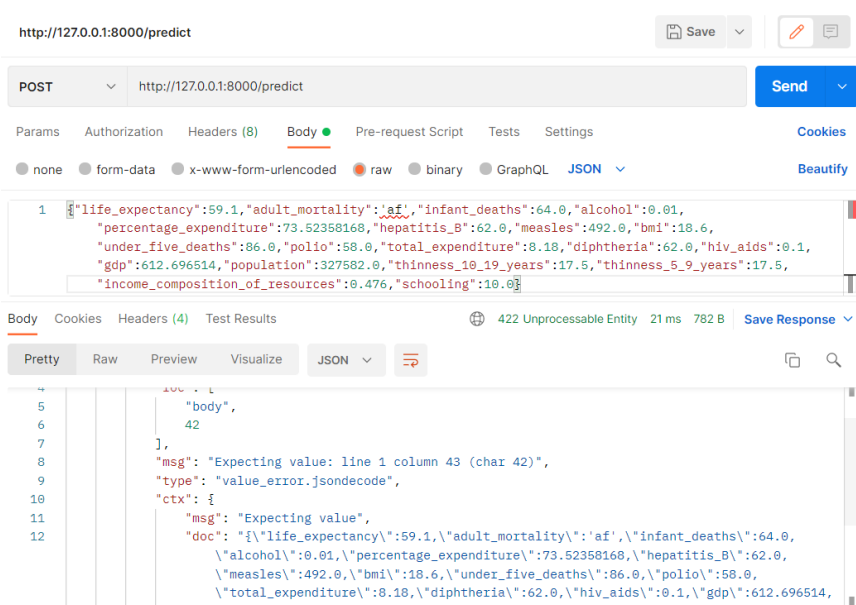
- Escenario 1 (Error)
- Escenario 2 (Error)
- Escenario 3 (Error)
- Escenario 4 (Ejecución correcta)
- Escenario 5 (Ejecución correcta)
- Escenario 6 (Ejecución correcta)

Escenario 1:

El primero de los escenarios que se quiere mostrar se encuentra con la especificidad que deben tener las entradas de los datos que ingresamos al pipeline. Al diseñar el pipeline se especificó que nuestras variables serían de tipo float. Sin embargo, dado que la API diseñada nos permitió facilitar el acceso a las predicciones a los usuarios, se tiene en cuenta el escenario en el que un usuario puede no saber el formato en el que debe ingresar los datos. En este caso, el pipeline va a retornar un error de tipo Error: Unprocessable Entity. Se ejecutaron escenarios con los siguientes valores:

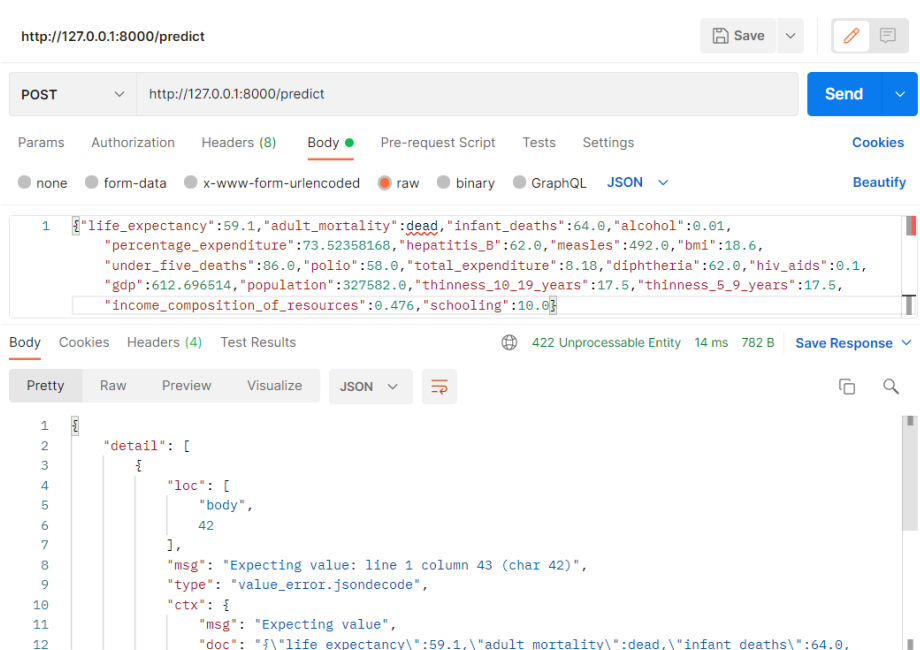
- Caso1:

```
{"life_expectancy":59.1,"adult_mortality":'af',"infant_deaths":64.0,"alcohol":0.01,"percentage_expenditure":73.52358168,"hepatitis_B":62.0,"measles":492.0,"bmi":18.6,"under_five_deaths":86.0,"polio":58.0,"total_expenditure":8.18,"diphtheria":62.0,"hiv_aids":0.1,"gdp":612.696514,"population":327582.0,"thinness_10_19_years":17.5,"thinness_5_9_years":17.5,"income_composition_of_resources":0.476,"schooling":10.0}
```



