"MACHINE LEARNING MODELS FOR PRICE PREDICTION OF PETROL "

Adhir Barman

st joseph's autonomous college

ABSTRACT

This study, which discusses forecasts for gasoline prices, falls under the purview of machine learning. The very limit of petrol, according to numerous research articles, would be 47 years, even with all the reserves that are readily accessible to the general population. the significance of crude oil, that has a big impact on how much petrol costs. The research approach, which makes use of particular libraries inside Python programming to highlight the value of gasoline through time, would also contain the research papers produced by other researchers, which will be listed in the reference table. Since we can estimate petrol prices but not with certainty what the future may hold, we will use the Y and X train test to arrive at an estimate.

INTRODUCTION

Forecasting the price of petrol can be done only when knowing where it is made from which would be crude oil which is its former stage. [1] Crude oil is very important for energy, policy-makers, market participants, energy risk management and portfolio diversification. Crude oil is used in several industries and petroleum is one of the bye-products. Crude oil prices are impacted by many factors, and the crude oil prices vary drastically based on the current economic and political issues. Therefore, variation of the price has nonlinear characteristics [2, 3]. In order to predict the crude oil prices, it is important to understand the factors that causes these variations, although predicting such issues accurately is difficult. Machine learning methods could be used to predict these indefinite influencing factors. The traditional econometrics model is not sufficient for crude oil price prediction. On the other hand, Recurrent Neural Network (RNN) and long short-term memory (LSTM) models can help in predicting the prices. The long-term trend of crude oil prices is part of commodity attributes, which is also linked to petroleum products, while short-term fluctuation is part of financial attributes. In this term paper, an effort has been made to estimate the petroleum price on the basis of crude oil from both historical and geometric perspectives.

The cost of crude oil, which is controlled by numerous elements, is uncertain. Time series analysis is one of known prediction model that can be used to project prices that use the historical information which we have so far acquired. The features used for prediction include seasonal and random elements which form trends and can be employed to anticipate prices, but they also rely on future trends since they are unpredictable without historical data [4,5]. Petrol does have a large yet crucial influence in our everyday lives because that permits various forms of human mobility. Predicting the fuel price is exceedingly difficult because there are so many different elements that affect it. The datasets are selected and classification analysis is carried out for the building of a predictive model [6] based on a number of variables that have datasets that are widely recognised. These models classify their goals that use the Random Forest method in order to forecast an increase in price range or percent over a predetermined period of time. The Random Forest algorithm, that can be used for both regression and classification is one of the most renowned supervised learning methods [7]. The decision tree's performance engineering are comparable to those of the random forest method [8]. Since the classifiers class can also be utilised for the random forest algorithm and the regressor can be used for tasks like regression, a combo of decision tree and bagging classifier is just not required [9].

The first pre-processed dataset that may be utilised for model training in the construction of prediction models is the petrol dataset. Both of these sets of data and experiment utilised to start the model's learning algorithm. The training is conducted using random forest methods, and an improvement in classifier accuracy is seen when predicting petrol prices [10].

REVIEW OF LITERATURE

The first step in forecasting petrol and crude oil prices is by determining the influencing factor. Since there are many factors which influence the prices many researcher's opinions are not usually consistent. The nonlinear properties of petrol and crude oil prices are characteristics in this term paper, namely commodities attributes and financial attributes. Examining these two features shows the variables that influence variations in petrol pricing [11].

Crude oil is a crucial source of energy, as well as the relationship between demand and supply in addition to the price of close substitutes reflect the commodity's significance. The trajectory in prices for crude oil is affected by the connection between supply and demand, a reflection factor for commodity. According to research by Gallo et al., supply variation is the primary cause of changes in crude oil prices, although price variations also affect demand [12].

Since demand and supply cannot constantly affect oil prices until the 20th century, their impact has exponentially increased [13].To also add on the fact that petroleum is the main product of crude oil within the supply and demand spectrum which affects the price of crude oil. Which also brings us another point mentioning the Price increases for other commodities have an impact on the cost of oil, which has an impact on the price of crude oil. [14] addressed the relationship between the price of silver and petroleum. The bilateral causation between oil prices and silver and gold prices was found by Bildirici and Turkmen [15], however the relationship between physical gold prices and crude oil price shocks is complex.

The collaboration between the European coal market and energy market was proven by Zang and Sun [16]. There was also long-term mixed association, especially with regard to natural gases and pricing [17]. Similarly, in terms of crude oil commodities allocations, the main factor influencing the dynamic of petrol and crude oil prices is indeed the link between demand and supply for these two goods as well as the price of substitute goods [18].

As the crude financial market was formed, the financial characteristics of crude oil continued to improve and grew rapidly. Due to the fact that crude oil is priced in US dollars, it functions as a currency with a value scale in trade transactions. A significant yet unfavourable correlation exists between both the price of crude oil and the dollar's exchange rate as a result of contractual terms on crude oil [19], and this unfavourable association has grown stronger over the past ten years [20].

By arguing that there is a nonlinear relationship between the price of crude oil and the currency value of the dollar, Kilian and Vigfusson [21] refuted this assertion. We can conclude that the dollar exchange rate is not the only variable impacting crude oil prices because Chen et al. [22] had shown us that this non - linear relationship is not noteworthy but that under the demand and supply focused price stocks, the dollar exchange rate has changed has caused different impacts on the crude oil prices. According to the research findings of Bouoiyour et al, crude oil and the Russian exchange rate are positively correlated [23]. The strong relationship between oil prices and rouble exchange rates was explained by Blokhina et al. [24]. Since crude oil could function as gold in the same way as the gold standard exchange system did, the price of crude oil is a barometer of the entire global economy. Even if the dollar and gold were no longer connected under the Jamaican system, investors still sought out physical security to safeguard against the risks associated with the dollar system's volatile value. Because investors utilise crude oil to allocate assets, a financial market for crude oil was established.

The similarities between gold and crude oil are not due to their composition or material, but rather to the fact that both are tangible assets that are accepted into the safe haven currency system, which links the markets for both commodities. Even though the connection between crude oil and gold only lasts a short time, Zhang and Wei [25] found that it is positive. Their interest rates affect the US dollar, which in turn influences the price of crude oil on a global scale [26].

We have already covered a number of elements, such as demand and supply alternate pricing, dollar exchange rates, money supply, and gold prices, which affect the dynamic variations in crude oil prices, which in turn affect petroleum prices. The components that have been chosen in this term paper's discussion to affect the price of crude oil internationally have been substituted into the model for an empirical study.

METHODS

Enhance Random Forest Algorithm (ERFA)

Highly supervised method with time consumption based on categorising the parameters of the model and applicability for a decision tree. It creates a forest where the traits are based on categories and clusters that can be utilized in regression analysis. A special algorithm called ERFA creates a block for every tree that is mapped as a feature in order to group the characteristics into subgroups. Classification techniques that prevent continually scanning the features by creating clusters inside the parameters of the decision trees. Following the development of an ERFA-based model that uses a prediction model and creates a subset in line with cluster splits from each node. The threshold value recommends the node to check the impending tree and suggest the price analysis based on the searching nodes. The data must be cleaned up during pre-processing to get rid of noise, separated commas, undesirable neutral values, etc. As a result of this procedure, the improved random forest quickly deleted those particular features from the dataset of petrol prices. Using the trees' branches, it is possible to extract the ultimate result. For best accuracy, machine learning models work best with categorization and feature mapping [10].

