1. Loading The Dataset

```
import pandas as pd
data = pd.read_csv("data.csv")
data.head()
```

	Time	Temperature (°C)	Pressure (kPa)	Bottom Oil Temperature (°C)	Top Oil Temperature (°C)	BDV (kV)	Health Index (%)	Health Classification
0	4/22/2024	50	112	45	51	65	88	Bad
1	4/22/2024	70	118	51	58	63	71	Bad
2	4/22/2024	48	128	44	49	73	89	Good
3	4/22/2024	49	128	41	47	74	95	Good
4	4/22/2024	55	117	45	58	66	74	Need_Maintenance

2. Understanding The Dataset

```
data.shape
data.info()
data.isnull()
data.isnull().sum()
data.dtypes
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5000 entries, 0 to 4999
Data columns (total 8 columns):
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   Time                         5000 non-null   object
 1   Temperature (°C)             5000 non-null   int64
 2   Pressure (kPa)               5000 non-null   int64
 3   Bottom Oil Temperature (°C)  5000 non-null   int64
 4   Top Oil Temperature (°C)     5000 non-null   int64
 5   BDV (kV)                     5000 non-null   int64
 6   Health Index (%)             5000 non-null   int64
 7   Health Classification        5000 non-null   object
dtypes: int64(6), object(2)
memory usage: 312.6+ KB
Time                          object
Temperature (°C)               int64
Pressure (kPa)                 int64
Bottom Oil Temperature (°C)    int64
Top Oil Temperature (°C)       int64
BDV (kV)                       int64
Health Index (%)               int64
Health Classification         object
dtype: object
```

```
data = pd.get_dummies(data)
data.columns
```

```
Index(['Temperature (°C)', 'Pressure (kPa)', 'Bottom Oil Temperature (°C)',
       'Top Oil Temperature (°C)', 'BDV (kV)', 'Health Index (%)',
       'Time_4/22/2024', 'Health Classification_Bad',
       'Health Classification_Good', 'Health Classification_Need Maintenance',
       'Health Classification_Need_Maintenance'],
      dtype='object')
```

3. Data Preprocessing

```
data['Temperature (°C)'] = data['Temperature (°C)'].astype('category')
data['Pressure (kPa)'] = data['Pressure (kPa)'].astype('category')
data['Bottom Oil Temperature (°C)'] = data['Bottom Oil Temperature (°C)'].astype('category')
data['Top Oil Temperature (°C)'] = data['Top Oil Temperature (°C)'].astype('category')
data['BDV (kV)'] = data['BDV (kV)'].astype('category')
data['Health Index (%)'] = data['Health Index (%)'].astype('category')
data.dtypes
data.describe().T
```

	count	unique	top	freq
Temperature (°C)	5000	24	47	892
Pressure (kPa)	5000	22	129	398
Bottom Oil Temperature (°C)	5000	14	43	646
Top Oil Temperature (°C)	5000	14	58	1461
BDV (kV)	5000	25	79	459
Health Index (%)	5000.0	41.0	95.0	722.0
Time_4/22/2024	5000	1	True	5000
Health Classification_Bad	5000	2	False	3913
Health Classification_Good	5000	2	False	2573
Health Classification_Need Maintenance	5000	2	False	3962
Health Classification_Need_Maintenance	5000	2	False	4552

```
y = data["Health Index (%)"]
X = data.drop("Health Index (%)", axis = 1)
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test=train_test_split(
    X,y,
    train_size = 0.80,
    random_state = 1)
from sklearn.linear_model import LinearRegression
lr = LinearRegression()
lr.fit(X_train,y_train)
```

```
▾ LinearRegression
LinearRegression()
```

```
lr.score(X_test, y_test).round(3)
```