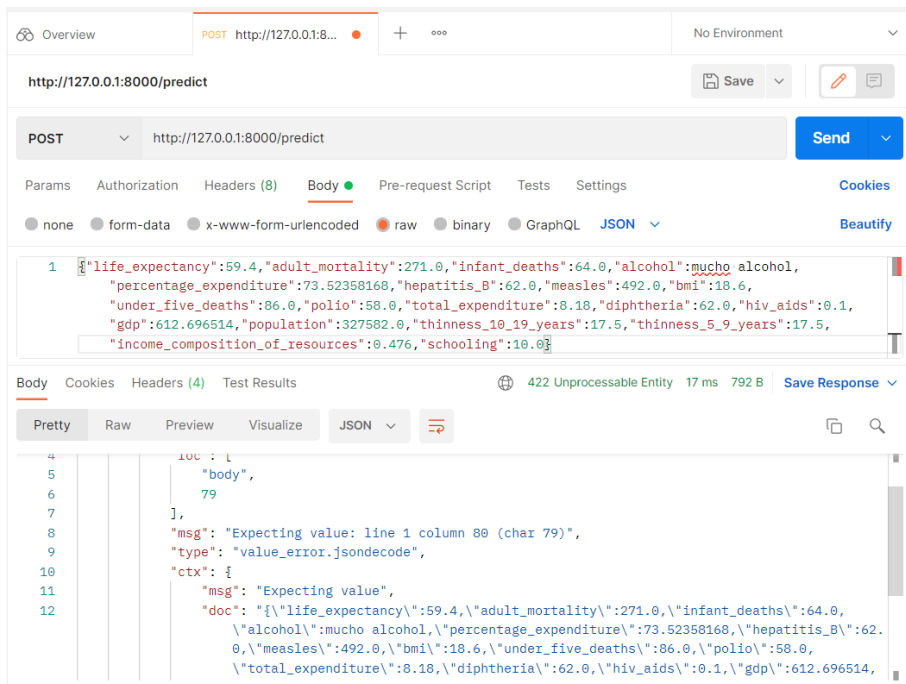
- Caso2:

```
{ "life_expectancy":59.1,"adult_mortality":dead,"infant_deaths":64.0,"alcohol":0.01,"percentage_expenditure":73.52358168,"hepatitis_B":62.0,"measles":492.0,"bmi":18.6,"under_five_deaths":86.0,"polio":58.0,"total_expenditure":8.18,"diphtheria":62.0,"hiv_aids":0.1,"gdp":612.696514,"population":327582.0,"thinness_10_19_years":17.5,"thinness_5_9_years":17.5,"income_composition_of_resources":0.476,"schooling":10.0}
```