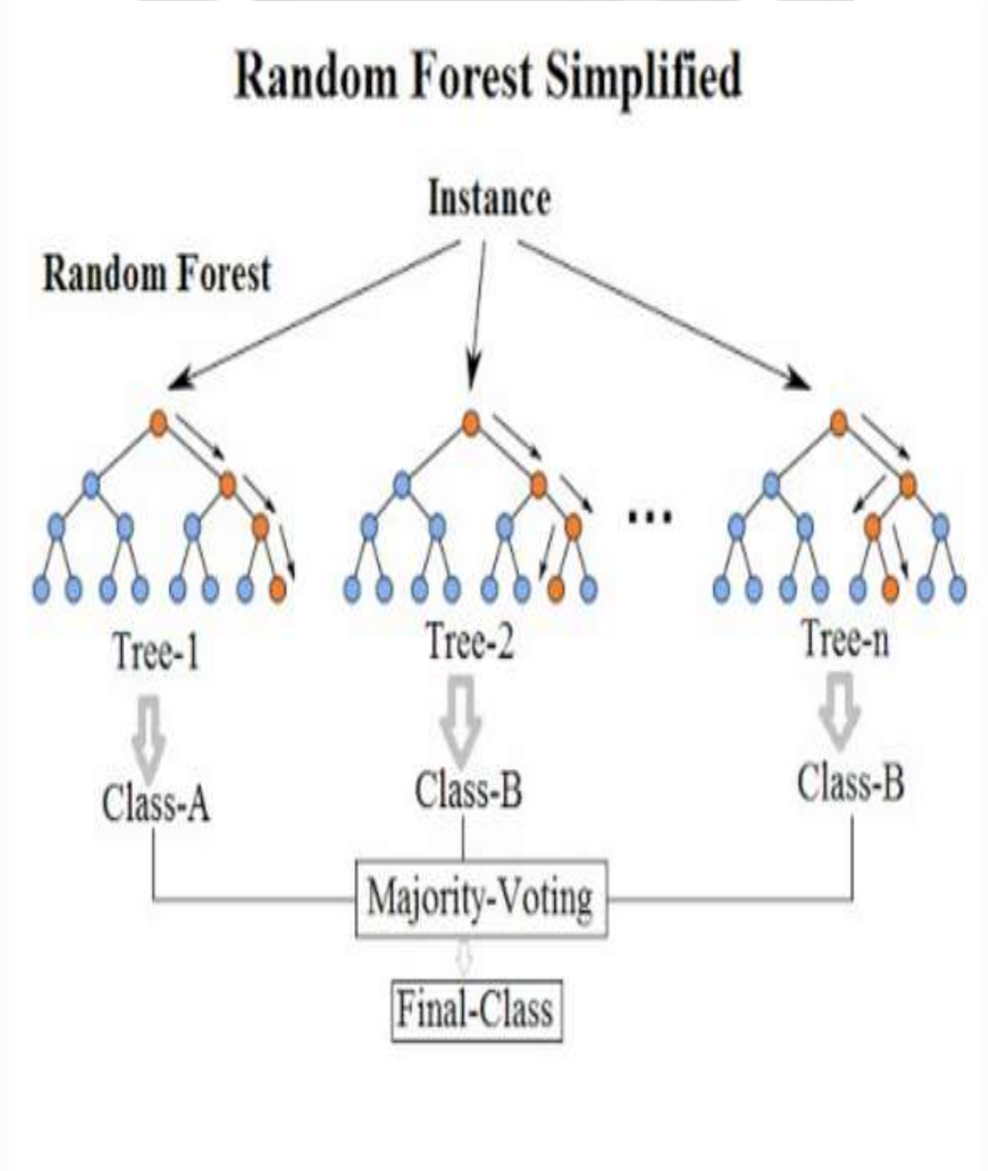


Figure 1. Random Forest Simplified

Proposed system

Petrol production always side tracks due to the heavy demand on fuel from the city Bangalore. Since the number of modes of transportation grow rapidly the need of petroleum sky rockets, and since Bangalore is one of the Metropolitan cities the need for petroleum would be much higher than most cities in India. Good deals and fuel cost consistency may result in greater profits for consumers. The value of petrol is not just chosen by the city; rather, the top executives in the global petroleum market will make the final decision and set the petrol price. The price for petrol will not drop it will only keep growing, thus we come to our main objectives of our research.

1) To create a model for forecasting fuel prices, which are variable and fluctuate on a daily basis.

2) To describe price fluctuations and establish a random forest method to estimate petrol prices with accuracy.

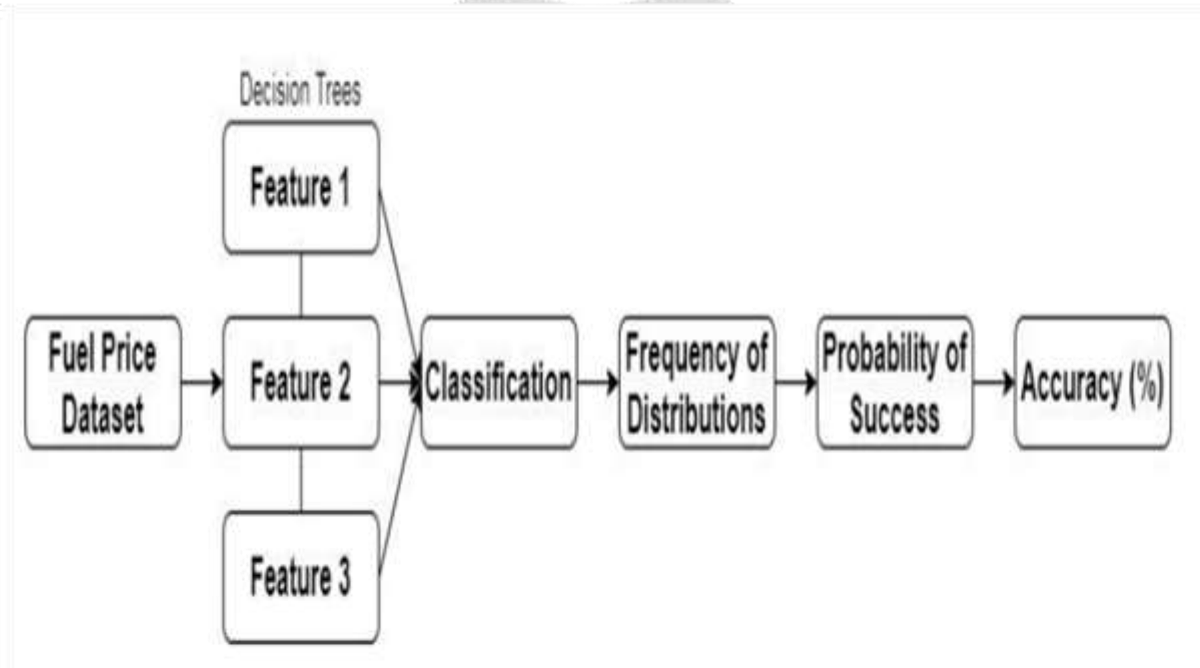


Figure 2. Proposed System [10]

Multiple decision trees will develop when there is a significant difference between the monthly price adjustments. Every time a different tree is used to generate an outcome, the result will vary in accordance with the highest prioritised optimum value being the best bargain.

The operation will be able to choose the optimal answer if all the valid results are represented as binary values of 0 and 1. By applying to every multiple decision trees, the categorization based on characteristics modified in percentage prediction would be sufficient. Through the data, we can see that every minor unstable values, such as underfitting or overfitting, will be reduced, resulting in a greater accuracy overall. The form and colour of a trees will be used to estimate the missing data in the data set. According on the duration of tree usage, the error margins and its order altered. As training dataset, which categorises as column in X variable which approved for predictions of price and Y variable as rows matching labels that meets specified pricing attributes are permitted to fit models, normalised according to the batch size to acquire the accurate price for fuel.