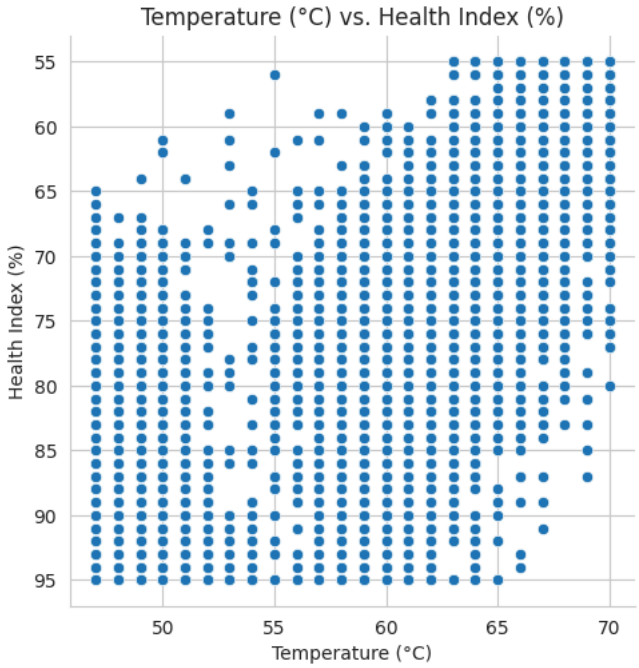
```
0.669
```

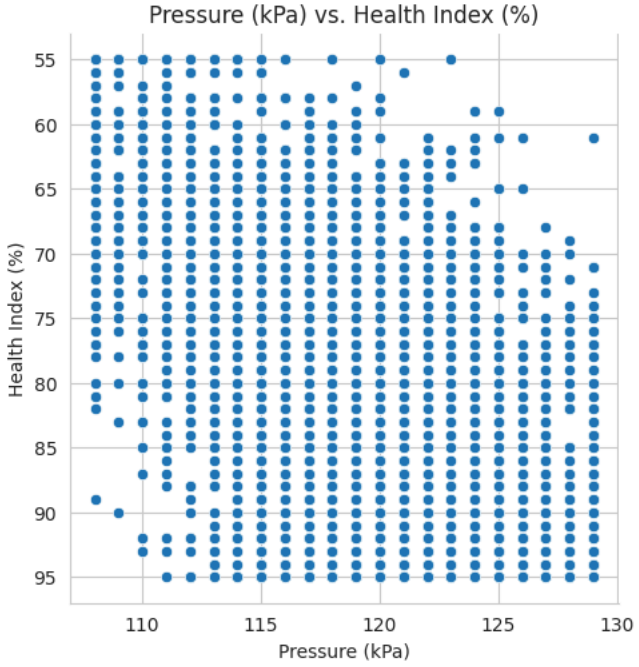
```
lr.score(X_train, y_train).round(3)
```