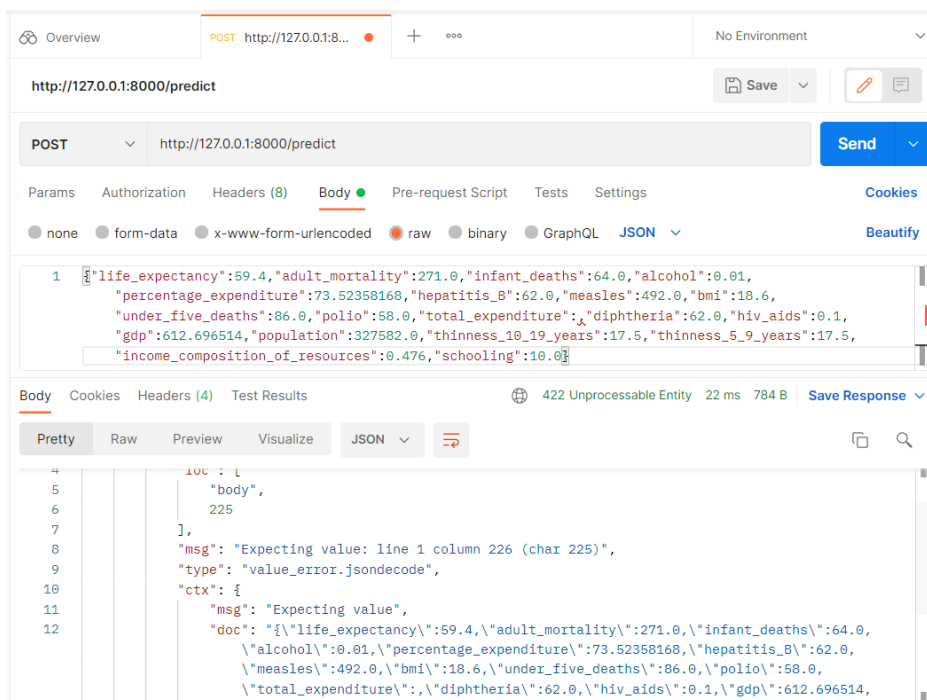
- Caso3:

```
{ "life_expectancy":59.4,"adult_mortality":271.0,"infant_deaths":64.0,"alcohol":mucho alcohol,"percentage_expenditure":73.52358168,"hepatitis_B":62.0,"measles":492.0,"bmi":18.6,"under_five_deaths":86.0,"polio":58.0,"total_expenditure":8.18,"diphtheria":62.0,"hiv_aids":0.1,"gdp":612.696514,"population":327582.0,"thinness_10_19_years":17.5,"thinness_5_9_years":17.5,"income_composition_of_resources":0.476,"schooling":10.0}
```



Caso4:

`{"life_expectancy":59.4,"adult_mortality":271.0,"infant_deaths":64.0,"alcohol":0.01,"percentage_expenditure":73.52358168,"hepatitis_B":62.0,"measles":492.0,"bmi":18.6,"under_five_deaths":86.0,"polio":58.0,"total_expenditure":,"diphtheria":62.0,"hiv_aids":0.1,"gdp":612.696514,"population":327582.0,"thinness_10_19_years":17.5,"thinness_5_9_years":17.5,"income_composition_of_resources":0.476,"schooling":10.0}`



Casos para los cuales, se obtuvo el error mencionado anteriormente.

Escenario 2:

Para el caso de este escenario se quisieron probar los rangos de los datos permitidos, dado que un usuario tiene total libertad de ingresar los datos que desee. Se considera que un usuario puede tener en cuenta el tipo de dato que debe ingresar (numérico), pero muchas veces los usuarios no son conscientes de las restricciones de cada uno de los tipos de datos, esto ocurre especialmente porque no saben con qué tipos de datos estamos trabajando por debajo. Se ejecutaron dos casos de prueba en los que se analizan valores extremadamente

[illegible]

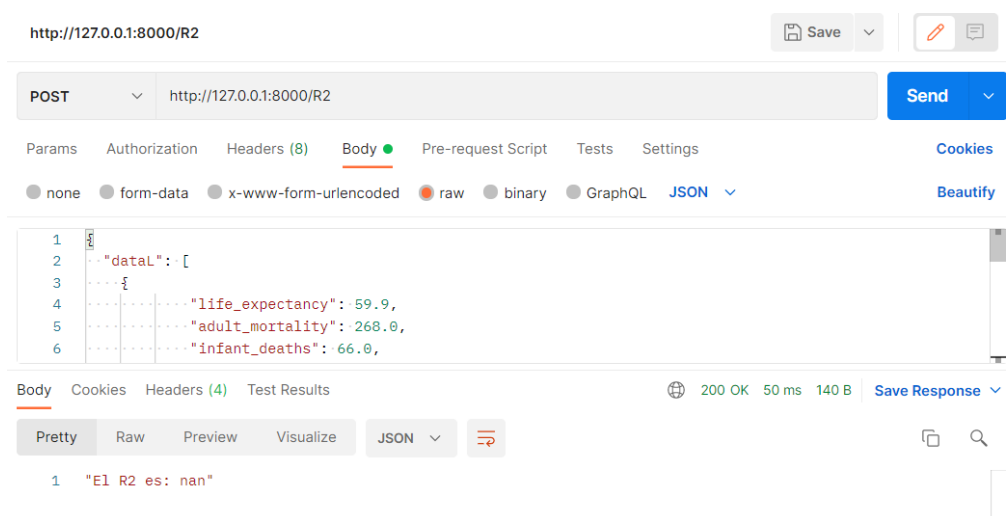
Escenario 3:

En este escenario se quiere analizar un caso interesante en el que el resultado dado es incoherente o no aporta mucho valor a cualquier persona que ejecute esta opción. En este caso se quiso ejecutar la segunda opción de nuestro modelo en la que se quiere saber el R^2 dados unos registros en formato JSON.