DATASET USED

Petroleum prices Dataset

No.	Country	Daily Oil Consumption (Barrels)	World Share	Yearly Gallons Per Capita
1	United States	19687287	20.30%	934.3
2	China	12791553	13.20%	138.7
3	India	4443000	4.60%	51.4
4	Japan	4012877	4.10%	481.5
5	Russia	3631287	3.70%	383.2
6	Saudi Arabia	3302000	3.40%	1560.2
7	Brazil	2984000	3.10%	221.9
8	South Korea	2605440	2.70%	783.4
9	Canada	2486301	2.60%	1047.6
10	Germany	2383393	2.50%	444.5
11	Mexico	2052607	2.10%	255.1
12	Iran	1803999	1.90%	347.6
13	France	1705568	1.80%	404.3
14	Indonesia	1623000	1.70%	95.1
15	United Kingdom	1583896	1.60%	366.2
16	Singapore	1357000	1.40%	3679.5
17	Thailand	1302000	1.30%	289.4
18	Italy	1236628	1.30%	312.5
19	Spain	1290063	1.30%	424.1
20	Australia	1114645	1.10%	704.3
21	Turkey	941861	1.00%	180.9
22	Taiwan	981203	1.00%	636.9
23	Netherlands	937098	1.00%	846
24	Egypt	877000	0.90%	142.3
25	Iraq	857000	0.90%	358.9
26	United Arab Emirates	896000	0.90%	1467.3

No.	Country	Daily Oil Consumption (Barrels)	World Share	Yearly Gallons Per Capita
27	South Africa	640000	0.70%	174.6
28	Argentina	709000	0.70%	249.8
29	Malaysia	708000	0.70%	353.7
30	Belgium	631522	0.70%	852.6
31	Pakistan	556000	0.60%	41.9
32	Poland	582161	0.60%	234.9
33	Venezuela	598000	0.60%	307.1
34	Vietnam	478000	0.50%	78.3
35	Nigeria	428000	0.40%	35.3
36	Philippines	429000	0.40%	63.4
37	Colombia	357000	0.40%	113.6
38	Algeria	429000	0.40%	162.2
39	Chile	351989	0.40%	296.3
40	Hong Kong	408491	0.40%	864.5
41	Kuwait	359000	0.40%	1390.9
42	Ukraine	244000	0.30%	83.7
43	Morocco	275000	0.30%	120
44	Peru	246000	0.30%	121.9
45	Ecuador	259000	0.30%	240.8
46	Kazakhstan	325000	0.30%	279.4
47	Greece	296101	0.30%	427.6
48	Austria	262352	0.30%	459.8
49	Sweden	322109	0.30%	502
50	Romania	200000	0.20%	154.9
51	Cuba	153000	0.20%	206.9
52	Hungary	155544	0.20%	244.5
53	Czech Republic (Czechia)	179956	0.20%	259.8
54	Lebanon	153000	0.20%	349.3
55	Portugal	236866	0.20%	351.7

No.	Country	Daily Oil Consumption (Barrels)	World Share	Yearly Gallons Per Capita
56	Turkmenistan	149000	0.20%	403.4
57	Switzerland	228194	0.20%	417.5
58	Denmark	158194	0.20%	424.6
59	Israel	236249	0.20%	446.6
60	Ireland	152404	0.20%	497.5
61	Libya	223000	0.20%	526.6
62	New Zealand	166913	0.20%	549.2
63	Finland	210030	0.20%	585.7
64	Panama	155000	0.20%	588.6
65	Norway	204090	0.20%	595.8
66	Oman	183000	0.20%	626.3
67	Qatar	172000	0.20%	993.4
68	Bangladesh	113000	0.10%	11
69	Myanmar	123000	0.10%	35.5
70	Kenya	114000	0.10%	35.6
71	Sudan	140000	0.10%	53.9
72	Angola	133000	0.10%	70.7
73	Sri Lanka	127000	0.10%	92.6
74	Syria	140000	0.10%	122.9
75	Tunisia	97000	0.10%	131.5
76	Jordan	114000	0.10%	182.9
77	Dominican Republic	133000	0.10%	196.1
78	Bulgaria	97000	0.10%	207.9
79	Belarus	137000	0.10%	222.3
80	Puerto Rico	96746	0.10%	451.7
81	Azerbaijan	96000	0.10%	151.2
82	Guatemala	93000	0.10%	86
83	Bolivia	90000	0.09%	125.1
84	Ghana	88000	0.09%	47.4
85	Slovakia	81587	0.08%	229.8

No.	Country	Daily Oil Consumption (Barrels)	World Share	Yearly Gallons Per Capita
86	Ethiopia	74000	0.08%	10.9
87	Serbia	74000	0.08%	128.1
88	Tanzania	71999	0.07%	20.8
89	Croatia	69000	0.07%	251.3
90	Bahrain	62000	0.06%	666.6
91	Lithuania	61612	0.06%	326.9
92	Yemen	60000	0.06%	33.9
93	Honduras	58000	0.06%	95.9
94	Trinidad and Tobago	57000	0.06%	634.3
95	Luxembourg	56194	0.06%	1487.2
96	Costa Rica	55000	0.06%	172.1
97	Jamaica	54000	0.06%	284.8
98	Uruguay	53000	0.06%	237.3
99	El Salvador	52000	0.05%	125.4
100	Slovenia	52298	0.05%	386.5
101	Cyprus	52000	0.05%	681.2
102	Côte d'Ivoire	51000	0.05%	32.8
103	Paraguay	51000	0.05%	115.4
104	Uzbekistan	49000	0.05%	23.9
105	Senegal	49000	0.05%	50.1
106	Cambodia	48000	0.05%	46.7
107	Malta	47000	0.05%	1652.2
108	Nepal	43000	0.04%	24.2
109	Mozambique	40000	0.04%	22
110	Cameroon	40000	0.04%	25.6
111	Kyrgyzstan	40000	0.04%	100.9
112	Latvia	37694	0.04%	292.7
113	Papua New Guinea	37001	0.04%	68.6
114	Nicaragua	37000	0.04%	90
115	Benin	36000	0.04%	50.8

No.	Country	Daily Oil Consumption (Barrels)	World Share	Yearly Gallons Per Capita
116	Afghanistan	35000	0.04%	15.2
117	Bosnia and Herzegovina	35000	0.04%	158.4
118	Georgia	33000	0.03%	126
119	Uganda	32001	0.03%	12.4
120	Estonia	28855	0.03%	336
121	Mauritius	28000	0.03%	340.2
122	Albania	27000	0.03%	143.4
123	Namibia	26000	0.03%	169
124	Zimbabwe	24000	0.03%	26.2
125	State of Palestine	24001	0.03%	79.4
126	Burkina Faso	23000	0.02%	18.9
127	Haiti	23000	0.02%	32.5
128	Mali	21999	0.02%	18.8
129	Zambia	22000	0.02%	20.6
130	DR Congo	21000	0.02%	4.1
131	Tajikistan	21000	0.02%	37.2
132	Mongolia	21000	0.02%	105.3
133	Botswana	21000	0.02%	149
134	North Macedonia	21000	0.02%	154.7
135	Gabon	21000	0.02%	160.3
136	Bahamas	20036	0.02%	812.7
137	New Caledonia	20000	0.02%	1118.2
138	Guinea	19001	0.02%	24.8
139	Iceland	19090	0.02%	880.9
140	North Korea	18000	0.02%	10.9
141	Madagascar	18000	0.02%	11.1
142	Laos	18000	0.02%	40.3
143	Moldova	18001	0.02%	67.9
144	Congo	17000	0.02%	52.3
145	Macao	17110	0.02%	428

No.	Country	Daily Oil Consumption (Barrels)	World Share	Yearly Gallons Per Capita
146	Fiji	16000	0.02%	281.2
147	Brunei	16000	0.02%	584.3
148	Togo	15000	0.02%	30.6
149	Guyana	14000	0.01%	278.2
150	Suriname	13000	0.01%	352.8
151	Maldives	11000	0.01%	354.6
152	Barbados	11001	0.01%	590.1
153	South Sudan	8001	0.01%	11.3
154	Liberia	8000	0.01%	26.7
155	Aruba	8000	0.01%	1169.4
156	Seychelles	7299	0.01%	1169.1
157	Montenegro	7000	0.01%	171.1
158	Rwanda	6700	0.01%	8.8
159	Sierra Leone	6500	0.01%	13.6
160	Malawi	6001	0.01%	5.3
161	Armenia	6000	0.01%	31.3
162	Somalia	5600	0.01%	6.1
163	Cabo Verde	5600	0.01%	161.6
164	Lesotho	5001	0.01%	36.9
165	Cayman Islands	4401	0.00%	1078.3
166	Belize	4001	0.00%	166.5
167	Gambia	3800	0.00%	27.1
168	Bhutan	3001	0.00%	62.4
169	Central African Republic	2800	0.00%	9.5
170	Grenada	2000	0.00%	278.1
171	Burundi	1499	0.00%	2.2
172	Comoros	1300	0.00%	25
173	Dominica	1301	0.00%	279.7
174	British Virgin Islands	1240	0.00%	647.6

No.	Country	Daily Oil Consumption (Barrels)	World Share	Yearly Gallons Per Capita
175	Vanuatu	1100	0.00%	60.6
176	Tonga	899	0.00%	136.3
177	Saint Pierre & Miquelon	660	0.00%	1705.1
178	Kiribati	400	0.00%	54.5
179	Montserrat	400	0.00%	1231.1
180	Saint Helena	70	0.00%	180.2
181	Niue	51	0.00%	484.4

METHODOLOGY

Petrol/Gas Price Worldwide (Program)

The tool used in this case study is Python. Python made it easier to interact with the data for cleaning, to perform a thorough analysis and create effective visualizations for useful insights. 3.2 Loading the R packages:

The following python packages were installed and loaded.