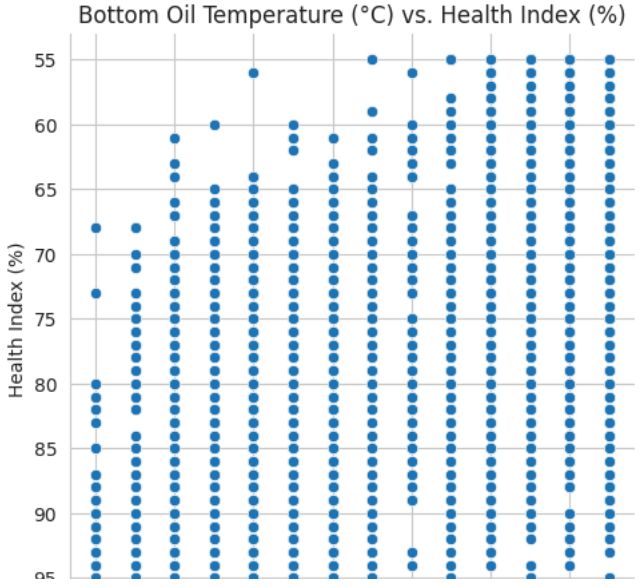
```
0.67
```

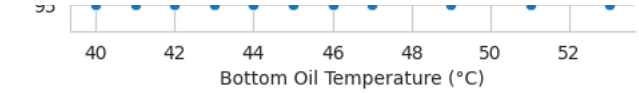
4. Data Visualization

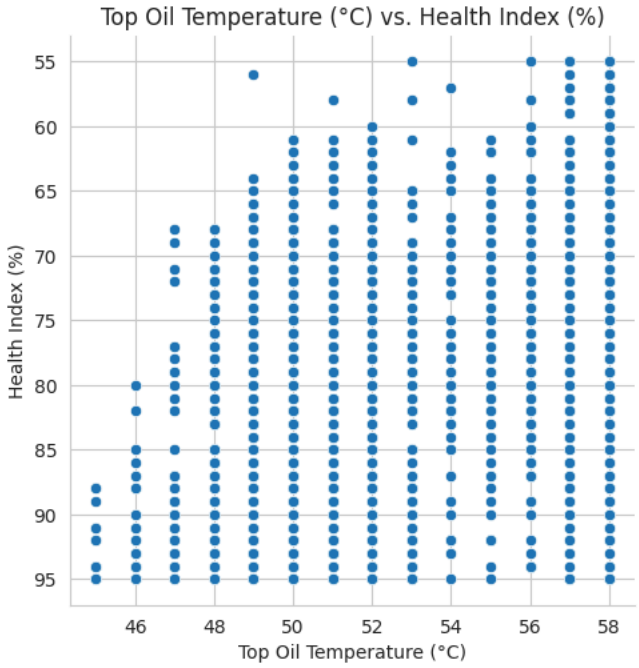
```
import seaborn as sns
import matplotlib.pyplot as plt
sns.relplot(x="Temperature (°C)", y="Health Index (%)", data=df)
plt.title("Temperature (°C) vs. Health Index (%)")
plt.xlabel("Temperature (°C)")
plt.ylabel("Health Index (%)")
plt.show()

sns.relplot(x="Pressure (kPa)", y="Health Index (%)", data=df)
plt.title("Pressure (kPa) vs. Health Index (%)")
plt.xlabel("Pressure (kPa)")
plt.ylabel("Health Index (%)")
plt.show()

sns.relplot(x="Bottom Oil Temperature (°C)", y="Health Index (%)", data=df)
plt.title("Bottom Oil Temperature (°C) vs. Health Index (%)")
plt.xlabel("Bottom Oil Temperature (°C)")
plt.ylabel("Health Index (%)")
plt.show()

sns.relplot(x="Top Oil Temperature (°C)", y="Health Index (%)", data=df)
plt.title("Top Oil Temperature (°C) vs. Health Index (%)")
plt.xlabel("Top Oil Temperature (°C)")
plt.ylabel("Health Index (%)")
plt.show()

sns.relplot(x="BDV (kV)", y="Health Index (%)", data=df)
plt.title("BDV (kV) vs. Health Index (%)")
plt.xlabel("BDV (kV) (°C)")
plt.ylabel("Health Index (%)")
plt.show()
```

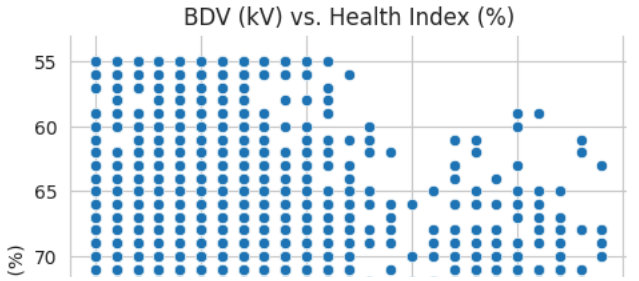

Temperature (°C) vs. Health Index (%)
Health Index (%)
Temperature (°C)
55
60
65
70
75
80
85
90
95
50
55
60
65
70


Pressure (kPa) vs. Health Index (%)
Health Index (%)
Pressure (kPa)
55
60
65
70
75
80
85
90
95
110
115
120
125
130


Bottom Oil Temperature (°C) vs. Health Index (%)
Health Index (%)
55
60
65
70
75
80
85
90


40
42
44
46
48
50
52
Bottom Oil Temperature (°C)


Top Oil Temperature (°C) vs. Health Index (%)
Health Index (%)
55
60
65
70
75
80
85
90
95
46
48
50
52
54
56
58
Top Oil Temperature (°C)


BDV (kV) vs. Health Index (%)
(%)
55
60
65
70