Encontramos que al presentar el caso mostrado en la parte posterior en el que solo se incluye 1 registro en la información dada al modelo, se presenta que la respuesta dada es un $R^2 = \text{NaN}$.

Caso1:

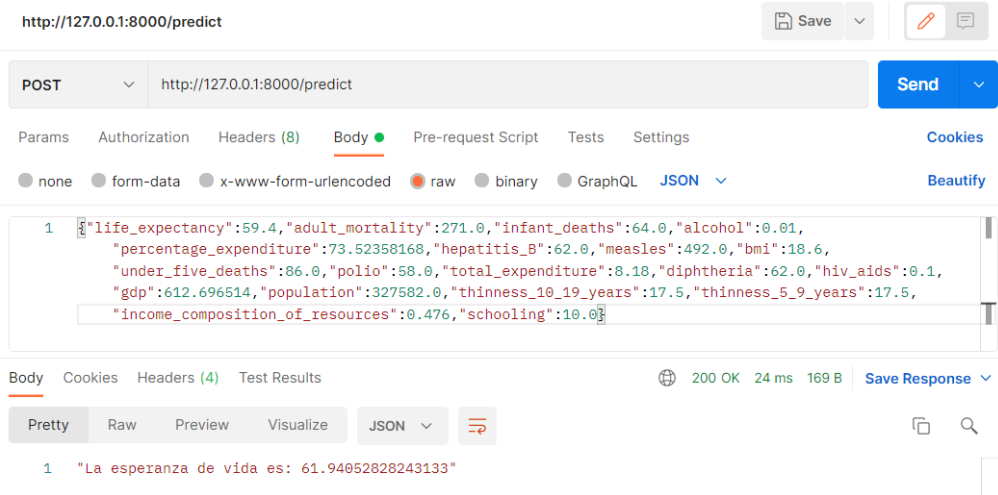
```
{"dataL":[{"life_expectancy":59.9,"adult_mortality":268.0,"infant_deaths":66.0,"alcohol":0.01,"percent age_expenditure":73.21924272,"hepatitis_B":64.0,"measles":430.0,"bmi":18.1,"under_five_deaths":89.0,"polio":62.0,"total_expenditure":8.13,"diphtheria":64.0,"hiv_aids":0.1,"gdp":631.744976,"population":31731688.0,"thinness_10_19_years":17.7,"thinness_5_9_years":17.7,"income_composition_of_resources":0.47,"schooling":9.9}]}
```



Escenario 4:

Este es uno de los escenarios de prueba que se ejecutaron satisfactoriamente. En este caso se ejecuta el pipeline con un solo registro. La URL especificada para esta tarea es /predict. Con la cual, al ingresar el JSON mostrado en la parte de abajo, se obtuvo que esa persona tiene una expectativa de vida de 61 años. Podemos notar que el modelo está arrojando muy buenos resultados dado que al revisar el “life_expectancy” real, notamos que según los datos, se tiene que sería de 59.4, lo cual es muy cercano al valor arrojado por nuestro modelo.

```
{"life_expectancy":59.4,"adult_mortality":271.0,"infant_deaths":64.0,"alcohol":0.01,"percentage_expenditure":73.52358168,"hepatitis_B":62.0,"measles":492.0,"bmi":18.6,"under_five_deaths":86.0,"polio":58.0,"total_expenditure":8.18,"diphtheria":62.0,"hiv_aids":0.1,"gdp":612.696514,"population":327582.0,"thinness_10_19_years":17.5,"thinness_5_9_years":17.5,"income_composition_of_resources":0.476,"schooling":10.0}
```