1. pandas
2. numpy
3. matplotlib
4. seaborn

```
In [1]: #import Libraries

        import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        import seaborn as sns
        import plotly.express as px
        import plotly.graph_objects as go
```

Import the datasets

The following CSV files containing data was imported for the analysis.

Petrolgas-prices-worldwide.csv

```
In [2]: df = pd.read_csv("Petrol_Price_.csv", encoding = 'latin-1')
        df.head()
```

	S#	Country	Daily Oil Consumption (Barrels)	World Share	Yearly Gallons Per Capita	Price Per Gallon (USD)	Price Per Liter (USD)	Price Per Liter (PKR)	GDP Per Capita (USD)	Gallons GDP Per Capita Can Buy	xTimes Yearly Gallons Per Capita Buy
0	1	United States	19,687,287	20%	934.3	5.19	1.37	289.97	63,414	12,218	13
1	2	China	12,791,553	13%	138.7	5.42	1.43	302.87	10,435	1,925	14
2	3	India	4,443,000	5%	51.4	5.05	1.33	281.93	1,901	376	7
3	4	Japan	4,012,877	4%	481.5	4.89	1.24	262.05	40,193	8,570	18
4	5	Russia	3,631,287	4%	383.2	3.41	0.90	190.56	10,127	2,970	8

```
In [
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 181 entries, 0 to 180
Data columns (total 11 columns):
 #   Column                              Non-Null Count  Dtype
---  ------                              --------------  -----
 0   S#                                  181 non-null    int64
 1   Country                             181 non-null    object
 2   Daily Oil Consumption (Barrels)     181 non-null    object
 3   World Share                         181 non-null    object
 4   Yearly Gallons Per Capita           181 non-null    float64
 5   Price Per Gallon (USD)              181 non-null    float64
 6   Price Per Liter (USD)               181 non-null    float64
 7   Price Per Liter (PKR)               181 non-null    float64
 8   GDP Per Capita ( USD )              181 non-null    object
 9   Gallons GDP Per Capita Can Buy      181 non-null    object
 10  xTimes Yearly Gallons Per Capita Buy 181 non-null   int64
dtypes: float64(4), int64(2), object(5)
memory usage: 15.7+ KB
```

```
In [4]: df.isnull().sum()
```

```
S#                                  0
Country                             0
Daily Oil Consumption (Barrels)     0
World Share                         0
Yearly Gallons Per Capita           0
Price Per Gallon (USD)              0
Price Per Liter (USD)               0
Price Per Liter (PKR)               0
GDP Per Capita ( USD )              0
Gallons GDP Per Capita Can Buy      0
xTimes Yearly Gallons Per Capita Buy 0
dtype: int64
```

The dataset has 181 rows and 11 columns with no null values. We also notice that columns like Daily Oil Consumption (Barrels) and GDP Per Capita (USD) are of datatype object which need to be converted to float for further analysis.

```
In [5]: df['Daily Oil Consumption (Barrels)']=df['Daily Oil Consumption (Barrels)'].apply(lambda x: x.replace(',',''))
        df['Daily Oil Consumption (Barrels)']=df['Daily Oil Consumption (Barrels)'].astype(float)
        df['GDP Per Capita ( USD )']=df['GDP Per Capita ( USD )'].apply(lambda x: x.replace(',',''))
        df['GDP Per Capita ( USD )']=df['GDP Per Capita ( USD )'].astype(float)
```

Statistical Summary

The statistical summary will further help to understand the minimum, maximum, standard deviation and central tendency of the dataset.

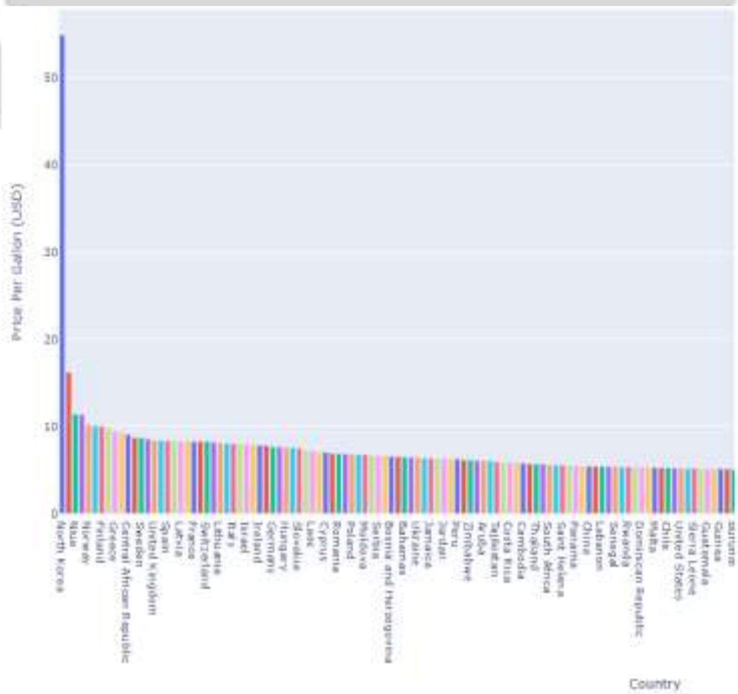
```
In [6]: df.describe()
```

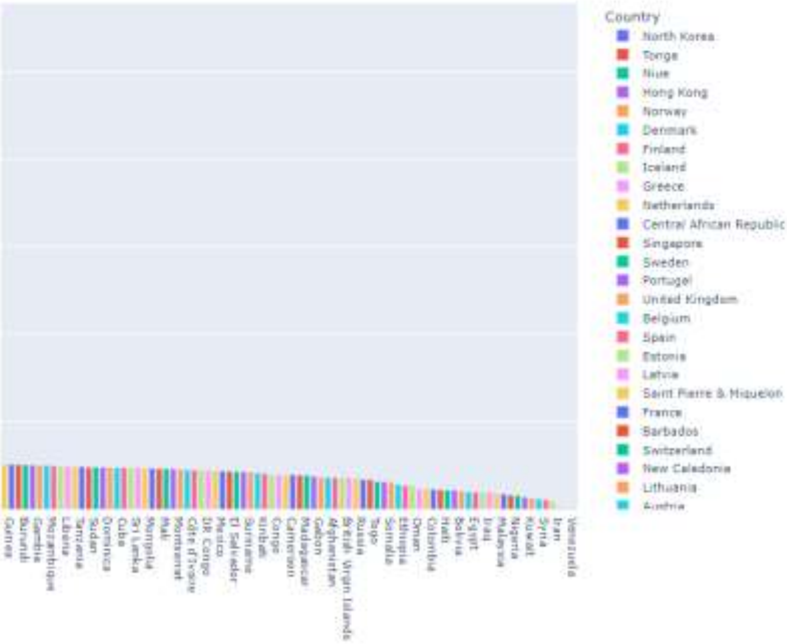
	S#	Daily Oil Consumption (Barrels)	Yearly Gallons Per Capita	Price Per Gallon (USD)	Price Per Liter (USD)	Price Per Liter (PKR)	GDP Per Capita (USD)	xTimes Yearly Gallons Per Capita Buy
count	181.000000	1.810000e+02	181.000000	181.000000	181.000000	181.000000	181.000000	181.00000(
mean	91.000000	5.335730e+05	332.006630	5.695691	1.505138	318.219227	15259.790055	14.204420
std	52.394338	1.858067e+06	436.558735	4.370484	1.154575	244.192081	20542.231615	48.613866
min	1.000000	5.100000e+01	2.200000	0.080000	0.020000	4.650000	274.000000	1.000000
25%	46.000000	2.003600e+04	53.900000	4.150000	1.100000	232.020000	2033.000000	6.000000
50%	91.000000	6.161200e+04	180.200000	5.280000	1.400000	295.040000	6127.000000	9.000000
75%	136.000000	2.623520e+05	424.600000	6.760000	1.790000	377.740000	20234.000000	12.000000
max	181.000000	1.968729e+07	3679.500000	54.890000	14.500000	3066.750000	115874.000000	654.00000(