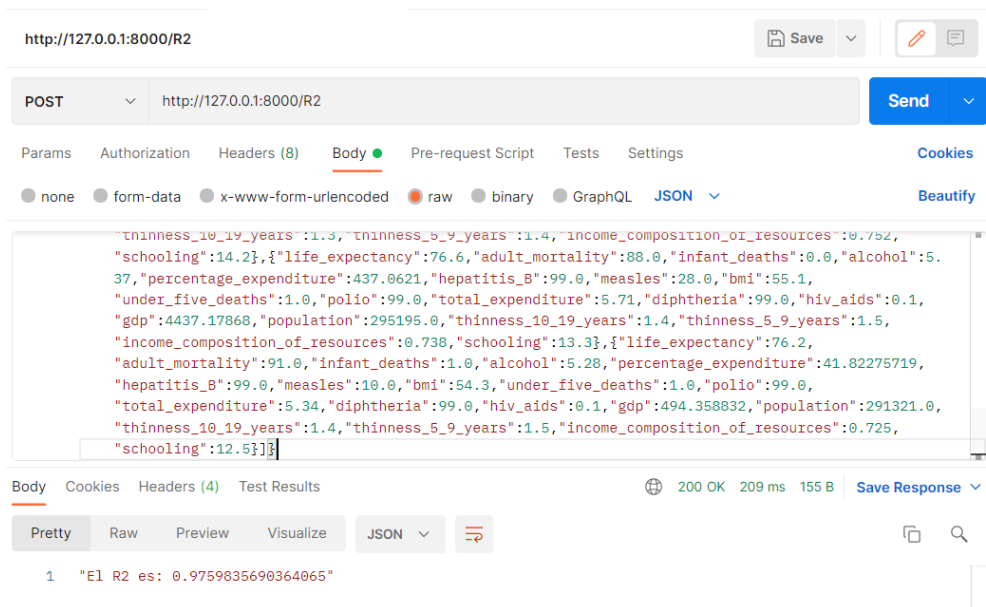


Escenario 5:

Consecuentemente, se analiza el escenario en el que se prueba la segunda opción del pipeline desarrollado. En este caso, se desea analizar el R^2 de la entrada mostrada en la parte posterior. Para este caso, la URL definida es /R2. Al finalizar la ejecución de la petición, Podemos notar que el resultado fue bastante bueno, ya que se obtuvo un R^2 de 0,97 y, ya que nosotros sabemos que buscamos R^2 cercanos a 1 que puedan describir una alta correlación de los datos, podemos concluir que el modelo está arrojando resultados con métricas bastante buenas.

```
{
  "dataL": [
    {
      "life_expectancy": 59.9,
      "adult_mortality": 271.0,
      "infant_deaths": 64.0,
      "alcohol": 0.01,
      "percent age_expenditure": 73.52358168,
      "hepatitis_B": 62.0,
      "measles": 492.0,
      "bmi": 18.6,
      "under_five_deaths": 86.0,
      "polio": 58.0,
      "total_expenditure": 8.18,
      "diphtheria": 62.0,
      "hiv_aids": 0.1,
      "gdp": 612.696514,
      "population": 327582.0,
      "thinness_10_19_years": 17.5,
      "thinness_5_9_years": 17.5,
      "income_composition_of_resources": 0.476,
      "schooling": 10.0
    },
    {
      "life_expectancy": 59.9,
      "adult_mortality": 268.0,
      "infant_deaths": 66.0,
      "alcohol": 0.01,
      "percent age_expenditure": 73.21924272,
      "hepatitis_B": 64.0,
      "measles": 430.0,
      "bmi": 18.1,
      "under_five_deaths": 89.0,
      "polio": 62.0,
      "total_expenditure": 8.13,
      "diphtheria": 64.0,
      "hiv_aids": 0.1,
      "gdp": 631.744976,
      "population": 317316.88,
      "thinness_10_19_years": 17.7,
      "thinness_5_9_years": 17.7,
      "income_composition_of_resources": 0.47,
      "schooling": 9.9
    },
    {
      "life_expectancy": 57.0,
      "adult_mortality": 293.0,
      "infant_deaths": 87.0,
      "alcohol": 0.02,
      "percent age_expenditure": 15.29606643,
      "hepatitis_B": 67.0,
      "measles": 466.0,
      "bmi": 13.8,
      "under_five_deaths": 120.0,
      "polio": 5.0,
      "total_expenditure": 8.79,
      "diphtheria": 5.0,
      "hiv_aids": 0.1,
      "gdp": 219.1413528,
      "population": 241189.79,
      "thinness_10_19_years": 19.5,
      "thinness_5_9_years": 19.7,
      "income_composition_of_resources": 0.381,
      "schooling": 6.8
    },
    {
      "life_expectancy": 56.7,
      "adult_mortality": 295.0,
      "infant_deaths": 87.0,
      "alcohol": 0.01,
      "percent age_expenditure": 11.08905273,
      "hepatitis_B": 65.0,
      "measles": 798.0,
      "bmi": 13.4,
      "under_five_deaths": 122.0,
      "polio": 41.0,
      "total_expenditure": 8.82,
      "diphtheria": 41.0,
      "hiv_aids": 0.1,
      "gdp": 198.7285436,
      "population": 2364.851,
      "thinness_10_19_years": 19.7,
      "thinness_5_9_years": 19.9,
      "income_composition_of_resources": 0.373,
      "schooling": 6.5
    },
    {
      "life_expectancy": 77.8,
      "adult_mortality": 74.0,
      "infant_deaths": 0.0,
      "alcohol": 4.6,
      "percent age_expenditure": 364.9752287,
      "hepatitis_B": 99.0,
      "measles": 0.0,
      "bmi": 58.0,
      "under_five_deaths": 0.0,
      "polio": 9.0,
      "total_expenditure": 6.0,
      "diphtheria": 99.0,
      "hiv_aids": 0.1,
      "gdp": 3954.22783,
      "population": 28873.0,
      "thinness_10_19_years": 1.2,
      "thinness_5_9_years": 1.3,
      "income_composition_of_resources": 0.762,
      "schooling": 14.2
    },
    {
      "life_expectancy": 77.5,
      "adult_mortality": 8.0,
      "infant_deaths": 0.0,
      "alcohol": 4.51,
      "percentage_expenditure": 428.7490668,
      "hepatitis_B": 98.0,
      "measles": 0.0,
      "bmi": 57.2,
      "under_five_deaths": 1.0,
      "polio": 98.0,
      "total_expenditure": 5.88,
      "diphtheria": 98.0,
      "hiv_aids": 0.1,
      "gdp": 4575.763787,
      "population": 288914.0,
      "thinness_10_19_years": 1.2,
      "thinness_5_9_years": 1.3,
      "income_composition_of_resources": 0.761,
      "schooling": 14.2
    },
    {
      "life_expectancy": 77.2,
      "adult_mortality": 84.0,
      "infant_deaths": 0.0,
      "alcohol": 4.76,
      "percentage_expenditure": 430.8769785,
      "hepatitis_B": 99.0,
      "measles": 0.0,
      "bmi": 56.5,
      "under_five_deaths": 1.0,
      "polio": 99.0,
      "total_expenditure": 5.66,
      "diphtheria": 99.0,
      "hiv_aids": 0.1,
      "gdp": 4414.72314,
      "population": 289592.0,
      "thinness_10_19_years": 1.3,
      "thinness_5_9_years": 1.4,
      "income_composition_of_resources": 0.759,
      "schooling": 14.2
    },
    {
      "life_expectancy": 76.9,
      "adult_mortality": 86.0,
      "infant_deaths": 0.0,
      "alcohol": 5.14,
      "percentage_expenditure": 412.4433563,
      "hepatitis_B": 99.0,
      "measles": 9.0,
      "bmi": 55.8,
      "under_five_deaths": 1.0,
      "polio": 99.0,
      "total_expenditure": 5.59,
      "diphtheria": 99.0,
      "hiv_aids": 0.1,
      "gdp": 4247.61438,
      "population": 2941.0,
      "thinness_10_19_years": 1.3,
      "thinness_5_9_years": 1.4,
      "income_composition_of_resources": 0.759,
      "schooling": 14.2
    }
  ]
}
```