Comparing price per Gallon per country

We will further try to understand the how petrol/gas prices vary across countries throughout theworld.

```
In [7]: country = df.sort_values("Price Per Gallon (USD)",ascending = False)
        fig = px.bar(country, x= "Country", y = "Price Per Gallon (USD)", color = "Country", title = "Price per gallon per country", hove
        fig.update_traces(textfont_size=12, textangle=0, textposition="outside", cliponaxis=False)
        fig.update_layout(
            height = 800,
            width=1600
        )
        fig.show()
```





Top 10 countries with highest petrol/gas price.

```
In [8]: df.sort_values("Price Per Gallon (USD)",ascending = False)[:10]
```

	S#	Country	Daily Oil Consumption (Barrels)	World Share	Yearly Gallons Per Capita	Price Per Gallon (USD)	Price Per Liter (USD)	Price Per Liter (PKR)	GDP Per Capita (USD)	Gallons GDP Per Capita Can Buy	xTimes Yearly Gallons Per Capita Buy
147	148	North Korea	18000.0	0%	10.9	54.89	14.50	3066.75	1300.0	24	2
180	181	Tonga	899.0	0%	136.3	16.20	4.28	905.22	4903.0	303	2
177	178	Niue	51.0	0%	484.4	11.43	3.02	638.73	15586.0	1,364	3
40	41	Hong Kong	408491.0	0%	864.5	11.35	3.00	634.29	46324.0	4,081	5
58	59	Norway	204090.0	0%	595.8	10.22	2.70	571.26	67390.0	6,594	11
53	54	Denmark	158194.0	0%	424.6	10.04	2.65	561.11	61063.0	6,082	14
63	64	Finland	210030.0	0%	585.7	10.01	2.64	559.21	48773.0	4,872	8
141	142	Iceland	19090.0	0%	880.9	9.83	2.60	549.48	59270.0	6,030	7
47	48	Greece	296101.0	0%	427.6	9.49	2.51	530.02	17623.0	1,857	4
22	23	Netherlands	937098.0	1%	846.0	9.33	2.47	521.35	52397.0	5,616	7

Top 10 countries with lowest petrol/gas price.

```
In [9]: df.sort_values("Price Per Gallon (USD)")[:10]
```

	S#	Country	Daily Oil Consumption (Barrels)	World Share	Yearly Gallons Per Capita	Price Per Gallon (USD)	Price Per Liter (USD)	Price Per Liter (PKR)	GDP Per Capita (USD)	Gallons GDP Per Capita Can Buy	xTimes Yearly Gallons Per Capita Buy
30	31	Venezuela	598000.0	1%	307.1	0.08	0.02	4.65	16056.0	200,700	654
49	50	Libya	223000.0	0%	526.6	0.12	0.03	6.56	3699.0	30,825	59
11	12	Iran	1803999.0	2%	347.6	0.20	0.05	11.21	2423.0	12,115	35
125	126	Brunei	16000.0	0%	584.3	0.83	0.22	46.53	27443.0	33,064	57
69	70	Syria	140000.0	0%	122.9	1.08	0.29	60.49	2033.0	1,882	15
36	37	Algeria	429000.0	0%	162.2	1.18	0.31	66.20	3310.0	2,805	17
37	38	Kuwait	359000.0	0%	1390.9	1.29	0.34	72.33	24812.0	19,234	14
68	69	Angola	133000.0	0%	70.7	1.39	0.37	77.83	1896.0	1,364	19
34	35	Nigeria	428000.0	0%	35.3	1.57	0.42	87.98	2097.0	1,336	38
57	58	Turkmenistan	149000.0	0%	403.4	1.62	0.43	90.52	7612.0	4,699	12

World's largest oil consumers

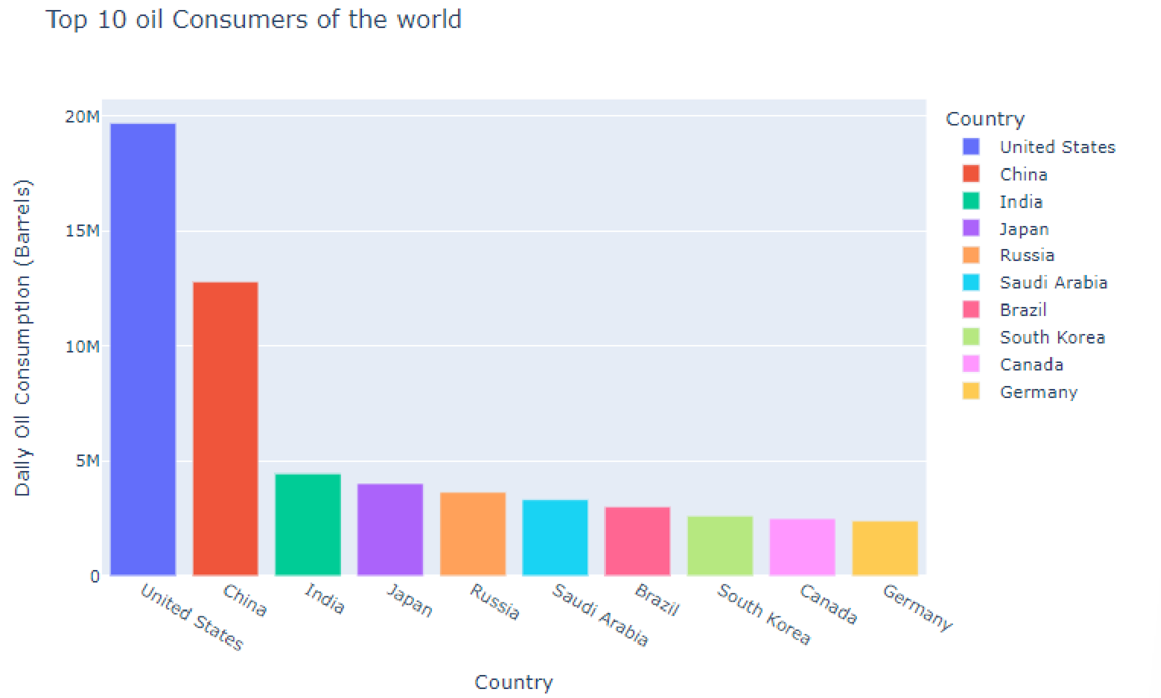
```
In [10]: consumption = df.sort_values("Daily Oil Consumption (Barrels)",ascending = False)[:10]
         consumption
```

Next let's compare the countries based on their daily oil consumption in barrels. Before that we need to convert Daily Oil Consumption (Barrels) into integer.

	S#	Country	Daily Oil Consumption (Barrels)	World Share	Yearly Gallons Per Capita	Price Per Gallon (USD)	Price Per Liter (USD)	Price Per Liter (PKR)	GDP Per Capita (USD)	Gallons GDP Per Capita Can Buy	xTimes Yearly Gallons Per Capita Buy
0	1	United States	19687287.0	20%	934.3	5.19	1.37	289.97	63414.0	12,218	13
1	2	China	12791553.0	13%	138.7	5.42	1.43	302.87	10435.0	1,925	14
2	3	India	4443000.0	5%	51.4	5.05	1.33	281.93	1901.0	376	7
3	4	Japan	4012877.0	4%	481.5	4.69	1.24	262.05	40193.0	8,570	18
4	5	Russia	3631287.0	4%	383.2	3.41	0.90	190.56	10127.0	2,970	8
5	6	Saudi Arabia	3302000.0	3%	1560.2	2.35	0.62	131.34	20110.0	8,557	5
6	7	Brazil	2984000.0	3%	221.9	5.36	1.42	299.27	6797.0	1,268	6
7	8	South Korea	2605440.0	3%	783.4	6.09	1.61	340.52	31632.0	5,194	7
8	9	Canada	2486301.0	3%	1047.6	6.76	1.79	377.74	43258.0	6,399	6
9	10	Germany	2383393.0	3%	444.5	7.65	2.02	427.44	46208.0	6,040	14

```
In [11]: fig2 = px.bar(consumption, x = "Country", y= "Daily Oil Consumption (Barrels)", color = "Country", title = "Top 10 oil Consumers
         fig2.update_traces(textfont_size=12, textangle=0, textposition="outside", cliponaxis=False)
         fig2.show()
```

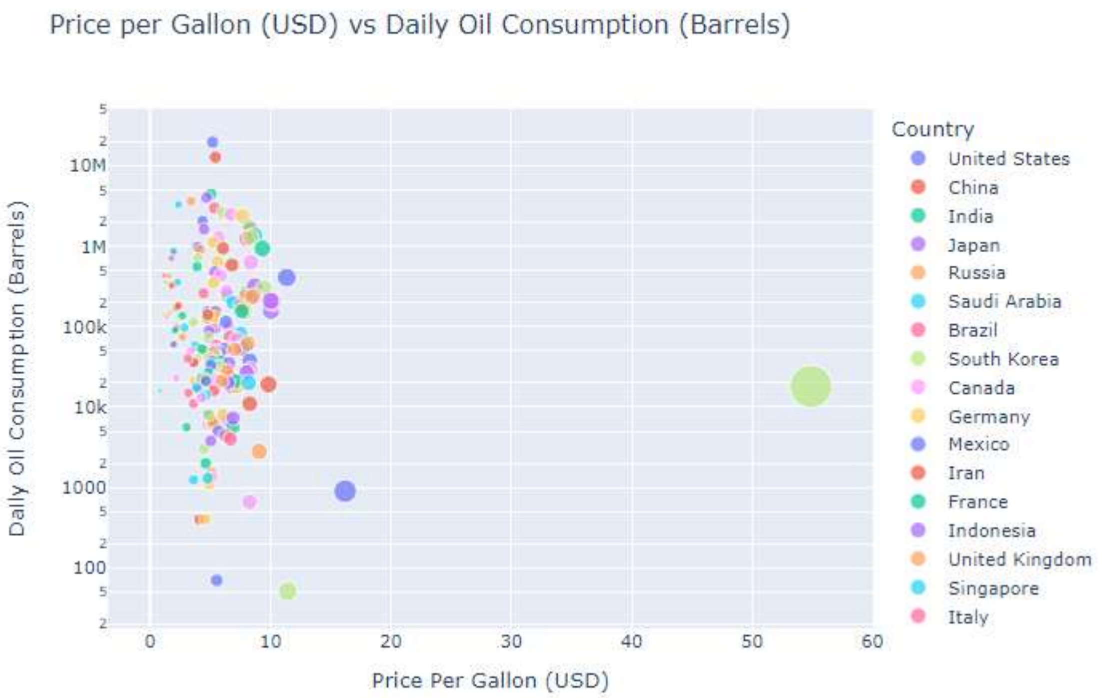
Looking for for top 10 oil consumers of the world.



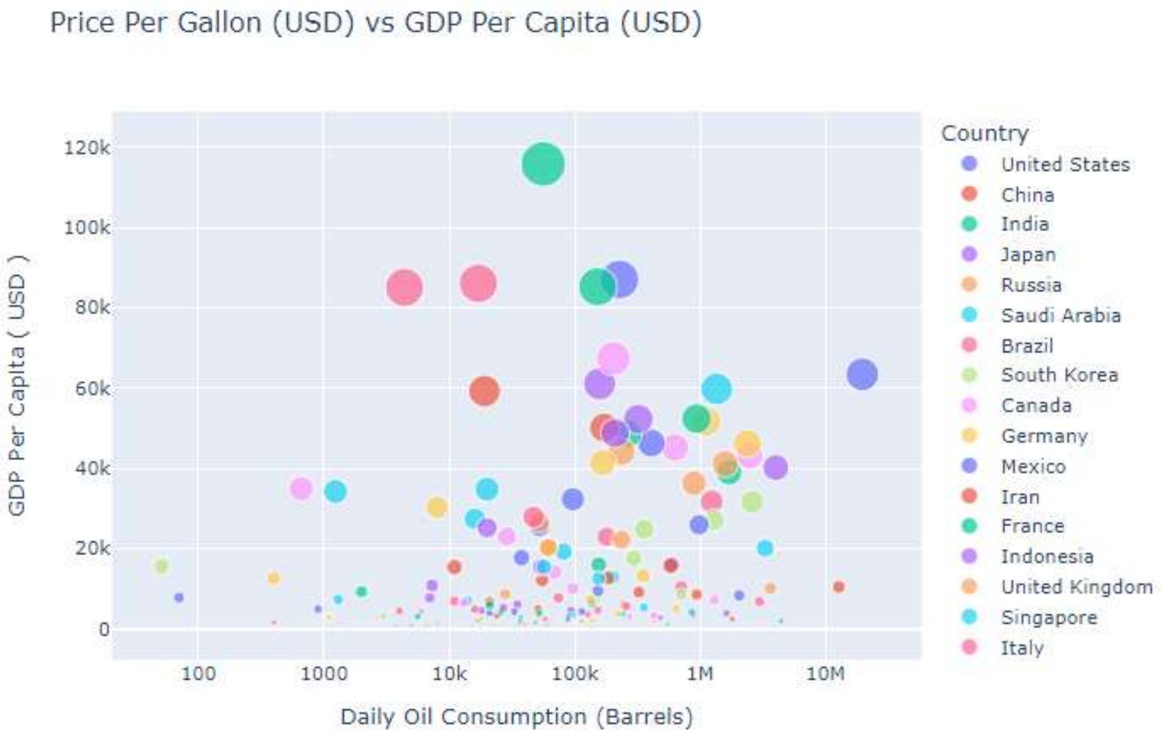

```
In [12]: fig3 = px.scatter(df, x= "Price Per Gallon (USD)", y= "Daily Oil Consumption (Barrels)", size = "Price Per Gallon (USD)", title =

         fig3.show()
```

Exploring relation between variables

```
In [13]: fig4 = px.scatter(df, x = "Price Per Gallon (USD)", y = "GDP Per Capita ( USD )", size = "Price Per Gallon (USD)", color ="Countr
         fig4.show()
```




```
In [14]: fig5 = px.scatter(df, x = "Daily Oil Consumption (Barrels)" , y = "GDP Per Capita ( USD )", size = "GDP Per Capita ( USD )", colo
         fig5.show()
```



Correlation matrix

```
In [15]: plt.figure(figsize= (20,10))
         sns.heatmap(df.corr(),annot = True, fmt = '.2f')
```

```
<AxesSubplot:>
```

Petroleum consumption Dataset (Program)

The bundled analytics libraries in this Python 3 environment are really useful. By the Python Docker container image: https://github.com/kaggle/docker-python For example, here's several helpful packages to load in

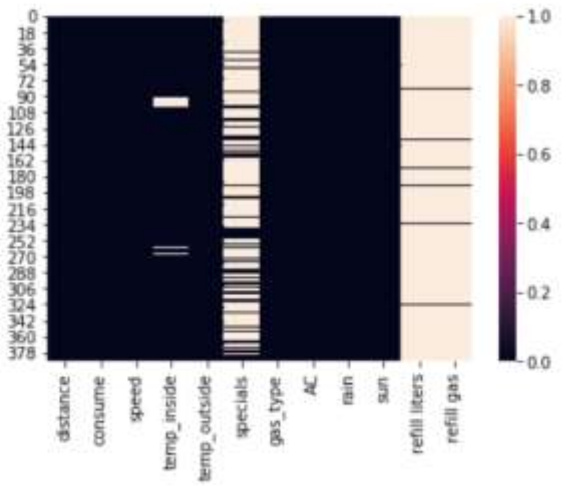
```
In [1]: import numpy as np
        import pandas as pd
        import os
        print(os.listdir("../input"))
        df=pd.read_excel("../input/measurements2.xlsx")
        print(df.head())
```

```
In [2]: import seaborn as sns
        sns.heatmap(df.isnull())
```