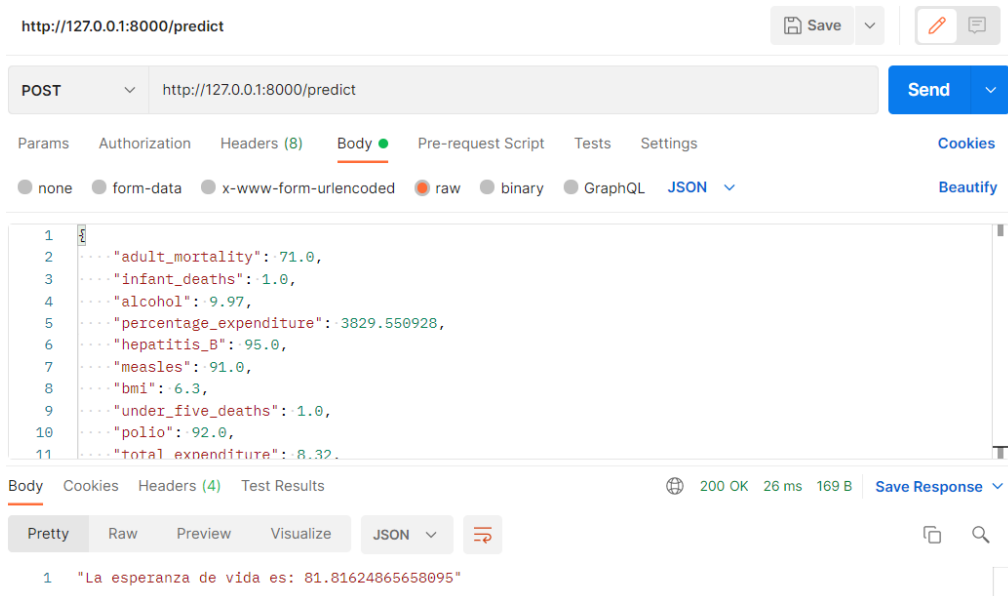
```
5_9_years":1.4,"income_composition_of_resources":0.752,"schooling":14.2},{
"life_expectancy":76.6,"adult_mortality":88.0,"infant_deaths":0.0,"alcohol":5.37,"percentage_expenditure":437.0621,"hepatitis_B":99.0,"measles":28.0,"bmi":55.1,"under_five_deaths":1.0,"polio":99.0,"total_expenditure":5.71,"diphtheria":99.0,"hiv_aids":0.1,"gdp":4437.17868,"population":295195.0,"thinness_10_19_years":1.4,"thinness_5_9_years":1.5,"income_composition_of_resources":0.738,"schooling":13.3},{
"life_expectancy":76.2,"adult_mortality":91.0,"infant_deaths":1.0,"alcohol":5.28,"percentage_expenditure":41.82275719,"hepatitis_B":99.0,"measles":10.0,"bmi":54.3,"under_five_deaths":1.0,"polio":99.0,"total_expenditure":5.34,"diphtheria":99.0,"hiv_aids":0.1,"gdp":494.358832,"population":291321.0,"thinness_10_19_years":1.4,"thinness_5_9_years":1.5,"income_composition_of_resources":0.725,"schooling":12.5}}}
```



Escenario 6:

Finalmente, se ejecutó otro escenario de prueba en el que se obtuvieron resultados coherentes. Para este caso se vuelve a ejecutar la predicción de la expectativa de vida para una persona dada la información dada en formato JSON. En este caso los resultados arrojados por el modelo muestran que la expectativa de vida predicha es de 81.6 años, para lo cual al compararla con la expectativa de vida real dada en los datos que es de 83 años, podemos notar que se pudo obtener un resultado coherente.

```
{"adult_mortality":71.0,"infant_deaths":1.0,"alcohol":9.97,"percentage_expenditure":3829.550928,"hepatitis_B":95.0,"measles":91.0,"bmi":6.3,"under_five_deaths":1.0,"polio":92.0,"total_expenditure":8.32,"diphtheria":93.0,"hiv_aids":0.1,"gdp":23465.38559,"population":198954.0,"thinness_10_19_years":0.7,"thinness_5_9_years":0.7,"income_composition_of_resources":0.905,"schooling":20.6}
```

Finalmente, para cada una de las problemáticas identificadas en la ejecución del pipeline se presentan dos soluciones a nivel de software: la primera consiste en una verificación de los tipos de datos que está ingresando el usuario, a través de condicionales saber qué tipo de dato asignar a cada una de las variables y de la misma manera en caso de que se esté esperando un dato numérico y se obtenga un string, se pueden manejar las excepciones para mostrar un mensaje al usuario indicando qué debe cambiar en los datos. En cuanto a los otros escenarios de error encontrados, la mejor solución encontrada se basa en el manejo de las excepciones. Es decir, en caso de que se deje algún dato vacío, manejar esta excepción para darle un mensaje más diciente para que pueda modificar los datos de entrada.