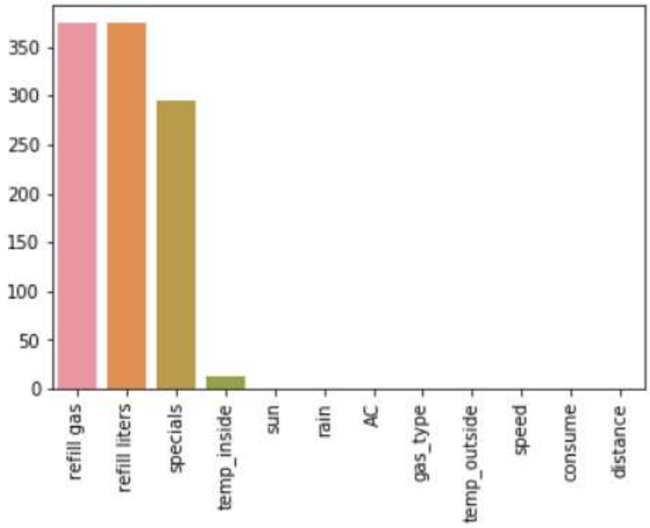
```
['gas_station_orig.jpg', 'measurements.csv', 'measurements2.xlsx']
   distance  consume  speed  ...  sun  refill liters refill gas
0      28.0      5.0     26  ...    0           45.0        E10
1      12.0      4.2     30  ...    0            NaN        NaN
2      11.2      5.5     38  ...    0            NaN        NaN
3      12.9      3.9     36  ...    0            NaN        NaN
4      18.5      4.5     46  ...    0            NaN        NaN

[5 rows x 12 columns]
```

```
In [3]: null_values=df.isnull().sum().sort_values(ascending=False)
        ax=sns.barplot(null_values.index,null_values.values)
        ax.set_xticklabels(ax.get_xticklabels(),rotation=90)
        import matplotlib.pyplot as plt
        plt.show()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f276a114be0>
```

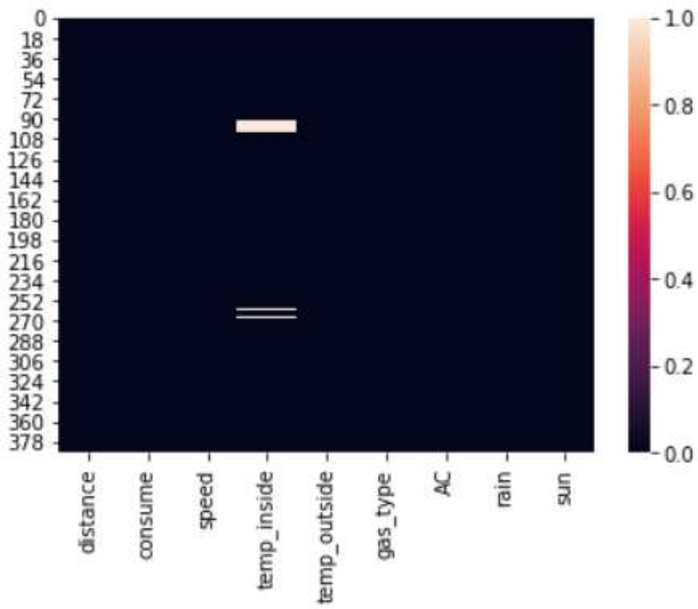



```
In [4]: df.drop(['refill gas','refill liters','specials'],axis=1,inplace=True)
        sns.heatmap(df.isnull())
```




```
In [5]: temp_inside_mean=np.mean(df['temp_inside'])
        print(temp_inside_mean)
```

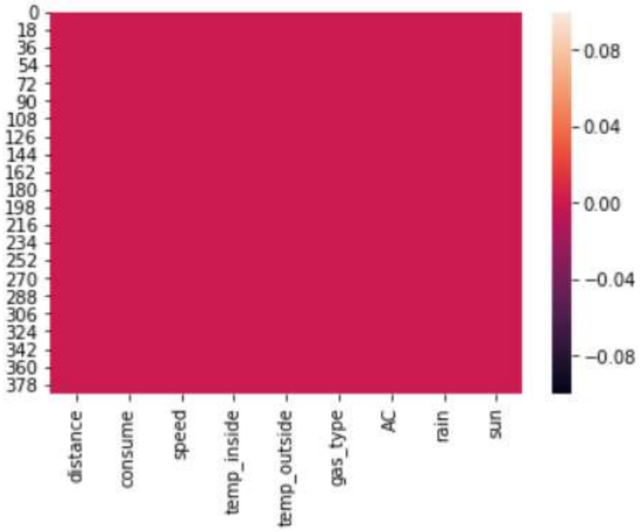
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f27646f4e80>
```




```
In [6]: df['temp_inside'].fillna(temp_inside_mean,inplace=True)
        sns.heatmap(df.isnull())
```

```
21.929521276595743
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f27646f4fd0>
```




```
In [7]: from sklearn.cross_validation import train_test_split
        from sklearn.linear_model import LinearRegression
        l=LinearRegression()
```

```
/opt/conda/lib/python3.6/site-packages/sklearn/cross_validation.py:41: DeprecationWarning: This module was deprecated in versi
on 0.18 in favor of the model_selection module into which all the refactored classes and functions are moved. Also note that t
he interface of the new CV iterators are different from that of this module. This module will be removed in 0.20.
  "This module will be removed in 0.20.", DeprecationWarning)
```

```
In [8]: x=df.drop(['consume','gas_type'],axis=1)
        y=df['consume']
        l.fit(x,y)
```

```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=1, normalize=False)
```

```
In [9]: x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3,random_state=42)
        l.fit(x_train,y_train)
```

```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=1, normalize=False)
```

```
In [10]: y_pred=l.predict(x_test)
         print(l.coef_,l.intercept_)
```

```
[ 0.00523674 -0.02371772 -0.14711979 -0.03724498  0.41456804  0.61676684
 -0.06407861] 9.389308142257121
```

```
In [11]: from sklearn import metrics
         print(metrics.mean_squared_error(y_test,y_pred))
         print(metrics.mean_absolute_error(y_test,y_pred))
         print(np.sqrt(metrics.mean_squared_error(y_test,y_pred)))
```

```
0.7424532609047082
0.6635761182069624
0.8616572757800565
```

```
In [12]: """from sklearn.preprocessing import LabelEncoder
         le=LabelEncoder()
         df['gas_type'] = le.fit_transform(df['gas_type'])"""
```

```
'"from sklearn.preprocessing import LabelEncoder\nle=LabelEncoder()\ndf[\'gas_type\'] = le.fit_transform(df[\'gas_type\'])'
```

```
In [14]: dum1=pd.get_dummies(df['gas_type'])
         print(dum1)
```

..	...	...
358	0	1
359	0	1
360	0	1
361	0	1
362	0	1
363	0	1
364	0	1
365	0	1
366	0	1
367	0	1
368	0	1
369	0	1
370	0	1
371	0	1
372	0	1
373	0	1
374	0	1
375	0	1
376	0	1
377	0	1
378	0	1
379	0	1
380	0	1
381	0	1
382	0	1
383	0	1
384	0	1
385	0	1
386	0	1
387	0	1

[388 rows x 2 columns]

	E10	SP98
0	1	0
1	1	0
2	1	0
3	1	0
4	1	0
5	1	0
6	1	0
7	1	0
8	1	0
9	1	0
10	1	0
11	1	0
12	1	0
13	1	0
14	1	0
15	1	0
16	1	0
17	1	0
18	1	0
19	1	0
20	1	0
21	1	0
22	1	0
23	1	0
24	1	0
25	1	0
26	1	0
27	1	0
28	1	0
29	1	0
..	...	...

In [15]:
```
df=pd.concat([df,dum1],axis=1)
df.drop('gas_type',axis=1,inplace=True)
x1=df.drop('consume',axis=1)
y1=df['consume']
from sklearn.cross_validation import train_test_split
from sklearn.linear_model import LinearRegression
l=LinearRegression()
x_train,x_test,y_train,y_test=train_test_split(x1,y1,test_size=0.3,random_state=42)
l.fit(x_train,y_train)
```

```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=1, normalize=False)
```

```
In [16]: y_pred_1=l.predict(x_test)
         print(y_pred_1)
```

```
[4.80398179 5.24631572 5.16373706 5.23299719 4.52776021 5.99062392
 5.73193936 5.23198354 5.8898096  4.94684204 4.0800537  4.78422755
 6.55357901 4.50083061 5.1268724  5.24267179 5.61167026 5.14823973
 5.48324723 5.36437201 4.13422549 5.30350959 4.94565881 5.23290799
 4.88631664 4.79418748 4.55506668 4.28205093 5.10144732 3.90735262
 4.97478302 5.29391251 4.75042548 4.56699402 5.53113778 5.02945576
 4.6453334  4.03415275 5.10287619 6.16080817 4.47545803 5.28255966
 5.37539962 4.41278157 4.69332325 4.39387259 5.10382269 5.1927726
 4.95992397 4.98995489 4.87121094 5.4268889  5.44648531 5.28120341
 4.61905757 4.90286809 6.70123899 5.3534319  4.71689758 4.78621524
 5.50574979 4.9290579  4.55311849 4.81518093 4.36022913 4.75672285
 5.55769604 4.34876836 4.82767226 4.91585314 4.28138845 4.6582407
 5.19170002 4.97280779 5.18528042 4.79819291 5.32165909 5.10687874
 5.38921307 5.15592614 5.26829591 5.45539801 4.47960294 5.3509791
 5.71243061 4.42243076 5.53113325 5.74565111 5.1678087  4.57634151
 4.81978083 4.50656632 5.10161474 3.96317992 4.30111744 5.47781482
 5.05321366 4.74406453 5.16373706 5.2337835  5.08221941 3.81421222
 4.58755104 4.49417409 5.39720411 4.50237128 4.34387901 4.53984859
 6.50203043 5.78353682 4.7085772  5.13955998 6.21742698 4.85512648
 4.7551128  5.46302901 4.8442509 ]
```

```
In [17]: from sklearn import metrics
         print(np.sqrt(metrics.mean_squared_error(y_test,y_pred_1)))
```

```
0.864693406954018
```

CONCLUSION

We were able to use an upcoming machine learning model to forecast petroleum and crude oil in this term paper. Modelling the non-linear variations in the price movement is facilitated by the deep short network and the recurrent neural networks. We were able to apply the random forest predictive algorithm to anticipate the price of gasoline because the price of gasoline is unpredictable because it just keeps going up. The prediction model had a higher level of accuracy than the current method, and it also predicted the price of petrol for any given time period. When the model was used to make predictions based on experimental findings, it was discovered to be reliable. By selecting the right model, this also aids in arranging the entire import process, which helps the importers in particular to minimise any significant financial losses.

REFERENCES

[1] Lei Yan, Yuting Zhu, and Haiyan Wang "Selection of Machine Learning Models for Oil Price Forecasting: Based on Dual Attributes of Oil", Published 21 Oct 2021.

[2] S. Degiannakis, G. Filis, and V. Arora, "Oil prices and stock markets: a review of the theory and empirical evidence," Energy Journal, vol. 39, no. 5, 2018.

[3] K. Lang and B. R. Auer, "The economic and financial properties of crude oil: a review," The North American Journal of Economics and Finance, vol. 52, Article ID 100914, 2020.

[4] Vijay Anand M, Devi.T, N. Poornima, Gnanavel R, "Accuracy Acquirement for Petrol Price Prediction Using Machine Learning Enhanced Random Forest Algorithm", Volume 12, Issue 3, June 2021: 2468-2474.

[5] Hamdi, Y.; Reem, A. A novel trend-based SAX reduction technique for time series. Expert Syst. Appl. 2019, 130, 113–123.

[6] Gao, X.; Fang, W.; An, F.; Wang, Y. Detecting method for crude oil price fluctuation mechanism under different periodic time series. Appl. Energy 2017, 192, 201–212.

[7] Zhang, M.; Jiang, X.; Fang, Z.; Zeng, Y.; Xu, K. High-order Hidden Markov Model for trend prediction in financial time series. Physica A 2019, 517, 1–12.

[8] Huang, P.; Ni, Y. Board structure and stock price informativeness in terms of moving average rules. Q. Rev. Econ. Financ. 2017, 63, 161–169.

[9] Brock, W.; Lakonishok, J.; Lebaron, B. Simple technical trading rules and the stochastic properties of stock returns. J. Financ. 1992, 47, 1731–1764.

[10] Vijay Anand M, Devi.T, N. Poornima, Gnanavel R, "Accuracy Acquirement for Petrol Price Prediction Using Machine Learning Enhanced Random Forest Algorithm", Volume 12, Issue 3, June 2021: 2468-2474.

[11] Lei Yan, Yuting Zhu, and Haiyan Wang "Selection of Machine Learning Models for Oil Price Forecasting: Based on Dual Attributes of Oil", Published 21 Oct 2021.

[12] A. Gallo, P. Mason, S. Shapiro, and M. Fabritius, "What is behind the increase in oil prices? Analyzing oil consumption and supply relationship with oil price," Energy, vol. 35, no. 10, pp. 4126–4141, 2010.

[13] A. Breitenfellner, J. C. Cuaresma, and C. Keppel, "Determinants of crude oil prices: supply, demand, cartel or speculation," Monetary Policy & the Economy, vol. 4, no. 4, pp. 111–136, 2009.

[14] W. Mensi, S. Hammoudeh, and S. H. Kang, "Precious metals, cereal, oil and stock market linkages and portfolio risk management: evidence from Saudi Arabia," Economic Modelling, vol. 51, pp. 340–358, 2015.

[15] M. E. Bildirici and C. Turkmen, "Nonlinear causality between oil and precious metals," Resources Policy, vol. 46, pp. 202–211, 2015.

[16] Y.-J. Zhang and Y.-F. Sun, "The dynamic volatility spillover between European carbon trading market and fossil energy market," Journal of Cleaner Production, vol. 112, pp. 2654–2663, 2016.

[17] M. Brigida, "The switching relationship between natural gas and crude oil prices," Energy Economics, vol. 43, pp. 48–55, 2014.

[18] Lei Yan, Yuting Zhu, and Haiyan Wang "Selection of Machine Learning Models for Oil Price Forecasting: Based on Dual Attributes of Oil", Published 21 Oct 2021.

[19] R. A. Lizardo and A. V. Mollick, "Oil price fluctuations and U.S. dollar exchange rates," Energy Economics, vol. 32, no. 2, pp. 399–408, 2010.

[20] M. I. Turhan, A. Sensoy, and E. Hacihasanoglu, "A comparative analysis of the dynamic relationship between oil prices and exchange rates," Journal of International Financial Markets, Institutions and Money, vol. 32, pp. 397–414, 2014.

[21] L. Kilian and R. J. Vigfusson, "Nonlinearities in the oil price-output relationship," Macroeconomic Dynamics, vol. 15, no. S3, pp. 337–363, 2011.

[22] H. Chen, L. Liu, Y. Wang, and Y. Zhu, "Oil price shocks and U.S. dollar exchange rates," Energy, vol. 112, pp. 1036–1048, 2016.

[23] J. Bouoiyour, R. Selmi, A. K. Tiwari, and M. Shahbaz, "The nexus between oil price and Russia's real exchange rate: better paths via unconditional vs. conditional analysis," Energy Economics, vol. 51, pp. 54–66, 2015.

[24] T. K. Blokhina, O. A. Karpenko, and A. V. Guirinskiy, “The relationship between oil prices and exchange rate in Russia,” International Journal of Energy Economics and Policy, vol. 6, no. 4, pp. 721–726, 2016.
[25] Y.-J. Zhang and Y.-M. Wei, “The crude oil market and the gold market: evidence for cointegration, causality and price discovery,” Resources Policy, vol. 35, no. 3, pp. 168–177, 2010.
[26] Y. S. Wang and Y. L. Chueh, “Dynamic transmission effects between the interest rate, the US dollar, and gold and crude oil prices,” Economic Modelling, vol. 30, pp. 792–798, 2013.
Figure 1. - https://www.researchgate.net/figure/Random-forest-simplified-representation_fig3_333152684
Figure 2. - Vijay Anand M, Devi.T, N. Poornima, Gnanavel R, “Accuracy Acquirement for Petrol Price Prediction Using Machine Learning Enhanced Random Forest Algorithm”, Volume 12, Issue 3, June 2021: 2468-2474.